Online Learning for Closed-Loop Delay-Constrained Metaverse Systems: A Two-Time-Scale Approach

Jorge Mirande*†, Tijani Chahed*, Salah Eddine Elayoubi†
* SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
† L2S, CentraleSupélec, Université Paris-Saclay, CNRS, 91190 Gif-sur-Yvette, France

***Abstract*—Wireless metaverse applications require tight coupling between haptic delay constraints and multimodal synchronization, creating a trade-off between immersion, reliability, and spectrum efficiency over time-varying channels. We propose a two-time-scale resource allocation framework for closed-loop remote operation. At the fast time scale, downlink bandwidth is adapted per cycle via constrained online convex optimization (COCO) to minimize spectrum cost under long-term delay constraints for haptic and multimodal traffic. At the slow time scale, visual payload is selected using an adversarial multi-armed bandit (EXP3) under non-stationary rewards. Numerical results show the proposed framework achieves near-benchmark spectrum efficiency while steering toward delay-feasible operating points that maximize the long-term immersive performance.**

***Index Terms*—Metaverse, 5G/6G, online learning**

I. Introduction

The metaverse enables real-time interaction with digital objects, remote devices, and users across cyber-physical systems, with applications spanning teleoperation, healthcare, and industrial automation [1]. A defining characteristic of these applications is their multimodal nature: visual, auditory, and haptic modalities must be jointly delivered to sustain immersion. Achieving consistent user perception therefore requires tight cross-modal synchronization under stringent latency and reliability constraints. In this context, the Tactile Internet (TI) supports closed-loop human-in-the-loop interaction, where users remotely control devices through bidirectional exchange of motion and force signals while receiving audiovisual feedback [2]. Unlike conventional multimedia systems, such closed-loop operation is highly sensitive to delay: haptic feedback typically requires round-trip latencies of 1–10 ms, while audiovisual streams must remain synchronized with haptic responses to preserve intermodal coherence [3].

Quality of experience (QoE) for metaverse applications is expressed in terms of immersive experience. In this context, higher visual fidelity offers better QoE, at the expense of increased payload and transmission time, whereas stringent haptic constraints demand additional bandwidth, increasing spectrum consumption. These challenges call for resource allocation mechanisms that jointly manage delay constraints and immersive performance under uncertainty.

In this work, we optimize the immersive experience of a metaverse user in terms of visual content by adapting it to system capacity and changing radio conditions. As the metaverse application runs at a human perception time scale, payload selection, accounting for different visual rendering, takes place at a slower time scale, e.g., seconds, than radio resource allocation. The latter is performed under uncertain radio conditions and unknown application processing times. We hence cast this problem as a two-time-scale online learning framework that couples fast time-scale bandwidth adaptation with slow time-scale visual payload selection.

Prior work within the TI literature has extensively characterized the stability and perceptual sensitivity of closed-loop haptic communication systems [2]–[4]. Teleoperation performance is highly susceptible to communication delay, with millisecond-level round-trip latency directly affecting both haptic perception and control stability [2], [3], while multimodal communication exhibits heterogeneous latency requirements, with haptic traffic demanding ultra-low latency and audiovisual streams tolerating larger delays [4]. To mitigate these effects, approaches such as predictive control and local rendering have been proposed [5], [6]. However, these works focus on end-device delay compensation rather than network-level resource allocation, and do not formulate the problem as delay-constrained closed-loop control with explicit bandwidth decisions under time-varying wireless conditions.

Recent studies [7]–[9] have explored learning approaches for resource management in immersive and extended reality (XR) systems, including latency-aware offloading, QoE-driven optimization, and network slicing. Related efforts consider virtualization and cross-modal transmission strategies to reduce latency and network load [10], [11]. However, these works primarily focus on open-loop QoE or throughput optimization, and do not enforce long-term delay constraints in closed-loop settings. In contrast, we formulate joint bandwidth and visual payload adaptation within an online closed-loop control framework that explicitly accounts for delay constraints over time.

Efforts to support ultra-low latency and high reliability for tactile applications include spectrum reservation and slicing mechanisms [12], [13], as well as analytical frameworks for modeling the impact of delay and signal distortion on haptic perception [14]. These approaches account for latency, reliability, and intermodal constraints, but remain limited to single-time-scale resource allocation or predefined traffic structures. They do not jointly address delay-constrained bandwidth control and adaptive visual payload selection across heteroge-

ISBN 978-3-903176-82-9

neous time scales, nor capture the impact of delay violations on immersive performance.

