Deployment Strategies for Ultra-Reliable and Low-Latency Communication in Factory Automation

Nadia Brahmi[1], Osman N. C. Yilmaz[2], Ke Wang Helmersson[3], Shehzad A. Ashraf[1], Johan Torsner[2]
Ericsson Research
[1] Aachen, Germany; [2] Jorvas, Finland; [3] Linköping, Sweden
{nadia.brahmi, osman.yilmaz, ke.wang.helmersson, shehzad.ali.ashraf, johan.torsner}@ericsson.com

Abstract— **Factory automation is one of the challenging use cases that the fifth generation, 5G, networks are expected to support. It involves mission-critical machine-type communications, MTC, with requirements of extreme low-latency and ultra-reliable communication to enable real-time control of automation processes in manufacturing facilities. In this paper, we discuss the deployment strategies for the 5G mission-critical MTC solution designed to meet the needs of factory automation applications. The paper analyzes the coverage and capacity aspects based on a series of system-level evaluations considering both noise-limited and interference-limited operations. It further analyzes the related trade-offs to provide insights on the network deployment strategies for a realistic factory scenario.**

Keywords—5G; Machine-type communication; Factory automation; Latency; Reliability; Coverage; Capacity; Deployment.

I. Introduction

While the fourth mobile generation has emerged as a response to the growing demand for mobile broadband traffic, the fifth generation (5G) is envisioned to drive the networked society vision where everything that benefits from being connected will be connected [1][2][3]. To achieve this vision, machine-type communication (MTC) will play a key role to support a variety of new use cases with diverse requirements and challenges. Based on these requirements, the MTC uses cases will range from massive MTC to mission-critical MTC. Massive MTC includes applications related to the Internet of things where a large number of low-cost and low-energy devices are connected and controlled via the wireless network. Mission-critical MTC, on the other hand, includes applications with the requirements of extreme low-latency and ultra-reliable communications to enable real-time control and automation of dynamic processes in various fields, such as industrial process automation and manufacturing, energy distribution, and intelligent transport systems. Hence, 5G should be developed to cope with all these new use cases satisfying a broad range of requirements and expectations.

The purpose of the paper is to further analyze the concept developed in [4] and [5] for mission-critical MTC focusing on the factory automation use case. It includes demanding applications, such as motion control and closed loop control applications for manufacturing purposes. Therefore, the in-time delivery of mission-critical messages and high transmission reliability (robustness) are very important to avoid interruptions in manufacturing processes.

Wired technologies, such as fieldbus and Ethernet-based systems, have been for a long time in use for factory automation and they still hold the major market share because of their reliability, availability and real-time guarantee. However, the interest for using wireless communications for factory automation has grown recently, thanks to the advantages that they bring in terms of flexible deployment and easy maintenance and thus reducing the overall cost of the communication network. To fully exploit the advantages of a wireless solution, it is important to evaluate the coverage and capacity demands for the assumed use case. Understanding the aspects of coverage and capacity will help devising suitable deployment strategies for a wireless network that needs to fulfil the reliability and latency requirements. Therefore, we present the paper with the following contributions:

- Analysis of the system coverage and capacity based on realistic modeling of factory automation,
- Trade-off analysis of the system deployment with respect to e.g., diversity order, number of cells, cell size and frequency planning.

The rest of the paper is structured as follows. Section II describes the selected factory automation scenario and its requirements with respect to reliability, latency and coverage. Section III gives the details of the simulation modeling and assumptions. The simulation results for coverage and capacity analysis are presented in Section IV with the detailed trade-off analysis in order to define deployment strategies. Finally, Section V concludes the paper and summarizes the main findings based on the simulation results.

II. Scenario Description and Requirements

The data traffic in factory automation can be categorized into three different types: i) Real-time cyclic (RTC) traffic, ii) Real-time acyclic (RTA) or sporadic traffic, and iii) Best effort (BE) traffic:

i. RTC data occurs typically in control loop applications; e.g., a logic controller evaluates periodic input data collected from the sensors to set commands for the actuators, accordingly.

ii. RTA data occurs for instance, if (sporadic) alarm signals are issued.

iii. BE data is transmitted during file transfers for machine configuration or for software updates.

