2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom) | 979-8-3503-3333-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/MetaCom57706.2023.00042

Challenges in Metaverse Research: An Internet of Things Perspective

Tarek Abdelzaher[1,*], Matthew Caesar[1], Charith Mendis[1], Klara Nahrstedt[1], Mani Srivastava[2], Minlan Yu[3]
[1]*Department of Computer Science, University of Illinois, Urbana, IL, USA*
[2]*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, USA*
[3]*Department of Computer Science, Harvard University, Cambridge, MA, USA*

***Abstract*—The paper describes research challenges arising from the increasing interest in supporting more immersive and more intelligent environments that enable the next generation of seamless human and physical interactions. These environments span the gamut from augmented-physical to virtual, and are referred to hereafter as the *Metaverse*. We focus on challenges that constitute a natural extension of *Internet of Things (IoT) research*. Among the key applications of IoT has always been the integration of physical and cyber environments to endow "things" with a better contextual understanding of their surroundings, and endow human users with more seamless means of perception and control, ranging from smart home automation to industrial applications. This IoT vision was based on the premise that the number of physical "things" on the Internet will soon significantly outpace humans. *Intelligent* IoT further envisions a proliferation of *edge intelligence* with which humans will interact. The paper elaborates the research challenges that extrapolate the above trajectory.**

***Index Terms*—Metaverse, IoT, Cyber-Physical Systems, AI, Machine Learning.**

I. INTRODUCTION

The rebranding of Facebook as Meta in 2021 [46], brought to the forefront debates over the timeliness and viability of the underlying technology roadmap.[1] Taking a step back to today's capabilities and demands, this paper develops a vision for emerging research challenges and applications in the space of cyber-mediated intelligent interactions between humans and their physical environment. We refer to these interactions as Metaverse applications. We argue that viable applications will build on two great successes of the preceding decade: (i) the proliferation of Internet of Things (IoT) technologies, and (ii) the popularization of practical machine intelligence. These two pillars combined will drive a next generation of content and usecases, centered around enabling new degrees of freedom in perception, interaction, and control.

Spurred by prospects of immersive virtual spaces, many Metaverse surveys emerged in the last few years. The most comprehensive Metaverse survey to date is likely by Lee *et al.* [51]. Several other surveys address aspects of the problem [17], [21], [70], [72], [95]. These aspects include policies, industries, and applications [70], immersion and interaction challenges [17], [21], [72], and privacy and security challenges [95]. Complementing the above efforts, this paper extrapolates a research trajectory that intersects *machine intelligence* and *IoT research* roadmaps, offering answers for what might be both technologically possible and likely, given the current impetus.

Over the last decade, significant and exciting advances in the IoT landscape have redefined its vision, frontiers, and challenges [6]. The natural extension of this roadmap leads to the emergence of a more intelligent, interactive, and distributed application ecosystem that bridges the cyber and physical realms. Said differently, while saturating human senses will ultimately take a bounded amount of bandwidth, it is the explosive growth of world data, generated by the physical environment, by software, and by machine intelligence that will truly drive future content expansion frontiers. This expansion will require innovative interfaces that enhance perception, immersion, and control, in the face of challenges at the edge, in the network, and at the back-end.

At the front end, emerging challenges will parallel two big trends in IoT research and development. The first trend has been the introduction of cloud-assisted machine intelligence at the edge for sensing, data processing, and control [1], [22]. The second trend has been the evolution of interaction modalities with IoT devices. These modalities evolved from utilizing simple APIs (such as structured environmental controls in smart home environments [42], [75]) to accepting human-like inputs, such as gestures [38], natural language commands [33] and visual likeness [73]. Extrapolating these trends, IoT research will continue to seek more natural and intelligent interfaces that allow individuals to interact with, influence, and perceive physical environments or their digital representations, thanks to new visual, acoustic, and haptic interfaces, paired with autonomous agents, machine learning, and digital twinning technology that significantly broaden the types of allowed cyber-physical interactions.

In the network, as more immersive content becomes more popular, applications will push the demand on *high-bandwidth low-latency networking* [90] and call for advances in a wide range of data services [78]. To create the abstraction of an integrated ecosystem, where user identity and (some aspects of) state persist across applications, application interoperability must be achieved, as opposed to creating application-specific

---

[*] Authors listed in Alphabetical Order.

[1] For an example article (of many), see https://wsacommunications.co.uk/blog/the-metaverse-is-dead-long-live-the-metaverse/ (March 2022)

979-8-3503-3333-6/23/$31.00 ©2023 IEEE
DOI 10.1109/MetaCom57706.2023.00042

experiences and vertical silos. The data plane of the new ecosystem will need to support decentralization, sharing, and persistent state [36]. The control plane will support resource allocation policies and control mechanisms for security, privacy, and financial transactions. These requirements will give rise to network architecture challenges, reminiscent of the exploration of design principles that informed today's layered Internet architecture [83].

At the back-end, as application-supported interactions become AI-heavy, it will become important to investigate solutions for accelerating the execution of machine learning primitives and democratize access to AI capabilities at scale. Cloud services in this space may include managing spatio-temporal content [53], efficient storage [13], approximate queries/search [59], distributed retrieval [63], and real-time content summarization [66].

The rest of this paper is organized as follows. Section II briefly reviews recent business drivers of Metaverse-like applications. Sections III, IV, V, VI, and VII overview the five key challenge areas: namely, the IoT front-end, network, back-end, application-inspired AI acceleration, and physical integration/automation, respectively. The paper concludes with Section VIII that summarizes the key points.

II. THE DRIVERS

According to a US-based survey, conducted by pwc.com in 2022, 82% of the surveyed business executives said they "expect Metaverse plans to be part of their business activities within three years"[2]. What are the drivers behind these predictions? Today's more optimistic Metaverse expectations are reminiscent of business predictions for IoT, roughly a decade ago, when a number of industry giants [24] projected an IoT device explosion to (up to) 50 billion by 2020, driven by the reduced cost of sensors, the growing availability of processing capacity and communication bandwidth, and technological advances such as IPv6 and ubiquitous wireless coverage. Many drivers were cited from personal wearables and home automation to office, healthcare, transportation, and smart city services [89], prompting organizational and market research pillars, such as McKinsey & Company, to predict the global IoT market to reach $11.1 trillion by 2025 [20]. What is the next step in this evolution?

Standardization: The prospective application domains of the Metaverse are perhaps best reflected in the membership of the *Metaverse standards forum*, founded in June 2022, with the goal of ensuring interoperability of Metaverse components. Current members include telecoms, search engines, chip manufacturers, defense contractors, software/OS developers, social media companies, the entertainment industry, game engines, sports associations, and consumer retail outlets. They feature such prominent names as Acer, Adobe, ARM, BBN, Blockchain, Comcast, Deloitte, T-mobile, Ericsson, Fujifilm, Google, Huawei, Ikea, Intel, Juniper, Lenovo, LG, Meta, Microsoft, the National Basketball Association (NBA), Nestle, Nokia, NVIDIA, Pramount, Qualcomm, Samsung, Siemens, SONY, Unity, Verizon, and W3C. As visions for Web 3.0 also gain ground, discussions for Mataverse support in Web 3.0 promise to resolve logistic hurdles, such as authentication, security, non-fungible tokens, and decentralized persistence (e.g., blockchain) protocols [47]. A recent market analysis report[3] predicts the market size for Metaverse-based applications to reach nearly one trillion USD by 2030, up from approximately $65B in 2022 (currently mostly attributed to revenues of today's gaming market, such as Epic Games, Roblox, and Minecraft).