The main contributions of this work are: (i) We develop a system model for wireless metaverse operation capturing closed-loop haptic delay constraints and intermodal synchronization under time-varying conditions (Section II), (ii) We propose a two-time-scale architecture with bandwidth adaptation at fast scale and visual payload selection at slow one (Section III), (iii) We formulate the problem within an online learning framework, combining constrained online convex optimization (COCO) for bandwidth allocation and a multi-armed bandit approach for payload selection (Section IV), and (iv) Numerical results demonstrate efficient bandwidth adaptation and delay-aware operation across varying channel conditions (Section V).

II. System Model

Enabling immersive metaverse experiences requires support for multiple sensory modalities, including visual and haptic information [1]. This multisensory interaction is supported by a bidirectional link between the metaverse users (MU) and the base station (BS). We consider in this work that the MU transmits commands via the uplink (UL), which the BS relays to the metaverse application. In response, the application generates feedback—consisting of haptic and visual responses, delivered to the MU on the downlink (DL).

To capture the cyclic nature of the interaction between the metaverse user and the metaverse application, we index the communication cycles by $t \in \mathbb{N}$. The corresponding sequence of events for each cycle t is illustrated in Fig. 1. It begins with the MU generating a command, which is transmitted to the BS within a duration denoted by $\tau_c(t)$. The BS then forwards the command to the metaverse application, where it is processed and a response is generated. We denote by $\tau_p(t)$ the duration including the wired transmission to the metaverse application, the command processing time, and the wired return of the response to the BS. In the DL communication from the BS to the MU, two transmissions occur. The first corresponds to the haptic response sent from the BS to the MU, with a duration denoted by $\tau_h(t)$. The second transmission primarily conveys visual information and requires a duration denoted by $\tau_v(t)$.

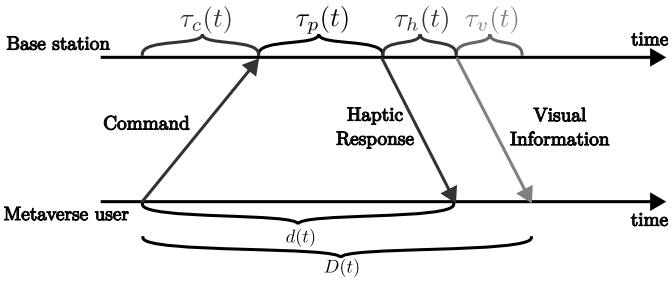

Fig. 1: Time evolution of the communication process between the metaverse user and the application. The base station first transmits the haptic information, followed by the visual information.

We define the closed-loop haptic delay at cycle t, denoted $d(t)$, as the time between the generation of a command and the reception of its haptic feedback. This delay includes the uplink transmission, processing and response generation at the metaverse application, and the downlink haptic transmission, and is given by: $d(t) = \tau_c(t) + \tau_p(t) + \tau_h(t)$. Assuming that the transmission steps within each cycle do not overlap in time, the total end-to-end delay at cycle t, denoted $D(t)$, can be expressed as the sum of the haptic delay $d(t)$ and the additional transmission time required to transmit the visual information: $D(t) = d(t) + \tau_v(t)$. This structure results in closed-loop communication for haptic interactions [2], whereas visual information is transmitted in an open-loop manner synchronized with the haptic feedback [15].

A. Downlink Transmission Model

At each communication cycle t, the base station allocates a downlink bandwidth $w(t) \in \mathcal{W}_0 = [\underline{w}, \overline{w}]$ to the MU, where $\underline{w} > 0$ denotes the minimum schedulable bandwidth unit and $\overline{w}$ represents the maximum allocable bandwidth. The bandwidth allocation $w(t)$ is determined at the radio level and may vary across communication cycles.

The haptic information generated in response to a command consists of a fixed payload of size l_h bits per cycle [2]. The visual information delivered to the MU consists of a payload of size ℓ bits. In the transmission model described in this section, ℓ is treated as constant across communication cycles; its adaptation over longer time horizons is addressed in the following section.

Let $\xi_h(t)$ and $\xi_v(t)$ denote the effective downlink spectral efficiencies (in bit/s/Hz) associated with the haptic and visual transmissions, respectively. These quantities depend on the instantaneous radio conditions at cycle t and are assumed to be strictly positive and bounded.

Under this model, the downlink transmission times for the haptic and visual information are given by: $\tau_h(t) = \frac{l_h}{w(t)\,\xi_h(t)}$ and $\tau_v(t) = \frac{\ell}{w(t)\,\xi_v(t)}$. These expressions capture the inverse dependence of transmission delay on the allocated bandwidth and the instantaneous channel quality, which is a priori unknown at the decision time.