Both RTC and RTA traffic types have to meet the real-time requirements, thus the data has to be transmitted within a specified latency bound. Therefore, the rest of the paper will focus on the industrial applications with stringent real-time traffic requirements, such as closed-loop control applications.

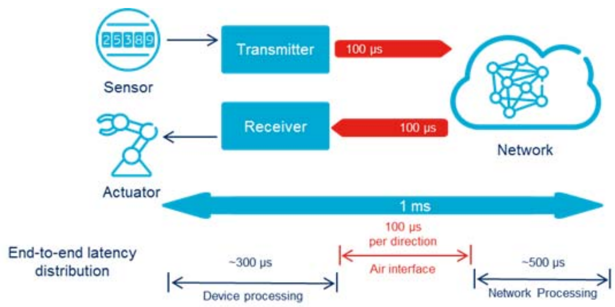

Fig. 1. Distribution of the end-to-end latency assuming a 100-μs air interface delay [7].

Closed-loop applications are usually executed at the lowest control system level (field level) by a master control unit and a set of slaves (devices) typically arranged in a logical star topology within a local area of few tens of meters, also known as automation cell. There can be multiple automation cells in close proximity to each other. The traffic follows a cyclic data processing model and depending on the controlled process the cycle may vary from hundreds of microseconds to hundreds of milliseconds. During one cycle each device has to transmit its data to the logical controller which has to process all the input data and calculate the new output data for the next cycle. Typical data sizes range from few bytes to few hundreds of bytes.

As an example, we assume that each 100-bit packet should be transmitted within a 1-ms end-to-end (e2e) latency budget. We aim at 100 μs for over-the-air transmission time which corresponds to 10% of the e2e latency budget as shown in Fig. 1. Furthermore, we consider a reliability requirement down to 10^{-9} as it is needed by some factory automation applications [6]. We note that the reliability is defined as the capability to deliver a data packet to the receiver within the delay bound and it is measured by the packet error rate (PER).

To evaluate the efficiency of the wireless deployment we define the coverage metric as the geographical locations or areas where the system is able to guarantee the reliability and the latency requirements. In addition, the capacity is evaluated in terms of load metric which refers to the percentage of frequency resources utilized for guaranteeing the reliability and the latency requirements at the system level.

III. Performance Evaluation Methodology

Diversity is considered as the key enablers of ultra-reliable communications. As shown in [4], large diversity orders are necessary to allow acceptable fading margins in case of fading channels, such as Rayleigh channel. Theoretically, this diversity could be achieved on the time, frequency and/or space domain. In order to achieve ultra-reliable communication within the strict low latency budget, exploiting the time diversity seems to be very challenging. On the other hand, to exploit the gains from frequency diversity, it is important to map the coded bits on the frequency resources having uncorrelated channel coefficients. Therefore, the required bandwidth would increase with the coherence bandwidth of the channel and thus making the exploitation of frequency diversity more bandwidth consuming. Therefore, we fully rely on the antenna diversity to achieve the required diversity order.

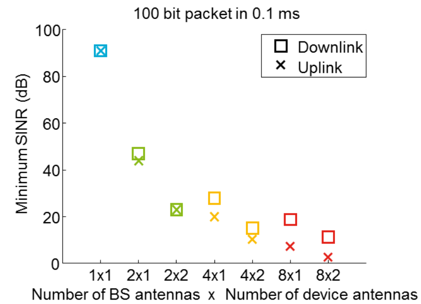

Fig. 2. Minimum required SINR to achieve the reliability and latency requirements for different antenna configurations assuming maximum ratio combining and QPSK modulation with a ½-rate convolutional code.