Gaming and Entertainment: While the most obvious near-future Metaverse value generators are perhaps VR gaming applications, other directions include revenue from 3D advertizing [43], online shopping [107], and social applications, including teleconferencing, recently brought to focus by the COVID-19 pandemic. The entertainment industry has also made plans to capitalize on virtual reality offerings. For example, Warner Music Group, in cooperation with Sandbox, have announced the creation of virtual theme park that features, among other things, virtual concerts and performances by top artists. Universal Studios in Japan integrated some of their rides with virtual experiences.

Training, Twinning, and Augmentation: Metaverse-based training and education have been cited as other examples of important applications [36], [92] with the potential to help learners improve attention span, build confidence, and generate a sense of community. This application space extends to life sciences and physical sciences, where virtual interactions may allow learners to explore and manipulate objects in novel ways. It is an instance of human enhancement that augments perception and manipulation capabilities with mediation from digital twins and novel physical manipulators [4].

Services: On the government side, investments are being made into using the Metaverse as a tool to improve public services. South Korea has recently announced a three year effort to create a virtual replica of Seoul [2]. Among other objectives, it will allow citizens' avatars to use tax services, libraries, and other public resources. Virtual reality has also been a proposed solution for preservation of cultural heritage [7], [41]. For example, the rising ocean levels have prompted the island nation of Tuvalu to embark on creating a Metaverse replica of their nation to preserve its heritage and art before they are fully submerged [41]. In a similar vein, a collaborative discussion has started on preserving the Australian Aboriginal (First Nation) culture by replication in the Metaverse [7], as well as preserving Japanese cultural elements such as Japanese tea ceremonies.

The Caveat: It remains to note one important difference between the explosive growth of the IoT ecosystem a decade ago and the prospective rise of the Metaverse today. IoT visions were driven by a proliferation of a new type of

---

[2] https://www.pwc.com/us/en/tech-effect/emerging-tech/metaverse-survey.html

[3] https://www.grandviewresearch.com/industry-analysis/metaverse-market-report

created content, namely, the creation of sensor data thanks to the abundance of smart sensors and low-cost embedded computing and network hardware. This new type of created content was then monetized via a range of IoT applications. This is similar to the way the growth of YouTube leveraged the ubiquitous creation of video content (thanks to camera phones) and the growth of Instagram leveraged the ubiquitous creation of images. In contrast, the Metaverse is driven by visions of immersive content *consumption* at a time when the creation of immersive content is not yet commonplace. While advances in such devices as 360 cameras [100] might soon change the status quo, successful expansion of the Metaverse remains contingent on the democratization of content creation. Moreover, due to the inherently compute-heavy nature of content services in the Metaverse, leveraging its full potential may also be contingent on democratizing access to machine learning at scale and democratizing the creation of digital twins for physical artifacts. The research roadmap presented in this paper discusses such infrastructure prerequisites.

III. THE FRONT END: ACCELERATION FOR IMMERSION

The *front end* of the Metaverse resides in the edge computing platforms that are either on one's person (e.g., head-mounted displays) or in one's physical vicinity (e.g., ambient displays), and realize a *Capture-to-Delivery* (C2D) pipeline that is at the core of cross- and mixed-reality interaction and immersion. The end-to-end C2D pipeline is a multi-stage one with functions that include *capture* (acquire signals carrying information about the user and their surrounding), *perception* (estimate from sensor data the state of the user and various entities in the environment), *cognition* (derive from the perceived state an understanding of the current situation and project its future trajectory), *memory* (isolate and store significant information relevant for future iterations of the C2D pipeline), *communication* (collaborate with other edge computing platforms and with premise, edge, and cloud based networked computing resources to optimize various dimensions of user experience, and, *presentation* (deliver to the users various multisensory experiences that fuse real and virtual information). The front end needs to provide these capabilities in diverse operational settings that range from well-instrumented indoor spaces (e.g., office, home, factory, store, hospital, etc.) to out in the wild (e.g., on the road, battlefield, forest, etc.).

Scene Capture & Perception: Essential to providing a user with a good cross-reality immersive experience is the ability of the system to know what is the user's state (not just physical but also physiological, cognitive, and other) and what is going on around the user, both in the real world and in the virtual world. This task of perception involves two steps: acquire using sensors physical signals relevant to the state information of interest and process those signals to predict the state information. To eventually immerse the user in a multisensory experience that is at least as rich as the real world we live in, Metaverse front end capture multimodal signals from the user and their environment, some corresponding to common human senses (e.g., vision and sound) and some beyond-human ones (e.g., RF and inertial). There are two key goals: capture high-fidelity information about the visual and acoustic aspects of the scene so that a digital twin can be constructed at a different time or place, and track detailed position and orientation of the user and their body parts relative to the scene so that sensory experiences can be suitably transformed as the user moves and interacts with the scene. Currently used sensing modalities include various types of cameras (2D, 360, RGBD), Lidars, acoustic arrays, inertial sensors, and wireless (WiFi, mmWave, UWB). While there have been considerable advances in both Inside-out and Outside-in tracking [49] – in the former, a body-mounted device (such as a head-mounted one) performs the necessary sensing while in the latter ambient devices such as in the environment take on the burden – the current state of the art leaves much to be desired. Outside-in approaches limit the user to a specific region of the physical world which is a priori instrumented, while Inside-out approaches present physical burden for the user due to the heavy weight of current devices and also fail to provide complete position and orientation of the entire body (e.g., it is hard to infer the state of the lower body from sensors in a head-mounted display).

Beyond position tracking, the Metaverse also requires sensing finer-grained and richer state about the user and their environment. For example, sensing what kind of floor is the user walking on [102] or sensing what gesture is the user making with their fingers or what is the force with which they are touching the object [74] or what is the expression on their face. To accomplish these sensing tasks, the captured physical signal must be suitably processed to derive the desired inferences. For example, Lidar point clouds require complex processing to isolate from the background [98], RGB and depth images need to be processed to create meshes for display, and 360 video cameras produce equirectangular format which then needs to be transformed into other representations such as navigation graphs [71]. While in the past, the sensor information was processed using first-principles algorithms based on physics and signal processing, in recent years deep neural networks with a large number of parameters and trained using data have shown much superior performance because of their ability to model complex and unknown physics [80], [81] as well as an ability to extract state information from unstructured hight dimensional sensor signals, but have also resulted in the challenge of implementing deep neural networks on low SWaP edge platforms without sacrificing performance [82].