The base station adjusts the downlink bandwidth to control transmission delays while accounting for the limited availability of radio resources. In this work, resource adaptation is performed only on the DL, which carries both the haptic feedback and the high-rate visual payload. The UL transmission, consisting of user command signals with comparatively small payload, is assumed to operate over a separately provisioned and fixed resource share. Moreover, the application processing delay is not subject to control and is treated as an exogenous bounded process. Consequently, both UL transmission time and processing delay are modeled as external components, while downlink bandwidth allocation constitutes the primary control variable for managing delay performance.

III. Two-Time-Scale Problem Formulation

We consider a two-time-scale control structure reflecting the different reconfiguration rates of radio scheduling and application rendering. The BS adapts the downlink bandwidth

$w(t)$ at the communication-cycle level to track fast channel and processing variations. In contrast, the visual payload ℓ induces rendering at the application layer and cannot be updated at this granularity. This motivates separating control decisions into a fast time scale (bandwidth) and a slow time scale (visual payload).

We partition the time horizon into epochs indexed by $n \in \{1, \dots, N\}$, where each epoch consists of T communication cycles. At the beginning of each epoch n, the application selects a visual payload size $\ell_n \in \mathcal{L}$, where $\mathcal{L} = \{\ell^{(1)}, \dots, \ell^{(K)}\}$ denotes the finite set of admissible visual payload levels. The selected payload remains fixed throughout the epoch. Within epoch n, for each communication cycle $t \in \{1, \dots, T\}$, the base station allocates a downlink bandwidth $w(t) \in \mathcal{W}_0$. Therefore, bandwidth is adapted at the fast time scale to track channel variations, while the visual payload evolves only at the slower epoch level.

A. Fast-Time-Scale Problem

At each cycle t, the BS selects a downlink bandwidth allocation $w(t)$ from the compact set $\mathcal{W}_0$, and incurs a spectrum allocation cost proportional to the allocated bandwidth. We represent this cost by an instantaneous loss function given by $f_t(w(t)) = \gamma\, w(t)$, where $\gamma > 0$ weights the cost of spectrum usage. The transmission delays at cycle t depend on the effective spectral efficiencies $\xi_h(t)$ and $\xi_v(t)$, as well as on the uplink transmission time $\tau_c(t)$ and the processing delay $\tau_p(t)$. These quantities are time-varying and unknown at the decision time, and are revealed only after $w(t)$ has been selected. Consequently, the realized delay performance and constraint values cannot be predicted with certainty beforehand, a learning process is thus required, as will be shown in the next section.

We consider the following two normalized constraint functions to capture both the visual and haptic delay violations,

$$g_{1,t}(w(t)) = \frac{d(t; w(t)) - \bar{d}}{\bar{d}},\; g_{2,t}(w(t)) = \frac{D(t; w(t)) - \overline{D}}{\overline{D}}. \tag{1}$$

The constraint $g_{1,t}$ enforces the requirement that the closed-loop haptic feedback $d(t; w(t))$ remains within the stringent delay budget $\bar{d}$, while $g_{2,t}$ ensures that the corresponding multimodal information $D(t; w(t))$ is delivered within the admissible intermodal delay $\overline{D}$. A communication cycle is therefore reliable only if both constraints are satisfied. Typically, $\bar{d}$ is on the order of 10 ms to ensure high-quality haptic interaction in the metaverse, whereas $\overline{D}$ is around 50 ms to preserve the perception of simultaneity across sensory modalities [15], [16].

The delay constraints cannot, in general, be enforced with certainty at each cycle due to the time-varying and unknown transmission conditions. We therefore impose the delay requirements in a long-term sense, allowing temporary violations while controlling their cumulative magnitude over the horizon T. The fast-time-scale problem is thus to minimize the cumulative spectrum cost over T cycles subject to long-term delay constraints.

B. Slow-Time-Scale Problem

The slow-time-scale problem governs the selection of the visual payload size ℓ_n at the beginning of each epoch n. The objective at the application layer is to enhance the immersive experience of the metaverse user while accounting for the delay performance induced by the fast-time-scale bandwidth allocation within the epoch.

