Digital Twins for Smart Spaces—Beyond IoT Analytics

Naser Hossein Motlagh, Martha Arbayani Zaidan, *Member, IEEE*, Lauri Lovén, *Senior Member, IEEE*, Pak Lun Fung, Tuomo Hänninen, Roberto Morabito, *Member, IEEE*, Petteri Nurmi, and Sasu Tarkoma, *Senior Member, IEEE*

Abstract—Smart spaces, physical spaces that are integrated with sensor-enabled Internet of Things devices, are a powerful paradigm for optimizing the operations of the space and improving its quality for the occupants. Managing the applications and services running in the space is a complex task as the operations of the devices and services are dependent on the physical characteristics of the space, the occupants of the space, and the technologies that are being integrated. Digital twinning, the combination of physical representations with a virtual counterpart, is a potential technology for facilitating the management of smart space devices and services. While digital twins are increasingly adopted in industry, their use in everyday environments remains low due to difficulties in creating and linking the virtual representation with the physical environment. In this article, we propose our vision for the adoption of digital twinning as a pathway to improve the functions of smart spaces. We derive a generic reference architecture that comprises four layers, covering the physical space, the sensing infrastructure, the network interfaces, and the underlying computational infrastructure. Next, we identify and address key requirements for the uptake of digital twins in smart spaces and assess their benefits using the ascendancy model of business analytics. Finally, to demonstrate the practicality of digital twinning, we present a proof-of-concept digital twin for the TellUs smart space at the University of Oulu in Finland and use it to highlight the potential benefits of different ascendancy levels.

Index Terms—Digital twin, Internet of Things (IoT), long range (LoRa) networks, sensor networks, smart spaces.

I. Introduction

DIGITAL twins refer to systems that couple physical entities with virtual counterparts, leveraging the strengths of both the virtual and the physical environments for the advantage of the entire system [1]. The virtual representation of the environment is referred to as the *digital twin*. A digital twin relies on sensor-enabled Internet of Things (IoT) devices that synchronize the state of the virtual object with that of the physical object. This requires monitoring the physical twin and its interactions and potentially integrating actuators that directly influence it [2]. A digital twin thus provides an endpoint for data acquisition from the physical counterpart and supports the efficiency of the physical part by optimizing its operations throughout its life cycle [3]. Digital twins are particularly relevant for industry and society [1] with examples of domains adopting digital twins including manufacturing, healthcare, and construction industry. Indeed, digital twins and the underlying IoT and other technologies are a key step in the next-generation transformation of industry [4], [5], [6].

Smart spaces–physical spaces that integrate sensor-enabled IoT devices–are emerging as a powerful solution to optimize operations and improve the quality of experience for occupants. Smart spaces build on the increasing availability of sensors in physical spaces, e.g., for monitoring energy use, thermal comfort of occupants, social distancing, air quality, and general well-being [7], [8], [9]. These sensor-enabled devices enable various applications, which help, among other things, to reduce electricity usage, optimize heating, ventilation, and air conditioning (HVAC) use, improve the safety of the space, and offer services that support the occupants [10], [11]. Indeed, smart spaces aim to enhance the functions of these environments and overall elevate the user experience [12], [13]. Currently, most smart spaces implement this functionality by relying on analytic platforms and hub-type IoT architectures. These offer a single point to collect information and to interact with the space [14] but lack a mechanism to evaluate and interact with services and devices from the outside. Digital twins can help overcome this bottleneck, offering a unified architecture for integrating and managing devices and services while at the same time offering a platform that supports the development and evaluation of new services. The potential and promise of digital twin technology in smart spaces are supported by existing research. However, this research has been mostly limited to building
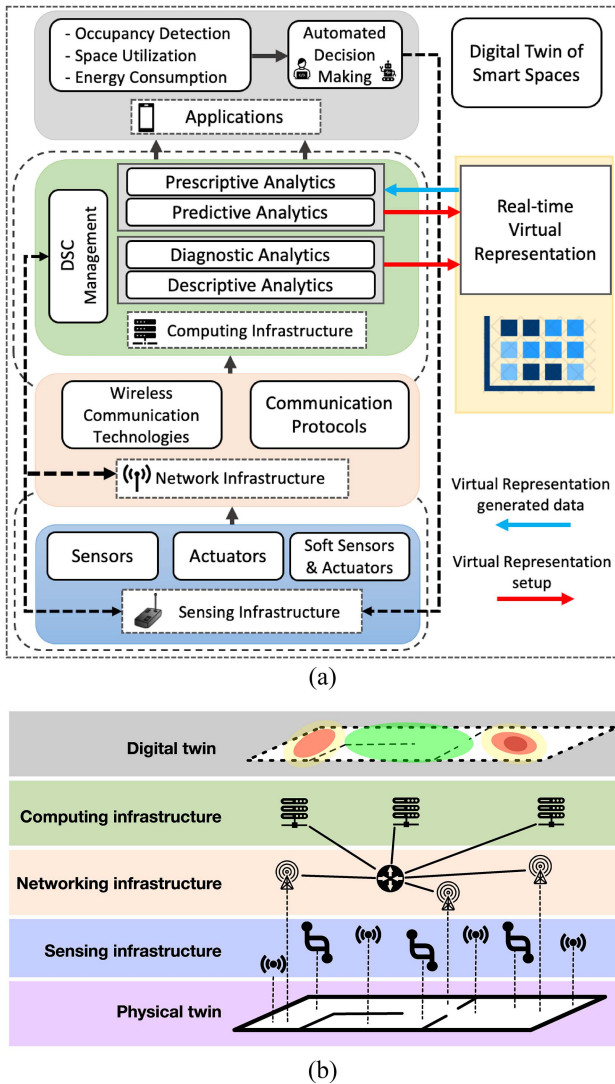
(a)

(b)

Fig. 1. From IoT analytics to digital twin of smart spaces. (a) Digital twin architecture for smart spaces. (b) Overview of the digital twinning of smart spaces.

management systems. It has not often focused on smaller, dedicated areas of space [15]. Improving the situation calls for a reinvestigation of the concept of digital twins for smart spaces together with architectures and technical solutions that can build and link virtual representations of the space with its physical counterpart.

This article contributes a vision for the adoption of digital twins in smart spaces. The overall vision is illustrated in Fig. 1 and provides a unified view of how to integrate sensors, actuators, network interfaces, and computing capabilities with the physical space and the occupants residing in it. Currently, these capabilities are mostly used for standalone applications that attempt to improve specific aspects of the space, but they can also be harnessed for digital twinning to offer a unified view that supports different applications and services and that facilitates management of the capabilities available for the space. Building on this vision, we first derive a generic reference architecture that comprises four layers: the physical space, the sensing infrastructure responsible for establishing the digital representation used by the digital

twin, the network infrastructure for linking the digital data with a virtual representation, and the computing infrastructure that controls the actuators in the space and offers an interface for applications and services; see Fig. 1(b). We then proceed to identify and address requirements for creating digital twins of smart spaces and highlight their potential benefits using the ascendancy business analytics model [16]. Finally, to demonstrate the practicality and the benefits of digital twinning, we present a proof-of-concept digital twin for the TellUs smart space at the University of Oulu in Finland [17]. Our proof-of-concept builds on a rich set of sensors from different modalities, as shown in Fig. 2. We also use our example to highlight the potential benefits of different ascendancy levels.