As mentioned in Section II, the traffic in factory automation comprises of small data packets in the order of few hundreds bits. For such small sizes, convolutional codes perform similar to the other complex low-density parity-check codes or turbo codes. Moreover, convolutional codes have an advantage of lower decoding complexity which results in lower processing delays [4]. Further details on the OFDM-based physical layer design and link-to-system modeling can be found in [4] and [5], respectively.

Fig. 2 shows the required signal-to-interference-and-noise ratio (SINR) for QPSK using a ½-rate convolutional code and different antenna configurations [4]. It is observed that increasing the diversity order decreases the required SINR. Moreover, the SINR requirements come out to be lower in uplink as compared to the downlink due to processing gains on the receiver side.

TABLE I. Simulation Assumptions

Packet size	100 bits
Average packet rate	0.1 packet/ms (per machine)
Machine density	0.5 device/m^2
PER	10^{-9}
Carrier frequency	2.4 GHz & 5.2 GHz
Uplink/Downlink bandwidth	From 1.5 MHz to 100 MHz
Air interface delay	100 μs (per direction)
Min. link distance	5 m
Transmit power BS / Device	30 dBm / 20 dBm
Antenna gain BS / Device	5 dBi / 0 dBi
Number of device antennas	[1 2]
Number of BS antennas	[2 4 8]
Noise figure BS / Device	5 dB / 9 dB
Modulation	QPSK, [16 64 256]-QAM
Coding	Convolutional coding with code rate ½ and constraint length 7
Overhead	30%
Fading model	**Slow fading:** Shadow fading with STD: 6 dB, Corr. distance: 10 m **Fast fading:** Rayleigh fading channel

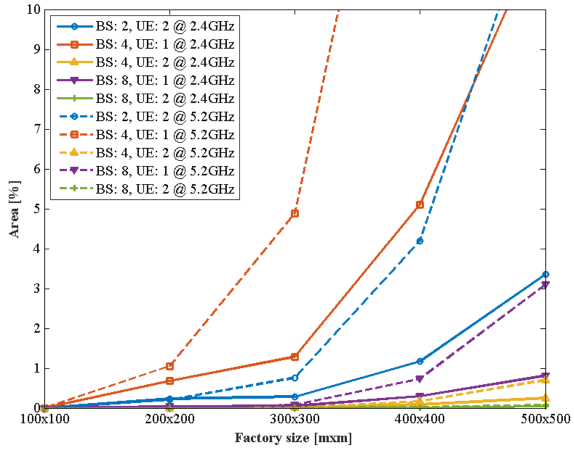

Fig. 3.a. Downlink out-of-coverage area assuming a single-cell deployment.

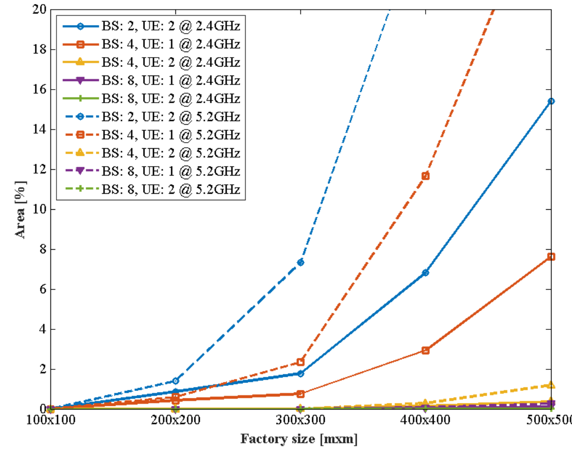

Fig. 3.b. Uplink out-of-coverage area assuming a single-cell deployment.

Coverage and capacity analysis is performed using a system-level simulator for which the main parameters are summarized in Table 1. In the performance evaluations, 2.4 GHz and 5.2 GHz are used as examples of carrier frequencies in the range of typical spectrum bands that could be used for factory automation. The simulations are based on path loss models for indoor factory hall measured in [8].

IV. Deployment Strategies and Capacity Analysis

The radio network planning and deployment should be done in such a way that the coverage is guaranteed for all the devices within the factory hall and the sufficient capacity is made available for the applications running during the manufacturing process. This section discusses different deployment options for the factory automation scenario described above.

