6G: The Personal Tactile Internet - And Open Questions for Information Theory

Gerhard P. Fettweis[1], *Fellow, IEEE*, Holger Boche[2], *Fellow, IEEE*
[1]Vodafone Chair TU Dresden, and Barkhausen Institute, Centers 6G-life & CeTI & 5G++Lab Germany & EKFZ
[2]Dept. Electrical and Computer Engineering, TU Munich, Centers 6G-life & CASA & MCQST

The initial vision of cellular communications was to deliver ubiquitous voice communications to anyone anywhere. In a simplified view, 1G delivered voice services for business customers, and only 2G for consumers. Next this also initiated the appetite for cellular data, for which 3G was designed. However, Blackberry delivered business smartphones, and 4G made smartphones a consumer device. The promise of 5G is to start the Tactile Internet, to control real and virtual objects in real-time via cellular. However, the hype around 5G is, again, focusing on business customers, in particular in the context of campus networks. Consequently, 6G must provide an infrastructure to enable remote controlled mobile robotic solutions for everyone – the Personal Tactile Internet. Which role can information and communication theory play in this context, and what are big challenges ahead?

Index Terms — **communication control co-design, trustworthiness, joint communication & sensing, resilience, Gearbox PHY**

I. PERSPECTIVE

For over 100 years, mankind has been dreaming of personal mobile robotic helpers that make our everyday lives more comfortable. There are many examples: As a child: Someone who cleans up our room or is an additional interactive companion. As an adult: Someone who helps with our daily chores, is a friendly companion, and cleans our room or carries the shopping. As a senior: Someone who makes life at home possible and delays the move to assisted living space.

This requires, based on innovations in electronics, a maturity in robotics technology that is foreseeable within the next 10 years. However, the missing piece is the wireless networking infrastructure to enable sensing and orchestration of control functions. That is the challenge to be tackled by 6G. Building on the previous five generations of mobile communications, we currently are in the process of learning what is missing for this vision. From 2030 on, with the launch of 6G, the old dream of mankind of robotic helpers easing our life will therefore become reality – the Personal Tactile Internet.

So far, every 10 years a new generation of mobile communications has been introduced. Here is a highly simplified overview that provides an insight into the trend of the application innovations.

The first application was voice telephony. With 1G, the analog networks as e.g. AMPS and NMT, telephony was made possible for professional customers. With 2G, e.g. GSM, mobile telephony for end-users (consumers) was heralded. Then came the innovative leap to mobile data transmission. With 3G, this was mainly introduced for professional customers (e.g. "Blackberry" and Nokia "Communicator"). With 4G came the broad acceptance of the smartphone for end-users. With 5G we are facing the introduction of the Tactile Internet [1], the innovative leap in enabling wireless remote control of collaborative robotics applications over the mobile network. Today this new remote-control and XR application is being introduced (in campus networks) for professional applications. It is revolutionizing e.g. the agricultural and health sectors, as well as the construction, logistics, and manufacturing industries. I.e. 5G is, just like 1G and 3G, a mobile phone technology whose new service mainly addresses professional users.

Note: It seems that every odd-numbered generation first "practices" a new leap innovation with professional applications before the following even-numbered generation makes it a mass application for end users.

The 5G learnings must be the basis for enabling 6G – the basis for broadening from 5G's professional to 6G's personal/consumer (collaborative) robotic and virtual helpers. In this paper we therefore first analyze 5G and some of its limitations to understand where performance limits need to be stretched far beyond the current capabilities. Then, we analyze requirements of 6G applications to understand new functionalities and new features necessary.

Already now a flurry of 6G vision papers has appeared, covering different aspects of how the new standard shall improve beyond 5G, as e.g. [2]-[7]. Here, we want to focus on a possible vision for 6G that specifically touches open challenges and questions which should be addressed by the information and communication theory community.

II. Obvious Improvements Over 5G

A. Data Rate

We see data rates increasing by 100x every 10 years, or doubling every 18 months. This follows Moore's Law of semiconductors [8][9]. For finding a reason why there continues to be a demand of increasing the data rate, let us focus on VR (virtual reality). Today we know that high resolution monitors should have 8K resolution, but a higher resolution seems less necessary. This is under the assumption of an approximately 30° horizontal and 17° vertical angle-of-view. A true 180°/180° VR experience, even under 1000x compression, therefore requires on the order of 1-10Gb/s streaming. A challenge, as VR glasses must be connected wirelessly. This example clearly shows a need for a further increase in rate.
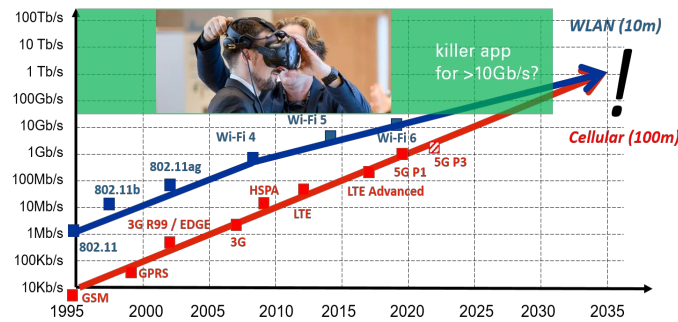

Fig. 1: The Wireless Roadmap, updated from [7]

Today's standard approach to achieve higher data rates is using classical linear modulation techniques like MIMO-OFDM and extending the cardinality of modulation as well as the number of antennas. This comes at a severe cost in terms of energy consumption per bit, not only due to requiring higher SNR but also due to reaching limits of the analog-to-digital converter (ADC).

⇨ This leads to the question: How can new ways be found to deliver the required service bandwidth by stepping beyond classic linear systems design, as e.g. by exploiting post-Shannon, and/or nonlinear methods supporting 1-bit ADCs (see Section IV)?

B. Latency

Many future Tactile Internet (distributed) remote-control and virtual or augmented reality "XR" applications require to run on tight and deterministic end-to-end latency constraints [1]. A packet arriving too early can always be delayed at the receiving end. However, a packet arriving too late is a lost packet. Hence, increasing the latency jitter can increase the packet loss rate.

Control systems acting on one or swarms of interactive mobile robotic devices (MRDs) can relax their bandwidth and latency requirements by orders of magnitude, using model predictive control (MPC), see Fig. 2 a&b. Stochastic MPC [10] has shown to be an effective measure to overcome stringent requirements and allow for systems to have probabilistic outcomes. An example of applying this in a wireless setting is platooning, where large gains are achievable to reduce the requirements on packet losses or latency constraints [11][12].

URLLC is costly, both from a capacity and energy point of view. Learning how to co-design communications and control from 5G based experience will be extremely beneficial.