Cognition & Memory: While scene capture and perception provide an awareness of the current state of the user and various entities in their surroundings, providing a meaningful multimodal immersive experience to the user requires processing the multidimensional state information to understand the significant spatiotemporal events and activities in the scene and projecting them into the future so as to both guide autonomous actions by the system (e.g., interacting on the user's behalf with other entities) as well as to provide the user with information relevant for their decisions and actions. Moreover, such understanding and projection can also guide the process of making memories, i.e., storing information that may be useful as context at a future moment in time. The primary challenge arises from the the the mismatch between the resources available on the edge computing platforms employed in the Metaverse systems and the large deep neural network models that are needed for good performance, such as those based on transformer architectures being used for image, video, speech, and natural language. Another challenges arise from the dynamic nature of the front end, for example with continual changes in sensor perspectives due to mobility, impairment of select sensing modalities due to environmental factors, and sensor occlusions due to body movement and ambient clutter. Lastly, in light of limited memory storage, the system needs to be intelligent about identifying the key change points that should trigger memory formation actions. A key to meeting these challenges will be to leverage the current or predicted user attention [69] to focus the limited processing resources on the most high value sensors streams or regions within a stream (e.g., in the case of 360 cameras), and to devise multimodal architectures that can adapt to the availability of sensor modalities by employing shared cross-modal latent representations [54], [99], and the availability of computing resources by employing approximate computations [104], neural network model compression [48], [56], [106], and input resolution adjustment [34]. As data memory consumes significant power, reducing memory footprint for both the models and the stored contextual data is important; for the former besides model compression methods such as weight virtualization [52] can help, while for the latter compression methods exploit semantics and latent representations are promising [105]. Lastly, important for addressing Cognition & Memory related challenges would be new hardware capabilities such as accelerators optimized not just for convolutional layers but newer transformer architectures, and tighter memory-processing integration such as compute-in-memory.

Edge Communication: Communication is core to the Metaverse, with communication taking place both in the virtual world (a user's digital twin exchanging information with digital twins of other users and with other entities) and the real world (user-worn and ambient devices exchanging information for sensing and presentation, and for off-loading computing to on-premise, edge, and cloud servers). The main challenge arises from the simultaneous requirements of high data rates due to rich sensors (such as cameras and Lidars) and high-resolution displays, as well as ultra low latency and tight time synchronization due to human perception requirements and to avoid adverse health effects such as nausea. This confluence of high-throughput (Gpbs) and low-latency (ms) is currently beyond the reach of low-power wireless technologies that can be incorporated into front end devices for use in the wild, and come with severe range limitations, essentially tethering the user to within a local area. The development of suitable wireless technologies is a barrier challenge that needs to be solved for Metaverse viability, and would involve addressing challenges across the entire wireless stack from antennas up. As high data rate would require the use of high carrier frequencies whose propagation is easily impaired by environment clutter and user's own body, it would be essential to take a joint sensing-communication approach whereby sensing of the world and user state informs the communication decisions. Also important would be compression that exploits semantics to significantly reduce data rate requirements. Semantic compression and decoding technology can represent complex concepts in multimedia streams concisely in low-dimensional latent representation spaces and decode efficiently from such spaces to generate realistic real-time environments for the user [15], [29], [103]. Such technologies can further be enhanced with asymmetric encoder/decoder design [14], [105] that places a lower computational burden on the headset, while compensating by using more computationally-heavy algorithms on the edge server. For example, deep compressive offloading [105] uses a lightweight encoder to compress outgoing data on the end device and a resource-intensive decoder to restore the data at the edge server. It further uses knowledge of downstream tasks to determine which data features could be compressed away and which features are more important for the specific task at hand, thereby achieving a higher overall compression ratio.

Display & Presentation: The final stage of the C2D pipeline is presentation of multimedia content to the user, and then closing the loop by sensing user's actions to continually adapt the presentation. Recent years have seen tremendous advances in audio-visual content presentation technologies such as AR/VR headsets that provide immersive visual experiences, lightfield-based parallel reality ambient displays that provide simultaneous viewer-specific content to a large number of users, and tiny earables with spatial audio that adapts to head position and room acoustics. For headsets and earables, communication bandwidth and battery life remain a painpoint. For example, Today's Meta Quest Pro headset[4] weighs over 700g. It comes equipped with 256GB storage, a 12GB RAM, and a SnapDragon processor that consumes several Watts of power, making cooling a potential concern when in protracted use. To ameliorate these challenges, solutions that decrease headset power consumption are needed. Such power-efficiency solutions might, for example, leverage architectural innovations that improve computational capacity per Watt [8], [101], or exploit protocols that balance computation between the headset and back-end support (e.g., edge or cloud servers) to minimize the computational needs of the former [62]. Besides audio-visual content, other innovative interfaces might emerge, such as haptic [18], olfactory [68] and cognitive [67] interfaces. For example, recent advances demonstrated the feasibility of communicating signals over a network from one user's brain to another [37]. Direct brain stimulation has also been shown to modulate cognitive performance [64] or induce certain affective states [84]. Combined with audio-visual stimuli, such interfaces could serve a wide variety of functions from stimulating productivity to enhancing immer-

---

[4] https://www.meta.com/quest/quest-pro/

sive entertainment. Olfactory and visual media must come together to provide a unique immersive experience. However, one of the challenges is synchronization of delivered media due to their different propagation delays to user's senses [68]. Haptics and immersive video have their own challenges such as delays introduced by force rendering, and overall reduction of disparity between haptic and visual data [10].

IV. NETWORK PROTOCOLS

Moving away from the edge network, a key challenge will be to reduce or mask end-to-end latency. In applications where users interact with each other or with a remote environment across long distances, latency may interfere with illusions of virtual presence. Even at the speed of light, it takes over 100ms to circle Earth once, which can be quite perceptible to an end user. Unfortunately, network latencies in local-, metro-, and wide-area settings are typically much larger than speed of light, due to suboptimality of Internet path selection (e.g., AS path inflation, hot potato routing), congestion, reliance on third-party services such as DNS, and protocol overheads. Even worse, the nature of these underlying properties is highly variable, leading to statistical effects such as jitter and non-stationarity, complicating the ability to counter them.

Combating Latency: Predictive algorithms may extrapolate from the current state to compute and render predicted future states ahead of data reception or to prefetch data ahead of user requests to offer the illusion of a more timely interaction with the remote environment. "Dead reckoning" has been widely used within video gaming environments to overcome latency, and such techniques could perhaps be extended to predict the more complex interactions found in metaverse environments. Some metaverse services may be replicated using content distribution networks or cloud services; the advent of edge computing may enable placement of services at the network edge close to users. The mechanisms underlying these services, such as anycast and network mapping, have been widely explored in the context of traditional Internet services (such as web hosting), and may need to be extended to support the novel low-latency and highly collaborative nature of multiverse services. Protocols will be needed to efficiently convey change. Receiver-side models could be used to predict future frames based on an activity model shared with the sender. The sender can then prioritize the transmission of only those bits that cannot be accurately predicted from the models.

Wide-area Scalability: The advent of cloud computing led to fundamental changes in the Internet's design, with the advent of massive private intra-cloud networks, edge computing CDNs, and a general "flattening" of the Internet's hierarchical structures. The metaverse will place new demands on the Internet to achieve real-time communications at scale. While low-latency communication mechanisms have pervaded cloud computing deployments with Infiniband and RDMA, we lack comparable technologies for the wide area. A challenge will be developing incentive-compatible mechanisms which can provide advanced quality of experience across multiprovider networks. Latency may be further reduced by exploring novel congestion control mechanisms optimized for real-time delivery, resizing router buffers and exploring novel router architectures to provide cost-effective guarantees on packet delivery at scale, and content distribution network architectures that decentralize metaverse systems, offloading core functions to edge computing. In recent years, intent-based networking has gained significant traction within industry, allowing operators to manage their networks through direct expression of high-level policies; however, much work in this space has focused on reachability policies, a challenge will be to develop frameworks to support customizable intents on user quality of experiences and metrics related to the metaverse experience. Programmable networks, including frameworks such as P4, may further creation of customizable packet processing logic that can adapt to changing workloads and operational constraints. Finally, operators may also benefit from more formal approaches to peering for low latency, leveraging machine learning and predictive technologies, as well as network surveys and structural observations to determine the best locations and approaches to peer.

