The Seven Worlds and Experiences of the Wireless Metaverse: Challenges and Opportunities

Omar Hashash, *Graduate Student Member, IEEE,* Christina Chaccour, *Member, IEEE,* Walid Saad, *Fellow, IEEE,* Tao Yu, *Member, IEEE,* Kei Sakaguchi, *Senior Member, IEEE,* and Mérouane Debbah, *Fellow, IEEE*

***Abstract*—The wireless metaverse will create diverse user *experiences* at the intersection of the *physical, digital, and virtual worlds*. These experiences will enable novel *interactions* between the *constituents* (e.g., extended reality (XR) users and avatars) of the three worlds. However, remarkably, to date, there is no holistic vision that identifies the full set of these worlds, constituents, and experiences, and the implications of their associated interactions on next-generation communication and computing systems. In this paper, we present a holistic vision of a *limitless, wireless metaverse* that distills the metaverse into an intersection of *seven worlds and experiences* that include the: *i) physical, digital, and virtual worlds*, along with the *ii) cyber, extended, live, and parallel experiences*. We then articulate how these experiences bring forth interactions between diverse metaverse constituents, namely, a) *humans and avatars* and b) *connected intelligence systems and digital twins (DTs)*. Then, we explore the wireless, computing, and artificial intelligence (AI) challenges that must be addressed to establish *metaverse-ready* networks that support these experiences and interactions. We particularly highlight the need for end-to-end synchronization of DTs, and the role of human-level AI and reasoning abilities for cognitive avatars. We conclude with a set of recommendations to deploy the limitless metaverse over future wireless systems.**

I. Introduction

The *metaverse*, that sits at the cross-roads of the *physical, digital,* and *virtual realms*, will enable a multitude of world *experiences* that allow users to travel across space and time. Metaverse user experiences are realized at the intersections of the metaverse worlds thereby enabling many socially impactful applications. For example, at the intersection of all three worlds, an extended reality (XR) user can be seamlessly teleported with their senses to visit the world using a multi-sensory avatar. Despite this promising potential of the metaverse, to date, the state-of-the-art [1]–[4] restricts its application space to the individual physical, digital, and virtual worlds, without exploring the opportunities offered by their intertwined experiences. Moreover, the current literature often neglects the role of key constituents that share the metaverse along with XR users and avatars. For example, connected intelligence systems (CISs) (e.g., autonomous vehicles, robots, etc.) actively intervene in the metaverse by acquiring digital twins (DTs) to proactively control their autonomous physical agents [5], a challenging aspect that is less understood in the metaverse literature.

Naturally, taking into account new experiences and constituents introduces unique challenges when deploying the metaverse in the real world. Chief among those challenges is the novel set of *interactions* between the constituents of the different worlds created by the metaverse experiences. Those include two distinct types of interactions: 1) between XR users and avatars and, 2) between CISs and DTs. As a byproduct of such interactions, the metaverse should be extended beyond its traditional human-oriented boundaries to embrace other constituents residing in the real world (e.g., CISs, physical assets, biological systems, etc). Indeed, engineering a *limitless metaverse* to cater for all types of experiences and interactions necessitates transitioning from *human-centric* to *everything-centric* designs. This, in turn, requires overcoming several unique wireless, computing, and artificial intelligence (AI) challenges. For example, the fact that avatars must be aware of their XR users introduces new AI challenges requiring human-level intelligence and cognition. Meanwhile, the need for real time configuration of CISs via DTs imposes stringent end-to-end (E2E) synchronization requirements such as near-zero latency and ultra high data rates to teleport massive physical entities into the metaverse.

Prior works [2]–[4] attempted to investigate the interactions in the metaverse, but they are mainly limited to physical modeling techniques of humans as avatars (e.g., [2] and [3]) and avatar construction methods [4]. Moreover, this prior art [2]–[4] does not design the digital world as a true replica of the real world. Instead, it conflates the virtual and digital paradigms. Consequently, these works do not fully capture the *human-to-avatar* and *CIS-to-DT* interactions between the worlds. *Evidently, there exists a gap in developing a rigorous metaverse vision that precisely recognizes its constituents, worlds, and experiences along with their implications on communication, computing, and AI system designs.*

The main contribution of this paper is to fill this gap by charting a *holistic vision of a limitless, wireless metaverse*, which unlocks the full set of metaverse worlds and their constituents. We particularly articulate how this boundless vision can generate novel experiences between worlds, driving interactions between the constituents over the network. Our contributions include:

- We transform the metaverse from a vague ensemble of worlds, as done in prior works [6], into an intersection of *seven worlds and experiences*, namely, *physical, digital, and virtual worlds*, with *cyber, extended, live, and paral-*

This research was supported by the U.S. National Science Foundation under Grants CNS-2210254 and CNS-2007635 and partially supported by the NICT JUNO project under grant 22404.

O. Hashash and W. Saad are with Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA, USA, Emails: omarnh@vt.edu, walids@vt.edu.