A. Coverage

As mentioned in Section II, the coverage is defined as the geographical locations or areas where the system is able to guarantee the reliability and latency requirements. This means that in each of these locations the device should receive a signal with an SINR higher than the minimum required value to achieve 10^{-9} PER within a 100-μs air interface delay in the selected factory automation scenarios. It is assumed that device positions are fixed and cannot be adapted to the radio conditions. Therefore, the deployment should guarantee the coverage needed to ensure safe and smooth operation of mission-critical applications.

As shown in Section III, the minimum required SINR largely depends on the antenna configuration for the given packet size, transmission format, reliability requirement and latency budget. Therefore, we may have different SINR requirements for the downlink and uplink communications. For instance, the transmission of a 100-bit data packet with QPSK modulation and ½ code rate, assuming antenna diversity with one base station (BS) and eight device antennas, would need around 1.5 MHz bandwidth provided that the received SINR is greater than or equal to 7.2 dB and 18.5 dB SINR for the uplink and downlink, respectively. If the number of transmit and receive antennas on the device side are increased to two, the minimum required SINR is decreased down to 2.7 dB and 11 dB for the uplink and downlink, respectively.

1) Single-cell deployment

A coverage analysis is made based on a finite number of samples in a set of factory floor layouts assuming a 0.5-m resolution. Different propagation models, line-of-sight and non-line-of-sight with heavy obstructions, and different carrier frequencies, such as 2.4 GHz and 5.2 GHz, are simulated based on the realistic propagation models for factories presented in [8]. A geographical point or location is considered out of coverage when the received SINR is below the minimum required SINR.

Fig. 3.a and Fig. 3.b show the percentage of area that is out of coverage for different factory floor layouts and antenna configurations in downlink and uplink, respectively. The simulations reveal that the coverage within the factory hall is limited by the downlink in particular due to the lower-order antenna receive diversity (i.e., less antennas at the device) compared to uplink. In addition, for typical factory floor layouts e.g., from 100 m x 100 m to 300 m x 300 m, a single BS placed at the center of the hall is sufficient to guarantee the full coverage (100%) assuming a bandwidth of around 1.5 MHz (for 100 bits) and a diversity order of eight.

It is important to note that the system bandwidth will be larger to be able to support simultaneous transmissions. More details on the capacity analysis will be discussed in the following subsection. Furthermore, it is shown that if we increase the number of device antennas from one antenna to two antennas, 99% of a 400 m x 400 m factory floor can be covered, yet in this case it is important to make sure that the devices do not move into the 1% of the area that is of no coverage. For medium and large factory halls the coverage is limited, therefore we need to add more BS's (i.e., multi-cell deployment).

2) Multi-cell deployment

For a multi-cell deployment, two options could be considered:

i. Frequency planning and coordination where the cells may use at least partially separate frequencies,

ii. Frequency reuse where neighboring cells fully operate on the same frequency bands or resources.

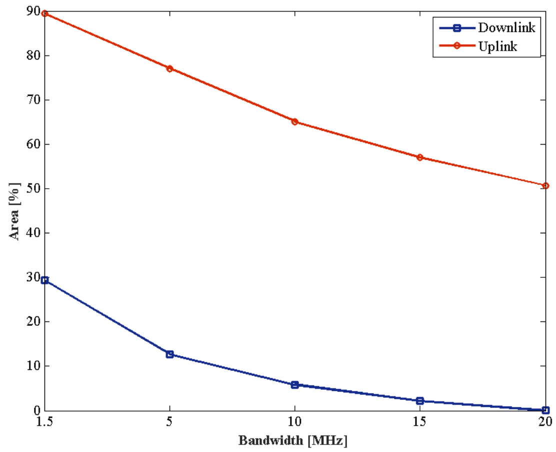

Fig. 4. Out-of-coverage area assuming a 2-cell deployment with frequency reuse-1 at 5.2 GHz.

