

AIMM

CELTIC-NEXT AIMM Project

WP3: Antenna Arrays and Reconfigurable Intelligent Surfaces

D3.2

Authors: Arman Shojaeifard, Deepa Jagyasi (InterDigital), Stephan ten Brink, Marc Gauger, Florian Euchner (University of Stuttgart), Norbert Schmidt, Daniel Martini (IMST), Fraser Burton, Yangyishi Zhang, Paul Botham (BT), Mark Beach, Zhihan Ren, Wei Wang (University of Bristol), Anton Menshov (CEMWorks)

Project Acronym: **AIMM**
Project Full Title: **AI-enabled Massive MIMO**
Project Coordinator: **Arman Shojaeifard (InterDigital)**
Project Duration: **24 months (Oct. 2020 – Sep. 2022)**
Submission Date: **30 September 2022 (M24)**
Dissemination Level: **External**

Abstract

This report provides a summary of activities from the work of the CELTIC-NEXT AIMM project Work-Package 3 (WP3) on “Antenna Arrays and Reconfigurable Intelligent Surfaces”. These include all results obtained on topics of work within WP3, namely on “Antenna Array Geometries”, “Reconfigurable Intelligent Surfaces”, and “Power Amplifier Enhancements”. Dissemination activities and future plans are also captured in this deliverable.

Disclaimer

This document contains material, which is copyright of certain PARTICIPANTS and may not be reproduced or copied without permission. The information contained in this document is the proprietary confidential information of certain PARTICIPANTS and may not be disclosed except in accordance with the regulations agreed in the Project Consortium Agreement (PCA).

All PARTICIPANTS have agreed to full publication of this document.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the PARTICIPANTS nor CELTIC-Plus warrant that the information contained in the report is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Executive Summary

This report describes the activities that have taken place within the CELTIC-NEXT AIMM project Work-Package 3 (WP3) on “Antenna Arrays and Reconfigurable Intelligent Surfaces”.

The focus of this work-package is on novel antenna configurations including collocated and distributed structures, use of Reconfigurable Intelligent Surfaces (RIS) as NR intelligent reflectors and low-complexity Multiple-Input Multiple-Output (MIMO) radios, and antenna array design and radio-frequency (RF) circuit development, all of which involve incorporating tools from Artificial Intelligence (AI) and Machine Learning (ML) against legacy approaches.

This deliverable provides information on the progress made against all these topics of work within WP3. Further, dissemination activities and future plans are provided.

Within Task 3.1 on “Centralised and Cell-less Antenna Arrays”, progress on the development of a decentralised channel sounder dubbed DICHASUS is reported by UST, including a description of several datasets collected in indoor, industrial and outdoor environments. In addition, some preliminary work on a low-cost WiFi-based Channel Sounder is outlined. Further, within Task 3.1, the viability of cell-less Massive MIMO for providing Fixed Wireless Access (FWA) services is investigated by BT, and University of Bristol has studied the lattice quantization for a centralized and scalable cell-free Massive MIMO network with realistic fronthaul. Cell-less Massive MIMO has drawn significant research interest. Its primary distinction from conventional MIMO aided cellular networks is the ability to eliminate the detrimental inter-cell interference (ICI), or to convert ICI into extra power for the intended signal via a multi-cell cooperation approach originated from network MIMO. Also as part of Task 3.1, CEMWorks report on the development of an accelerated electromagnetic simulation platform for large-scale antenna array virtual prototyping.