C. Chaccour is with Ericsson, Inc., Plano, Texas, USA, Email: christina.chaccour@ericsson.com.

T. Yu and K. Sakaguchi are with the Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, Japan, Emails: yutao@mobile.ee.titech.ac.jp, sakaguchi@mobile.ee.titech.ac.jp.

M. Debbah is with the KU 6G Research Center, Khalifa University of Science and Technology, Abu Dhabi, UAE, and also with CentraleSupélec, University Paris-Saclay, 91192 Gif-sur-Yvette, France, Email: merouane.debbah@ku.ac.ae.

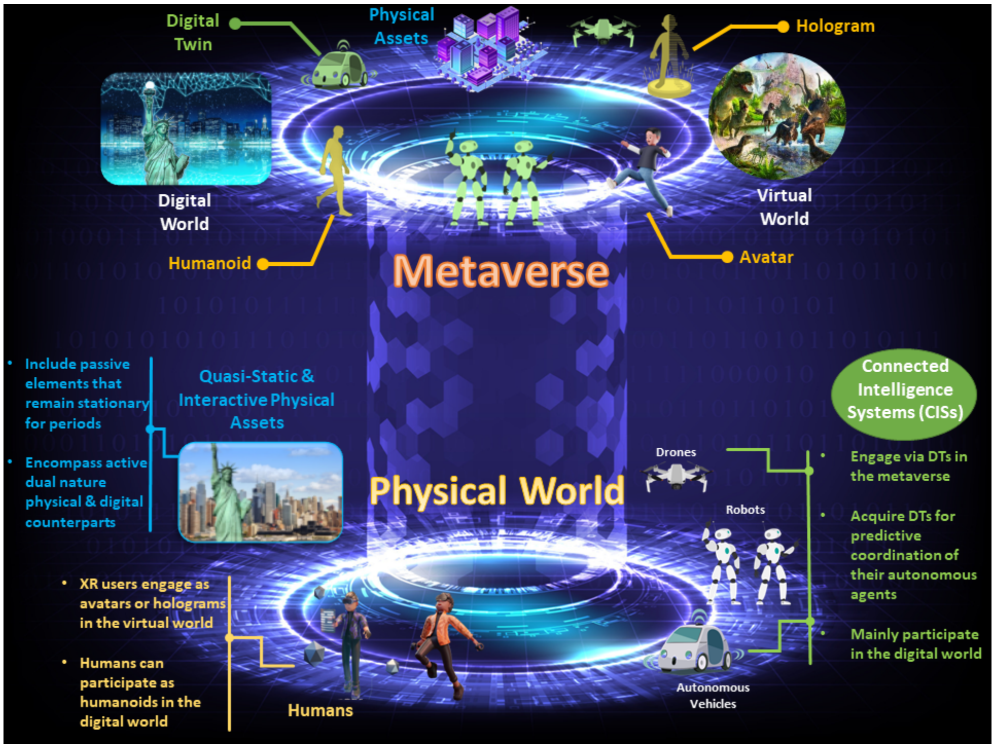

Fig. 1: Illustration of the limitless wireless metaverse comprising different worlds and diverse constituents.

lel experiences (see Fig. 2), that have unique applications and challenges.

- From these experiences, we identify major challenges pertaining to the *interactions* between the constituents of the different worlds. We rigorously define the *human-to-avatar* and *CIS-to-DT* interactions from the physical world into the virtual and digital worlds, respectively.
- We explore the wireless and computing challenges needed to create metaverse-ready networks that support the identified experiences and interactions. We delineate key challenges such as achieving E2E synchronization between XR users and avatars on the network, as well as providing ultra-synchronization high-rate low latency communications to support metaverse interactions.
- We identify key AI challenges including the desynchronization of DTs and the need for resilient designs that accounts for it.
- Finally, we conclude with key recommendations to deploy the wireless metaverse.

II. Metaverse Vision: Breaking Down the Worlds, Constituents, and Experiences

To capture the peculiarity of the metaverse, we unfold its *worlds*, and specify their corresponding *constituents* (see Fig. 1), and delineate the *experiences* arising at their intersection (see Fig. 2).

A. The Metaverse Worlds

1) Physical World: This world is a subset of the real world that encompasses its human (biological), machine, and system fabrics.

2) Digital World: The extent of the digital world broadly entails a duplicate of the real world *in the digital domain*. The digital world mirrors the physical world to enable an alternative *digital reality*. Hence, to acquire a high fidelity digital representation of the real world, real world elements are massively scaled with sensing abilities to confidently replicate their characteristics digitally. The duplicated elements constitute synchronous twin-like representations, however, *in the soft sense of bits and bytes*.

3) Virtual World: In compliance with the *parallel universe theory*, the virtual world is a *multiverse*, i.e., a collection of artificially synthesized hyperspaces, fabricated on a fictional plane of imaginary elements. This world is composed of *visualizable* elements that resemble those of the real world, *in shape*, however, they are purely synthetic, *in nature*. Accordingly, one can conceive that this world includes the set of a) enhanced online platforms (e.g., Roblox), b) social working environments (e.g., Meta Horizon), and c) supplementary assets for enterprise services (e.g., lands, stores, etc).