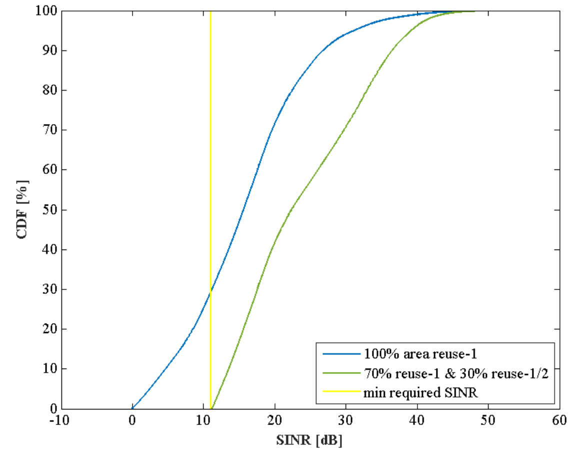

Fig. 5.a. Downlink SINR CDF of frequency partitioning where 70% of the area operates on reuse-1 and 30% on reuse-1/2.

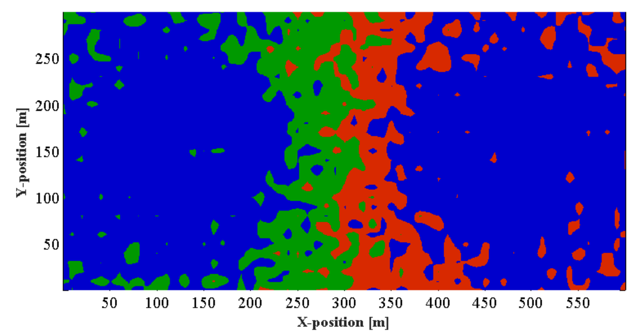

Fig. 5.b. Illustration of downlink frequency partitioning assuming a 2-cell deployment where 70% of the area operates on reuse-1 (blue region) and 30% on reuse-1/2 (green and red regions).

With separate frequency channels used for each cell, the bandwidth required for the full coverage will depend on the factory hall size and the number of BS's. For instance, if we need four BS's to cover the factory hall then the total amount of bandwidth needed to guarantee the full coverage may be quadrupled (i.e., 6 MHz) with a noise-limited operation. On the other hand, the analysis of the simulation results for a 200 m x 200 m factory floor shows that with 8x2 antenna configuration (i.e., eight BS and two device antennas) and without interference, the worst-coverage case devices have around 10 dB SINR gain compared to the minimum required SINR. This can be used either for reducing the transmission time (with higher modulation order) and thus increasing the number of supported devices (i.e., the system capacity); or for reducing the diversity requirements by lowering the number of antennas.

With the same frequency channels used in all cells (i.e., frequency reuse-1), the simulations, which is performed for a 600 m x 300 m factory floor assuming a 2-cell deployment at 5.2 GHz, show that the interference is larger than the noise and the system is interference limited thereof. Since the power spectral density of interference depends on the bandwidth, increasing the bandwidth helps lowering the interference level and thus increasing the SINR level up to the minimum required SINR. However, it is found out that the total amount of bandwidth needed to compensate the effect of the interference is very high. For instance, assuming the worst case interference without power control, even with a bandwidth of 20 MHz we have less than 50% coverage in uplink (see Fig. 4). On the other hand, if we could use frequency planning, full coverage could be easily achieved with only a 3-MHz bandwidth (i.e., 2 x 1.5 MHz). Based on this observation, the frequency planning is considered beneficial and it helps reducing the amount of bandwidth needed for the system compared to the reuse-1 case. However, this observation is only valid assuming a perfect control of the interference from adjacent systems. More precisely, if the system is not designed to operate with the frequency reuse, it becomes very sensitive to other systems being deployed independently by other factory owners. In this case, the system might not provide the robustness and reliability required by the mission-critical applications. Therefore, it is crucial to provide techniques that can address inter-cell interference (ICI) and ensure full coverage.