II. DIGITAL TWINS OF SMART SPACES

Digital twinning of a smart space starts from the sensor infrastructure that operates in the physical space and processes sensor data following a general model of IoT applications that integrates networking and computing on top of the sensor layer serving as input; see Fig. 1(b). First, sensors must be installed into the space, and actuators need to be fitted to allow control over the environment. Second, the connectivity of sensors and actuators must be ensured by deploying appropriate networking infrastructure. Finally, computational resources (either cloud, local, or edge/fog-based) must be provided to process sensor data. Fig. 1(a) details our envisioned digital twin architecture for smart spaces. The architecture includes four layers, corresponding to those on Fig. 1(b), and each is responsible for a specific task. In the sections below, we take a detailed look at each of the layers, survey the most relevant technologies, and highlight the functionality provided by different types of digital twins.

A. Sensing Infrastructure

The sensing infrastructure includes physical devices, such as sensors that observe the environment and generate data, actuators that trigger and actuate the operations of the sensors, as well as software-based (i.e., virtual) sensors and actuators. A variety of low-cost sensing solutions are available for building digital twins for smart spaces. However, while many of these sensors can be considered as viable solutions for smart space deployments, they may still have technological and methodological limitations. Hence, it is often recommended to deploy more than one type of sensor so that they can complement each other and offset these limitations. Such heterogeneous deployments also allow for a better understanding of the events occurring in the environment. In the following, we review and address some of the most popular sensor solutions, covering their limitations and suitability for creating a digital twin of a smart space.

1) Passive Infrared Sensors: These sensors focus on the infrared radiation changes caused by movements in the spaces. While the main advantage of these sensors includes their low-cost and low-energy requirements, the limitations of passive infrared (PIR) sensors is their low accuracy and their inability to directly detect people. PIRs are widely used in real implementations, and they are considered as an important candidate

in physical infrastructure. Indeed, digital twin-based analytics of PIR sensors allow occupancy counting and identifying mobility patterns such as the direction of movements.

2) Environmental Sensors: These sensors provide information, e.g., on temperature, humidity, pressure, CO_2, $PM_{2.5}$, and the presence of volatile organic compounds (VOCs). Thus, these sensors provide an overall view of the conditions in the space and can be used to provide indications of the air quality and potential problems inside the data. While environmental sensors are becoming cheaper and easier to deploy, the less expensive variants often have lower accuracy and require either periodic manual recalibration or software-based recalibration, which can be achieved using machine learning methods. Environmental sensors can also provide indications of the presence, movements, and even activities of occupants even if these can only be detected with a delay [11].

3) Cameras: Cameras, including those focusing on specific wavelengths, such as thermal cameras monitoring infrared radiation, can be used to visually monitor the events in the spaces. The main drawback with the use of cameras is their invasiveness in terms of privacy. Deploying infrared cameras in large quantities in smart spaces can also be expensive. Lower resolution sensors, such as thermal array sensors, can be used as an alternative to capture data that is useful for analyzing the presence of people without violating privacy [10].

4) Light and Noise Sensors: Ambient light and sound levels are essential for ensuring visual and hearing comfort. These sensors are often affordably priced, allowing deploying them in high numbers. However, their sensing accuracy may decrease over time. Both types of sensors can also be used as proxies for occupancy counting. Additionally, noise sensors can be used to identify the types of activities occurring in indoor environments.

5) Wireless Sensing: Wireless sensing takes advantage of the wireless channel to detect activities taking place in space. In its simplest form, wireless sensing can simply estimate occupancy by counting the number of active connections, whereas in the more complex case, the wireless channel can be used to monitor vital signals by looking at fluctuations in the wireless channel between a transmitter and receiver [18]. As most devices have wireless interfaces, the devices can technically be of any type, including laptops, mobile phones, and tablets, and the wireless technology can be IEEE 802.11 (WiFi), Bluetooth, or another short-range technology. However, advanced wireless sensing can only be run on devices that provide detailed information about the wireless channel (so-called channel state information) and typically require dedicated devices that can capture the relevant information.

6) Other Sensing Technologies: These include, for example, electricity smart meters, geophone sensors, and micro-switches. Electricity smart meters can be utilized to monitor energy consumption levels. Moreover, by training energy consumption models specific to the smart space, they can also be used to estimate the number of users there. Moreover, geophone sensors can be used to detect vibration on the floors caused by stepping, and micro-switches can be attached to seats to identify which seat is in use.

B. Networking Infrastructure

Networking infrastructure includes all the necessary components to interface physical infrastructure and computing infrastructure and transfer data between them. These include those directly embedded in and used by the physical infrastructure components (e.g., devices' radio access technologies and communication protocols), as well as those that form the backbone of our architecture (e.g., access points, switches, routers, etc.). Sensor data as well as actuator instructions are transmitted through the networking infrastructure with appropriate wireless technologies and IoT communication protocols.

1) Wireless Communication Technologies: Various wireless communication technologies may be utilized by sensors and actuators within the smart space, facilitating their interaction with the computing infrastructure via the network infrastructure. A range of technologies can provide reliable communication. In addition to reliability, several factors must be considered when determining the most appropriate connectivity method: 1) application requirements; 2) communication range; 3) bandwidth; 4) power consumption; and 5) security. Given these requirements, three wireless technologies stand out for meeting these needs and for their broad compatibility with diverse sensors: Bluetooth low energy (BLE), long range (LoRa), and cellular IoT (LTE-M and NB-IoT). We emphasize that these are not the sole solutions; individual sensors may employ other methods. For instance, wireless M-bus is often utilized by smart meters for measurement transmission. As wireless M-Bus, and other related technologies, are focused on specific devices or sensors we omit them since they cannot offer a generic interface for integrating all computing aspects of the space into a digital twin. We also note that most of these technologies operate within the ISM frequency band which means that using multiple different technologies can cause significant cross-technology interference and degrade network performance. The choice of network technology is also significant because it affects where the computations are expected to reside. Specifically, BLE usually assumes connecting to a separate device that resides in the same space, whereas cellular IoT and LoRa connect to a hub or base station that allows edge-type of computing without having the computing support reside inside the space. While most smart spaces come equipped with WiFi networks, sensor-enabled IoT devices rarely use WiFi due to their high-power usage, and thus other technologies are usually used to connect to a device which can then take advantage of the WiFi or other communications infrastructure available in the space.

BLE is a widely used short-range wireless communication technology particularly suited for being used in the IoT landscape because of its low-power requirements, low-installation costs, and high pervasiveness. It enables data exchange using a 2.4-GHz license-free (ISM) frequency band, ensuring a nominal max range of above 100 m in open space. The bit rate is 1 Mbit/s (with an option of 2 Mbit/s in Bluetooth 5), and the maximum transmit power is 10 mW (100 mW in Bluetooth 5). The BLE's design decision results in low-energy consumption, cost, and dimensions of the chipset, making this technology especially popular for sensors that interact with a higher end device, e.g., wearables interacting with a smartphone or computer.

Cellular IoT (NB-IoT, LTE-M) technologies have been defined in the context of 3rd Generation Partnership Project (3GPP) and designed to enable more streamlined machine-type communication (MTC). The main advantages of this type of technology is its seamless coexistence with 5G access technology and the support of IP-based end-to-end traffic. The suitability of cellular IoT deployments stems from various aspects, including the high interoperability with mobile telecommunications standards and the lower power usage compared to broadband cellular technologies (e.g., conventional LTE or 5G) when related to the performance in terms of data rate (0.2–1 Mb/s) and range (up to several kilometers).