Within Task 3.2 on “Reconfigurable Intelligent Surfaces”, assessment of RIS, physical-layer technologies, and trade-offs is provided by BT. We consider the performance comparison of both sub-6 GHz and mm-Wave reconfigurable reflective surfaces compared to conventional phased array systems. Relative performance in both Massive MIMO and relay networks is investigated. In general, RIS shows strong potential wherever energy consumption is a high priority and green solutions are necessary. Further, within Task 3.2, the use of RIS as intelligent reflectors for enhancing coverage performance of 5G NR and beyond is discussed by InterDigital, including activities around development of a link-level simulator (LLS) and use of AIML (unsupervised learning) for RIS-aided communications. CEMWorks, as part of Task 3.2, provide virtual prototyping and simulation of metasurfaces for RIS. University of Bristol provides the basic ray tracing simulation of the urban scenario and the electromagnetic simulation of a 14x14 array, then proposed a Deep Reinforcement Learning (DRL) based optimization method for RIS-assisted multi-receiver communication with the DRL and verification results.

Within Task 3.3 on “Massive MIMO Pre-distortion Architectures for Power Amplifiers”, IMST provides algorithmic design of classical digital pre-distortion (DPD) algorithms for MIMO arrays, as well as for AI-based solutions, including validation through simulation activities. Both approaches are validated and optimized using measurement data from work package 6.1. .

The dissemination activities within AIMM WP3 include a press release in the 6GWorld website on RIS, a closed RIS workshop organised in Jan. 2021, open RIS workshops jointly organized between AIMM, RISE-6G, and ARIADNE project consortiums at the IEEE ICT'21 conference in June 2021 and at the IEEE PIMRC'21 conference in September 2021. A public newsletter has also been issued based on the work within AIMM WP3. Furthermore, AIMM WP3 participants were instrumental in the foundation of a new Industry Specification Group (ISG) on RIS at ETSI, the very first pre-standards group on a 5G-Adv/6G topic aiming to bridge the gap between research and standards. Results from the AIMM WP3 have already been disseminated into ISG RIS through ETSI technical contributions. Other dissemination activities from AIMM WP3 have also taken place to date including disclosure filings, paper publications, and industry talks/panels.

In summary, AIMM WP3 has delivered significant results in accordance with the project plan, and the partners involved will accordingly utilize the findings through exploitation activities.

Table of Contents

Executive Summary	3
Table of Contents	4
Abbreviations	5
1 Introduction	7
2 Technical Work Produced	8
2.1 Centralised and Cell-Less Antenna Arrays	8
2.1.1 Distributed MIMO Channel Sounder Development.....	8
2.1.2 Cell-less Massive MIMO for Fixed Wireless Access.....	11
2.1.3 Virtual Prototyping of Large-Scale Antenna Arrays	21
2.1.4 Centralized and Scalable Cell-Free Massive MIMO with Realistic Fronthaul	22
2.2 Reconfigurable Intelligent Surfaces	26
2.2.1 RIS – Physical-layer Options & Trade-offs	27
2.2.2 RIS-Integrated NR with AIML.....	44
2.2.3 Virtual Prototyping and Simulation of Reconfigurable Intelligent Surfaces	48
2.2.4 Simulation and Optimization on Reconfigurable Intelligent Surface assisted Multiple-receiver Communication.....	51
2.3 Power Amplifier Enhancements	55
3 Summary	74
References	75