A related issue is architecture design principles for the Metaverse that allow interoperability and give rise to a common ecosystem, as opposed to a set of individual applications that share in common the use of immersive content.

V. THE BACK-END: CONTENT PROCESSING SERVICES

A growing fraction of today's global computational demands stems fundamentally from the need to *process* networked world data [23], [88]. Since the Metaverse application ecosystem will give rise to new more immersive data formats, novel challenges arise in back-end data processing.

Access, Indexing, and Search: Among the largest drivers for increased computational needs in recent history has been the *democratization of at-scale content sharing*, driven by Web and social network platforms. Today, the most visited websites are those that offer global content access, indexing, and search, such as Google, YouTube, Facebook, and Twitter. The Metaverse application ecosystem offers the next step in immersion and thus advances the modalities of shared content. It will popularize a next generation of content that evolves from text, images, sound, and video to 3D environments with haptic audio-visual modalities [19] and possibly cognitive influence interfaces. Platforms for content sharing will need to evolve to support operations on the new content modalities. At present, scalable search of complex content modalities, such as video, is still relatively new [40]. While initial systems exist for scalable video analytics using machine-learning approaches [39], [79], the interfaces for specifying visual concepts for search purposes and algorithms for ranking potential matches remain an active research topic. Searching immersive 3D multimedia content for specific types of activity, experiences, or artifacts is an important future challenge.

Summarization: A related issue is to re-imagine the very role of search services in the Metaverse age. The interface of today's search engines has not substantially changed in 30 years; a list of matching entries, such as videos, Webpages, or Tweets, is returned the way it was when the first browser appeared [5]. In the meantime, the number of Internet users has grown by orders of magnitude. Metaverse content consumers should be empowered with mechanisms that facilitate understanding the content beyond the top matching items. For example, imagine what it might be like to retrieve past tourist experiences at a particular landmark. Technologies are needed to generate *summaries* of *representative* experiences, possibly from thousands of distinct specific experiences stored. Means for summarizing both individual content items and content collections will become necessary. The emergence of the Metaverse may precipitate the development of next-generation algorithms for organizing and summarizing unstructured multimedia data in an unsupervised manner. Hierarchical categorization and summarization will help consumers understand the totality of content and efficiently navigate to shared clips, memories, or experiences, much the way they can navigate review categories on Yelp or Travelocity today. An example of an unstructured content summarization service today (for the most parsimonious content modality – text), based on pre-trained language models [76] (as opposed to knowledge-base grounding), is ChatGPT [93]. It is likely to be the first of many more AI-based services that will increasingly target multimedia and more resource-intensive content to generate more engaging and immersive summaries.

Generative Foundation Models: A related direction might lie in the creation of immersive experiences based on pre-trained foundation models (in AI) [9]. Foundation models are large deep neural networks that are pre-trained using a self-supervised approach on extensive amounts of data and are customizable to a variety of downstream content generation tasks. For example, ChatGPT [93] is a foundation model that can generate text in response to a prompt. Foundation models are also being explored in other fields as a way to summarize common knowledge, somewhat analogously to large-scale simulations. An example is foundation models for networking [50]. The 3D immersive environment of the envisioned Metaverse suggests that future foundation models might assimilate text, audio, and visual content, offering customized multimedia experiences, prompted by user input, or synthesize multimedia summary highlights based on user queries.

VI. AI ACCELERATION AND APPLICATION SERVICES

The computational challenges posed by Metaverse applications on the computing infrastructure will drive advances in hardware and compiler technology. Much recent work has already focused on accelerating different AI-inspired computational kernels. Future advances are needed to support the computational needs of creating, summarizing, searching, or otherwise processing Metaverse content.

Compiler Infrastructure and Optimizations: First, machine learning models and methods used in metaverse applications tend to have sparse components and hence we can leverage advances in domain specific hardware and compiler techniques in sparse tensor algebra to accelerate such workloads. For example, a collaborative filtering content recommendation system might compute its recommendations from observing who liked what in the past, as well as their similarity relations to the item being recommended. However, "who liked what" is a very sparse matrix, as each user likely interacts with only a small subset of all available items. Recent advances in domain specific compilation techniques targeting sparse tensor algebra [45], [85] and graph computations [109] can potentially be used to automatically generate fast code for these applications. Different compiler techniques focusing on eliminating redundancies [111] and workspaces [44] have been suggested to further improve upon these compiler infrastructures. Additionally, we can gain even more runtime performance by specializing the optimization strategies specific to the sparsity pattern of the sparse matrices. Some works focus on selecting the best sparse storage format [110]. Others propose adaptive optimizations such as input-sensitive load balancing schemes [97], adaptive sparse tiling [32] and cache-aware graph segmentation schemes [108]. These techniques use both heuristic-based compiler algorithms as well as data-driven auto-tuning techniques.

Also, it is important to carefully select and invent compiler and program optimization techniques for machine learning models that run on hardware that are designed for metaverse applications. For example, heavily pruned [27], quantized machine learning models [30] may dominate inference tasks in metaverse. Therefore, the applications should leverage compiler techniques that focuses on applying those approximate techniques and optimizes for those. Works that focus on training neural networks with low memory requirements [57], works that focus on applying approximate optimization techniques [87], works that focus on edge devices [12] and infrastructures for federated learning [35] will be of relevance in this front. Moreover, in processing streaming content, such as immersive multimedia, there are ample opportunities to optimize computations. There have been many efforts at leveraging compiler optimizations to speedup these multimedia computations including vectorization [16], specialized code generation through frameworks such as FFTW [28] and SPIRAL [26], early work on optimizing streaming computations such as Streamit [91] and work on optimizing compression and decompression workloads [77]. These approaches should be adapted to work with hardware platforms that will drive content delivery and computation in metaverse applications.

Distributed Acceleration: Training large AI models, such as foundations models for multimodal/3D immersive content, will require distributed computation. Metaverse applications will thus push the envelope of distributed learning, motivating and advancing the development of frameworks that significantly accelerate distributed AI [65]. Different from conventional distributed training, we envision future Metaverse learning will ingest a diverse set of data (texts, images, and videos) and run a DAG of heterogeneous models. For example, the DAG may start with models targeting a specific format of data and later get additional training across data types. This brings new challenges for acceleration because we may incur different delay at different training stages for different data types. We can address this challenge by leveraging new algorithms and accelerators to data preprocessing, compression, and feature extraction (e.g., by leveraging hardware-based video decoder and in-network computing). Moreover, we need to dynamically allocate various computing resources (accelerators, GPUs) to different training tasks based on the workloads in the runtime. We need to allocate more resources to parallelize those bottleneck tasks more.