B. Real World Constituents and their Representations

Decomposing the metaverse into different worlds demands pinpointing the *physical* world constituents and mapping them to their *digital* and *virtual* representations:

1) Humans: To engage in the metaverse, humans are represented in multiple forms depending on the world where they reside:

- **Avatars:** One common human surrogate in the virtual world is in the form of avatars. An avatar is basically a 3D human-like bot that allows users to interact and attain visualization from peers using XR devices. However, due to the real-time synchronization and embodiment of senses and feelings vital to preserve the human-to-avatar duality, avatars must inevitably use *human-driven AI or control systems*.
- **Humanoids:** We define *humanoids* as *massively sensed matterless humans*. Unlike avatars, who embody the XR

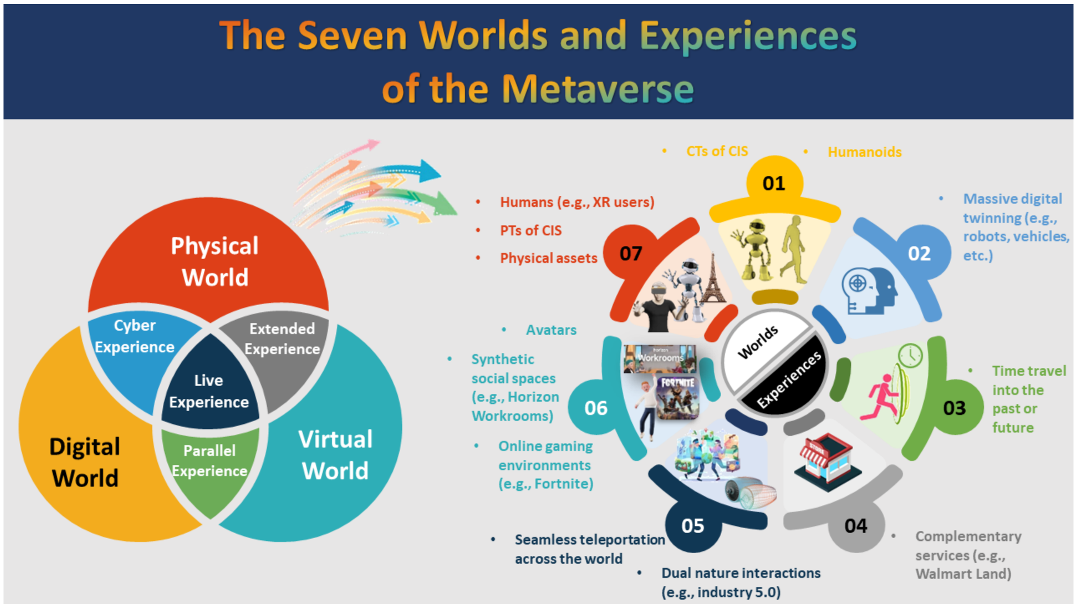

Fig. 2: Illustration of our envisioned seven worlds and experiences of the wireless metaverse.

user in the virtual world, humanoids capture the human existence, in the digital world, without their involvement. In other words, humanoids are digital representations of humans composed of a void nature. Hence, humanoids play a critical role in defining the mechanism of digital applications (e.g., DT configuration mechanism).

- **Holograms:** Human holograms will be a key metaverse constituent. Holograms are 3D projected images synthesized by capturing the light reflected over human entities. Using holographic technology, humans can become omnipresent in the virtual world, once their impinged light beams are captured and reconstructed.

2) CISs: In our envisioned metaverse, complex cyber-physical agents (e.g., autonomous vehicles, drones, robots, etc.) will inevitably share the real world with humans. CISs require DTs to proactively configure their autonomous physical agents and optimize their decision making process. We emphasize that, in the metaverse, *the role of DTs is far beyond a mere replication of physical world elements*. In contrast, DTs – in the massive twinning context – serve as sophisticated, *large-scale, bidirectional operational AI models* that can proactively control, predict, and configure the states of numerous autonomous systems [7]. These CISs make decisions based on their interactions with the environment (e.g., through reinforcement learning or other means). Meanwhile, a cyber twin (CT) is simulated in the digital world to mimic the functionalities of its physical twin (PT). For providing real-time predictive coordination, advanced AI tools are needed to forecast the future state of the CT. Based on this predicted state, PT configurations are proactively adjusted, enabling seamless blending between physical and digital worlds.