We model epoch-level immersive reward associated with visual payload ℓ_n as

$$r_n = \frac{\varphi(n)\, \ln\!\left(\frac{\ell_n}{\underline{\ell}}\right)}{\ln\!\left(\frac{\overline{\ell}}{\underline{\ell}}\right)}, \tag{2}$$

where the logarithmic term captures diminishing returns with respect to visual payload size, i.e., increasing payload when video quality is already high will only increase the QoE marginally [17]. We denote by $\underline{\ell}$ the baseline payload level that generates no immersive gain, and $\overline{\ell} = \max \mathcal{L}$, $\varphi(n) \in [0, 1]$ represents the reliability of the multimodal interaction during epoch n, given by the impact of delay violations within an epoch through a violation rate:

$$\varphi(n) = 1 - \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{g_{1,t}(w(t)) > 0 \;\vee\; g_{2,t}(w(t)) > 0\}, \tag{3}$$

where a violation is declared whenever either the closed-loop haptic delay constraint or the inter-modal delay constraint is not satisfied, and $\mathbb{1}\{\cdot\}$ denotes the indicator function.

This formulation explicitly couples the two time scales: increasing the visual payload enhances visual fidelity but also increases the likelihood of delay violations at the fast time scale, thereby reducing the effective reward through $\varphi(n)$. The mapping between the selected visual payload ℓ_n and the resulting reward r_n is unknown a priori, since $\varphi(n)$ depends on stochastic channel realizations and the fast-time-scale bandwidth adaptation, and it is only observed at the end of each epoch. The slow-time-scale problem is therefore to sequentially select $\ell_n \in \mathcal{L}$ so as to maximize the cumulative reward over N epochs, $\max_{\{\ell_n\}} \sum_{n=1}^{N} r_n$.

IV. Online Solution Methodology

A. Fast-Time-Scale Problem: COCO

To address the bandwidth-allocation problem under uncertainty and long-term constraints, we formulate the fast-time-scale control as a COCO problem. At each cycle t, the BS selects $w(t) \in \mathcal{W}_0$, after which the loss function $f_t(\cdot)$ and constraint functions $g_{m,t}(\cdot)$ are revealed.

Within this framework, performance can be evaluated relative to a dynamic benchmark that may vary across cycles. For each cycle t, define the instantaneous feasible set $\mathcal{W}(t) = \{w \in \mathcal{W}_0 \mid g_{m,t}(w) \le 0,\; m = 1, 2\}$. The dynamic comparator is then defined as

$$w_t^\star = \arg \min_{w \in \mathcal{W}(t)} f_t(w). \tag{4}$$

The sequence $\{w_t^\star\}_{t=1}^T$ represents the best cycle-wise feasible decisions with full knowledge of the realized loss and constraint functions. This dynamic comparator characterizes an oracle policy that perfectly tracks the time-varying system conditions while satisfying the instantaneous delay constraints. Following the COCO framework [18], performance at the fast time scale is evaluated through regret minimization under long-term constraints. The dynamic regret is defined over $t \in \{1, \dots, T\}$ within an epoch as

$$\mathrm{Reg}_f(T) = \sum_{t=1}^{T} \big(f_t(w(t)) - f_t(w_t^\star)\big), \tag{5}$$

where $w_t^\star$ denotes the cycle-wise optimal feasible decision.

Constraint satisfaction is quantified through the cumulative constraint violation measure defined as

$$\mathrm{Vio}_f(T) = \sum_{m=1}^{2} \sum_{t=1}^{T} [g_{m,t}(w(t))]_+, \tag{6}$$

where $[\cdot]_+ = \max\{\cdot, 0\}$. This metric captures the aggregate magnitude of delay violations and does not permit compensation across cycles: a violation at a given communication cycle contributes positively to the cumulative measure regardless of subsequent system behavior. In the context of immersive metaverse interaction, this reflects the fact that delay violations degrade user experience and cannot be compensated by improved performance in later cycles.

To construct the online bandwidth allocation policy, we adopt the COCO framework of [18] and specialize the associated primal–dual methodology to our delay-constrained setting. Within this approach, constraint violations are controlled through virtual queues that act as time-varying dual variables. At each cycle t, the queues associated with the normalized delay constraints evolve as

$$Q_m(t) = \max\Big\{(1-\eta)\, Q_m(t-1) + \big[g_{m,t}\big(w(t)\big)\big]_+,\ \beta\Big\} \tag{7}$$

for $m \in \{1, 2\}$, where $\eta \in (0, 1)$ and $\beta \in (0, \frac{G}{\eta})$ are algorithm parameters, and $G > 0$ is the uniform bound on $|g_{m,t}(w)|$.