LoRa is a wireless communication technology specifically designed for IoT. The data rates of LoRa are lower compared to the other technologies we have introduced (0.3 Kbit/s to 27 Kbit/s) but it enables a greater coverage (due to operation in sub-GHz bands) allowing transmission range of up to dozens of kilometers in line of sight and very low-power usage and installation costs. Due to its design, the technology is well-fitted for the IoT applications that mostly report measurements, i.e., transfer data predominately in uplink, and which can tolerate packet losses.

2) Communication Protocols: Communication protocols for IoT usually build on either a publish–subscribe model where clients publish data that applications or devices can subscribe to or a request–response model where a server or a proxy queries devices for information. Popular examples of these types of protocols are the message queuing telemetry transport protocol (MQTT) for the former and the constrained application protocol (CoAP) for the latter.

MQTT is an IoT protocol originally designed to work on top of TCP, which follows the publish/subscribe model. MQTT client publishes messages to an MQTT broker, which are subscribed by other clients or may be retained for future subscriptions. Every message is published to an address, known as a topic. Clients can subscribe to multiple topics and receive every message published on each topic. The TCP nature of MQTT, which also uses TLS for securing the data transfer, makes this protocol connection oriented. Still related to its reliability capabilities, MQTT allows using three different levels of QoS. With the lowest level of QoS, MQTT operates in a best-effort message delivery fashion. A QoS of one guarantee that a message is delivered at least one time to the receiver, while the highest QoS guarantees that each message is received only once by the intended recipients.

COAP is a lightweight IoT protocol that has originally been defined in the context of the constrained restful environments (CoREs) working group of Internet Engineering Task Force (IETF). CoAP is developed to interoperate with RESTful systems and protocols (e.g., HTTP), through an architecture that can alternatively follow both request/response and resource/observe paradigms. Different from MQTT and although inspired by HTTP, the original CoAP standard uses UDP as a transport protocol and DTLS for security. Despite being connectionless datagrams protocol, reliability and QoS definition are ensured through the use of "confirmable" messages (which must be acknowledged by the receiver with an ACK packet) and "nonconfirmable" messages (which do not require to be acknowledged by the receiver). Unlike MQTT, CoAP uses universal resource identifier (URI) instead of topics. The publisher publishes data to the URI and the subscriber subscribes to a particular resource indicated by the URI. When a publisher publishes new data to the URI, all subscribers are notified about the new value as indicated by the URI.

We highlight these two application layer protocols in favor of alternatives (e.g., Hyper Text Transport Protocol-HTTP and Advanced Message Queuing Protocol-AMQP) due to the favorable tradeoffs that CoAP and MQTT can offer in terms of power consumption versus resource requirements, bandwidth versus latency, and message size versus message overhead.

On top of these application layer protocols, we use lightweight M2M (LwM2M). LwM2M is a REST-based protocol from the open mobile alliance (OMA) for M2M and IoT device management that defines the application layer communication protocol between an LwM2M server and an LwM2M client running on an IoT-embedded device. Although LwM2M was originally built to work on top of CoAP, its latest versions (from 1.2 on) started to support additional application layer protocols, including MQTT and HTTP. The main advantage of using this kind of device management protocol on top of different application layer protocols is the possibility of ensuring interoperability and addressing all the challenges raised by the heterogeneous nature of IoT devices and of the applications that are executed on top of them. The possibility of managing a plethora of devices in a unified fashion brings several advantages, further emphasized by the fact that LwM2M's device management capabilities include, inter alia, remote provisioning of security credentials, firmware updates, connectivity management (e.g., for cellular and WiFi), remote device diagnostics and troubleshooting.

C. Computing Infrastructure

Computing infrastructure, placed in the cloud, locally, or at the edge of the network, performs analytics of the data received from the network infrastructure, provides real-time virtual representation from the events in the smart space, and generates virtual data as input for analytics. As described previously, the location of the computing infrastructure can also depend on the underlying network technology as BLE effectively assumes a hub that resides inside the space, whereas cellular IoT and LoRa communicate with base stations that can be outside of the space. The hub can either integrate processing directly on itself or rely on data centers accessed through cloud interfaces. While the cloud provides access to significant computational capacity, concerns about latency, privacy, and bandwidth may require using computational capacity in closer proximity to the physical twin. Indeed, the architecture should consider the computational resources as a continuum, ranging from the lightweight sensor and actuator devices, through edge and fog nodes placed in nearby computing hotspots or network hubs, all the way to cloud-based data centers [19]. Deployed upon this computing infrastructure are the components, often encapsulated into microservices, of the application providing the digital twin. These microservices, as well as the tasks running upon them further require orchestration and

life-cycle management. These services are often provided by middleware, such as Kubernetes or Docker Swarm [19].

The devices, services, and connectivity (DSC) management module gathers data from the sensing infrastructure using LwM2M. The aim is to gain insights into the sensing and actuation operations on the devices, as well as other telemetry tasks. The DSC management module also controls the accuracy of the sensed data, and the amount of data to be sent in uplink, aiming to reduce data redundancy and unnecessary data transmission. It additionally collects telemetry information from each sensing infrastructure device (e.g., device ID, device model, running service, percentage of remaining battery, memory, and CPU/MCU utilization) and of the connectivity technology used for the data upload (e.g., Signal to Noise Ratio—RSSNR, Reference Signal Received Power—RSRP, Received Signal Strength Indicator—RSSI, and Cell Identity—NCI).

The set of information collected is used for performing AI-powered analytics to infer, for example: 1) how devices and running applications are using energy and under what circumstances; 2) how to improve the sensing capabilities through AI-powered sensors' recalibrations; and 3) how to optimize the end-to-end data transfer in an energy-efficient fashion. Some of the analytics results performed by the computing infrastructure are then sent back to the sensing infrastructure devices, which execute the analytics actionable insights. Analytics results can enable the devices to improve their battery life and sensing capabilities and ensure more optimized and energy-efficient end-to-end communication.

D. Applications

The applications block presents the applications developed using analytics data. These applications can be for instance used for occupancy detection, space utilization, energy consumption, and air quality monitoring in order to improve the environment or optimize resource utilization.

The analytics required by the application can be categorized according to Gartner's analytics ascendancy model [16]. The model identifies four different analytics levels, ordered by the value of their results as well as the complexity of the methods. Descriptive analytics offers a view into the current status of the observed system, as well as its recent history, answering the question "what happened." Diagnostic analytics looks for the causes and effects behind the system status, answering the question "why did it happen." Predictive analytics projects system status into the future, finding out "what will happen next." Finally, prescriptive analytics looks for means of affecting future outcomes: "how can we make it happen?"

The ascendancy levels describe also digital twins, providing a measure of capability and utility. In particular, in the context of smart spaces, a digital twin can help facility managers to know (*descriptive*) and understand (*diagnostic*) the past and current status of the space, assess potential future trajectories (*predictive*), or even find ways of changing those trajectories or selecting the most beneficial (*prescriptive*).

Different types of smart space digital twins, categorized with Gartner's analytics ascendancy model [16], along with sample use cases, are listed below.