3) Quasi-static and interactive physical assets: The metaverse's physical world also includes *physical assets*. Unlike avatars and CISs, the assets representing physical elements and equipment cannot be replicated into the metaverse with the same level of sophistication. While DTs of CISs utilize AI models to actively control the state of autonomous systems, and avatars can be human-driven AI models, physical assets broadly lack this interactivity. This limits the presence of assets to *unidirectional simulation streams.* Physical assets are broadly grouped into two categories:

- **Quasi-static physical assets:** These are representations of physical elements that remain *stationary* for a periodic duration (e.g., buildings, statues, etc). In other words, their rate of change is minimal with respect the rate of change of other constituents. Generally, these elements require massive sensing abilities to be teleported into the metaverse. Quasi-static physical assets are also *passive* (e.g., any change on the digital replica of the statue of liberty in the metaverse does not reflect back to the real world).
- **Interactive physical assets:** Interactive assets comprise the set of *active* elements that have a *dynamic* dual nature of physical and digital counterparts. These assets require a continuous, real-time information pipeline between the counterparts to keep them synchronized. A key example would an interactive machine in Industry 5.0 that is physically present in the real world to be controlled via human intervention. Simultaneously, this machine is digitally teleported into the metaverse for remote interactions with XR users. Here, any minute change from one counterpart is promptly reflected onto the other.

C. On the Experiences Between the Worlds

As the worlds collide to create a novel set of metaverse applications, different *experiences* are realized at every intersection (see Fig. 2):

1) Cyber Experience: The overlap between physical and digital realms creates a new cyber experience to enable CISs and cyber-physical systems that admit DTs throughout their life-cycle. In this experience, PTs and human interventions occur in the physical world, while the corresponding CTs reside in the digital world.

2) Extended Experience: In an extended experience, the physical and virtual boundaries interlapse to extend the real world, thereby providing supplementary virtual services or assets. Prime examples here lie within the industrial and enterprise metaverse. For example, enterprise stores (e.g., Walmart Land) can include merchandise attributed to the real world, however, they may be commercialized in a virtual store that extends the geospatial store from the physical world. This is also the ground for the bulk of state-of-art virtual applications that represent the XR user as a virtual world avatar.

3) Live Experience: The crossroad of all worlds represents a focal point for an online experience that recreates real life in different spatial settings. Live experiences enable seamless holographic teleportation across different locations worldwide. Another example of live experiences includes Industry 5.0 applications. For instance, a physical machine can be duplicated into the digital world and visualized as a virtual element accessible to XR users for intervention.

4) Parallel Experience: The intersection of the digital and virtual worlds promises a parallel experience for XR users to *witness* the world in different time and spatial frames. By leveraging the power of generative AI tools (e.g., diffusion models), XR users can visualize the world in a time travel experience, to the past or future, while also bridging the gap with distant spatial locations.

III. Human-to-Avatar and CIS-to-DT Interactions in the Metaverse

A. Human-to-Avatar Interactions: Towards Cognitive Avatars

Realizing the affinity between XR users and avatars hinges on the human-to-avatar interactions that allow avatars to faithfully *embody* their respective XR users when interacting with peers in the virtual world. For instance, to maintain a robust duality, the avatar should promptly mimic its XR user. This is carried out in two modes: 1) *forward mode* from XR user to avatar and 2) *backward mode* from avatar to XR user, as shown in Fig. 3. The avatar should be able to announce this duality in emulating the XR user actions (forward mode) and actuating the repellent feedback (backward mode). Embodying the XR user in the avatar requires more than just *blindly copying* the user position and movements as in [2]. While this may partially suffice in the forward mode [3], it fails to account for the backward mode in which the feedback from the avatar to the XR user should be *in sync* as well. To address this challenge, the avatar must become *cognizant* of the XR user actions, by *understanding the underlying logic stemming from the sensory inputs that initiated them.* Indeed, a *cognitive avatar* should *learn the persona of its underlying XR user* and *mimic human intelligence*. This is accompanied with the need to transfer the actions and feedback in between XR user and avatar.

To perfectly embody the XR user in the forward mode, it is vital to conserve the *synchronization, accuracy, and precision* in duplicating the actions to the avatar. One way to tackle this duality is through a *mirror game* [8] – a powerful framework for investigating the social motor coordination between two human players. This concept can be extended by substituting one of the players with an *AI-driven cognitive avatar*. Accordingly, the players adhere to a *leader-follower* strategy, where the avatar (follower) learns the unique *kinematic fingerprint* that characterizes how the XR user (leader) exhibits movements and actions in response to their sensory and tracking information (e.g., through imitation learning). The ultimate goal in this forward mirror game is acquiring an AI-driven avatar that minimizes the mismatch in replicating the instantaneous XR actions in terms of accuracy and synchronization.

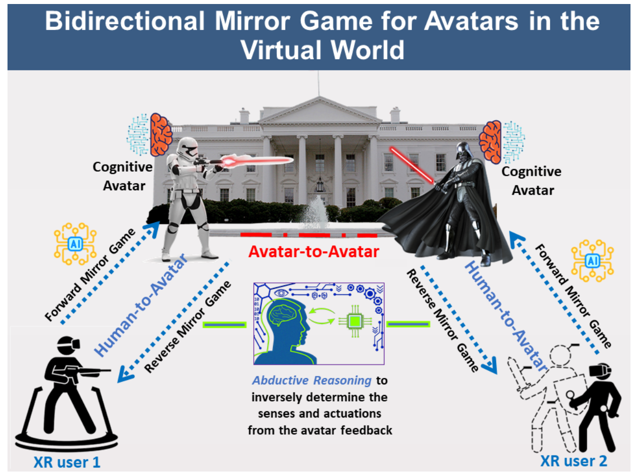

Fig. 3: Illustration of the bidirectional mirror game for avatar interactions in the virtual world.