The virtual queues accumulate positive delay violations and increase the effective penalty imposed on constraints that are persistently violated, thereby biasing subsequent decisions toward more conservative bandwidth allocations. At each cycle t, the BS computes the bandwidth allocation by minimizing a surrogate objective constructed using information available at the end of cycle $t-1$. Specifically, [18] solves

$$\min_{w \in \mathcal{W}_0} \gamma\, w + \alpha_{t-1}(w - w(t-1))^2 + \sum_{m=1}^{2} Q_m(t-1)\, \big[g_{m,t-1}(w)\big]_+ \tag{8}$$

where $\alpha_{t-1} > 0$ is a non-decreasing regularization parameter.

As observed from Eq. (8), a constraint contributes to the surrogate objective at cycle t only if it was violated at the previous cycle $t-1$. Consequently, the structure of the first-order optimality condition depends on the active set of violated delay constraints at cycle $t-1$. For a given active-set configuration, let $\tilde{w}$ denote the stationary point of the surrogate objective, obtained by differentiating Eq. (8) with respect to w and equating the derivative to zero, as follows

$$\gamma + 2\alpha_{t-1}(\tilde{w} - w(t-1)) - \frac{C(t-1)}{\tilde{w}^2} = 0, \tag{9}$$

where the coefficient $C(t-1)$ aggregates the contributions of the active delay constraints at cycle $t-1$. Its exact expression depends on which of the constraints $g_{1,t-1}$ and $g_{2,t-1}$ are violated, and is detailed in Algorithm 1. For each cycle t, the candidate solution is projected onto the compact feasible set $\mathcal{W}_0$ via the Euclidean projection operator $\mathcal{P}_{\mathcal{W}_0}(\cdot)$. For any unconstrained candidate $\tilde{w} \in \mathbb{R}$, the projection is given by $\mathcal{P}_{\mathcal{W}_0}(\tilde{w}) = \arg\min_{w \in \mathcal{W}_0} |w - \tilde{w}|^2 = \min\{\overline{w},\ \max\{\underline{w}, \tilde{w}\}\}$.

Algorithm 1 Fast-Time-Scale via Virtual-Queue COCO [18]

Input: Epoch index n and fixed payload ℓ_n; set $\mathcal{W}_0 = [\underline{w}, \overline{w}]$; parameters $\eta \in (0,1)$, $\{\alpha_t\}$, spectrum-cost weight γ; initial bandwidth $w(1) \in \mathcal{W}_0$; initial queues $Q_m(1) = \beta$, $m \in \{1, 2\}$.
Output: Bandwidth sequence $\{w(t)\}_{t=1}^T$.
for $t = 2, \dots, T$ **do**
  $C(t-1) \leftarrow 0$
  if $g_{1,t-1}\big(w(t-1)\big) > 0$ **then**
    $C(t-1) \leftarrow C(t-1) + \frac{Q_1(t-1)}{\bar{d}}\, \frac{l_h}{\xi_h(t-1)}$
  end if
  if $g_{2,t-1}\big(w(t-1)\big) > 0$ **then**
    $C(t-1) \leftarrow C(t-1) + \frac{Q_2(t-1)}{\overline{D}} \Big(\frac{l_h}{\xi_h(t-1)} + \frac{\ell_n}{\xi_v(t-1)}\Big)$
  end if
  Find $\tilde{w}$ such that $\gamma + 2\alpha_{t-1}(\tilde{w} - w(t-1)) - \frac{C(t-1)}{\tilde{w}^2} = 0$
  $w(t) \leftarrow \mathcal{P}_{\mathcal{W}_0}(\tilde{w})$
  $g_{1,t}(w(t)) \leftarrow \dfrac{d(t) - \bar{d}}{\bar{d}}$; $g_{2,t}(w(t)) \leftarrow \dfrac{D(t) - \overline{D}}{\overline{D}}$
  $Q_m(t) \leftarrow \max\Big\{(1-\eta)\, Q_m(t-1) + [g_{m,t}(w(t))]_+, \beta\Big\}$
end for
return $\{w(t)\}_{t=1}^T$

B. Slow-Time-Scale Problem: Bandits

We model the sequential feedback structure as a bandit learning problem. At each epoch n, the controller selects a visual payload level $\ell_n \in \mathcal{L}$ and observes only the realized reward r_n associated with the chosen level, without access to the rewards of the other payload options. Recall that, by construction, the reward satisfies $r_n \in [0, 1]$ for all admissible payload levels.

Importantly, the reward sequence need not be stochastic or stationary. The reliability factor $\varphi(n)$ is generated by the interaction between the slow-time-scale visual payload selection and the fast-time-scale COCO dynamics under time-varying channel and processing conditions. As a result, the induced reward process may exhibit non-stationary or adversarial behavior. This motivates the use of adversarial bandit algorithms, such as EXP3, which provide regret guarantees without requiring distributional assumptions [19].