Another challenge is latency (especially tail latency). We need to continuously collect data and make fine-tuning decisions for the training models. Any delay in Metaverse significantly affect user experiences or decisions. Therefore, we need to introduce streaming algorithms that process images and videos even before they fully arrive at the servers. Our runtime system needs to incorporate deadline-aware scheduling, which prioritize those operations and data on the critical path. We also need to dynamically select the right models based on the latency requirements. Moreover, we need to introduce approximate solutions that explore the tradeoffs between latency and accuracy. We also need to mitigate stragglers while retaining accuracy. Multi-tenancy is another challenge for accelerating distributed training. Different users may provide different amounts and types of data and may need different training models. We need to identify new ways to share GPU resources across tenants with minimal context switching overhead. Moreover, we need to find ways to enforce end-to-end sharing policies into individual components in the data processing pipeline (e.g., NICs, CPUs, and GPUs).

VII. PHYSICAL AUTOMATION

With advances in autonomy and robotics, pervasive automation will increasingly delegate physical functions to drones, robotic assistants, and autonomous machines. At the same time, a growing number of IoT devices will automate increasingly many applications from smart home energy control to city-wide traffic management. These applications will offer control APIs in the Metaverse, thus allowing Metaverse users to perform tasks in the physical environment.

Users in the metaverse will be empowered to interact with a set of agents in the physical world, an evolution of today's IoT applications, including various forms of control agents (e.g., HVAC), mobility devices (e.g., autonomous cars, and drones), and delegated physical functions (e.g., household and industrial robotics). Support for automation will become embedded in increasingly more everyday items, giving them the appearance of intelligence on a budget, within limits on their embedded memory and computational capacity. Significant optimizations will be needed to offer these capabilities while respecting the constraints of their resource-limited computational environment.

Safety and Security: Task execution that impacts the physical environment brings about a myriad of concerns, commonly associated with computational systems that interact with their physical environment, or *cyber-physical systems* [86]. For example, safety and security assurances will become more important [96]. Solutions will be needed that vet the safety of actions in real-time before they are committed to the environment. Conflicting actions and actions with conflicting policy goals should be avoided [60], [61].

Digital Twins: To ameliorate safety problems and offer an opportunity for safe "what-if" analysis, a proliferation of digital twins is expected for networked "things" [11]. Digital twins are thus envisioned to become commonplace Metaverse citizens [25], [58]. Several companies including BMW, Coca-Cola, and NASCAR have recently announced partnerships around building replicas of their products in the Metaverse. For example, BMW's digital twin allows it to explore different configurations of factory automation that optimize the manufacturing workflow. Challenges in maintaining digital twins include reliable low-latency communication [94], synchronization [31], [31], networking [3], security/privacy considerations [25], low-resource operation, and edge-cloud coordination [55], with applications in both industrial and social contexts. Digital twinning through the metaverse will transform the way operators and developers create and manage cyber physical infrastructures. When designing a new infrastructure, the developer will leverage the metaverse to rapidly mock up and prototype their designs without the constraints of physical hardware. To test their designs, developers will create physical "unit tests" and other testing frameworks by manipulating the metaverse environment, running the device under test through diverse environmental scenarios, and exposing it to unexpected contingencies. ensuring the device will perform as expected. The self-contained properties of the digital twin will provide a natural location to apply verification and synthesis technologies, providing further assurance that system software and protocols are properly implemented with bounded behaviors. The operator can then "push" their implementations from the cloud out into the real physical environment, ensuring that their pre-tested implementation and properly-tuned configuration is what is running in the real environment. After deployment, the operator can make use of advanced ML and anomaly detection algorithms to study behavior of the deployed device, as well as to dynamically tune operational settings.

Modeling the Environment: An additional challenge will be in twinning the physical environment surrounding the deployed system. Many simulators exist today for both software systems (computer networks, microcontroller architectures, etc.) as well as physical media (mechanical stress, hydrology and fluid flow, etc), but several things are lacking. First, we lack comprehensive environmental simulators that can "stitch together" the joint behaviors of objects and software across different objects and materials within an environment. Second, simulators often lack mechanisms to represent environments with the finer levels of granularity that may be important for effective twinning. For example, network simulators such as ns-3 and CORE provide highly effective environments to simulate wireless protocol behaviors, yet lack the ability to simulate specific environmental considerations that may greatly affect wireless operations (e.g., suppose there are three trees between the sender and receiver – how is multipath affected?) To address this, it may be possible to leverage advances in high fidelity wireless simulations (e.g., by using ray tracing). Third, we lack effective ways to quickly prototype and instrument dynamic environments within simulations. The emergence of LIDAR and mmWave to map physical environments may provide effective technologies to assist in rapidly prototyping metaverse environments that mirror existing physical locations. Prior work in simulating systems of mechanical objects (e.g., stress vectors within buildings, sound propagation) may further improve realism.

VIII. CONCLUSIONS

We described challenges in implementing the Metaverse application ecosystem as observed through the lens of IoT research topics. These challenges will enable a world, where the boundaries between cyber, physical, and social realms are blurred, new immersive experiences are fed by billions of data points from multitudes of human, AI-generated, and physical sources, machine intelligence extracts value for a growing list of novel applications, and novel social media platform concepts seamlessly integrate humans, avatars, physical devices, and digital twins, redefining what IoT looks like. This paper invites IoT, machine learning, and Metaverse researchers to a collaborative roadmap to generate tomorrow's systems and applications.

ACKNOWLEDGEMENTS

This work is supported in part by ARL Cooperative Agreement W911NF-17-2-0196, NSF grants 20-38817, 18-35834, 21-06592, and 22-11301, IBM-IIDAI institute, and Boeing Inc. It is also supported in part by ACE, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

REFERENCES

[1] T. Abdelzaher, Y. Hao, K. Jayarajah, A. Misra, P. Skarin, S. Yao, D. Weerakoon, and K.-E. Årzén. Five challenges in cloud-enabled intelligence and control. *ACM Transactions on Internet Technology (TOIT)*, 20(1):1–19, 2020.

[2] Z. Allam, A. Sharifi, S. E. Bibri, D. S. Jones, and J. Krogstie. The metaverse as a virtual form of smart cities: Opportunities and challenges for environmental, economic, and social sustainability in urban futures. *Smart Cities*, 5(3):771–801, 2022.

[3] P. Almasan, M. Ferriol-Galmés, J. Paillisse, J. Suárez-Varela, D. Perino, D. López, A. A. P. Perales, P. Harvey, L. Ciavaglia, L. Wong, et al. Network digital twin: Context, enabling technologies, and opportunities. *IEEE Communications Magazine*, 60(11):22–27, 2022.

[4] M. Aloqaily, O. Bouachir, F. Karray, I. Al Ridhawi, and A. El Saddik. Integrating digital twin and advanced intelligent technologies to realize the metaverse. *IEEE Consumer Electronics Magazine*, 2022.

[5] M. Andreessen and E. Bina. Ncsa mosaic: A global hypermedia system. *Internet Research*, 4(1):7–17, 1994.

[6] S. Bagchi, T. F. Abdelzaher, R. Govindan, P. Shenoy, A. Atrey, P. Ghosh, and R. Xu. New frontiers in iot: Networking, systems, reliability, and security challenges. *IEEE Internet of Things Journal*, 7(12):11330–11346, 2020.

[7] B. Barba, D. V. Lee-Ah Mat, A. Gomez, and J. Pirovich. Discussion paper: First nations' culture in the metaverse. 2022.

[8] A. Basu, J. Acharya, T. Karnik, H. Liu, H. Li, J.-S. Seo, and C. Song. Low-power, adaptive neuromorphic systems: Recent progress and future directions. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 8(1):6–27, 2018.