In the backward mode, the feedback impinging from peer avatars (and virtual elements) should be synchronously reflected to the XR user. This interaction is translated into *senses* and *actuations*. For example, if an avatar is punched on their arm, then this “punched” arm should move *in sync* with that of the XR user. To perform this process, the feedback to the XR user is reflected via a *reverse mirror game*. Flipping the roles, the avatar (leader) in this backward game will use its acquired knowledge, from the forward mode, to *reason for and execute the feedback*, and further pass the corresponding senses and actuations to the XR user (follower). Hence, the avatar requires *abductive reasoning* capabilities to *inversely* reach the senses and actuation inferences that the XR user would most likely feel and experience due to the feedback. Therefore, the overall interaction should be modeled as a *bidirectional mirror game* that integrates the synergies in forward and backward games.

Embodiment requires not only overcoming the mismatch challenges, but also promoting technologies such as the *Internet of Senses*, *Internet of Feelings*, and *affective computing*. In fact, enabling somatosensation for cognitive avatars requires harmonizing the senses to achieve *semantic congruence*, i.e., agreement between the meanings of senses. Thus, synchronization (in time) of the senses at the XR user level is necessary to provide the desired harmony. Moreover, an agreement between the *meanings* of perceived senses and their effective overall stimuli is necessary to ensure the true reception of senses. Moreover, cognitive avatars should be able to express the feelings of the XR users in their interaction with peers. Here, affective computing and *emotion AI* [9] allow the avatar to reflect the physiological state of the XR user – a cornerstone for enabling viable avatar interactions.

B. CIS-to-DT Interaction: A Multi-View Generative AI Approach

Another key metaverse interaction occurs among CISs and DTs. The CIS-to-DT interaction involves the DT prediction mechanism that proactively configures the (physical) autonomous CIS agents in real time, as explained in Section II. However, predicting the future state of a CT is a complex process governed by multiple factors.

A key property of the CIS-to-DT interaction is faithfully predicting the future states of the CTs so as to perfectly optimize the states of the PT proactively. Unlike other simplistic works that predict the future states of CISs on an individual DT basis [10], we consider predicting the future under more practical considerations. In particular, beyond predicting the future state of PT directly from the current state, we also consider looking into the *correlations* existing between the constituents of the world. In other words, our solution considers that the future states of a CIS are also dependent on the current states/actions of the other CISs and the physical assets in the environment. For instance, consider the example in Fig. 4 with autonomous vehicles as CISs. Each vehicle's future state depends on its current state as well as the actions and states of neighboring vehicles (e.g., a speed decrease by one vehicle impacts the speed of other vehicles). In addition, the action taken by the PT is contingent on the state of the physical assets that directly affect the PT. This will implicitly require considering the state of the CT's counter environment in terms of *physical assets* (e.g., road) in the prediction process. Furthermore, CISs will blend with humans in the physical world. Thus, the actions taken by the PTs are affected by the states of humans in their proximity. The future states of the CTs resulting from those will therefore depend on the *situational state of the humans* (e.g., if humans are crossing the road, then the speed of the vehicles will be impacted in the future state). Henceforth, faithfully predict the future states of the CTs, it is pivotal to provide a prediction framework that collectively integrates the: *i) states of PTs, ii) states of the physical assets, and iii) situational states of humans.*

Nevertheless, predicting the future states of these correlated constituents can be seen as equivalent to predicting the future state of the environment. That said, the environment is primarily composed of the humans, CISs, and assets. One approach to address this issue is by considering the aforementioned states as three viewpoints describing our physical system setting at time t. Then, these viewpoints are combined to predict the next setting using the framework of *multi-view learning* – a rigorous AI framework that allows the fusion of data generated from *multiple views of the same subject* [11]. In this framework, each data stream particularly describes the subject from one viewpoint, whereas their union provides a *complementary* overview of the subject in hand. In our setting, the different viewpoints are analogous to the descriptions of the current states of PTs, physical assets, and humans in the physical setting at time t, as in Fig. 4. Firstly, it is necessary to capture these viewpoints from the data. In the first view, the CISs must utilize their collected PT data to describe their states. In the second view, states of the physical assets could be captured through a distributed sensing architecture that uses the massive numbers of sensors in the physical environment [12]. In the third view, the network must capture the human presence via wireless sensing of their situational states. Thereby, their presence will be captured in the form of *humanoids* in the digital world. To provide high resolution modeling of humanoids, joint sensing and communications at (sub-)terahertz (THz) and millimeter wave (mmWave) frequencies is a promising candidate [13].