Abbreviations

Abbreviation	Definition
3G	Third generation cellular
3GPP	Third Generation Project Partnership
4G LTE/LTE-A	Fourth generation cellular Long Term Evolution/Long Term Evolution Advanced
5G NR	Fifth generation cellular New Radio
A1	O-RAN interface between Non-RT RIC and Near-RT RIC
AARX	Antenna array as receiver
AAS	Active antenna system
AI/ML	Artificial Intelligence/Machine Learning
BS	Base station
CAPEX	Capital expenditure
CoMP	Coordinated multipoint
CPRI	Common Public Radio Interface
CQI	Channel Quality Indicator
CSI	Channel State Information
CS-RS	Cell-Specific Reference Signal
CU	Centralised unit
DCI	Downlink Control Indicator
DMRS	Demodulation Reference Signal
DPB	Dynamic Point Blanking
DPC	Dirty-paper-coding
DPD	Digital Pre-Distortion
DPS	Dynamic Point Selection
DRL	Deep Reinforcement Learning
DSP	Digital signal processing
DU	Distributed unit
E2	O-RAN interface between Near-RT RIC and CUs/DUs
eCPRI	Enhanced Common Public Radio Interface
EM	Electromagnetic
eNB	eNodeB (4G LTE/LTE-A base station)
EVM	Error vector magnitude
F1	3GPP interface between CU and DU
FD MIMO	3D/full-dimension MIMO
FR1	Frequency range 1
FR2	Frequency range 2
FWA	Fixed wireless access
gNB	gNodeB (5G NR base station)
HBF	Holographic Beamforming
HLS	Higher-layer-split
ICI	Inter-Carrier Interference
IPR	Intellectual property rights
IRS	Intelligent Reflecting Surface
KPI	Key-performance-indicator
L#	Layer number # on the protocol stack
LLS	Link Level Simulation
LMS	Least Mean Squares
LOS	Line-of-sight
MAC	Medium Access Control
MDT	Minimisation of drive test
MIMO	Multiple-input multiple-output
MOBTX	Mobile transmitter
MRT	Maximum-ratio-transmission
M-TRP	Multi transmission/reception points
Near-RT	Near-real-time
Non-RT	Non-real-time
Open RAN	Ecosystem for open standardised interfaces implementation
OPEX	Operational expenditure

O-RAN	O-RAN Alliance
PA	Power amplifier
PBCH	Physical Broadcast Channel
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PHY	Physical Layer
PRACH	Physical Random Access Procedure
PSS	Primary Synchronisation Signal
PUSCH	Physical Uplink Shared Channel
QoE	Quality-of-experience
QoS	Quality-of-service
RAN	Radio access network
rApp	An application designed to run on the Non-RT RIC
REFTX	Reference transmitter
RIC	O-RAN RAN Intelligent Controller
RIS	Reconfigurable Intelligent Surfaces
RIT	Radio Interface Technology
RLC	Radio Link Control
RLS	Recursive Least Squares
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RT	Real-time
RU	Radio unit
SE	Spectral efficiency
SINR	Signal-to-interference-plus-noise ratio
SISO	Single-input single-output
SON	Self-organising-network
SRIT	Set of Radio Interface Technologies
SSB	System synchronisation block
SSS	Secondary synchronisation Signal
TXRU	Transceiver chain
UE	User equipment
vRAN	Virtualised RAN
X2	3GPP interface between eNBs
xApp	An application designed to run on the Near-RT RIC
Xn	3GPP interface between gNBs
ZF	Zero-forcing

1 Introduction

MIMO is a key air-interface technology in nearly all modern communications systems [1]. MIMO, through utilization of multiple antennas at the radios, can provide several benefits including enhancing spectral efficiency and quality-of-service [2]. Despite significant performance improvements achieved through MIMO to date, there exists a significant gap between the theoretical versus practical performance of multi-antenna systems [3].

Motivated by the above, the AIMM project targets radical performance improvements and efficiency dividends for 5G and beyond MIMO systems through adoption of AI/ML capabilities in both link-level and system-level RAN domains, as well as alternative deployment methods including radio intelligent surfaces and cell-less antenna systems. To achieve the set targets, the AIMM project work is divided between six tightly coupled work-packages, as illustrated in Fig. 1 below.