1) Descriptive Twins: Present the current status and situation in the smart space, relying on data from physical sensors and software systems, such as meeting room reservation tools, linked to the space. Presented data can include, for example, the occupancy or reservation status of spaces, the temperature or CO_2 level at sensor locations, or the number, location, and possibly even identity of the people in the smart space. Techniques used at this level may include data aggregation, visualization, and summary statistics, such as calculating the average temperature, humidity, or CO_2 levels for each zone.

2) Diagnostic Twins: Further analyze the descriptive data. Results from the analysis can, for example, find the correlation between the CO_2 level and the occupancy of a room, or the temperature and the number and location of people, or possible discrepancies between meeting room reservations and their actual occupancy. Methods employed at this stage may include correlation analysis, root cause analysis, and data mining, such as assessing the relationship between CO_2 levels and occupancy (using PIR sensors) and investigating reasons behind unusually high humidity levels in a particular zone.

3) Predictive Twins: Provide insights into the future status of the smart space. Potential use cases include, for example, the expected occupancy of meeting rooms, or projected energy costs of the smart space, based on expected occupancy, energy price projections, and external weather forecasts. Approaches utilized at this tier may include machine learning algorithms, such as random forests or support vector machines, time-series analysis techniques, such as ARIMA or exponential smoothing, and forecasting models to predict future conditions in the smart space.

4) Prescriptive Twins: Build upon the insights offered by analytics on lower ascendancy levels, prescriptive twins provide means for the operator on how to control the space, or even control the smart space autonomously. Examples of use cases include, for example, controlling the heating or ventilation of the space, based on predictive analysis of occupancy, temperature, and CO_2 levels, and projected energy costs. Procedures used at this stage may include optimization algorithms to determine optimal settings for lighting and HVAC systems, decision trees, and simulation models to test the impact of different control strategies on overall energy consumption and air quality, recommending the most efficient strategy for managing the smart space.

Our proposed digital twin architecture (shown in Fig. 1) offers automated decision making and enables actuating sensors and adjusting systems in the smart space to meet the users' needs. Indeed, at the application layer, the automated decision making can for example use the data analytics results of the occupancy detection to adjust the operations of the ventilation and lighting systems to optimize energy consumption as well as to provide visual and thermal comfort for the space users.

III. EXPERIMENT

We use measurements of the wireless sensor network deployed in the TellUs smart space at the University of Oulu, Finland, to explain the benefits of the digital twin of smart spaces.
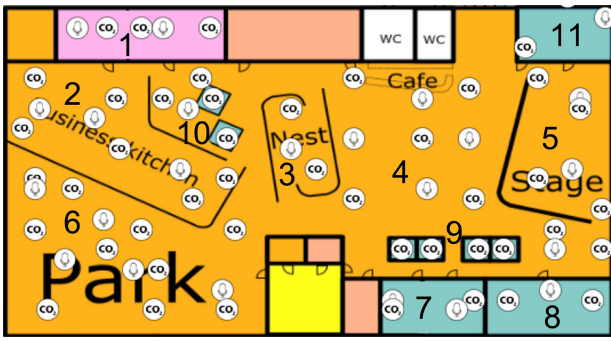
Fig. 2. Sensor deployment at the TellUs smart space at the University of Oulu, Finland. For simplicity, the space names are labeled with numbering 1–11 (i.e., different zones). The CO_2 and microphone icons show the locations where, respectively, the multisensor devices and the noise sensors are installed.

Physical Twin: The floor plan of the TellUs space is shown in Fig. 2. The TellUs space is located on the first floor of the University campus and has an area size equal to approx 66.856 m $\times$ 36.446 m. TellUs includes different spaces planned for various purposes. These spaces include the closed spaces used as meeting rooms shown by numbers 1, 7, 8, and 11, large meeting rooms (shown by 5) are designed for students' collaborative works, and open study areas are depicted by numbers 2, 6, and 10. The Nest space (shown by 3) is a relaxation and thinking space, Cafe TellUs shown by 4 provides cafeteria facilities, and the small closed spaces (at 9, and partly at 10) that host up to four people are used for individual use or small group meetings [17].

Sensing and Networking Infrastructure: In order to design our sensor deployment, we conducted a test phase experiment prior to the current sensor deployments. During this phase, we deployed 352 LoRa wide area network (LoRaWAN) sensor nodes at the TellUs smart space [8]. To deploy sensors permanently and in order to identify the optimal number of sensors and find proper locations for installing sensors, in our experiment, we implement a test phase by deploying a total of 352 LoRaWAN of the same sensor nodes (Elsys ERS sensors [20]) in Tellus space. We collected a total of 9 917 848 lines of data for 410 continuous days of measurements from June 2017 to November 2018. We further use the data that includes temperature, humidity, CO2, motion, and light to test the reliability of measurements and sensor operations. We carried out a comprehensive data analysis using different visualizations, such as sensor measurements during weekdays and weekends, time-series measurements, and correlation studies between PIR and CO2 concentrations. Indeed, the test phase allowed us to identify the hotspots and areas necessitating continuous monitoring. Through the insights gained from this experiment and expert advice, we managed to optimize the sensor count to 68 devices, a number deemed appropriate for the TellUs space, even in terms of calibration and maintenance. Therefore, in the current operation phase, we deployed 68 sensor nodes in the TellUs space. While 23 units of these sensors are noise sensors (shown with microphone icons in Fig. 2) that can only measure sound, the other 45 sensor units shown by CO_2 can measure temperature, humidity, CO_2, motion, and light. These sensors have been calibrated in the factory by the manufacturer prior

to their deployment [21]. Before we embarked on the sensor deployment, we calibrated each sensor in a laboratory setting using a reference sensor. This was done to ensure the highest possible accuracy in our measurements. Each sensor is powered by two 3.6 V-AA lithium batteries. Based on experts' suggestions, the sensors are attached to the ceiling frames of the TellUs space with a specific minimum distance between each other. Using the LoRaWAN technology, every 15 min, the sensor units transmit data on the 868 MHz ISM band to a LoRAWAN gateway manufactured by Multitech. While the performance of the LoRaWAN deployment is documented in [22], in our case, none of the sensors had more than 25% of their packets lost, while for some of the nodes less than 0.5% of packets were lost. The gateway is connected to an external biconical D100–1000 antenna with a gain of 2dBi, and transfers the data to the ThingWorx commercial cloud platform using the MQTT protocol.

Computing Infrastructure and Platform: Data is stored on a local campus server via Python scripts querying the ThingWorx commercial cloud platform.[1] The local campus server is equipped with an RDBMS PostgreSQL server for data storage, R, Shiny Server, and Django REST Framework. In-depth technical details regarding deployment setup and analysis are provided elsewhere [23]. Moreover, the computing infrastructure allows implementing management tools and functions and deploying virtual counterparts. These tools, available online,[2] include the following: a "Device management" function, which allows managing the installations and adding new devices; "Bootcamp," which allows us to check the status of the installed sensors; a "map view," that provides location information about the installed sensor devices; "API key," that provides real-time data stream from sensors on request for API and obtaining a key. "Open data" is also another virtual element offered by our system, providing open data freely under the CC BY 4.0 license in the form of staticcsv-files. Furthermore, the system provides real-time "data visualization" using Grafana which is an open-source software platform for visualizing time-series data.