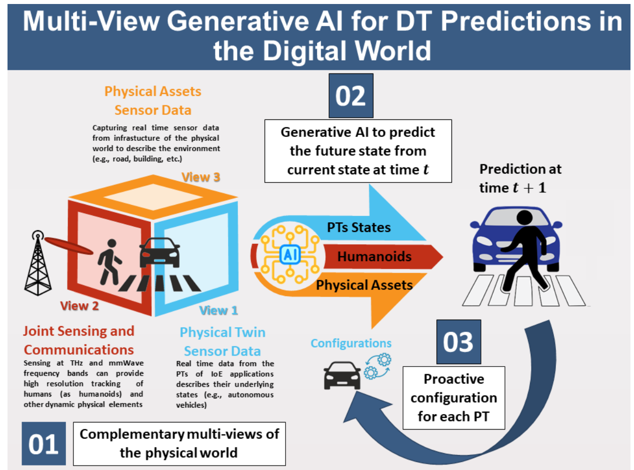

Fig. 4: Overview of the multi-view generative AI process for the DT configuration mechanism in the digital world.

Secondly, after acquiring multiple views of data at time t, the future system setting at $t+1$ (or any additional states, i.e., $t+n$) is predicted by leveraging generative AI abilities. Notably, one promising model can be in the form of large multi-modal models that can both predict the future with multiple modes of data and capture the necessary correlations between the constituents [14]. Thirdly, according to the predicted state of each CT, the corresponding PTs are configured.

IV. Wireless, Computing, and AI Challenges

In this section, we identify a plethora of wireless, computing, and AI challenges that should be addressed to create a metaverse-ready wireless network that supports all metaverse worlds and experiences.

A. Metaverse Architecture

To enable the cyber experience, we must deploy a synchronous digital world over the network. Nevertheless, replicating the real world in a centralized, cloud-based manner could incur high delays that can jeopardize this synchronization. To alleviate this challenge, a shift towards a *decentralized, edge-enabled* digital world is indispensable. Indeed, in [12], we showed that the proper path to *digital reality* is accomplished through decentralizing the digital world into *sub-metaverses*, i.e., digital representations of spaces in the physical world. These sub-metaverses are orchestrated along with their constituents at the wireless edge. In contrast, deploying the *virtual world* is driven by the requirements of the extended, live, and parallel experiences. On the one hand, the extended experience needs a centralized architecture for its applications (e.g., cloud gaming and social networking). On the other hand, the live and parallel experiences require the opening of a *pipeline from the decentralized, digital world* into the virtual world. Therefore, a slice of the virtual world must be confounded with the digital world at the edge. In other words, the virtual world should be deployed in a *semi-distributed* fashion, split between cloud and edge.

B. Wireless and Computing Challenges in the Metaverse

1) Preserving Synchronization and Homogeneity of the Digital World: Conserving the synchronization between the real world and its decentralized sub-metaverses is perhaps the most pressing challenge for maintaining reliable predictions in the CIS-to-DT interaction as well as an immersive live experience. However, this decentralized metaverse needs precise orchestration to properly function. Particularly, it is crucial to: *1) synchronize each sub-metaverse with its counter physical space*, and *2) preserve the homogeneity of the digital world* by keeping its different sub-metaverses in sync with one another. Hence, a critical challenge here is to minimize two types of synchronization: a) *inter-synchronization time* between the physical and digital worlds to ensure a high fidelity replica of all physical assets and, b) *intra-synchronization time* between the sub-metaverses themselves that is pivotal to conserve homogeneity. This requires new probabilistic or stochastic techniques that can effectively model and distribute the physical world under stringent wireless and computing resource budgets. We took a first step towards addressing this challenge in [12]. Therein, we showed how one can model the physical world through a combination of spatial and sensing distributions. Subsequently, we decomposed the physical world into sub-metaverses at the edge through an optimal transport theory technique that yields an optimal synchronization performance. As shown in Fig. 5, this results in a non-uniform distribution over 4 different edge servers at base stations (BSs) with DTs of different synchronization intensity μ, as it exclusively incorporates the available edge resources and the synchronization intensity of DTs. Here, we consider DT applications having three values of μ, where μ captures the maximum allowable time for the DT to replicate the action executed by the PT, and is specific to a certain autonomous application (e.g., robots, drones, etc). In particular, we showed in [12, Fig. 3] that our proposed method achieves up to 25.73% reduction in inter-synchronization time in comparison to the uniform signal-to-noise ratio (SNR)-based association method.

2) Synchronization of Cognitive Avatars on the Network: Since avatars are mainly AI models, one wonders where the model of the cognitive avatar must reside. To maintain high synchronization with the XR user, the avatar model should remain in its proximity. Thus, it is apropos to deploy the *avatars at the edge* for computing purposes. Nevertheless, avatars interact in the virtual world (at the cloud or at another edge) within the extended, live, and parallel experiences. However, moving an avatar away from the edge to the desired destination (cloud or edge) will increase the synchronization mismatch with its corresponding XR user. This can severely degrade the quality-of-experience (QoE). Thus, enabling the avatar interactions without jeopardizing the synchronization and QoE of the XR user is a major challenge. One can investigate the potential of *semantic telepresence* to aid cognitive avatars, where the avatar can still reside at one edge and send its *semantic clone* (or impact) to the other edge to interact with the elements there [15]. Here, the clone will return semantic feedback to the edge where the avatar is located, triggering a response in the human-to-avatar interaction. This feedback contains the sensory actions and actuations that the XR user would have experienced in this interaction.