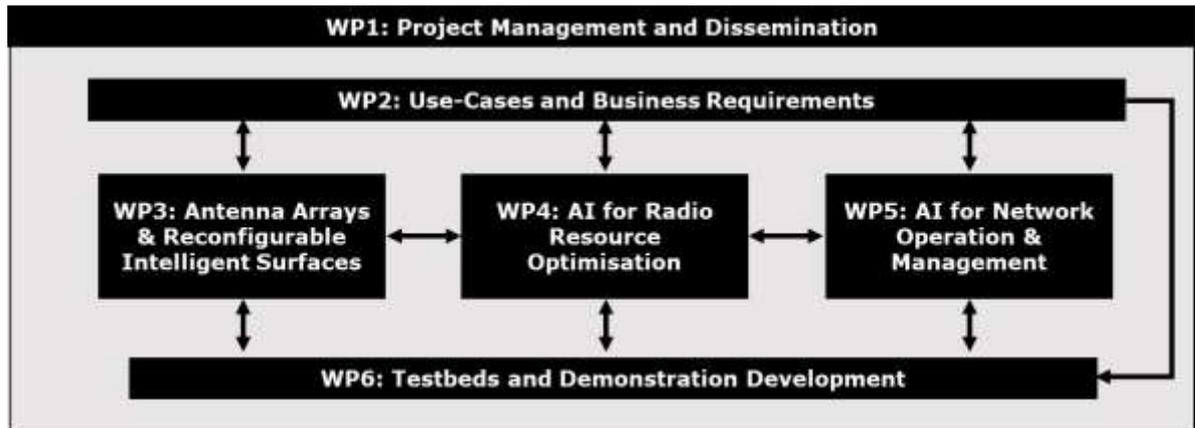


Fig. 1: AIMM project work-package structure.

The AIMM project WP3, titled “Antenna Arrays and Reconfigurable Intelligent Surfaces”, focuses on the benefits and practical realisation of extremely large antenna arrays, the use of RIS as intelligent reflectors as well as low-complexity MIMO radios, and design of DPD techniques for Massive MIMO power amplifiers. The use cases, key-performance-indicators (KPIs), and initial value assessment for technical solutions being developed within WP3 are provided by WP2 (please see, AIMM WP2 deliverables). Further, verification of some of the technical solutions will be facilitated through testbed activities, including those in WP6.

This deliverable provides a summary of the work that has been carried out to date within this AIMM project WP3. Further, dissemination activities and future work plan for this work-package are accordingly highlighted in this report.

2 Technical Work Produced

This section provides the summary of the technical work by the AIMM WP3 participants carried out within the three main tasks of this work-package.

2.1 Centralised and Cell-Less Antenna Arrays

Massive MIMO is in the state of being deployed in the first commercialized products, and already considered to be one of the major technologies to cover the required increase of area capacity.

During the last few years, it has been shown that the antenna geometry is crucial for separating closely spaced users [4]. As most of these practical implementations and investigations work with a limited number of antennas (64) the question remains whether a very large number of antennas (>1024) is more advantageously deployed in a concentrated or distributed fashion. Moreover, the best antenna geometry for different types of cellular systems is yet to be found, with cell sizes ranging from femto to macrocell.

A large number of antennas results in a huge amount of data, and hardware cost is increased significantly. Therefore, advanced techniques for base station design as well as channel state information data acquisition, as outlined in [5], need to be studied and improved, to further reduce the amount of hardware costs per chain, while keeping all the advantages of Massive MIMO. Moreover, fast and accurate electromagnetic simulation of antenna arrays with large number of antennas is a challenging, but crucial task within the search for cost-effective optimal antenna design that satisfies given specifications.

Arrays with many antenna elements can be created through holographic antennas [6], or by using large buildings structures a "MULIBU"-array, i.e., a distributed MULti-Linear antenna array on BUilding. Holographic Massive MIMO creates a 3D array based on simple 2D array planes, and therefore promising higher spatial diversity. Yet, it is anticipated that this only works above 6 GHz bands, and, therefore, cannot be used for macrocells. Thus, MULIBU deployments are needed for extending this technology to sub-6 GHz frequency bands. The main challenges to overcome can be found in connecting several large linear antenna arrays to a concentrated processing unit, without losing receiver sensitivity. Therefore, new high data rate RF and/or optical connectivity solutions and advanced antenna technologies need to be investigated to overcome this practical challenge.

Task 3.1 within WP3 looks at the simulation and applications of extremely large antenna arrays, considering both flexible antenna geometries and user-centric (cell-less) network operation.