Performance at the slow time scale is evaluated through the cumulative regret with respect to the best fixed visual payload level in hindsight. For each $\ell^{(k)} \in \mathcal{L}$, let $r_n(k)$ denote the reward that would have been obtained at epoch n had payload level $\ell^{(k)}$ been selected.

The cumulative regret over N epochs is defined as

$$\mathrm{Reg}_s(N) = \max_{k\in\{1,\dots,K\}} \sum_{n=1}^{N} r_n(k) - \mathbb{E}\left[\sum_{n=1}^{N} r_n\right], \qquad (10)$$

where the expectation is taken with respect to the internal randomization of the bandit algorithm. This regret notion measures the performance gap between the learning policy and an oracle that, in hindsight, commits to the single best visual payload level over the entire horizon.

V. Numerical Analysis

In this section we evaluate the proposed two-time-scale framework. For this, we emulate the uplink and downlink signal-to-noise ratio (SNR) processes that evolve according to temporally correlated first-order Gauss–Markov models, capturing the discrete-time dynamics of wireless fading channels [20]. The processing delay follows a correlated lognormal process clipped to a bounded interval. The spectral efficiencies are computed from the instantaneous SNR using the finite-blocklength approximation [21] for $\xi_h(t)$ and Shannon's formula for $\xi_v(t)$. The set of admissible visual payload levels is $\mathcal{L} = \{25.0, 62.5, 100.0, 156.2\}$ KBytes, consistent with reported values for metaverse applications [22]. The remaining simulation parameters are summarized in Table I.

Param.	Value	Param.	Value
$\underline{w}$	$1.8 \cdot 10^5$ Hz	β	1
$\overline{w}$	$9 \cdot 10^6$ Hz	$\bar{d}$	10 ms
γ	$3.33 \cdot 10^{-9}$	$\overline{D}$	50 ms
α_t	$6.67 \cdot 10^{-15}\sqrt{t}$	l_h	5 KBytes
η	$1/T$	$\underline{\ell}$	10 KBytes

TABLE I: Simulation parameters.

A. Evaluation of fast-time-scale solution for a single epoch

We first evaluate the behavior of the fast-time-scale algorithm within a single epoch of $T = 1000$ communication cycles, for a fixed visual payload level of $\ell = 62.5$ KBytes, as depicted in Fig. 2. It shows the realized bandwidth allocation $w(t)$ together with two reference policies: the dynamic benchmark $w_t^\star$ (see Eq. (4)) and a static optimal solution defined as $w^* = \max_{t\in[1,\dots,T]}\{w_t^*\}$. We observe that the proposed controller closely tracks the dynamic benchmark while remaining within the feasible interval $[\underline{w}, \overline{w}]$. It adapts to fast stochastic variations in channel and processing conditions without incurring excessive spectrum consumption. In contrast, the static allocation systematically over-provisions resources and therefore does not constitute a meaningful performance reference in a time-varying environment.

We next examine the average operating point of the fast-time-scale controller and its dependence on the selected visual payload. To this end, we generate 5000 independent epoch realizations. Fig. 3 shows the mean downlink bandwidth allocation $w(t)$ for different payload levels $\ell \in \mathcal{L}$. As expected, increasing the visual payload shifts the steady-state operating point toward higher bandwidth values, since the end-to-end delay constraint becomes more restrictive due to the larger visual transmission time. For the reference case of $\ell = 62.5$ KBytes, we additionally include percentile envelopes to characterize the variability induced by stochastic channel and processing dynamics.

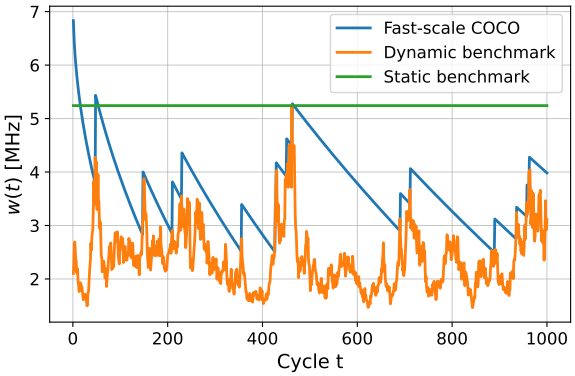

Fig. 2: Fast-time-scale performance over $T = 1000$ communication cycles. (a) Bandwidth allocation $w(t)$ (COCO) compared with the dynamic benchmark $w_t^\star$ and a static benchmark $w_t^\star$.