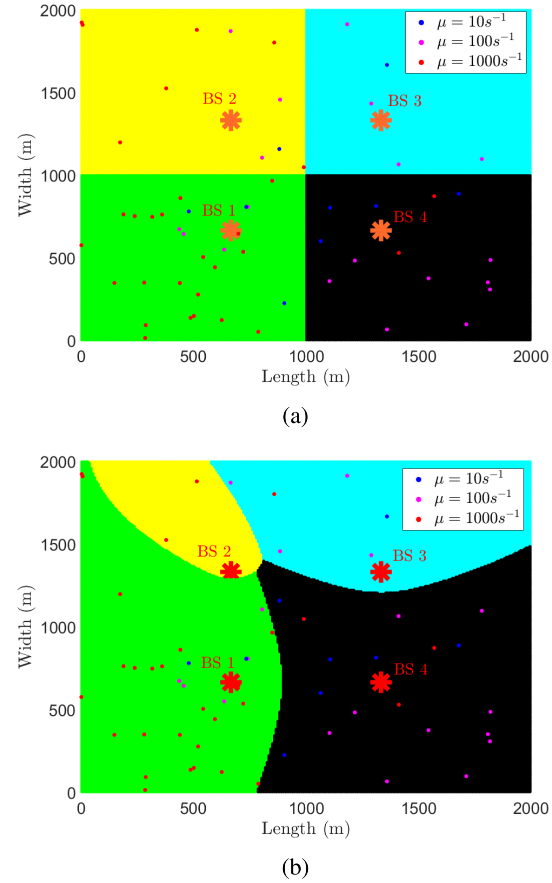

Fig. 5: a) SNR method and b) optimal transport method for sub-metaverse and DT associations.

3) Ultra Synchronization, High Rate, Low Latency Communications for Human-to-Avatar Interactions: Existing wireless technologies that can deliver high rate and low latency for XR users [13] (e.g., THz networks) cannot support our envisioned human-to-avatar interaction. In particular, while existing solutions can possibly guarantee a seamless XR service in terms of HD frames and haptic feedback, they cannot sustain a synchronization of 1) senses and, 2) movements and actions between XR users and their avatars. Indeed, integrating XR for avatars in the metaverse needs an additional dimension of *ultra synchronization* over the network. This is different from the previously discussed synchronization in that it is related to the *delay gap* between the senses/actions themselves. By achieving ultra-synchronization, the network can minimize the *delay between the senses* which is not possible to guarantee with existing ultra reliable low latency communication approaches. Moreover, ultra synchronization is necessary to guarantee the E2E synchronization of movements and actions between XR users and their avatars. This can be crucial for the real time fusion of actions in the *avatar-to-avatar* interaction, which requires the precision of multi-dimensional actions (e.g., through arms, legs, etc.) to forth come together. Therefore, designing such ultra synchronization paradigm is major a network challenge.

C. AI Challenges in the Metaverse

1) Resilience of DTs to de-synchronization: It is evident that CIS-to-DT interactions occur in a *non-stationary* open world setting. Thus, a CT can encounter out-of-distribution data shifts with every change in the physical world of the PT. This data shift can degrade the accuracy of the CT model in twinning the PT and eventually distort its predictions. To ensure the *reliability* of this interaction, the CT must adapt to the data shift by updating its underlying AI model. This could yield a *de-synchronization gap* between the twins [7]. Since DTs must be *hisory-aware*, this gap will continue to incrementally increase with each update phase. It is thus challenging to adapt the CT model to reach utmost accuracy, while limiting this increasing gap. Hence, the *resilience* of DTs to de-synchronization will play a critical role in providing a swift return of the DTs into sync. To achieve such resilience, we need new AI schemes to adaptively and rapidly update the DT models to minimize the de-synchronization gap. Those AI schemes must possess unique properties such as the ability to incorporate incremental knowledge. One promising approach is using continual, lifelong learning [7]. In fact, continual lifelong learning can incorporate a swift update that allows the CT to even generalize over its history. In addition, one can develop a semantic language between the PT and CT that allows the transmission of efficient representations describing the situation of the PT. If properly designed, this language can be robust to the variational data shifts, hence eliminating de-synchronization [15].

2) Higher Order Reasoning for Cognitive Avatars: One main challenge in the human-to-avatar interaction lies in the abilities of cognitive avatars to exploit the knowledge about their human persona, from the forward mode, to facilitate abductive reasoning, in the backward mode. To carry this out, the avatar should primarily determine the relationships between the user sensory input and their actions and movements. Thus, the effectiveness of abductive reasoning hinges on the clarity and consolidation of such relationships. To address this problem, avatars must be endowed with higher order reasoning capabilities. One possible approach is to investigate the *relational reasoning* between the sensory inputs and actions to draw stronger conclusions. Such reasoning incurs a formal, higher order form of relationships, beyond statistical ones. This could involve the use of *relational abstractions* for finding advanced formal conclusions.