2.1.1 Distributed MIMO Channel Sounder Development

To study massive MIMO channels, in particular different antenna geometries (linear, holographic, along buildings), antenna node connection technologies, and multiple antenna calibration algorithms, a channel sounder dubbed DICHASUS (Distributed Channel Sounder by University of Stuttgart) is being developed as part of the AIMM project. In contrast to an earlier channel sounder by our institute, DICHASUS is based on a decentralized (i.e., distributed) structure, rather than channeling all measurements through a central node. The measured channel state information (CSI) will be used in, e.g., AIMM-WP4 to study machine learning algorithms for various applications, such as user localization, reducing UL CSI feedback for DL precoding, and others.

DICHASUS covers the frequency range from 70MHz to 6 GHz with a maximal bandwidth of 50 MHz. It consists of a) a mobile transmitter (MOBTX) with one transmit antenna in the field, b) a multiple antenna array (AARX) as receiver (mimicking a base station), and c) and a reference transmitter (REFTX) for calibration placed right next to the antenna array. The MOBTX continuously transmits pilot symbols using a configurable OFDM-modulation, enabling the extraction of CSI per subcarrier at the receiver antenna array. In the meantime, "ground truth" position information (e.g. GNSS coordinates) are collected, so that position tags are available alongside channel coefficients, which is important for various machine learning tasks. The reference transmitter is placed next to the antenna array with line-of-sight propagation paths to all array antenna elements. It also transmits OFDM based pilot signals, but at a frequency orthogonal to the mobile transmitter in the field, to minimize mutual interference. At all times, the AARX can receive both the signal from the MOBTX (the channel supposed to be measured) as well as the signal from the REFTX. The AARX is

distributed, without a central clock, and thus relies on regular (or continuous) phase, amplitude, and timing calibration over-the-air using the REFTX's signal. First experiments show that, despite the AARX's distributed architecture, a frequency stability of below 50 ppb can be achieved. The study and further improvement of this kind of calibration was subject of the AIMM project.

After a synchronization phase, each antenna element of the AARX individually dumps its captured I/Q samples to its own USB storage device. This decentralized sample storage avoids memory bottlenecks, which is a prerequisite for scalability. The recordings of all antenna elements are then combined and turned into CSI later using off-line post-processing. With this architecture, the set-up is quasi-linearly scalable to a large number of antenna elements without changing the architecture, i.e., without changing AARX, MOBTX or REFTX. Also, the decentralized RX concept of the AARX allows arbitrary RX array antenna geometries, e.g., along building fronts and the like. Optionally, individual receivers can be connected either through a cable-bound or a wireless link for controlling the measurements, e.g. for carrier frequency and bandwidth configuration and real-time monitoring. Due to the individual quartz oscillators of the RX elements, each measured RX antenna signal needs to be post-processed with respect to estimating, compensating and tracking phase offset, frequency offset, and timing offset, using the REFTX signal. To offload the burden of compensating all of those effects using post-processing, an important task is to perform a coarse frequency calibration of all elements prior to the measurement over-the-air by controlling the individual PLLs.



Fig. 2: Left: Setup for channel measurement campaign in industrial environment (research factory “Arena 2036”) with robot-mounted mobile transmitter and 32-antenna array in the background.

Since late 2021, the channel sounder is up and running with a total of 32 antenna elements and, in mid-2022, it has been extended to 64 antenna elements. In addition to datasets collected in an indoor office environment, datasets generated in the “Arena 2036” research factory, representative of a typical industrial environment, have been made public and analyzed. As expected, the wireless channel measured in the factory environment exhibits a greater delay spread (and hence lower coherence bandwidth) compared to datasets measured in indoor office environments. Due to general activity inside the factory, the channel is less stable over time. Thanks to a newly acquired tachymeter (total station), the position of the robot-mounted transmitter can be precisely tracked, information that is invaluable for machine learning-type applications as well ground truth data for positioning experiments.