⇨ Hence, it will be of utmost importance to learn from Tactile Internet applications which are being tested and rolled out within 5G within the next 5 years.

⇨ Only then will we be precisely able to learn which improvements on jitter, packet error rate, as well as latency are truly necessary for 6G.

⇨ What are key learnings? For sure we need experience with 5G applications to learn.

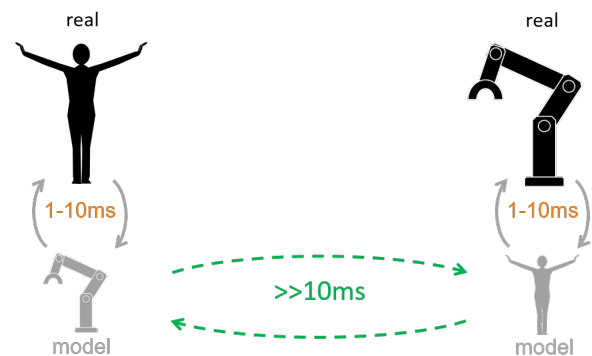Fig. 2a: Direct remote control of a robot with strict ultra-reliable low-latency (URLLC) requirements

Fig. 2b: Network controlled stochastic MPC for human-in-the-loop operations, enormously relaxing latency and reliability over the wireless link (green).

C. Consensus Processing

In the following, mobile robotic devices or XR devices will jointly be referred to as MRDs. A main objective of operating MRDs is to guide them safely and according to their intended purpose within their ambient environment and/or within a swarm/constellation of MRDs. Each MRD will therefore be equipped with many sensors to gather information of its ambient environment, however, it also will require information beyond its own sensing capabilities. I.e., installed fixed sensors and/or sensors on board of (surrounding) MRDs must be accessed to gain a full picture of the ambience and to prepare the trajectory and action of the MRD. This can be carried out at three levels of exchange of information:

1. Exchange of raw data, where the network and its edge cloud can also serve as a raw data collector and sink. The MRD then can access this data to analyze and confer its decision based on its own analytics capabilities.

2. Exchange of preprocessed data, e.g. the radar or spectroscopic imaging result of each sensor. It can be collected in the mobile edge cloud to generate an adaptive imaging map of the ambient scenario. Again, the MRD then can access this imaging data to analyze and confer its decision based on its own analytics capabilities.

3. Exchange of cognitive preprocessed sensor data, i.e. the signal processing analysis of each sensor has identified objects, e.g. a person, bicycle, or bird, and these objects and their position and motion trajectory are sent to the surrounding MRDs, e.g. via the edge cloud.

Levels 2 and 3 will heavily rely on machine learning (ML) preprocessing, and will be merged into a consensus finding stage of ML processing in an MRD to carry out its (motion) control. As we can expect a rising number of these Tactile Internet applications, we can expect that the exchange of ML-preprocessed objects over future 6G networks will start dominating the traffic load, see Fig. 3.
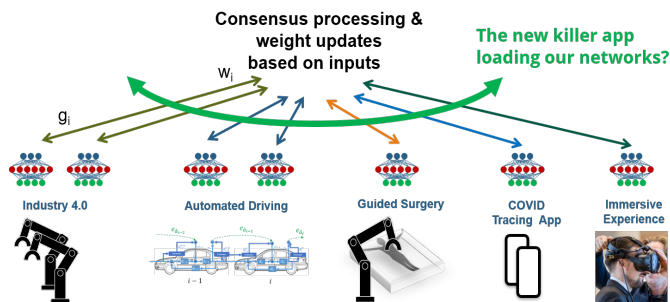

Fig. 3: The Tactile Internet applications require artificial intelligence (AI) based sensor processing and control of objects. Federated learning and consensus-based AI processing therefore is to be expected to become prominent, with the 6G network carrying the load for the exchange of preprocessed objects.

Therefore, many open (information and communication theory) questions need to be addressed, as e.g.:

⇨ What is the best tradeoff in terms of energy efficiency versus required network capacity? Does it make sense to use Level-2 for static environments and Level-3 for dynamic mobile objects, i.e. a combination of the two levels?

⇨ From a rate and energy point of view would it be better to consolidate the control consensus ML processing in the edge cloud or in the MRDs?

⇨ While targeting an acceptable sensing/imaging precision for an application, what is the tradeoff between more signal processing at the sensors resulting in less consensus based preprocessed sensor fusion necessary, or less sensor processing and more effort at the point of sensor fusion?

⇨ What if legal constraints are considered as well? How does this change boundaries and optimization criteria?

⇨ Can concepts presented below in Section III.C help in ensuring to minimize the energy and rate effort?

⇨ How does federated learning add to the complexity of the design challenge?

D. Extended Coverage

Every cellular network, so far, has increased the delivered data rate not only by improving the modulation and coding, but has heavily relied on improving the link budget, mainly by reducing the cell radius. Therefore, the number of base stations has continued to grow over time. For affluent nations or densely populated areas this is economically feasible. However, for nearly half of the world population and the dominant part of the earth's land surface this has posed a challenge. Delivering what

is defined as mobile broadband is a moving target, due to the exponentially increasing data rate (Section II.A). This requires a continued growth in number of installed base stations. As a result a large population of the earth has not been able to keep up in terms of mobile broadband internet access.

Hence, new methods must be found to overcome the digital divide between nations and regions [13].

⇨ Can satellite networks truly deliver broadband data rates, knowing that this is a moving target, increasing 10-fold every 5 years. Is this the most energy efficient approach? As satellites on their trajectory around the earth rarely fly past areas where their service is required, is this resource efficient and the environmentally best approach?

⇨ With massive MIMO on the horizon, can earth-based systems deliver a boost in link budget to deliver the gain needed for the continued data rate race? How can an enormous link budget increase be implemented via massive MIMO? Is the additional channel resource required for pilots to realize beam searching and beam tracking not countering the gain?

⇨ Can an information & communication theoretic framework be found for solving this challenge that includes synchronization, channel estimation, pilot resources, unequal service delivery over the earth's surface, as well as answering questions of sustainability?

E. Scalability

To cover more advanced applications it seems that with every generation of digital cellular the communications data rate is to be ever increasing, and latency requirements are becoming tighter. This can be nicely captured by a rate-latency plane, see Fig. 4 [14]. However, not every application requires the "top right corner case", i.e. the most stringent requirements to be available. This would then be an over-designed solution and lead to a cost and energy burden far and above the needed. Already in 4G and now in 5G this has been realized and subsets have been defined within this plane, as e.g. the existing standards "4G NB-IoT" or "5G RedCap".
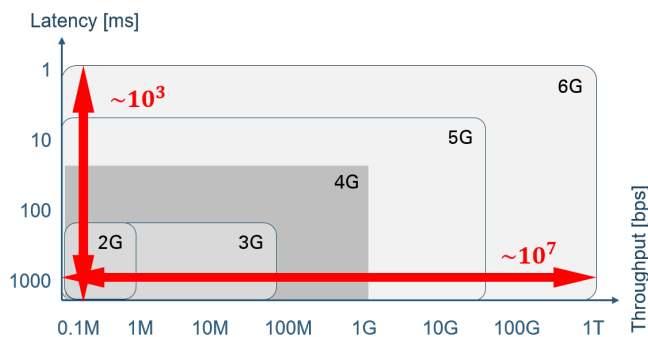

Fig. 4: Anticipated 6G rate-latency plane [14].