[9] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[10] A. Borresen, C. Wolfe, C.-K. Lin, Y. Tian, S. Raghuraman, K. Nahrstedt, B. Prabhakaran, and T. Annaswamy. Usability of an immersive augmented reality based telerehabilitation system with haptics (artesh) for synchronous remote musculoskeletal examination. *International Journal of Telerehabilitation*, 11(1), June 2019.

[11] D. M. Botín-Sanabria, A.-S. Mihaita, R. E. Peimbert-García, M. A. Ramírez-Moreno, R. A. Ramírez-Mendoza, and J. d. J. Lozoya-Santos. Digital twin technology challenges and applications: A comprehensive review. *Remote Sensing*, 14(6):1335, 2022.

[12] H. Cai, J. Lin, Y. Lin, Z. Liu, H. Tang, H. Wang, L. Zhu, and S. Han. Enable deep learning on mobile devices: Methods, systems, and applications. *ACM Trans. Des. Autom. Electron. Syst.*, 27(3), mar 2022.

[13] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos. Iot-based big data storage systems in cloud computing: perspectives and challenges. *IEEE Internet of Things Journal*, 4(1):75–87, 2016.

[14] B. Chen, Z. Yan, and K. Nahrstedt. Context-aware compression optimization for visual analytics offloading. In *ACM Multimedia Systems (MMSys)*. ACM, 2022.

[15] S. Chen, P. Xia, and K. Nahrstedt. Activity-aware adaptive compression: A morphing-based frame synthesis application in 3dti. In *ACM International Conference on Multimedia*. ACM, October 2013.

[16] Y. Chen, C. Mendis, M. Carbin, and S. P. Amarasinghe. Vegen: a vectorizer generator for SIMD and beyond. In T. Sherwood, E. D. Berger, and C. Kozyrakis, editors, *ASPLOS '21: 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Virtual Event, USA, April 19-23, 2021*, pages 902–914. ACM, 2021.

[17] R. Cheng, N. Wu, S. Chen, and B. Han. Will metaverse be nextg internet? vision, hype, and reality. *IEEE Network*, 36(5):197–204, 2022.

[18] W. Dangxiao, G. Yuan, L. Shiyi, Y. Zhang, X. Weiliang, and X. Jing. Haptic display for virtual reality: progress and challenges. *Virtual Reality & Intelligent Hardware*, 1(2):136–162, 2019.

[19] F. Danieau, A. Lécuyer, P. Guillotel, J. Fleureau, N. Mollet, and M. Christie. Enhancing audiovisual experience with haptic feedback: a survey on hav. *IEEE transactions on haptics*, 6(2):193–205, 2012.

[20] J. Deichmann, M. Roggendorf, and D. Wee. Preparing it systems and organizations for the internet of things. 2015.

[21] J. M. D. Delgado, L. Oyedele, P. Demian, and T. Beach. A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, 45:101122, 2020.

[22] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet of Things Journal*, 7(8):7457–7469, 2020.

[23] P. Dhar. The carbon impact of artificial intelligence. *Nat. Mach. Intell.*, 2(8):423–425, 2020.

[24] D. Evans. The internet of things: How the next evolution of the internet is changing everything. *CISCO white paper*, 1(2011):1–11, 2011.

[25] S. B. Far and A. I. Rad. Applying digital twins in metaverse: User interface, security and privacy challenges. *Journal of Metaverse*, 2(1):8–16, 2022.

[26] F. Franchetti, T.-M. Low, T. Popovici, R. Veras, D. G. Spampinato, J. Johnson, M. Püschel, J. C. Hoe, and J. M. F. Moura. SPIRAL: Extreme performance portability. *Proceedings of the IEEE, special issue on "From High Level Specification to High Performance Code"*, 106(11), 2018.

[27] J. Frankle and M. Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2019.

[28] M. Frigo. A fast fourier transform compiler. In *Proceedings of the ACM SIGPLAN 1999 Conference on Programming Language Design and Implementation*, PLDI '99, page 169–180, New York, NY, USA, 1999. Association for Computing Machinery.

[29] A. Golinski, R. Pourreza, Y. Yang, G. Sautiere, and T. S. Cohen. Feedback recurrent autoencoder for video compression. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[30] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding. In Y. Bengio and Y. LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[31] Y. Han, D. Niyato, C. Leung, C. Miao, and D. I. Kim. A dynamic resource allocation framework for synchronizing metaverse with iot service and data. In *ICC 2022-IEEE International Conference on Communications*, pages 1196–1201. IEEE, 2022.

[32] C. Hong, A. Sukumaran-Rajam, I. Nisa, K. Singh, and P. Sadayappan. Adaptive sparse tiling for sparse matrix multiplication. In J. K. Hollingsworth and I. Keidar, editors, *Proceedings of the 24th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2019, Washington, DC, USA, February 16-20, 2019*, pages 300–314. ACM, 2019.

[33] M. B. Hoy. Alexa, siri, cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, 37(1):81–88, 2018.

[34] Y. Hu, S. Liu, T. Abdelzaher, M. Wigness, and P. David. On exploring image resizing for optimizing criticality-based machine perception. In *2021 IEEE 27th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, pages 169–178. IEEE, 2021.

[35] D. Huba, J. Nguyen, K. Malik, R. Zhu, M. Rabbat, A. Yousefpour, C.-J. Wu, H. Zhan, P. Ustinov, H. Srinivas, K. Wang, A. Shoumikhin, J. Min, and M. Malek. Papaya: Practical, private, and scalable federated learning. In D. Marculescu, Y. Chi, and C. Wu, editors, *Proceedings of Machine Learning and Systems*, volume 4, pages 814–832, 2022.

[36] G.-J. Hwang and S.-Y. Chien. Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, 3:100082, 2022.

[37] L. Jiang, A. Stocco, D. M. Losey, J. A. Abernethy, C. S. Prat, and R. P. Rao. Brainnet: a multi-person brain-to-brain interface for direct collaboration between brains. *Scientific reports*, 9(1):6115, 2019.

[38] S. Jiang, P. Kang, X. Song, B. P. Lo, and P. B. Shull. Emerging wearable interfaces and algorithms for hand gesture recognition: A survey. *IEEE Reviews in Biomedical Engineering*, 15:85–102, 2021.

[39] D. Kang, P. Bailis, and M. Zaharia. Blazeit: optimizing declarative aggregation and limit queries for neural network-based video analytics. *PVLDB*, 2018.

[40] D. Kang, P. Bailis, and M. Zaharia. Challenges and opportunities in dnn-based video analytics: A demonstration of the blazeit video query engine. In *CIDR*, 2019.

[41] N. Kelly and M. Foth. An entire pacific country will upload itself to the metaverse. it'sa desperate plan–with a hidden message. *The Conversation*, 2022.

[42] M. Khan, B. N. Silva, and K. Han. Internet of things based energy aware smart home control system. *Ieee Access*, 4:7556–7566, 2016.

[43] J. Kim. Advertising in the metaverse: Research agenda. *Journal of Interactive Advertising*, 21(3):141–144, 2021.

[44] F. Kjolstad, P. Ahrens, S. Kamil, and S. Amarasinghe. Tensor algebra compilation with workspaces. *International Symposium on Code Generation and Optimization*, February 2019.