V. Conclusion and Recommendations

This paper presented a vision of a limitless, wireless metaverse while identifying its key constituents, worlds, experiences, interactions, and challenges. Building on the developed roadmap, we conclude with several recommendations:

1) Towards a Digital World: Given that XR and virtual world technologies are already underway, a first step toward a limitless metaverse should be to implement a complementary, scalable digital world.

2) Advanced Immersive System: The metaverse will not be only an application of XR and DTs. Instead, it represents an alternative parallel reality, driven by interactions between versatile constituents that should all be highly immersed to build this metaverse (see Fig. 1).

3) Prominent Metaverse-Ready AI: Realizing the metaverse vision requires deploying a new breed of AI algorithms with desirable properties such as reasoning, resilience, and generalization. Developing such metaverse-ready AI frameworks, while building on emerging tools (e.g., large language models), is necessary for a real world deployment of the metaverse.

References

[1] Y. Wang *et al.*, "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 319–352, 2022.

[2] K. Y. Lam *et al.*, "Human-avatar interaction in metaverse: Framework for full-body interaction," in *Proceedings of the 4th ACM International Conference on Multimedia in Asia*, Dec. 2022, pp. 1–7.

[3] L. U. Khan *et al.*, "Metaverse for wireless systems: Architecture, advances, standardization, and open challenges," *arXiv preprint arXiv:2301.11441*, 2023.

[4] J. Wang *et al.*, "Wireless sensing data collection and processing for metaverse avatar construction," *arXiv preprint arXiv:2211.12720*, 2022.

[5] L. U. Khan *et al.*, "Digital-twin-enabled 6G: Vision, architectural trends, and future directions," *IEEE Communications Magazine*, vol. 60, no. 1, pp. 74–80, Jan. 2022.

[6] M. Ali *et al.*, "Metaverse communications, networking, security, and applications: Research issues, state-of-the-art, and future directions," *arXiv preprint arXiv:2212.13993*, 2022.

[7] O. Hashash *et al.*, "Edge continual learning for dynamic digital twins over wireless networks," in *Proc. of the IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, Jul. 2022, pp. 1–5.

[8] M. Lombardi *et al.*, "Using learning to control artificial avatars in human motor coordination tasks," vol. 37, no. 6, pp. 2067–2082, 2021.

[9] G. Bernal *et al.*, "Emotional beasts: visually expressing emotions through avatars in VR," in *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 2395–2402.

[10] Y. Lu *et al.*, "Adaptive edge association for wireless digital twin networks in 6G," *IEEE Internet of Things Journal*, vol. 8, no. 22, pp. 16 219–16 230, Nov. 2021.

[11] C. Xu *et al.*, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[12] O. Hashash *et al.*, "Towards a decentralized metaverse: Synchronized orchestration of digital twins and sub-metaverses," in *Proc. of the IEEE International Conference on Communications (ICC)*, Rome, Italy, May 2023, pp. 1–6.

[13] C. Chaccour *et al.*, "Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 967–993, 2022.

[14] S. Xu *et al.*, "Large multi-modal models (lmms) as universal foundation models for ai-native wireless systems," *arXiv preprint arXiv:2402.01748*, 2024.

[15] C. Chaccour *et al.*, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.

Omar Hashash is a Ph.D. student at the Electrical and Computer Engineering Department at Virginia Tech. His research interests include 6G, artificial intelligence (AI), digital twins, metaverse, and edge intelligence.

Christina Chaccour is a Network Solutions Manager at Ericsson, Inc. She received her Ph.D. from Virginia Tech in 2023. She is the recipient of the Best Paper Award at IEEE IFIP NTMS 2019. She is an Ericsson delegate and editor in influential bodies like 5G Americas, Next G Alliance, and WinnForum.

Walid Saad is a Professor at the Department of Electrical and Computer Engineering at Virginia Tech and an IEEE Fellow. He works at the intersection of wireless networks, machine learning, and game theory. He received many major awards including the 2023 IEEE Marconi Prize Paper Award.

Tao Yu is a specially appointed Associate Professor at Tokyo Tech Academy for Super Smart Society. His research interests include sensor networks, wireless power transmission, and UAV comunications. He is a IEEE and IEICE member.

Kei Sakaguchi is a Dean in Tokyo Tech Academy for Super Smart Society and a Professor at Tokyo Institute of Technology. His research interests include mmWave communications and wireless power transmission. He received many best paper awards, and is a IEEE Senior Member.

Mérouane Debbah is a Professor at Khalifa University of Science and Technology in Abu Dhabi and the founding Director of the KU 6G Research Center. He is an IEEE Fellow, a WWRF Fellow, a EurasipFellow, an AAIA Fellow, an Institut Louis Bachelier Fellow, and Membre émérite SEE. He has received more than 20 best paper awards.