The rate-latency-plane anticipated for 6G covers 10 orders of magnitude in total, 7 on rate and 3 on latency; this cannot be implemented efficiently on one hardware, where efficiency is e.g. measured by cost and energy. It would be of utmost importance to have a scalable framework that can be parameterized to its requirement and implemented efficiently in

silicon by a "push-button". This way resource efficient solutions could be provided even for markets not large enough for a full-blown custom chip-set development. Hence:

⇨ How can we build a communications modem that is optimal under an information and communication theoretic condition for (nearly) each point in the rate-latency plane, but allows for a scalable implementation solution?

⇨ Are signal processing concepts that scale in terms of hardware and signal processing load without requiring recompilation of firmware a viable possibility, e.g. proposed in [14]? Or do we need to find concepts going far beyond?

To address this rate-latency plane from an even more extreme energy point of view, we propose the new concept of a "Gearbox PHY" in Section IV.

III. POSSIBLE TRUE 6G INNOVATIONS

We know that many applications of the Tactile Internet have extremely high latency and reliability requirements (URLLC: "ultra-reliable low latency communications"). 5G offers network slicing for this purpose to guarantee the required quality of service. However, this is available statically and must be requested from the network for the possible most demanding case, although moving objects rarely remain in the worst-case state. Adaptivity is required here. Also, we have learned from applications (Section II.B) that mobile networked control that uses model prediction and incorporates probabilistic stochasticity of outcomes, i.e. "stochastic model predictive control", leads to a significant reduction in requirements by at least an order of magnitude in control data rate, latency requirement, as well as packet error rate.

Herein lies the opportunity to relax requirements ("key performance indicators") to achieve the democratization of Tactile Internet applications, making it available for consumer end-customers. Joint research on solutions, such as communications control codesign, can uncover enormous potential here. This requires joint optimization of AI-based control algorithms as well as communications technology.

A. Trustworthiness

In 1450 J. Gutenberg's invention of letterpress printing from Mainz revolutionized society and heralded 300 years of renaissance. Due to the production rate of the machine, information could be distributed regionally at the speed of a horse-drawn carriage. Savonarola and others invented fake news and populism, Luther was able to spread his ideas, and the monopoly of knowledge of the monks was overthrown [15].

In 1989 the World Wide Web was invented in Geneva, sometimes referred to as "digital printing". Not only is the rate of producing copies and speed of distribution now determined by electronics, but any person can "print digitally" by "posting" (mis)information and distribute it instantly to the world population. We are at the beginning of a "second Renaissance".

The biggest challenge is, as in the 1st Renaissance, the loss of trustworthiness. Restoring this must be understood as a basic societal challenge. Technically, it requires a solution based on the interlocking of processor hardware, operating system, and radio communication, under application-specific real-time latency constraints. To anchor devices of the Internet of Things

(IoT) in a "trustworthy" way requires new procedures and paradigms [16], which will have to be extended by AI and quantum-based approaches in the future. Based on this, the entire network architecture for 6G must be newly developed.

A current common understanding of security and privacy research is to uphold these two security properties against outside attacks, as outlined in [17]. However, we must not only design systems that are robust against attacks from the outside, but also from within! For this goal we have to develop new information theoretic tasks like oblivious transfer, information masking, and secure computing [18]. Quantum Communication in combination with classical communication and these new information theoretic tasks offer additional advantages for achieving trustworthiness under real world communication conditions and quantum computing based on quantum computing or any other future computing technology. We must be able to deliver systems according to spec, even when subsystems have been included from sources which might not be fully trusted:

⇨ Can we fully trust that engineers are developing code to spec and not going beyond while obeying outside interests?

⇨ What are theoretical measures of trust? How can we design systems that can be certified to measurable trust levels?

⇨ Can we ensure that online checking happens, i.e. we continuously automatically monitor the code base and recertify such that we can guarantee that the code is designed to spec and no other functionality is supported?

B. Joint Radio Communications & Sensing

A key 6G technology is the symbiosis of radio communication and radio sensor technology. The use of Joint Communication & Sensing (JC&S [19]) in deploying mobile robotics with 6G will revolutionize entire industries and enable completely new end-user applications. This can send innovation shock waves through (end-user) markets, such as hobby, sports and fitness, gaming, kitchen appliances and white goods, mobility, logistics, gardening equipment, and tools.

Mobile robots and XR, "MRDs", must use radio-wave based sensing (e.g. radar, spectroscopy, and positioning) to detect their environment in 3D, and they must communicate with each other via a radio interface. At anticipated MRD densities today's radar spectrum is neither sufficient for stepping up towards 3D radar, nor can it deliver the reliability and precision levels required for future applications (including drones, autonomous cars, etc.), nor can it deliver the spectroscopy-based services. However, the spectral efficiency of radar and of radio sensor technology could be improved significantly beyond the state of the art if it were coordinated in the radio access. The mobile radio spectrum, on the other hand, is not sufficient for the 100Gb/s data transmission rates envisaged for the future. Both, sensing as well as communications, require further spectrum allocation. Today, sensing and communications have separate/orthogonal assigned spectra. The good news is that large radio resources are rarely simultaneously needed for broadband communication and radio sensing. This suggests the idea of running both communication and sensing together in one RF system/spectrum, referred to as "JC&S" [19]. It also opens the possibility of the implementing only one transceiver

system and its antennas.

The development of JC&S requires expertise in radio sensing and communications. This leads to the question: how can radio sensing technologies be integrated "as a service" into the 6G RAN (radio access network) jointly with "communications as a service" by using a new suitable radio access system design? For radar, e.g., the static as well as dynamic environment must be captured in a 3-dimensional space, not as a constant process, but dynamically to the needs. Radio terminals are activated on request to improve accuracy and capture dynamics. This would enable physical maps with *cm*-accuracy and temporal resolution in the *ms*-range - a significant step beyond Internet maps of today. Many new applications can be envisioned that allow for entrepreneurism. Even bird flight could be detected and drones could avoid collisions to ensure bird protection.