[45] F. Kjolstad, S. Kamil, S. Chou, D. Lugato, and S. Amarasinghe. The tensor algebra compiler. *Proc. ACM Program. Lang.*, 1(OOPSLA):77:1–77:29, Oct. 2017.

[46] S. Kraus, D. K. Kanbach, P. M. Krysta, M. M. Steinhoff, and N. Tomini. Facebook and the creation of the metaverse: radical business model innovation or incremental transformation? *International Journal of Entrepreneurial Behavior & Research*, 2022.

[47] N. Kshetri. Web 3.0 and the metaverse shaping organizations' brand and product strategies. *IT Professional*, 24(02):11–15, 2022.

[48] N. D. Lane and P. Georgiev. Can deep learning revolutionize mobile sensing? In *Proceedings of the 16th international workshop on mobile computing systems and applications*, pages 117–122, 2015.

[49] H. Langley. Inside-out v outside-in: How vr tracking works, and how it's going to change. *URL: https://www. wareable. com/vr/inside-out-vs-outside-in-vr-tracking-343*, pages 1–7, 2017.

[50] F. Le, M. Srivatsa, R. Ganti, and V. Sekar. Rethinking data-driven networking with foundation models: challenges and opportunities. In *Proceedings of the 21st ACM Workshop on Hot Topics in Networks*, pages 188–197, 2022.

[51] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *arXiv preprint arXiv:2110.05352*, 2021.

[52] S. Lee and S. Nirjon. Fast and scalable in-memory deep multitask learning via neural weight virtualization. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 175–190, 2020.

[53] S. Li, S. Hu, R. K. Ganti, M. Srivatsa, and T. F. Abdelzaher. Pyro: A spatial-temporal big-data storage system. In *USENIX Annual Technical Conference*, pages 97–109, 2015.

[54] Y. Li, J.-Y. Zhu, R. Tedrake, and A. Torralba. Connecting touch and vision via cross-modal prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10609–10618, 2019.

[55] W. Y. B. Lim, Z. Xiong, D. Niyato, X. Cao, C. Miao, S. Sun, and Q. Yang. Realizing the metaverse with edge intelligence: A match made in heaven. *IEEE Wireless Communications*, 2022.

[56] J. Lin, W.-M. Chen, Y. Lin, C. Gan, S. Han, et al. Mcunet: Tiny deep learning on iot devices. *Advances in Neural Information Processing Systems*, 33:11711–11722, 2020.

[57] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han. On-device training under 256kb memory. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

[58] Z. Lv, L. Qiao, Y. Li, Y. Yuan, and F.-Y. Wang. Blocknet: Beyond reliable spatial digital twins to parallel metaverse. *Patterns*, 3(5):100468, 2022.

[59] H. Ma and W. Liu. A progressive search paradigm for the internet of things. *IEEE MultiMedia*, 25(1):76–86, 2017.

[60] M. Ma, S. M. Preum, and J. A. Stankovic. Cityguard: A watchdog for safety-aware conflict detection in smart cities. In *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, pages 259–270, 2017.

[61] M. Ma, S. M. Preum, W. Tarneberg, M. Ahmed, M. Ruiters, and J. Stankovic. Detection of runtime conflicts among services in smart cities. In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 1–10. IEEE, 2016.

[62] P. Mach and Z. Becvar. Mobile edge computing: A survey on architecture and computation offloading. *IEEE communications surveys & tutorials*, 19(3):1628–1656, 2017.

[63] I. Mehmood, A. Ullah, K. Muhammad, D.-J. Deng, W. Meng, F. Al-Turjman, M. Sajjad, and V. H. C. de Albuquerque. Efficient image recognition and retrieval on iot-assisted energy-constrained platforms from big data repositories. *IEEE Internet of Things Journal*, 6(6):9246–9255, 2019.

[64] M. Meinzer, D. Antonenko, R. Lindenberg, S. Hetzer, L. Ulm, K. Avirame, T. Flaisch, and A. Flöel. Electrical brain stimulation improves cognitive performance by modulating functional connectivity and task-specific activation. *Journal of Neuroscience*, 32(5):1859–1866, 2012.

[65] P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, M. Elibol, Z. Yang, W. Paul, M. I. Jordan, et al. Ray: A distributed framework for emerging {AI} applications. In *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*, pages 561–577, 2018.

[66] K. Muhammad, T. Hussain, M. Tanveer, G. Sannino, and V. H. C. de Albuquerque. Cost-effective video summarization using deep cnn with hierarchical weighted fusion for iot surveillance networks. *IEEE Internet of Things Journal*, 7(5):4455–4463, 2019.

[67] C. Mühl, B. Allison, A. Nijholt, and G. Chanel. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces*, 1(2):66–84, 2014.

[68] N. Murray, Y. Qiao, B. Lee, and M. G.-M. Karunakar, A.K. Subjective evaluation of olfactory and visual media synchronization. In *Proceedings of ACM Multimedia Systems Conference (MMSys)*. ACM, February 2013.

[69] A. Nguyen, Z. Yan, and K. Nahrstedt. Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1190–1198, 2018.

[70] H. Ning, H. Wang, Y. Lin, W. Wang, S. Dhelim, F. Farha, J. Ding, and M. Daneshmand. A survey on metaverse: the state-of-the-art, technologies, applications, and challenges. *arXiv preprint arXiv:2111.09673*, 2021.

[71] J. Park and K. Nahrstedt. Navigation graph for tiled media streaming. In *ACM International Conference on Multimedia*. ACM, October 2019.

[72] S.-M. Park and Y.-G. Kim. A metaverse: taxonomy, components, applications, and open challenges. *IEEE access*, 10:4209–4251, 2022.

[73] K. Patel, H. Han, and A. K. Jain. Secure face unlock: Spoof detection on smartphones. *IEEE transactions on information forensics and security*, 11(10):2268–2283, 2016.

[74] S. Pei, P. Chari, X. Wang, X. Yang, A. Kadambi, and Y. Zhang. Forcesight: Non-contact force sensing with laser speckle imaging. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–11, 2022.

[75] R. Piyare and S. R. Lee. Smart home-control and monitoring system using smart phone. *ICCA, ASTL*, 24:83–86, 2013.

[76] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897, 2020.

[77] J. Ray, A. Brahmakshatriya, R. Wang, S. Kamil, A. Reuther, V. Sze, and S. P. Amarasinghe. Domain-specific language abstractions for compression. In A. Bilgin, M. W. Marcellin, J. Serra-Sagristà, and J. A. Storer, editors, *31st Data Compression Conference, DCC 2021, Snowbird, UT, USA, March 23-26, 2021*, page 364. IEEE, 2021.

[78] P. P. Ray. A survey of iot cloud platforms. *Future Computing and Informatics Journal*, 1(1-2):35–46, 2016.

[79] F. Romero, J. Hauswald, A. Partap, D. Kang, M. Zaharia, and C. Kozyrakis. Optimizing video analytics with declarative model relationships. *Proceedings of the VLDB Endowment*, 16(3):447–460, 2022.

[80] S. S. Saha, S. S. Sandha, L. A. Garcia, and M. Srivastava. Tinyodom: Hardware-aware efficient neural inertial navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–32, 2022.

[81] S. S. Saha, S. S. Sandha, S. Pei, V. Jain, Z. Wang, Y. Li, A. Sarker, and M. Srivastava. Auritus: An open-source optimization toolkit for training and development of human movement models and filters using earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–34, 2022.