Different approaches have been proposed in the literature, e.g., [9] and [10], to optimize the robustness of the system and mitigate the effect of interference between neighboring cells. For the factory automation scenario, we selected frequency partitioning [10] due to its simplicity and the limited exchange of information needed between the neighboring cells. It is based on a device categorization into cell-edge devices and cell-center devices. This categorization could be done during the initial planning and deployment phase or periodically based on the measurements performed by the devices, such as the ratio between the desired signal power and the interference generated by other cells. The cell-edge devices of each cell are restricted to transmit in certain parts of the frequency band so that the transmissions do not overlap with the frequency resources used by the cell-edge devices of the neighboring cells. Since the devices that are close to the cell-center generate less interference to other cell-center devices of the neighboring cells, this scheme allows reuse-1 for some frequency resources used for the cell-center devices.

Fig. 5.a illustrates the potential coverage gain, which could be achieved with the frequency partitioning in downlink for a 600 m x 300 m factory floor (at 5.2 GHz) depicted in Fig. 5.b. In fact, the out-of-coverage area could be decreased from 30% (blue curve in Fig. 5.a) to 0% if the neighboring cells could ensure, by means of coordination, that the cell edge devices (located at 30% of the area) are transmitting using separate

frequencies. The rest of the devices located at the cell cell-centers (i.e., 70% of the area) reuse the same frequency without any coordination signaling between the neighboring cells. The same effect can also be obtained using an intelligent scheduling scheme where the neighboring cells fully coordinate in order to schedule the cell edge devices on separate time instants while still reusing the same frequency. In addition, there is a large potential to enlarge the reuse-1 region by applying transmission power control algorithms and/or other coordination techniques.

B. Capacity

The system capacity of the wireless network largely depends on the application requirements (e.g., reliability target and delay budget), the traffic characteristics (e.g., message length and interval), as well as the traffic distribution and density. Therefore, the capacity evaluations should carefully address these variables for the selected factory automation scenarios.

To analyze the capacity boundaries, we consider a single-cell deployment with the antenna configuration of two device antennas and eight BS antennas. We assume 100 MHz bandwidth with a 5.2 GHz carrier in order to provide full coverage within the respective setup as discussed in the previous section for the coverage analysis.

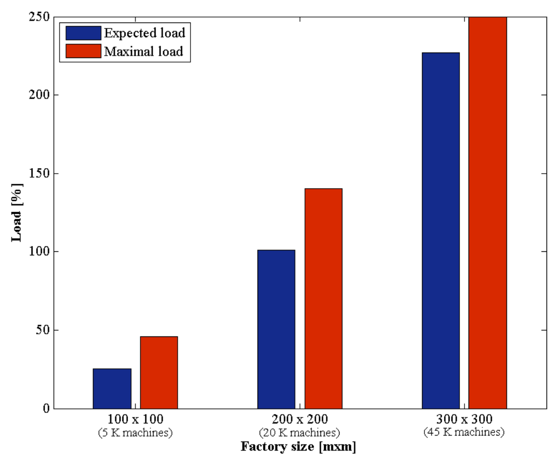

Fig. 6.a. Downlink system capacity when interference-mitigated environment is assumed for different factory floor layouts.

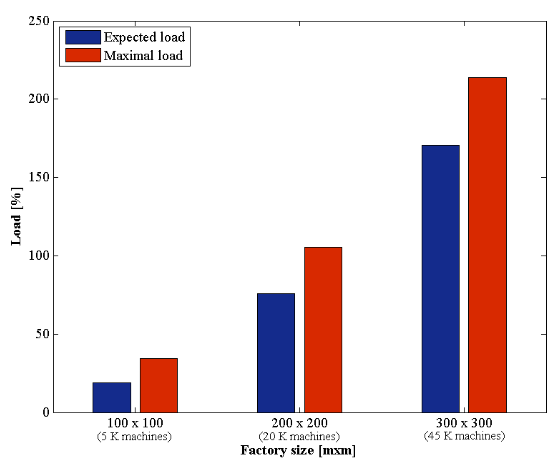

Fig. 6.b. Uplink system capacity when interference-mitigated environment is assumed for different factor floor layouts.