The same conceptual idea can be applied/translated to spectroscopy, enabling spectroscopic 3D maps: Not only could environmental pollution be detected and located. Based on material recognition, MRDs could make a distinction between e.g. leaf swirling in the wind and animals crossing the road, and the MRD's trajectory is then controlled accordingly.

The technical basis for this vision is to design one radio interface for JC&S. The RAN must combine sensing as well as data packets in a joint radio access protocol. If possible, without bearing the costs of a full-duplex transceiver. The hardware and signal design challenge is to achieve extremely high transmit-receive isolation without resorting to highly directive antennas, as they create a challenge when optimizing for coverage. JC&S requires a joint design optimization including the antennas, the RF transceiver, the digital signal processing and modulation, and the MAC (radio access protocol).

The key point is that MRDs need to detect their ambient environment via radio-based sensing and need to connect to each other to "see around corners" and gain additional insight. Today, radar, spectroscopy, and geolocation as well as communications is mostly solved via separate systems. With the latest developments in the standardization of UWB (IEEE 802.15.4z) we can see first steps of integration to a joint physical radio system for communications and sensing.

New research approaches enable developing a new 6G RAN enabling JC&S. In this system, communication and sensing can be "called as a service". A first set of open questions is:

⇨ How can we define capacity-performance for active/passive radar imaging? Does distortion theory allow us to understand the energy-performance tradeoff?

⇨ What are the "best" modulation techniques for communications in a framework where in particular pilot signals must serve for channel state estimation, passive radar, active radar, spectroscopy, and positioning?

⇨ We are used to understanding the transmission channel for data communications. How is the "reflection channel" for radar and spectroscopy defined? How correlated is it with the transmission channel, as both are influenced by an overlapping set of reflectors?

⇨ How to define the "passive radar" channel [20] and how does this fit into the picture?

⇨ How can we make use of learnings on the "reflection channel" to optimize e.g. maximum-a-posteriori guided

massive MIMO beam searching, reaching far beyond ideas for improving beam searching in a classical way (e.g. [21])?

Remark: Again, the transition from an odd to an even generation cellular requires a fundamental redesign of the physical layer: 1G → 2G from analog to digital/CDMA; 3G → 4G from CDMA to OFDM; 5G → 6G OFDM to what?

C. Extending Beyond Shannon

The transmission of messages is determined by the theory of Shannon. The transmission of status states is considered "beyond Shannon" and obeys other theories with higher capacity potential. Exploiting this and embedding it in 6G must be the aim.

Post-Shannon information theory has developed initial coding methods for the transmission and storage of data, e.g. for identification as defined in [22], which achieve exponential gains compared to the Shannon and Turing approaches and thus have much better scaling behavior in terms of necessary energy and hardware components. According to the analysis of SRC [17], the data generation rate grows faster than the data transmission rate, and the amount of data generated is already larger than the amount of data that can be transmitted. For the identification problem post-Shannon information theory can be used to develop new transmission methods that have much better scaling behavior of the amount of data that can be transmitted than Shannon's approach, which means that, on the one hand, the trend between the amount of data generated and the amount of data that can be transmitted observed in [17] no longer occurs and, on the other hand, latency can be significantly reduced. Post-Shannon transmission methods achieve further enormous gains in terms of achievable transmission rates compared to the classical Shannon approach to message transmission through additional resources such as quantum entanglement, common randomness, synchronization and feedback. This can further reduce latency. Moreover, quantum entanglement and shared randomness enable full compensation of active jamming attacks on the 6G system for both Shannonian message transmission and the first post-Shannon transmission schemes studied [22], thus achieving resilience by design. This is particularly interesting since the successful execution of jamming attacks by an attacker cannot be detected by Turing machines and thus not by digital hardware and protocols [23].

Besides, these initial post-Shannon transmission methods allow a secure transmission of information, which cannot be broken even by quantum computers of arbitrary complexity. It is particularly interesting to note that the respective optimal transmission rates for secure post-Shannon transmission methods are equal to the optimal transmission rates without security requirements. Thus, the post-Shannon transmission schemes already studied achieve security by design and one does not pay a price for security in terms of rate performance [23]. Furthermore, these techniques can be combined with

oblivious transfer, information masking, and secure computing to achieve trustworthiness by design. The introduction of quantum technologies in combination with new communication tasks offer additional gains in terms of rates, security, complexity, trustworthiness, new services, and functionality.

⇨ The potential of information theory needs to be unleashed to improve energy efficiency for communications, protocols and hardware platforms, reduce latency, achieve post-quantum security [24] and resilience by design.

⇨ This will require the development of entirely new methods for transmission and computing and corresponding protocols. Furthermore, methods for distillation of quantum entanglement and of classical information such as shared randomness must be developed. These new communication resources must be researched and developed as part of post-Shannon transmission methods.

⇨ The aim must be to couple the development of quantum communication networks with 6G development so that, on the one hand, quantum communication can meet practically relevant latency and resilience requirements and, on the other hand, quantum communication can later be integrated into the 6G network.

D. Integrity and Resilience

Today's cellular networks operate stably as everyone naively assumes that no adverse opponents will blind a base station or radio terminal with "radio guns". So far only bits and bytes could be harmed which limits the criminal energy on attacks. If "mission-critical" robotic helpers can be disturbed, this looks different. We need to develop 6G in such a way that an integrity management system ensures resilient radio access [25]:

1. *Prepare* for a possible attack by implementing measures of robustness, as e.g. use multiple antennas with physically separated beams (sectors) as well as multi-connectivity.
2. Permanently *monitoring* the condition of the network to identify possible interferers, as e.g. radio scanning to identify and locate possible interferers.
3. *Adapt* according to the situation and remain able to deliver (basic) services, e.g. by quickly reconnecting to a different base station or different antenna beams/sectors.
4. *Counter* by initiation of possible countermeasures, such as RF spoofing of the interfering object in such a way to attract its jamming target away from its true target, or jamming the interferer such that it cannot connect to its control point.

This presents a very large challenge, and very little research has been carried out so far for resilient system design for commercial systems. Commercially available standardized systems whose specifications are public are posed with a major challenge ahead. A jammer can synchronize to a system and intelligently focus its energy on one or few symbols of a packet. Some first examples of open questions to be addressed are:

⇨ What are information theoretic methods and frameworks allowing deepen the understanding of the system problem?

⇨ What are possible bounds on system performance?

⇨ How can this insight be exploited to find answers to design system solutions?

First insight on research in this direction is e.g. in [26][27].