Data Sets and Preanalysis: The sensor network's measurements were collected from July 1, 2020 to May 31, 2021, which yielded a total of 8040 data points, recorded hourly. The data is collected on a cloud-based service and is openly available for further study.[3] We also use our previous data set collected from June 2017 to November 2018. Then, first, we process both data sets for the test phase ($\mathcal{T}$) and implementation phase ($\mathcal{I}$) by removing outliers and anomaly data points, and then we perform a preanalysis. We carry out this step to ensure that the data collected by sensors are reliable and can be further utilized for further analytics. Table I summarizes the key statistics, which are mean and standard deviation (STDV), of the data sets for the test phase ($\mathcal{T}$) and the implementation phase ($\mathcal{I}$).

Our observations indicate that the area is frequently occupied, as suggested by the mean values of the variables (CO_2

---

[1]"Alvar," 2018. [Online]. Available: http://alvar.erve.vtt.fi.

[2]https://smartcampus.oulu.fi/manage/

[3]https://etsin.fairdata.fi/dataset/98ff83ec-96fb-45d2-af82-c041200a3fb2

TABLE I
KEY STATISTICS OF THE DATA SETS FOR THE TEST PHASE ($\mathcal{T}$)
AND THE IMPLEMENTATION PHASE ($\mathcal{I}$)

Measurement	Mean		STDV	
	$\mathcal{T}$	$\mathcal{I}$	$\mathcal{T}$	$\mathcal{I}$
CO_2	459.05	444.59	161.17	100.28
Humidity	30.60	41.93	16.25	4.63
Light	215.41	361.90	302.41	342.36
Temperature	20.65	23.57	8.71	15.56
PIR	44.76	39.95	37.79	41.30

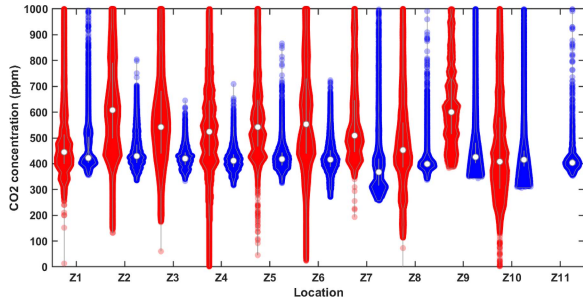
Fig. 3. Violin graphs—the comparison of CO_2 concentrations at 11 zones during the test phase ($\mathcal{T}$, red) and the implementation phase ($\mathcal{I}$, blue).

TABLE II
DIAGNOSTIC TWIN, PROVIDING READINGS IN THE 11 ZONES IN TELLUS

	Metric	Temp (°C)	RH (%)	CO_2 (ppm)	PIR	Lux	Noise
Z_1	med.	21	25	422.5	84	115	34
	std.	1.31	15.73	73.33	40.68	130.47	3.34
Z_2	med.	19.6	28	429.5	85	150	34
	std.	1.71	12.98	58.52	40.07	189.33	1.53
Z_3	med.	20	27	420	86.5	190	34
	std.	1.58	13.26	36.84	40.78	275.18	0.92
Z_4	med.	19.85	28	411.5	49.5	166	37
	std.	1.77	13.79	38.43	32.65	173.31	3.35
Z_5	med.	19.85	28	418	76.5	115.5	34
	std.	1.55	14.38	40.24	36.59	64.16	2.39
Z_6	med.	20	28	416	78	139	34
	std.	1.89	12.49	57.17	37.94	131.07	1.78
Z_7	med.	21	26	367.5	81	143	41
	std.	1.47	15.13	143.25	40.45	99.09	5.38
Z_8	med.	20.1	29	399	84.5	163.5	34
	std.	1.74	15.61	52.49	39.46	60.9	3.03
Z_9	med.	20.15	30	426	4	80.5	NaN
	std.	1.58	13.50	622.05	4.39	35.74	NaN
Z_{10}	med.	20.5	28	415.5	75.5	162	34
	std.	1.40	13.06	488.71	31.06	71.35	1.60
Z_{11}	med.	20.95	25	404	82	162	35
	std.	1.54	16.14	51.97	40.94	42.72	3.30

improvement over $\mathcal{T}$. Despite the optimized number of sensors installed in the space, the quality of spatial coverage remains consistent.

IV. RESULTS

In this section, we present the outcomes of our experiment in the TellUs smart space, concentrating on the construction of a digital twin. Our study illustrates three distinct types of digital twins: 1) descriptive; 2) diagnostic; and 3) predictive. We also use Table II and Fig. 4 to present the results of our experiment. Finally, we discuss possible avenues for a prescriptive twin.

A. Descriptive Twin

A descriptive digital twin provides information on the current and past state of the physical twin. An implementation of a descriptive twin, Table II describes the current status in the TellUs smart space in terms of temperature, humidity, air quality, occupancy, light, and noise. Table II also shows the median (med) and standard deviation (std) of the measurements from the sensors in 11 zones in the TellUs smart space. The measurements include temperature (Temp in °C), relative humidity (RH in %), carbon dioxide (CO_2 in ppm), PIR sensor, Lux, and Noise (in dbA). The findings displayed in Table II affirm that the environmental conditions within the TellUs smart space align with typical indoor environments. Furthermore, these results actualize a descriptive twin, elucidating the events within the smart space and offering insight into its dynamics. For example, the median temperature varies between 19 °C and 21 °C (with std. hovering around 1.5 °C), and the median of RH varies between 25% and 29% (with the average of std of 15%). These are typical indoor conditions for buildings in Finland.

The median values for the three highest concentrations of CO_2 align with the maximum PIR measurements, observed in zones Z_1 to Z_3. However, the median CO_2 concentration for Z_9 is also high, while the Z_9 is small—this is a cubic closed space where the CO_2 is trapped and the effect of the ventilation system is almost none. Consequently, the measurements from the PIR sensor, which detects movement, are notably small.

mean exceeds 400, and PIR mean is above zero). The mean temperature, humidity, and light values conform to the standard thermal and comfort levels for indoor environments. For example, the typical indoor temperature in Finland is defined to be 20 °C [24]. As demonstrated in Table I, key statistics, namely, mean and STDV, exhibit similarity between $\mathcal{T}$ and $\mathcal{I}$. This consistency validates the reliability of the collected data sets.