In the simulation scenario, industrial automation machines (devices) are uniformly spread out on the factory floor. For the evaluations at least one device is always assumed to be located at the worst coverage location (i.e., the lowest SINR), yet all the devices can potentially be located at the bad-coverage locations e.g., due to mobility, which is taken into account for maximal load computations. Furthermore, we assume that the industrial application requirements dictate that only one message out of one billion packets may be lost or delayed within a 100-µs transmission time interval. The capacity evaluations follow the simulation assumptions for the traffic generation, such as the density of the devices, packet size and rate, as given in Table I. For the packet arrival modeling, Poisson process, commonly used in radio traffic simulations, e.g., in [11], is assumed provided that the required statistical confidence is met for the evaluation of rare events e.g., large number of simultaneous transmissions.

Fig. 6.a – Fig. 7.b illustrate the load measures, which denote what percentage of the frequency resources is needed to serve the factory automation machines reliably for downlink and uplink, respectively. In the figures, expected load and maximal load refer to the mean and largest capacity demands respectively depending on the packet arrival and channel condition.

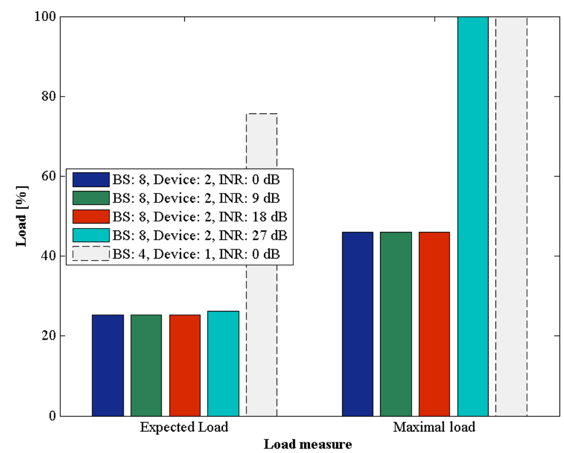

Fig. 7.a. Downlink system capacity when interference-present environment is assumed for a 100 m x 100 m factory floor.

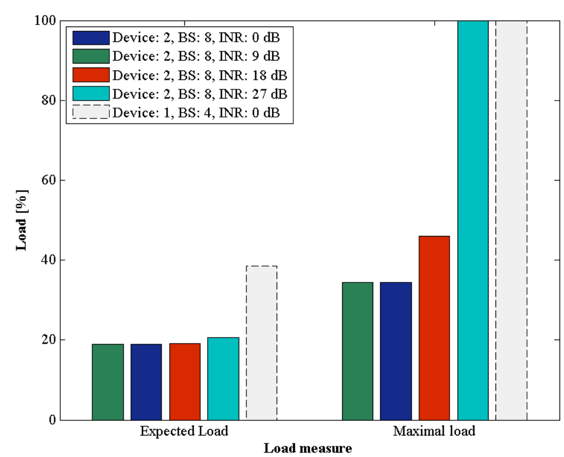

Fig. 7.b. Uplink system capacity when interference-present environment is assumed for a 100 m x 100 m factory floor.

As shown in Fig. 6.a and Fig. 6.b, when the factory floor area is expanded assuming an interference-mitigated scenario, the expected load increases nearly proportionally to the number of the machines. This means that the system capacity is mainly impacted by the traffic demand rather than the factory floor dimensions, in the range of 100 m x 100 m and 300 m x 300 m, as the assumed antenna diversity is able to guarantee the coverage robustness for the selected factory automation scenarios. It is also observed that the capacity of the single-cell deployment is sufficient to serve 5 K devices spread over a 100 m x 100 m factory floor, yet it becomes rather challenging when the number of devices is 20 K or more. Therefore, a larger system bandwidth allocation in both downlink and uplink may be needed to support denser scenarios i.e., 20 K devices or more within a single cell coverage.