TABLE I
THE "FIGURE OF CHALLENGE"

Cellular Gen.	Bandwidth	A/D resolution per I/Q dimension	# MIMO multiplexing streams	A/D res. per I/Q dim,
2G	0.2 MHz	6 b → 64 qt	1	26M qt/s
3G	5 MHz	8 b → 256 qt	2	5G qt/s
4G	20 MHz	10 b → 1K qt	4	164G qt/s
5G	200 MHz	12 b → 4K qt	8	13T qt/s
Av. Incr.	*10x*	*4x*	*2x*	*80x*
⇒ 6G*)	2 GHz	14 b → 16k qt	16	7P qt/s

The "Figure-of-Challenge" defined as the total number of quantization bins per second. This is calculated by: (I-phase + Q-phase) × (bandwidth) × (ADC in quantization bins: qt) × (# MIMO transceivers). It is measured in quantization bins per second (qt/s). The numbers assumed here are typical numbers at time of the cellular standard's launch, where different chip implementations can clearly differ from the numbers given.

*) Extrapolation

IV. GEARBOX PHY

Every five years the data rate in mobile communications increases tenfold (Section II.A). The resulting tenfold increase in power consumption could previously be counteracted by densifying the base station deployment and exploiting the tenfold increase in the energy efficiency of electronics every five years. Both are no longer possible to the same extent. In addition, the "softwareization" within 5G networks is replacing hardware accelerators with general purpose computers with a power consumption that is orders of magnitude higher [17]. In addition, radio access must be redesigned to address both, improve coverage and reduce exposure. Hence, we have an energy challenge ahead, which becomes even worse when analyzing the analog-to-digital converter (ADC) requirements.

As we will move towards 100Gb/s and beyond, the ADC cannot deliver the scaling in power consumption needed [17][28]. Table I lists a very approximate view on the ADC development (terminal implementation) and shows the increase in decision bins "qt" per second needed (qt/s), referred to as the ADC "Figure-of-Challenge".

It is important to note that ADC sampling rates f_s of 5G have reached, for the first time, the 200-300 MHz "knee" of the ADC figure-of-merit [28]. This means that increasing f_s beyond 300 MHz comes at a high price: the ADC's power consumption increases not linearly, but with f_s^2 – a dramatic change! When extrapolating the requirements for 6G, see Table I, f_s is projected to reach 2GHz, far beyond the knee, which leads to a 100-fold power consumption increase by the ADC alone. Combining this with the projected parameters shown in Table I results in nearly 3 orders of magnitude increase in ADC power consumption. This would drive the total ADCs power consumption (assuming 1 fJ/qt [25]) from around 10mW (best case) for current 5G to a staggering 10W for 6G (at 10 fJ/qt).

The improvement of the ADC figure-of-merit defined in terms of "energy per qt" has been approximately 38-fold over the last decade [25]. However, the "knee" at 200-300 MHz has

not nudged over the last 12 years. Note, for the purpose of simplifying the problem description, we here discuss the easier unit qt than the "decision thresholds" that are found in [25].

Projecting another (unlikely) 38-fold advance to be available for 6G, due to the f_s^2 behavior, the 3 orders-of-magnitude challenge mentioned above would reduce to 25x, still making the ADC the dominant contributor to the transceiver's power consumption. It will become an energy bottleneck [17].

A. Power Versus Data Rate

Let us now try understanding the dominant power consumption contribution to a transceiver implementation when trying to increase the data rate *R*. Recalling Shannon's simplified capacity formula for the special case of gaussian signaling, a MIMO channel with nearly equal Eigenvalues, and a frequency-flat AWGN channel:

$$C = M \cdot W \cdot ld(1 + m \cdot SNR), \text{ where}$$

C is the channel capacity;
M is the number of antennas used for MIMO multiplexing;
m is the number of antennas used for beamforming gain;
W is the channel bandwidth.

TABLE II
THE IMPACT OF PARAMETERS ON RATE (AWGN CASE)

	Parameter	Low *SNR*	High *SNR*
1.	$SNR \rightarrow R$	$P_S \sim R$ $P_{ADC} \sim 2^R$	$P_S \sim 2^R$ $P_{ADC} \sim 2^R$
2.	$m \rightarrow R$	$P_{ADC} \sim 2^R$	
3.	$M \rightarrow R$	$P_{ADC} \sim R^2$	$P_{ADC} \sim R^2$
4.	$W \rightarrow R$	$P_{ADC} \sim R^2$	$P_{ADC} \sim R^2$
(4.) ($f_s \leq 200$ MHz)	$W \rightarrow R$	$P_{ADC} \sim R$	$P_{ADC} \sim R$

Assumption: $f_s \geq 200$ MHz is at or above the "knee", except for last row.

It is well-known that the above capacity formula also has a similar curve progression in fading channels [29]. Without loss of generality we can therefore examine the AWGN case which can be approximated for two corner cases:

High SNR (lower bound): $C = M \cdot W \cdot ld(m \cdot SNR)$
Low SNR (upper bound): $C = M \cdot W \cdot m \cdot SNR$

For simplicity let us assume the rate of communications *R* to be at capacity $R = C$. This allows for drawing the following conclusions on the power consumption impact of increasing the rate *R* by changing each of the four parameters *SNR*, *m*, *M*, *W* individually, see Table II cases 1. – 4.

1. When considering the SNR alone, an increase in transmit power PS results in which rate increase, and what is the impact on the power consumption PADC of the ADC due to a rate increase via an exponential increase in modulation cardinality?

2. What is the power consumption impact of increasing the rate R by increasing the massive MIMO analog beamforming gain m? The improved the link budget is then exploited by increasing the bit-resolution of the ADC by R.

3. What is the energy impact of R by exploiting multipath propagation by adding M MIMO antennas for achieving rate increase through multiplexing M data streams? As long as the channel has full rank, the rate increases linearly with M. However, the cardinality of the summed transmitted signal space increases with M, and therefore the ADC resolution of each of the M transceivers also needs to increase by ld(M) bits (best case: if all eigenvalues of the channel matrix had equal weight).

4. What is the energy impact of changing W? Up to a value of W = 200 MHz this resulted in

$$W \sim R \quad \Rightarrow \quad P_{ADC} \sim R$$

This is good news, as long as W remained below the "knee". However, once a sampling rate exceeds 300 MHz, this changes dramatically to a quadratic dependency.

The result is interesting. Every decade we have enjoyed generations of digital cellular standards with 100x increase in data rate from 2G to 3G, and 3G to 4G, as well as 4G to 5G, without paying too much cost in increased power consumption. The reasons are that we have:

1. Increased the link budget by reducing the cell size;
2. Exploited advancements of semiconductor technology.in energy improvements;
3. Exploited the fact that ADC sample rates were below "the knee" ($\leq$ 200 MHz), and that the ADC energy consumption in fJ/qt has improved by 38 per decade below "the knee";
4. Exploited innovations in power efficient circuit design/implementation.

There has been a fine balance in place to manage the power challenge which has led to the parameter set in every column of Table I, to carefully orchestrate a beautiful and balanced next generation solution every time.