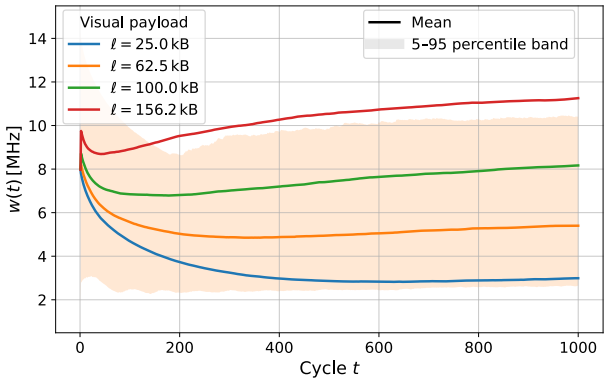

Fig. 3: Visual payload dependent spectrum allocation over $T = 1000$ communication cycles. Mean downlink bandwidth $w(t)$ for different payload modes $\ell_n \in \mathcal{L}$.

Finally, Fig. 4 depicts the fraction of delay violations over the epoch for each visual payload level. We observe that higher payload values result in increased violation rates, reflecting the reduced feasibility region under heavier visual traffic. In contrast to the instantaneous bandwidth allocation, the variability of the violation fraction decreases as the number of communication cycles increases, yielding progressively more stable reliability estimates. This behavior is consistent with the virtual-queue dynamics of the proposed COCO algorithm, which accumulates delay violations and adaptively adjusts the bandwidth allocation to steer the system toward delay-feasible operating points, thereby progressively reducing the variability of the long-term violation.

B. Evaluation of slow-time-scale visual payload adaptation

We next evaluate the performance of the slow-time-scale bandit algorithm responsible for selecting the visual payload level ℓ_n across epochs, as illustrated in Fig. 5. It depicts the

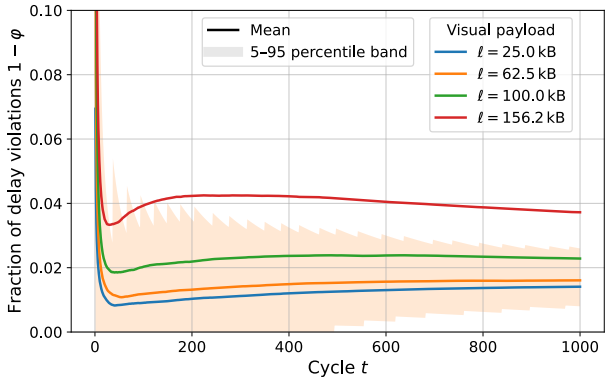

Fig. 4: Fraction of delay violations over $T = 1000$ communication cycles for different visual payload modes $\ell_n \in \mathcal{L}$.

instantaneous reward r_n and $\varphi(n)$ at each epoch. Selecting lower visual payload levels yields high reliability but limited immersive gain. Conversely, selecting higher visual payload levels under unfavorable radio conditions increases transmission load and degrades reliability. The selection progressively concentrates on the payload level that maximizes long-term reward, reflecting the transition from exploration to exploitation.

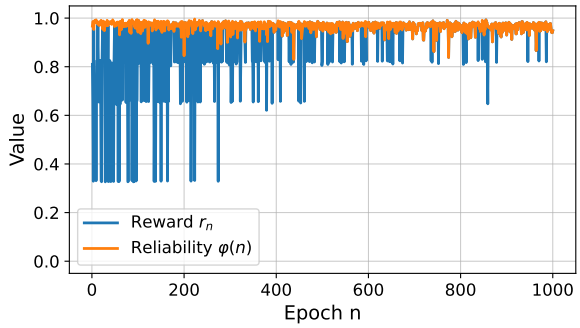

Fig. 5: Performance of the slow-time-scale bandit-based visual payload adaptation, reliability and reward evolution.

VI. Conclusion

This paper investigated two-time-scale resource allocation for wireless metaverse systems supporting closed-loop haptic interaction and multimodal content delivery under stringent delay requirements. We proposed a two-time-scale online framework that captures the coupling between round-trip haptic delay, intermodal coherence, and spectrum efficiency over time-varying wireless channels. At the fast time scale, downlink bandwidth allocation was formulated as a COCO problem with virtual queues to enforce long-term delay constraints. At the slow time scale, visual payload selection was modeled as an adversarial multi-armed bandit problem using EXP3, relying only on epoch-level reward feedback. Numerical results demonstrated stable bandwidth adaptation, controlled delay violations, and effective immersion–reliability trade-offs across different visual payload regimes.