Fig. 7.a and Fig. 7.b depict how large the interference, in terms of the interference-to-noise ratio (INR) as discussed in [5], may impact the system capacity in downlink and uplink, respectively. Despite the expected capacity loss in the presence of interference, it was found that a high-order antenna diversity, which comprises two device antennas and eight BS antennas in this particular example, not only improves the coverage but also extends the capacity and keeps the system robust against unexpected or unmitigated interference sources; therefore we highlight the diversity as an essential tool also for the capacity optimization and robustness.

V. CONCLUSIONS

In this paper, we have looked at the factory automation as one of the challenging mission-critical use cases targeted by future cellular networks, 5G. It has been shown that with a proper physical layer design exploiting diversity gain it is possible to guarantee ultra-reliable communications with extreme low-latency down to sub-millisecond. With such a design, full coverage can be provided for a 300 m x 300 m factory floor. For larger factory halls, where more base stations need to be deployed, interference becomes a limiting factor with reuse-1. Simulations have shown that if it is possible to keep the deployments under control, partial frequency reuse or frequency separated system can be more spectral efficient as it requires less bandwidth than a system in which the same frequency channels are fully reused among neighboring BS's. Capacity evaluations have shown that it is possible to serve nearly 20 K devices with a reasonable antenna configuration and a bandwidth allocation. The system capacity is mainly affected by the diversity and the system bandwidth. Furthermore, to improve the system availability in terms of coverage and capacity, interference management techniques (e.g., ICI coordination) could be used.

REFERENCES

[1] A. Osseiran, et al., "The foundation of the Mobile and Wireless Communications System for 2020 and beyond Challenges, Enablers and Technology Solutions," In proceedings of IEEE Vehicular Technology Conference (VTC2013-Spring), Dresden, Germany, June 2013.

[2] METIS, Deliverable D1.1. Scenarios, requirements and KPIs for 5G mobile and wireless system, May 2013.
https://www.metis2020.com/documents/deliverables

[3] Ericsson, Networked Society Essentials, March 2013.
http://www.ericsson.com/res/docs/2013/networked-society-essentials-booklet.pdf

[4] N. A. Johansson, Y.-P. E. Wang, E. Eriksson, M. Hessler, "Radio Access for Ultra-Reliable and Low-Latency 5G Communications," IEEE International Conference on Communications (ICC) Workshop on 5G & Beyond - Enabling Technologies and Applications, June 2015.

[5] O. N. C. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, J. Sachs "Analysis of Ultra-Low Latency and Ultra-Reliable 5G Communication for a Factory Automation Use Case," IEEE International Conference on Communications (ICC) Workshop on 5G & Beyond - Enabling Technologies and Applications, June 2015.

[6] ETSI TR 102 889-2 V1.1.1, Electromagnetic compatibility and Radio spectrum Matters (ERM); System Reference Document; Short Range Devices (SRD); Part 2: Technical characteristics for SRD equipment for wireless industrial applications using technologies different from Ultra-Wide Band (UWB).

[7] ITU-T, The Tactile Internet, Technology Watch Report, August 2014.
http://www.itu.int/en/ITU-T/techwatch/Pages/tactile-internet.aspx

[8] E. Tanghe, W. Joseph, L. Verloock, L. Martens, H. Capoen, K. V. Herwegen, and W. Vantomme, "The industrial indoor channel: Large-scale and temporal fading at 900, 2400, and 5200 MHz," IEEE Transactions on Wireless Communications, July 2008.

[9] A. Hamza, S. Khalifa, H. Hamza, K. Elsayed , "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," IEEE Communications Surveys & Tutorials, March 2013.

[10] S. S. Han, J. Park, T.-J. Lee, H. G. Ahn, K. Jang, "A new frequency partitioning and allocation of subcarriers for fractional frequency reuse in mobile communication systems," IEICE Transactions on Communications, August 2008.

[11] O. N. C. Yilmaz, J. Hämäläinen, S. Hämäläinen, "Self-optimization of random access channel in 3rd Generation Partnership Project Long Term Evolution," Wiley Wireless Communications and Mobile Computing, December 2011.