Furthermore, a crucial benefit of our current sensor implementation phase ($\mathcal{I}$) is the use of an optimized quantity of sensors, which, compared to the test phase ($\mathcal{T}$), does not compromise the quality of spatial coverage. Fig. 3 presents the comparison of CO_2 concentrations in the form of violin graphs at eleven zones between $\mathcal{T}$ (red on the left) and $\mathcal{I}$ (blue on the right). A violin graph depicts distributions of CO_2 concentration data for eleven zones using density curves. The white dots represent the median while the tails represent dispersion (i.e., the confidence intervals) in CO_2 concentration data in different zones. The width of each curve corresponds with the approximate frequency of data points in each zone. Note that, in the test phase ($\mathcal{T}$) our data set does not have the CO_2 data for Z_{11}, as we had not deployed CO_2 sensor during the test measurement for that zone. The consistency of measurements between both phases, $\mathcal{I}$ and $\mathcal{T}$, substantiates the viability and rationality of our implementation ($\mathcal{I}$). There is a greater dispersion in $\mathcal{T}$ than in $\mathcal{I}$ due to the larger volume of sensory data in the $\mathcal{T}$ phase. Nevertheless, the medians of each violin plot (the white dots) are in close proximity between $\mathcal{T}$ and $\mathcal{I}$. The medians for both phases are within the acceptable ranges of indoor CO_2 concentration [25]. This observation demonstrates that the installation of an optimized number of sensors at hotspot locations can effectively provide coverage for each zone within the TellUs Smart Space. In conclusion, the illustrated results suggest that the sensor deployment in $\mathcal{I}$ is an
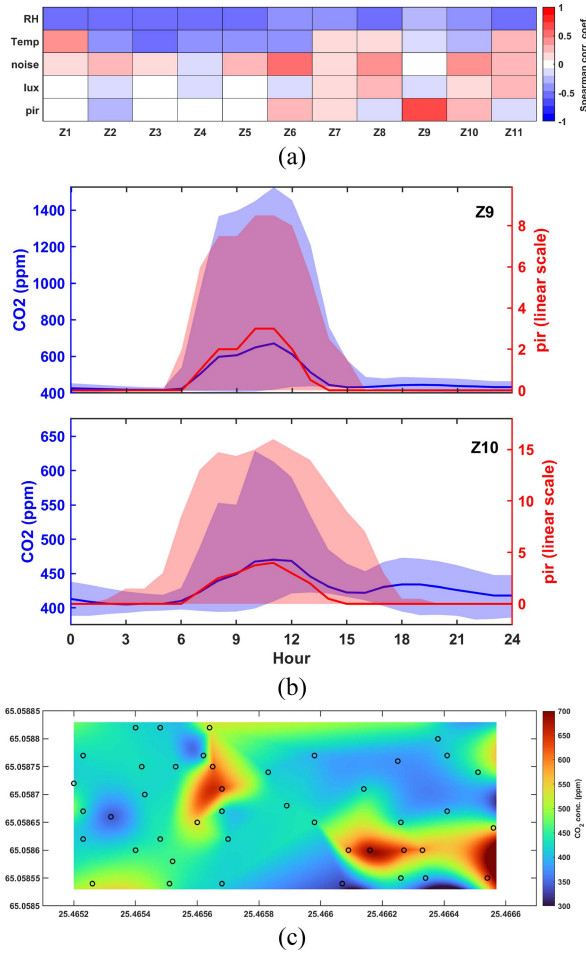
Fig. 4.   Diagnostic and Predictive twins: analyzing and estimating different sensor modalities. (a) Heatmap of Spearmann correlation coefficients between CO_2 and other parameters in different zones. (b) Median (line) and standard deviation (shaded area) of the diurnal cycle of CO_2 and pir in Z_9 and Z_{10}. (c) Predictive twin, interpolating CO_2 sensor observations to locations where sensor data is not available.

As such, the descriptive results imply a correlation between CO_2 and PIR. The max and min of the medians of light sensor readings are 190 and 163.5 lux, respectively. These values are considered acceptable for the TellUs smart space, given that the lights are switched off during the night and during periods of inactivity. Based on the American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE), for open plan office spaces, the acceptable noise level range between 49–58 dBA. In TellUs, in most of the zones, the noise level is equal to 34 dbA. Z_4 and Z_7, corresponding to the cafeteria and the meeting room 7, show the highest median noise (at 37 and 41 dbA). The cafeteria accommodates numerous visitors daily, and the meeting room records the highest occupancy rate. Despite this, the maximum median noise level stands at 41 dbA, which remains below the threshold defined by ASHRAE.

B. Diagnostic Twin

Diagnostic analytics examine the correlations, causative factors, and effects underlying the phenomena observed at TellUs. Fig. 4(a) shows the heatmap of the Spearman correlation coefficients between CO_2 and all variables in all zones. The heatmap plot aids in comprehending the relationships between high-CO_2 concentration levels and other variables. For example, PIR and noise levels are positively correlated to the CO_2 concentration, especially in Z_6, Z_8, and Z_{10}. This correlation is particularly evident in Z_9, which showcases a high correlation between CO_2 and PIR, given its small, enclosed spaces devoid of external influences like ventilation systems. Moreover, Fig. 4(b) shows the diurnal cycles of CO_2 concentration (blue) and the values of the PIR sensors (red) for Z_9 and Z_{10}. The blue and red shaded areas indicate the standard deviation of CO_2 and PIR diurnal cycles, respectively. In the figure, the overlap in patterns of blue and red lines and the shaded area confirms that the two variables are indeed correlated in Z_9 and Z_{10}. These results suggest that the presence of humans (indicated by high values of PIR and noise) influences the amount of CO_2 concentration in TellUs. Therefore, by controlling the number of human movements in different zones in the office, the level of CO_2 may also be controlled.

C. Predictive Twin

Predictive analytics offer insights into the potential future state of the smart space, or the anticipated value of specific variables in areas where no corresponding sensors are deployed. For example, Fig. 4(c) interpolates the CO_2 sensor readings to cover the whole TellUs smart space, including areas with no CO_2 sensors. The small enclosed spaces at Z_9 and Z_{10} are clearly visible as areas with high-CO_2 concentration. Elsewhere, CO_2 levels across the TellUs space remain relatively low, suggesting effective ventilation or sparse occupancy.

Moreover, another predictive model can estimate the amount of CO_2 concentration in places where CO_2 sensors are not available, but PIR sensors are. In fact, as indicated by descriptive and diagnostic analytics, a correlation exists between CO_2 concentration and PIR values, the latter serving as a proxy for the number of individuals present in the observed area [26]. A predictive model of CO_2 concentration can thus take inputs of PIR sensors and estimate CO_2 by, for example, a linear model, with $|CO_2| = \alpha + \text{PIR}\,\beta + \epsilon$.

In this formulation, $|CO_2|$ is the concentration of CO_2 (in ppm) and PIR is the number of movements recorded. β is the model coefficient, α is the bias, and ϵ is the model error.

D. Prescriptive Twin

Prescriptive analytics offer mechanisms for regulating the TellUs smart space to sustain a healthy and productive environment. For example, according to diagnostic and predictive analytics, CO_2 may be caused by the number of people present. Thus, a prescriptive digital twin can calculate the maximum occupancy for a meeting room at a given time, based on the highest estimated CO2 values in that room, and accordingly adjust room reservations.

V. LESSONS LEARNED

We now turn to discuss the lessons, limitations, challenges, and crucial considerations we have encountered in deploying

sensors in smart spaces for digital twin creation, drawing on our own experiences.

Traditional Set-Up Versus Digital Twin: When transitioning from our previous TellUs setup, which lacked digital twin capabilities, to the newly established digital twin-enabled model, we evaluated our system from several key perspectives. These aspects include *resource utilization*, *energy efficiency*, *latency*, and *scalability*. We hypothesize that integrating digital twin technology can lead to improvements in resource utilization, as it enables real-time monitoring and control over various resources, ensuring their efficient use. In contrast, traditional smart spaces without digital twins may rely on manual adjustments or predefined schedules, which often do not align with actual usage patterns and requirements. Additionally, we believe that the digital twin model has the potential to significantly enhance energy efficiency by intelligently controlling systems based on occupancy levels, ambient conditions, and other factors, while traditional smart spaces may lack the necessary data and control mechanisms to achieve the same level of optimization. Transitioning to a digital twin can also potentially reduce latency, as real-time data processing and decision making are facilitated. Traditional smart spaces may experience delays in processing and respond to events due to the lack of a unified data and control platform. We also suggest that the digital twin simplifies the integration of new devices and services, making it easier to scale the smart space as needed. In comparison, traditional smart spaces may face challenges in adapting to changing requirements and integrating new technologies.