Going forward, for 6G the situation is different: Cell sizes can hardly be reduced, and the semiconductor roadmap does not promise the same prospects in reducing ADC power consumption. And finding new ideas on further circuit improvements are becoming a challenge.

⇨ Hence, are we caught in the power challenge, just as semiconductor chips?
⇨ For more than a decade, the resulting heat challenge due to power consumption per unit chip area has stopped processor clock rates from increasing, referred to as "the power wall". Is an equivalent power wall hitting 6G cellular for *R*?

Of course, this discussion is based on link analysis only and not assuming any further analog and RF impairments. We know that wireless networks are interference limited, and cell-edge performance depends on experienced interference and achievable link budget, resulting in the measured SINR. To overcome this, coordinated multipoint [30] has been introduced, which however does increase the linearity requirement of the RF/analog receiver as well as the ADC resolution requirements! Hence, detailed research on building a theoretic framework is necessary to build a thorough understanding of the impact of methods and implemented ideas on the power consumption of networks.

Summarizing the results of Table II, increasing R by changing *SNR, m, M, W* we are faced with the big challenge: at least a quadratic power consumption dependency on the rate increase incurred by the ADC. Therefore, it is of utmost importance to focus on the ADC challenge.

B. The Analog to Digital Interface

Historically, our system design has been based on the understanding of linear system theory. We sample at a rate f_s to adhere to the Nyquist sampling theorem, and quantize such that the quantization noise is negligible when compared to the interference-plus-noise floor, hence hardly impacting the SINR.

On the one hand advanced CMOS semiconductor processes enable designing circuits that have a very high time resolution due to possessing a cutoff frequency f_T between 300GHz and 400GHz [31]. Time resolution circuits at $\frac{1}{4}f_T$ are easily viable, i.e. amounting to 75-100 GHz time resolution. On the other hand, amplitude resolution is becoming more challenging, as the threshold voltage remains to be at $0.2 - 0.4$ V and the supply voltage has been reduced over time. Time resolution RF transceiver design [32] therefore could be a solution. What could this imply for the future ADC interface?

The most energy efficient ADC per quantization bin "qt" is the 1-bit converter. This particular ADC only detects zero crossings, and therefore little-to-no power-consuming AGC (automatic gain control) circuit needs to be implemented. To shed some light onto the efficiency of this design let us assume 4-fold oversampling with respect to bandwidth with a 1-bit ADC, which generates 8 quantization bins (8 qt). Due to band limitation of the input signal only 8 possible value constellations of the 4-bit long sequence are possible. A 3-bit quantized ADC also has 8 qt, with 8 possible values. However, in contrast to the 1-bit oversampled ADC, the 3-bit ADC has to possess a linear input amplitude range of operation. Therefore, generating qt in time versus qt in amplitude results in a very significant difference in circuit energy consumption.

From an information theoretic rate perspective, oversampled 1-bit conversion of band-limited signals gives room for deriving fascinating results [33][34]. However, one must understand that the spectral efficiency is compromised in relation to the SNR point of operation. It is clear, but too often forgotten, that high spectral efficiency has its price in reduced energy efficiency, as discussed in the previous subsection. And, if we project the availability of future spectrum allocation at mm-wave frequencies (reaching even beyond 80 GHz), spectral efficiency is less of an issue as energy efficiency.

⇨ Therefore, could focusing on energy efficiency be a solution and therefore possibly allowing for new nonlinear modulation techniques e.g. requiring only 1-bit conversion and oversampling?

⇨ But for which point of operation in terms of data rate delivery is this a good idea?

C. Possible Way Forward: The "Gearbox PHY"

It becomes clear from the above analysis that for achieving maximum energy efficiency we must focus on rate versus spectral efficiency [5]. If we have abundant spectrum and are in a range of 0.1 bit/s/Hz and below, impulse modulation with a 1-bit ADC will be best in terms of energy efficiency [35]. There is a crossover point at which an increase in spectral efficiency

TABLE III
A POSSIBLE GEARBOX PHY SCENARIO

Spectral Avail-ability:	Abundant	Low	Medium	High
PHY/ "Gear":	e.g.: Impulse Radio	e.g.: ZXM or CPM	e.g.: MIMO-OFDM	e.g.: MIMO-OFDM or OTFS [40] or WH-OFDM [41]

Basic idea of changing the complete transceiver chain depending on the point of operation in the category plane spectral efficiency to achieve maximum energy efficiency. This could be extended by a second dimension to a matrix by including the time/frequency channel selectivity. How to categorize the gears, and which "gear" i.e. which box to use is an open research topic.

leads to other methods as e.g. ZXM (zero crossing modulation [34]) outperforming impulse modulation. Then again, a further increase in spectral efficiency leads to continuous phase modulation, and finally, to QAM.

Jointly with the *SINR* and *m*, as well as the number *M* of MIMO transceivers, the points change when to cross over from impulse modulation to ZXM to CPM to QAM. Within each of these classes of modulation, each "gear" of a "Gearbox PHY", adaptive modulation and coding [36] is to be applied. However, at these crossover points more fundamental changes should happen, as the complete transceiver must change. We refer to this as a "change of gear". Depending on the channel state characteristic and its time/frequency variation (coherence time, coherence bandwidth, delay spread, and Doppler spread) the crossover point of when to "switch gear" will have to adapt.

Taking this idea forward, now a medium access control (MAC) protocol needs to be designed. Today the physical air interface (PHY) is designed around the maximum data-rate operation point. A typical MAC is to subdivide a time-frequency resource grid into "resource blocks" (RBs) which are allocated to users to fulfil the communication requests. However, this entails that in case of a not fully loaded system, many unused resource blocks are kept free, and used RBs are operated at a spectral efficiency beyond the necessary value, resulting in a loss in energy efficiency.

A conventional way to overcome this is to use CDMA spread spectrum. However, this requires a full-resolution ADC at full system bandwidth. Instead, the Gearbox PHY idea is to fully switch the transceiver to the "gear" of operation and only use the bandwidth and hence spectral efficiency which provides the best rate versus energy tradeoff.

The reason why this approach could be a major step in advancing towards energy efficiency is due to two points:

1. A fully loaded system is a rare event over a 24/7-time schedule.
2. An area pixel in a cell which receives maximum rate is a rare pixel in a cell.

For understanding the second point let us recall the path loss in a cellular system to find out how often a good link is achievable that allows for delivering high data rates. There is a clear gap between the typical theoretical deployment in a hexagonal cellular grid and reality, see e.g. [38]. Also new

measurements for industrial indoor settings confirm this observation [37]. Hence, in reality a terminal seldom experiences a link budget good enough to achieve peak rates. What can we learn from this? This finding can be translated to a new way of designing the PHY.