[82] S. S. Saha, S. S. Sandha, and M. Srivastava. Machine learning for microcontroller-class hardware-a review. *IEEE Sensors Journal*, 2022.

[83] J. H. Saltzer, D. P. Reed, and D. D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems (TOCS)*, 2(4):277–288, 1984.

[84] R. Sellaro, M. A. Nitsche, and L. S. Colzato. The stimulated social brain: effects of transcranial direct current stimulation on social cognition. *Annals of the New York Academy of Sciences*, 1369(1):218–239, 2016.

[85] R. Senanayake, C. Hong, Z. Wang, A. Wilson, S. Chou, S. Kamil, S. Amarasinghe, and F. Kjolstad. A sparse iteration space transformation framework for sparse tensor algebra. *Proc. ACM Program. Lang.*, 4(OOPSLA), Nov. 2020.

[86] L. Sha, S. Gopalakrishnan, X. Liu, and Q. Wang. Cyber-physical systems: A new frontier. In *2008 IEEE international conference on sensor networks, ubiquitous, and trustworthy computing (sutc 2008)*, pages 1–9. IEEE, 2008.

[87] H. Sharif, P. Srivastava, M. Huzaifa, M. Kotsifakou, K. Joshi, Y. Sarita, N. Zhao, V. S. Adve, S. Misailovic, and S. Adve. Approxhpvm: A portable compiler ir for accuracy-aware optimizations. *Proc. ACM Program. Lang.*, 3(OOPSLA), oct 2019.

[88] M. A. B. Siddik, A. Shehabi, and L. Marston. The environmental footprint of data centers in the united states. *Environmental Research Letters*, 16(6):064017, 2021.

[89] J. A. Stankovic. Research directions for the internet of things. *IEEE internet of things journal*, 1(1):3–9, 2014.

[90] F. Tang, X. Chen, M. Zhao, and N. Kato. The roadmap of communication and networking in 6g for the metaverse. *IEEE Wireless Communications*, 2022.

[91] W. Thies, M. Karczmarek, and S. P. Amarasinghe. Streamit: A language for streaming applications. In R. N. Horspool, editor, *Compiler Construction, 11th International Conference, CC 2002, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2002, Grenoble, France, April 8-12, 2002, Proceedings*, volume 2304 of *Lecture Notes in Computer Science*, pages 179–196. Springer, 2002.

[92] A. Tlili, R. Huang, B. Shehata, D. Liu, J. Zhao, A. H. S. Metwally, H. Wang, M. Denden, A. Bozkurt, L.-H. Lee, et al. Is metaverse in education a blessing or a curse: a combined content and bibliometric analysis. *Smart Learning Environments*, 9(1):1–31, 2022.

[93] E. A. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting. Chatgpt: five priorities for research. *Nature*, 614(7947):224–226, 2023.

[94] D. Van Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong. Edge intelligence-based ultra-reliable and low-latency communications for digital twin-enabled metaverse. *IEEE Wireless Communications Letters*, 11(8):1733–1737, 2022.

[95] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, 2022.

[96] M. Wolf and D. Serpanos. Safety and security in cyber-physical systems and internet-of-things systems. *Proceedings of the IEEE*, 106(1):9–20, 2017.

[97] J. Won, C. Mendis, J. S. Emer, and S. P. Amarasinghe. WACO: learning workload-aware co-optimization of the format and schedule of a sparse tensor program. In T. M. Aamodt, N. D. E. Jerger, and M. M. Swift, editors, *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2, ASPLOS 2023, Vancouver, BC, Canada, March 25-29, 2023*, pages 920–934. ACM, 2023.

[98] J. Wu, Z. Wang, A. Sarker, and M. Srivastava. Acuity: Creating realistic digital twins through multi-resolution pointcloud processing and audiovisual sensor fusion. In *The 8th ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI 2023)*. ACM and IEEE, 2023.

[99] T. Xing, S. S. Sandha, B. Balaji, S. Chakraborty, and M. Srivastava. Enabling edge devices that learn from each other: Cross modal training for activity recognition. In *Proceedings of the 1st International Workshop on Edge Systems, Analytics and Networking*, pages 37–42, 2018.

[100] M. Xu, C. Li, S. Zhang, and P. Le Callet. State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):5–26, 2020.

[101] Q. Xu, T. Mytkowicz, and N. S. Kim. Approximate computing: A survey. *IEEE Design & Test*, 33(1):8–22, 2015.

[102] Z. Yan, Y. Lin, G. Wang, Y. Cai, P. Cao, H. Mi, and Y. Zhang. Lasershoes: Low-cost ground surface detection using laser speckle imaging. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.

[103] Z. Yang, Y. Cui, B. Yu, Z. Anwar, R. Bocchino, N. Liyanckar, K. Nahrstedt, R. Cambell, and W. Yurcik. Real-time 3d video compression for tele-immersive environments. In *SPIE/ACM Multimedia Computing and Networking (MMCN)*. SPIE, January 2006.

[104] S. Yao, Y. Hao, Y. Zhao, H. Shao, D. Liu, S. Liu, T. Wang, J. Li, and T. Abdelzaher. Scheduling real-time deep learning services as imprecise computations. In *2020 IEEE 26th international conference on embedded and real-time computing systems and applications (RTCSA)*, pages 1–10. IEEE, 2020.

[105] S. Yao, J. Li, D. Liu, T. Wang, S. Liu, H. Shao, and T. Abdelzaher. Deep compressive offloading: Speeding up neural network inference by trading edge computation for network latency. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 476–488, 2020.

[106] S. Yao, Y. Zhao, A. Zhang, L. Su, and T. Abdelzaher. Deepiot: Compressing deep neural network structures for sensing systems with a compressor-critic framework. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–14, 2017.

[107] K. Yoo, R. Welden, K. Hewett, and M. Haenlein. The merchants of meta: A research agenda to understand the future of retailing in the metaverse. *Journal of Retailing*, 2023.

[108] Y. Zhang, V. Kiriansky, C. Mendis, S. P. Amarasinghe, and M. Zaharia. Making caches work for graph analytics. In J. Nie, Z. Obradovic, T. Suzumura, R. Ghosh, R. Nambiar, C. Wang, H. Zang, R. Baeza-Yates, X. Hu, J. Kepner, A. Cuzzocrea, J. Tang, and M. Toyoda, editors, *2017 IEEE International Conference on Big Data (IEEE BigData 2017), Boston, MA, USA, December 11-14, 2017*, pages 293–302. IEEE Computer Society, 2017.

[109] Y. Zhang, M. Yang, R. Baghdadi, S. Kamil, J. Shun, and S. Amarasinghe. Graphit: A high-performance graph dsl. *Proc. ACM Program. Lang.*, 2(OOPSLA), oct 2018.

[110] Y. Zhao, J. Li, C. Liao, and X. Shen. Bridging the gap between deep learning and sparse matrix format selection. *SIGPLAN Not.*, 53(1):94–108, feb 2018.

[111] T. Zhou, R. Tian, R. A. Ashraf, R. Gioiosa, G. Kestor, and V. Sarkar. React: Redundancy-aware code generation for tensor expressions. In *Proceedings of the International Conference on Parallel Architectures and Compilation Techniques*, PACT '22, page 1–13, New York, NY, USA, 2023. Association for Computing Machinery.