User Satisfaction in the Loop: In the future, we plan to focus on empirically validating the hypotheses mentioned in the previous point and further enhancing user satisfaction by leveraging methodologies that extend the digital twin capabilities with the possibility of relying on fast-feedback user satisfaction. This approach will allow us to create an even more personalized and comfortable environment for users, addressing individual preferences and needs more effectively. Moreover, we advocate for the need to introduce well-defined and standardized metrics that enable researchers and practitioners to fairly evaluate and compare their digital twin based systems with traditional smart spaces across all key aspects. The development, for example, of a digital twin benchmarking suite would facilitate more accurate assessments and encourage further advancements in the field.

Sensing Accuracy: Any sensor that is used for measuring environmental variables should be calibrated at the factory. For our experiment, we performed an additional calibration of the sensors prior to deployment to ensure the capture of reliable data. However, in real-life sensor deployments, one-time calibration of sensors before deploying them does not guarantee data accuracy as these low-cost sensors drift over time and generate anomalous data. Thus, an alternative solution to ensure data accuracy is the periodic calibration of individual sensors or the sensor network in an automated fashion using reference sensors or implementing an automated method that calibrates the sensors opportunistically.

Sensor Deployment: To generate sufficient and appropriate data from smart spaces for digital twin creation, we must strategically deploy the right number of sensors, ensuring they are properly spaced at designated locations. In our experiment, we have deployed 68 LoRaWAN sensor nodes in our TellUs smart space. To achieve this number, we carried out an earlier test by deploying 352 of these sensors in TellUs space and carried out one year of measurements (2017–2018) [8]. Thus, based on our earlier data analysis and the engineers' new sensor deployment design, the number of sensor deployments was optimized and the unnecessary sensors were removed.

Data Management: A digital twin necessitates a real-time data stream. Therefore, data needs to be consumed in real-time and also stored for further analytics. Typically, the data generated by the sensors in smart spaces do not require storage capacity compared to other applications, e.g., hyperspectral imaging. For example, in our experiment, the sensors measure every 15 min and produce 8040 lines of data for each sensor. We collected data from 1 July 2020 to 31 May 2021 (11 months of data), and our data set has a *.csv* file size of 780 MB. Using MQTT protocol, we transmitted the data to the ThingWorx commercial cloud platform and also queried the data by Python scripts and stored it on local servers. Hence, deploying a large number of sensors of even different varieties may not challenge data storage, thus any form of a data storage system, including on the edge or cloud may be an appropriate solution. Integrating new sensors into the digital system presents another challenge due to potential differences in data formats. However, to establish interoperability, thus, the use of unique standards, such as ONEDM or IPSO smart objects, is necessary to obtain a similar data format.

Network Management: Within our TellUs smart space, we set up a wireless sensor network consisting of 68 sensor devices, each transmitting data packets on the 868 MHz ISM band to a remote server via LoRaWAN radio access network technology. To collect data from the sensor nodes, we used a LoRa gateway that is connected to an external biconical D100–1000 antenna which has a gain of 2dBi. Then, using the MQTT protocol the data was relayed from the LoRa gateway to the ThingWorx cloud platform and also stored the data on our local servers. Thanks to the advances in communication technologies that offer a wide variety of networking solutions that are appropriate for IoT deployment in smart spaces as sufficiently addressed in earlier sections.

Security and Privacy: Typically, the sensors which are used in smart spaces, such as motion detectors, environmental and thermal array sensors, and wireless sensing systems due to their application purposes do not capture information that includes people's identities. In our deployment, we used environmental and PIR sensors which did not involve any sort of security and privacy concerns. However, in case cameras are used for occupancy detection, recent studies introduce methods (e.g., by reducing the video frame resolution) that mitigate possible security and privacy concerns that might threaten people's privacy.

Lifecycle Management: Sensors may decay, break, or be dislodged from their locations due to indoor human activities. This mandates continuous monitoring of the operation of the sensors. In our deployment, each sensor device is powered by two 3.6 V-AA lithium batteries. With a sampling frequency

of 15 min and two batteries for each sensor, the devices are expected to have power for about 24 months. In theory, the batteries are depleted after prolonged use powering the sensors. Thus, it is necessary to ensure a continuous power supply for sensors. Indeed, one important advantage of Digital Twin is its automated management which enables the detection of silent sensors that do not transmit data, enabling fixing the problem. Moreover, life-cycle management also concerns the Software-over-the-air (SOTA) and firmware-over-the-air (FOTA) capabilities of IoT deployment. Our deployment did not encompass any SOTA/FOTA capability. This is a big limitation because, when SOTA/FOTA are needed, we are forced to perform such operations device by device or by a group of devices (usually from the same manufacturer). We aim to overcome such limitations by relying on the FOTA/SOTA capabilities of LwM2M, which is already used in our system for performing device bootstrapping and device management.

VI. Conclusion

Smart spaces are progressively becoming commonplace, as new sensor-enabled devices grow more affordable, easier to deploy, maintain, and operate. The growing pervasiveness of IoT-enabled devices is amplifying the potential to build services and applications that benefit occupants, optimize the space's functionality, and manage various aspects of the space. Currently, these applications and services are primarily implemented using dedicated IoT analytics platforms, to which sensor-enabled devices connect, providing a single interface for applications and services. Regrettably, this approach is not scalable and tends to result in a siloed solution where capabilities are optimized for individual use cases, rather than offering a unified view that could better manage, maintain, and leverage capabilities across a broad range of applications and services.

In this article, we argue that smart spaces have matured to a point where dedicated IoT analytics platforms alone are no longer sufficient. Instead, digital twinning, the process of linking the physical space with a virtual representation, serves as a more fitting paradigm for managing and supporting the space. Indeed, we propose that the sensor, communication, and computing infrastructure have reached sufficient maturity to integrate the operations of the space through digital twinning. We presented a generic reference architecture for implementing digital twins for smart spaces using a layered architecture that integrates four different levels (physical space, sensing infrastructure, communications, and computations). We also provided an overview of the key technologies that are currently available and used the analytic ascendancy model to highlight benefits for different stages of implementation. We also presented a proof-of-concept implementation using the TellUs smart space at the University of Oulu in Finland to highlight the benefits digital twins can bring to smart spaces, as well as how different levels of technological maturity affect these benefits. In sum, our work paves the way for transitioning beyond IoT analytics platforms, harnessing digital twin technology to improve smart space quality and offer a unified approach to accessing computing capabilities, thereby enhancing the benefits these spaces offer to the occupants.

References

[1] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the digital twin: A systematic literature review," *CIRP J. Manuf. Sci. Technol.*, vol. 29, pp. 36–52, May 2020.

[2] R. Saracco, "Digital twins: Bridging physical space and cyberspace," *Computer*, vol. 52, no. 12, pp. 58–64, Dec. 2019.

[3] R. Eramo, F. Bordeleau, B. Combemale, M. van Den Brand, M. Wimmer, and A. Wortmann, "Conceptualizing digital twins," *IEEE Softw.*, vol. 39, no. 2, pp. 39–46, Mar./Apr. 2022.