Massive MIMO base station antennas can address "pixels" in a coverage area of a cell (coined here "cell-pixel"). Depending on the SINR in every cell-pixel and depending on service requirements, i.e. communication data rate, sensing function, or identification function, it is possible to optimize the PHY to match the local load and SINR. In particular, this can be done such that the energy-optimal "gear" is engaged.

We are used to exploiting orthogonality, splitting the radio access in time, frequency, and with massive MIMO we extend this to space within each cell, this we refer to as a MAC-pixel. However, little has been done to not only pack service requirements (e.g. differentiated according to the rate-latency plane) into different MAC-pixels. But this should also be done according to the link budget and the MAC-pixels' local SINR. E.g. pixels with a low rate-demand and low spectral load should use spike modulation that achieves way better energy performance [39]. Today we optimize complete RANs for the few "pixels" where high rates are achievable.

Coordinated by a MAC that controls MAC-pixel access, the resources in a radio cell can be adjusted to use the energetically optimal choice of modulation and multi-antenna transmission. In contrast to today's 5G Adaptive Modulation & Coding [36] where only the cardinality of the QAM constellation and error correction coding is adapted, the Gearbox PHY shall achieve large energy efficiency through changing gears, notwithstanding using adaptive modulation and coding within each gear. This way adaptivity includes the whole transceiver including the resolution of A/D conversion as well as antenna count. It requires a joint system design, spanning from radio planning over information/ communications theory all the way to hardware and MAC design.

D. Five key learnings can be made:

1. The cellular network installed must be designed to carry the peak traffic. However, when looking at the 24/7 weekly traffic, the peak occurs only during a couple hours a day. Hence, a typical load is far lower and therefore the modulation should be adapted to this.
2. When introducing massive MIMO, the average cellular area pixel carries far less load than peak pixels. Optimizing modulation and coding for heavily loaded pixels which require very high spectral efficiency results in a very suboptimal choice for the majority of other pixels.
3. As we are approaching mm-wave spectrum above 80GHz, and if the new spectrum can be used jointly for communication and sensing, vast new amounts of spectrum can be used. Hence, instead of today's approach of using an OFDMA MAC that leaves vast amounts of spectrum unused when traffic demand is low, we shall rather adapt the modulation and coding to less spectrally efficient schemes to think in terms of Shannon: A smaller spectral efficiency can lead to high gains in energy efficiency! And, we should adapt the modem hardware accordingly as its ADC requirements

can be relaxed to achieve a large gain in hardware energy efficiency as well!

4. Hence, the idea might be to move from a "gear" of spectrally very inefficient spike modulation during very low traffic & sensing demand [35], to a gear of ZXM during low demand, to continuous-phase modulation and to QAM and MIMO-OFDM as demand increases. This way the analog-to-digital interface is matched to energy optimization under semiconductor constraints and opportunities, exploiting the high cutoff frequency f_T of current CMOS technologies. Not mentioning the adaptation and change of gears to address the wide range of differing sensing service needs. And not to mention the necessity of gears to adapt to very different channel situations, possibly even requiring OTFS or WH-OFDM [40]-[42].
5. Spectral efficiency comes at the cost of paying additional energy, and is only needed for those MAC-pixels that experience heavy traffic demands. And, in view of the mm-wave spectrum becoming available, energy efficiency is the priority and not spectral efficiency anymore.

V. Bigger Picture & Conclusions

Every odd generation of cellular technology has provided mankind with a new level of connected service. Every even-numbered generation has democratized this, making the new service available to the broad consumer base, see Fig. 5. However, the consumer was never unprepared. Parallel to 1G, cordless telephony became popular and prepared the consumer to see the benefit of untethered voice communication anywhere anytime. Parallel to 3G, WiFi became popular and convinced the consumer of untethered Internet access. Parallel to 5G, UWB standardization is currently getting prepared to specify joint communication and sensing for local applications, i.e. untethered gesture detection and ambient sensing.
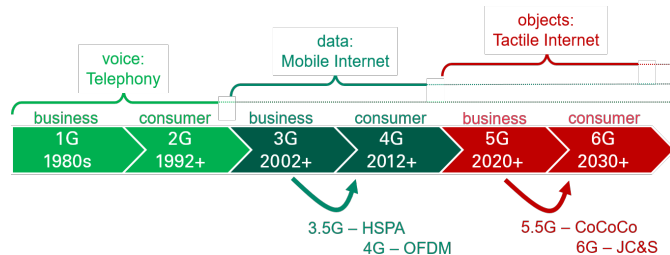

Fig. 5: The simplified cellular generation roadmap. CoCoCo: communications control codesign to relax the latency requirements for consumer Tactile Internet applications.

It is also to be noticed that cellular standardization is a continuous process, with updates being released typically every year, with generational changes approximately every 10 years. After 5 years of experiencing the new application step of odd generations a "half generation update" typically enabled a first roll-out for consumers, which then were happily ready for the next generation to appear to make the application truly successful. Hence, we can expect to see e.g. resilience features as well as communications control codesign to be specified for "5.5G" as an updated 5G standard, but the new topics discussed

above as the Gearbox PHY with JC&S to have to wait for 6G.