Acknowledgments

This work was carried out in the context of 5GMetaverse, a project funded by the French government as part of the economic recovery plan, namely “France Relance”, and the investments for the future program.

References

[1] H. Wang *et al.*, “A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges,” *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 671–14 688, 2023.
[2] J. Sachs *et al.*, “Adaptive 5G low-latency communication for tactile internet services,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325–349, 2019.
[3] E. Steinbach *et al.*, “Haptic communications,” *Proceedings of the IEEE*, vol. 100, no. 4, pp. 937–956, 2012.
[4] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, “5G-enabled tactile internet,” *IEEE JSAC*, vol. 34, no. 3, pp. 460–473, 2016.
[5] C. Schuwerk, R. Chaudhari, and E. Steinbach, “Delay compensation in shared haptic virtual environments,” in *IEEE Haptics Symposium*, 2014.
[6] X. Xu, H. Singh, Q. Liu, M. Panzirsch, T. Hulin, and E. Steinbach, “A novel energy compensation scheme for quality enhancement in time-delayed teleoperation with multi-dof haptic data reduction and communication,” *IEEE Transactions on Haptics*, pp. 936–944, 2021.
[7] W. Yu, T. J. Chua, and J. Zhao, “Asynchronous hybrid reinforcement learning for latency and reliability optimization in the metaverse over wireless communications,” *IEEE JSAC*, vol. 41, no. 7, 2023.
[8] L. Feng *et al.*, “Resource allocation for metaverse experience optimization: A multi-objective multi-agent evolutionary reinforcement learning approach,” *IEEE TMC*, vol. 24, no. 4, pp. 3473–3488, 2025.
[9] D. Minovski, N. Ögren, K. Mitra, and C. Åhlund, “Throughput prediction using machine learning in LTE and 5G networks,” *IEEE TMC*, pp. 1825–1840, 2023.
[10] Z. Xiang *et al.*, “Reducing latency in virtual machines: Enabling tactile internet for human-machine co-working,” *IEEE JSAC*, vol. 37, no. 5, 2019.
[11] X. Wei, J. Liao, L. Zhou, H. Sari, and W. Zhuang, “Toward generic cross-modal transmission strategy,” *IEEE Transactions on Communications*, vol. 72, no. 10, pp. 6059–6072, 2024.
[12] Z. Hou, C. She, Y. Li, T. Q. S. Quek, and B. Vucetic, “Burstiness-aware bandwidth reservation for ultra-reliable and low-latency communications in tactile internet,” *IEEE JSAC*, vol. 36, no. 11, 2018.
[13] G. Kokkinis, A. Iosifidis, and Q. Zhang, “Deep reinforcement learning-based video-haptic radio resource slicing in tactile internet,” in *IEEE ICC*, 2025, pp. 3821–3826.
[14] J. Schulz, C. Dubslaff, P. Seeling, S.-C. Li, S. Speidel, and F. H. P. Fitzek, “Negative latency in the tactile internet as enabler for global metaverse immersion,” *IEEE Network*, pp. 167–173, 2024.
[15] Z. Shi, S. Hirche, W. X. Schneider, and H. Muller, “Influence of visuomotor action on visual-haptic simultaneous perception: A psychophysical study,” in *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2008, pp. 65–70.
[16] M. Di Luca and A. Mahnan, “Perceptual limits of visual-haptic simultaneity in virtual reality interactions,” in *2019 IEEE World Haptics Conference (WHC)*, 2019, pp. 67–72.
[17] H. Du, J. Liu, D. Niyato, J. Kang, Z. Xiong, J. Zhang, and D. I. Kim, “Attention-aware resource allocation and QoE analysis for metaverse xURLLC services,” *IEEE JSAC*, vol. 41, no. 7, pp. 2158–2175, 2023.
[18] J. Wang, B. Yan, and Y. Liu, “Doubly-bounded queue for constrained online learning: Keeping pace with dynamics of both loss and constraint,” *AAAI*, 2025.
[19] T. Lattimore and C. Szepesvári, *Bandit Algorithms*. Cambridge University Press, 2020.
[20] A. Goldsmith, *Wireless Coms.* Cambridge Univ. Press, 2005.
[21] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Block-fading channels at finite blocklength,” in *ISWCS*, 2013.
[22] C. Schiavo, M. Favero, A. Buratto, and L. Badia, “Bidirectional age of incorrect information: A performance metric for status updates in virtual dynamic environments,” in *MetaCom*, 2025.