[4] S. Huang, B. Wang, X. Li, P. Zheng, D. Mourtzis, and L. Wang, "Industry 5.0 and society 5.0—Comparison, complementation and co-evolution," *J. Manuf. Syst.*, vol. 64, pp. 424–428, Jul. 2022.

[5] F. Dembski, U. Wössner, M. Letzgus, M. Ruddat, and C. Yamu, "Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany," *Sustainability*, vol. 12, no. 6, p. 2307, 2020.

[6] B. Yang, Z. Lv, and F. Wang, "Digital twins for intelligent green buildings," *Buildings*, vol. 12, no. 6, p. 856, 2022.

[7] M. Sulzer, A. Christen, and A. Matzarakis, "A low-cost sensor network for real-time thermal stress monitoring and communication in occupational contexts," *Sensors*, vol. 22, no. 5, p. 1828, 2022.

[8] N. H. Motlagh et al., "Monitoring social distancing in smart spaces using infrastructure-based sensors," in *Proc. IEEE 7th World Forum Internet Things (WF-IoT)*, 2021, pp. 124–129.

[9] L. Lovén, M. Pakanen, E. Gilman, and S. Pirttikangas, "Wellbeing in smart environments: Definition, measurement, prediction and control," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Wearable Comput.*, 2018, pp. 693–697.

[10] M. Rinta-Homi, N. H. Motlagh, A. Zuniga, H. Flores, and P. Nurmi, "How low can you go? Performance trade-offs in low-resolution thermal sensors for occupancy detection: A systematic evaluation," *Proc. ACM Interact. Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 3, p. 126, Sep. 2021.

[11] A. Zuniga, N. H. Motlagh, H. Flores, and P. Nurmi, "Smart plants: Low-cost solution for monitoring indoor environments," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 23252–23259, Nov. 2022.

[12] D. J. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intell. Syst.*, vol. 27, no. 1, pp. 32–38, Jan./Feb. 2012.

[13] S. Balandin and H. Waris, "Key properties in the development of smart spaces," in *Proc. HCI*, 2009, pp. 3–12.

[14] J. Mineraud, O. Mazhelis, X. Su, and S. Tarkoma, "A gap analysis of Internet-of-Things platforms," *Comput. Commun.*, vols. 89–90, pp. 5–16, Sep. 2016.

[15] F. J. Villanueva, O. Aceña, J. Dorado, R. Cantarero, J. F. Bermejo, and A. Rubio, "On building support of digital twin concept for smart spaces," in *Proc. IEEE Int. Conf. Human-Mach. Syst. (ICHMS)*, 2020, pp. 1–4.

[16] M. Maoz, "How IT should deepen big data analysis to support customer-centricity," Gartner, Stamford, CT, USA, document G00248980, 2013.

[17] E. Gilman et al., "Internet of Things for smart spaces: A university campus case study," *Sensors*, vol. 20, no. 13, p. 3716, 2020.

[18] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2020.

[19] H. Kokkonen et al., "Autonomy and intelligence in the computing continuum: Challenges, enablers, and future directions for orchestration," 2022, *arXiv:2205.01423*.

[20] "The smart building sensor for indoor environment monitoring." Accessed: Oct. 28, 2022. [Online]. Available: https://www.elsys.se/en/ers/

[21] R. Yasmin, J. Petäjäjärvi, K. Mikhaylov, and A. Pouttu, "On the integration of LoRaWAN with the 5G test network," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, 2017, pp. 1–6.

[22] R. Yasmin, J. Petäjäjärvi, K. Mikhaylov, and A. Pouttu, "Large and dense LoRaWAN deployment to monitor real estate conditions and Utilization rate," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2018, pp. 1–6.

[23] R. Yasmin et al., "Combining IoT deployment and data visualization: Experiences within campus maintenance use-case," in *Proc. 9th Int. Conf. Netw. Future (NOF)*, 2018, pp. 101–105.

[24] N. H. Motlagh et al., "Low-cost air quality sensing process: Validation by indoor-outdoor measurements," in *Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, 2020, pp. 223–228.

[25] B. Du, M. C. Tandoc, M. L. Mack, and J. A. Siegel, "Indoor CO2 concentrations and cognitive function: A critical review," *Indoor Air*, vol. 30, no. 6, pp. 1067–1082, 2020.

[26] N. H. Motlagh et al., "Indoor air quality monitoring using infrastructure-based motion detectors," in *Proc. IEEE 17th Int. Conf. Ind. Inform. (INDIN)*, vol. 1, 2019, pp. 902–907.

Naser Hossein Motlagh received the D.Sc. degree in networking technology from the School of Electrical Engineering, Aalto University, Espoo, Finland, in 2018.

He is a Senior Researcher with the Department of Computer Science, Helsinki Institute for Information Technology, University of Helsinki, Helsinki, Finland. His research interests include the Internet of Things, wireless sensor networks, environmental sensing, and unmanned aerial and underwater vehicles.

Martha Arbayani Zaidan (Member, IEEE) received the Ph.D. degree in automatic control and systems engineering from Sheffield University, Sheffield, U.K., in 2014.

He is a Senior Researcher with the Department of Computer Science and Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki, Finland. His research interests include AI/ML for intelligent control systems, health monitoring technologies, applied physics, atmospheric, and environmental sciences.

Lauri Lovén (Senior Member, IEEE), received the D.Sc. (Tech.) degree from the University of Oulu, Oulu, Finland, in 2021.

He is a Postdoctoral Researcher with the 6G Flagship Research Program, Center for Ubiquitous Computing, University of Oulu. He has worked as a Faculty Universitätsassistent with the Distributed Systems Group, TU Wien, Wien, Austria, in 2022. His current research concentrates on edge intelligence.
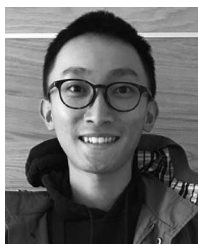
Pak Lun Fung received the Ph.D. degree in atmospheric science from the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, Helsinki, Finland, in 2022.

He is a Postdoctoral Researcher with INAR. His research interests include air quality proxy derivation and low-cost sensor measurements.

Tuomo Hänninen received the D.Sc. degree in telecommunications engineering from the University of Oulu, Oulu, Finland, in 2018.

He has been coordinating several applied research programs and activities with the University of Oulu, where he is a Research Director with the Centre for Wireless Communications. His research interests include IoT, spectrum sharing, critical communication, and energy and unmanned aerial vehicles.

Roberto Morabito (Member, IEEE) received the Ph.D. degree in networking technology from Aalto University, Espoo, Finland, in 2019.

He is a Senior Researcher with the Department of Computer Science, University of Helsinki, Helsinki, Finland. His research interests lie at the intersection of IoT, edge computing, and distributed artificial intelligence.

Petteri Nurmi received the Ph.D. degree in computer science from the Department of Computer Science, University of Helsinki, Helsinki, Finland, in 2009.

He is an Associate Professor of Distributed Systems and Internet of Things with the Department of Computer Science, University of Helsinki, Helsinki, Finland. His research interests include distributed systems, pervasive data science, and sensing systems.

Sasu Tarkoma (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Helsinki, Helsinki, Finland, in 2006.

He is a Professor of Computer Science with the University of Helsinki. He is a Visiting Professor with the 6G Flagship with the University of Oulu. He holds ten granted U.S. patents. He has authored four textbooks and has published over 250 scientific articles. His research interests include Internet technology, distributed systems, data analytics, and mobile and ubiquitous computing.