More than half a century after Shannon's fundamental papers we still have a wide-open area of untouched research challenges when addressing the open issues on the way towards 6G. Alone enabling the Tactile Internet for consumer applications and addressing the energy challenge ahead under semiconductor technology and circuit restrictions gives us a seemingly unbounded number of open challenges. Interestingly, these mostly cannot be solved by mathematical understanding alone, but require to consider the reality of networks and of hardware. I.e., true systems understanding is a challenge of the future for information theory's role towards 6G.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] G. Fettweis, "The Tactile Internet: Applications and Challenges." *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, March 2014, pp. 64-70.
[2] Oulu 6G paper repository: https://www.oulu.fi/university/news/6g-white-papers-2020
[3] "Samsung's 6G White Paper Lays Out the Company's Vision for the Next Generation of Communications Technology," https://bit.ly/2C4ZKNi, July 14, 2020.
[4] X. You, et al., "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," in Science China, Information Sciences, January 2021, Vol. 64, doi 110301:1–110301
[5] S. Buzzi, C.-L. I, et al. "A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead," IEEE Journal on Selected Areas in Communications, Volume: 34, Issue: 4, April 2016, pp. 697-709.
[6] *6G The Next Horizon*, W. Tong and P. Zhu, eds., Cambridge, 2021.
[7] M. Uusitalo et al., "European Vision for the 6G Network Ecosystem," White Paper by the 5G Infrastructure Association, to be published at EuCNC 2021 and at https://5g-ia.eu/, June 2021.
[8] Reducing Emissions in Our Operations. 2021. N.p.: Vodafone, https://www.vodafone.com/our-purpose/planet/reducing-emissions-in-our-operations.
[9] G. Fettweis and S. Alamouti, "5G - Personal Mobile Internet Beyond What Cellular Did for Telephony," IEEE Communications Magazine, Volume: 52, Issue: 2, February 2014, pp. 140 – 145.
[10] A. Mesbah, "Stochastic Model Predictive Control: An Overview and Perspectives for Future Research." IEEE Control Systems Magazine, vol. 35, no. 6, December 2016, pp.30-44.
[11] A. Gonzales, et al., "String Stable CACC under LTE-V2V Mode 3: Scheduling Periods and Transmission Delays." in Proceedings of IEEE 2nd 5G World Forum - 5GWF, 2019.
[12] A. Villamil, A. González, et al., "Feedforward Filter Design for CACC with N-Vehicle Look Ahead: A Frequency Domain Approach for Further Optimization of Radio Resource Usage," In 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), pp. 1-5.
[13] *Information and Communication Technology for Development for Africa: Second International Conference,* F. Mekuria, E. Nigussie, T. Tegegne eds., Vol. 1026, Springer 2019.
[14] G. Fettweis, E. Matus, "Scalable 5G MPSoC Architecture," in *51st Asilomar Conf. on Signals, Systems, and Computers*, 2017. pp. 613-618.
[15] I. Goldin and C. Kurtana, "*Age of Discovery: Navigating the Storms of Our Second Renaissance,*" 2nd edition, Bloomsbury, September 2017.
[16] G. Fettweis, T. Hentschel, P. Grünberg, "Das neue Internet muss aus Ankern des Vertrauens bestehen," published by Atlantik Brücke, in German, https://www.atlantik-bruecke.org/das-neue-internet-muss-aus-ankern-des-vertrauens-bestehen/

[17] J. Ang, D. Apalkov, F. Assaderaghi, et al., Decadal Plan for Semiconductors – Full Report, Semiconductor Research Corporation (SRC), Jan. 2021, url: www.src.org/about/decadal-plan/.
[18] U. Pereg, Ch. Deppe, et. al. "Classical State Masking over Quantum Channels", arXiv:2109.12647, January 2021.
[19] *Proceedings of the 1st IEEE International Online Symposium on Joint Communications & Sensing*, February 2021, IEEEXplore, https://jcns-symposium.org/
[20] H.D. Griffiths, C.J. Baker, "*An introduction to passive radar*," Artech House; 2017.
[21] M. Khalili, et al., "Evaluation of Detection Accuracy and Efficiency of Considered Beam Alignment Strategies for mmWave Massive MIMO Systems," in *Proc. of the Asilomar Conference* 2021.
[22] R. Ahlswede and G. Dueck, "Identification via Channels," in *IEEE Trans. Information Theory*, vol. 35, no. 1, Jan. 1989, pp. 15-29.
[23] J. A. Cabrera Guerrero, H. Boche, et al., "6G and the Post-Shannon Theory," in *Shaping Future 6G Networks: Needs, Impacts and Technologies*, Wiley, Dec. 2021.
[24] J. Renner, S. Puchinger, and A. Wachter-Zeh, "LIGA: A cryptosystem based on the hardness of rank-metric list and interleaved decoding," Des. Codes Cryptogr. Vol. 89, April 2021, pp. 1279-1319.
[25] "Disaster Resilience: A National Imperative," Washington, DC: *The National Academies Press*, 2012. doi: 10.17226/13457
[26] H. Boche, R.F. Schaefer, H.V. Poor, "Denial-of-Service Attacks on Communication Systems: Detectability and Jammer Knowledge," in *IEEE Trans. on Signal Processing*, Vol. 68, May 2020, pp. 3754-3768-
[27] N. Franchi, T. Hößler, et al., "Selected Aspects and Approaches on Improving Dependability in Industrial Radio Networks," in *Wireless Networks and Industrial IoT: Applications, Challenges and Enablers*, Springer International Publishing, 2021, pp. 21-38.
[28] B. Murmann, "ADC Performance Survey 1997-2020," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html.
[29] M.S. Alouini and A. Goldsmith, "Capacity of Raileigh Fading Channels Under Different Adaptive Transmission and Diversity-Combining Techniques," *IEEE Trans. Veh. Techn.*, Vol. 48, No. 4, pp 1165-1181.
[30] *Coordinated Multipoint in Mobile Communications*, P. Marsch and G. Fettweis, eds., Cambridge University Press, 2011
[31] J. Watts et al., "RF-pFET in Fully Depleted SOI demonstrates 420 GHz FT," in *Proc. of IEEE RFIC*, 2017, pp. 84-87.
[32] R. B. Staszewski, "Digitally Intensive Wireless Transceivers," *IEEE Design & Test of Computers*, vol. 29, no. 6, December 2012, pp. 7-18.
[33] S. Shamai, "Information rates by oversampling the sign of a bandlimited process," *IEEE Transactions on Information Theory*, Volume: 40, Issue: 4, Jul 1994, pp. 1230 - 1236.
[34] G. Fettweis, M. Dörpinghaus, et al., "Zero crossing modulation for communication with temporally oversampled 1-bit quantization." In *53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 207-214.
[35] F. Roth, M. Dörpinghaus, et al., "Rate Analysis of a Spike-Based Signaling Scheme for Energy-Efficient Low Rate Communications," in 2021 *1st IEEE International Online Symposium on Joint Communications & Sensing (JC&S)*, Feb. 2021.
[36] A. Goldsmith, "Adaptive Modulation and Coding for Fading Channels," in Proc. *1999 IEEE Information Theory and Communications Workshop*.
[37] F. Burmeister, N. Schwarzenberg et al., "Measuring Time-Varying Industrial Radio Channels for D2D Communications on AGVs," *IEEE Wireless Communications and Networking Conference*, March 2021.
[38] P. Marsch, S. Khattak, et al., "A Framework for Determining Realistic Capacity Bounds for Distributed Antenna Systems," *2006 IEEE Information Theory Workshop,* DOI: 10.1109/ITW2.2006.323698
[39] M. Win, R. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Transaction on Communications*, vol. 48, no. 4, 2000, pp. 679-689.
[40] R. Hadani, S. Rakib, et al., "Orthogonal Time Frequency Space Modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6.
[41] R. Bomfin, et al., "A Robust and Low-Complexity Walsh-Hadamard Modulation for Doubly-Dispersive Channels," in *IEEE Communications Letters*, Nov 2020, DOI:10.1109/LCOMM.2020.3034429.
[42] R. Bomfin, M. Chafii, et al., "A robust baseband transceiver design for doubly-dispersive channels," in *IEEE Transactions on Wireless Communications*, Mar 2021, pp. 1-16.