Furthermore, measurements with distributed antenna setups (two and four separate arrays) have been carried out, both in an indoor environment and in an industrial environment, confirming that distributed operation of the channel sounder is feasible.

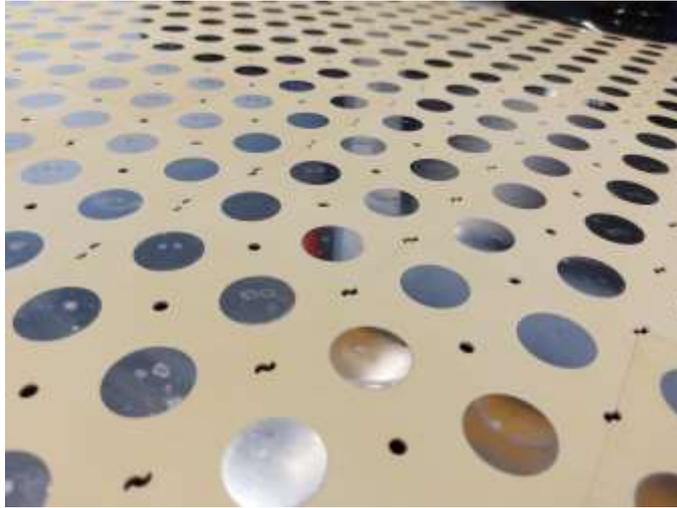


Fig. 3: Newly developed 3.4GHz patch antenna array.

DICHASUS supports arbitrary carrier frequencies from 100MHz to 6GHz (and potentially up to 12GHz with reduced bandwidth). Most datasets have been measured at a carrier frequency of 1.27GHz, due to the availability of a suitable patch antenna array. To also be able to capture datasets at a frequency of 3.4GHz, which lies in an amateur radio band and should exhibit similar propagation characteristics as frequencies allocated to 5G networks at around 3.7GHz, a new patch antenna array for this frequency has been developed and manufactured.

In addition to DICHASUS, development of a realtime-capable, ultra-low-cost WiFi-based channel sounder has also been initiated. DICHASUS and the WiFi-based channel sounder have very different objectives though: DICHASUS is meant to produce large, high-quality datasets for scientific purposes, such as research into Channel Charting, channel estimation, feedback reduction, cell-free massive MIMO and channel modelling. Researchers do not need to operate DICHASUS themselves, since DICHASUS datasets are published for use by the scientific community. The WiFi-based channel sounder, on the other hand, is meant for applications like real-time demonstrators and hands-on research, where the quality of data is less important than the ability to easily capture data. For example, WiFi-based Human Activity Recognition is one such field where the ability to quickly capture and analyze data for rapid iteration cycles is important. Furthermore, algorithms and methods developed and fine-tuned on DICHASUS datasets can be deployed to the WiFi-based system for demonstration purposes.

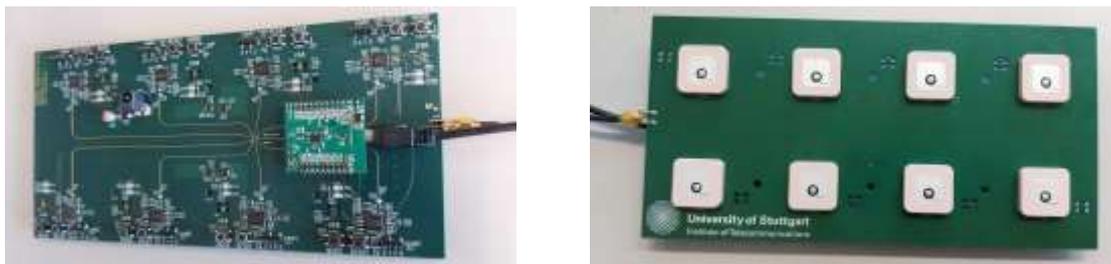


Fig. 4: Front and back of the main circuit board of the WiFi-based channel sounder.

In contrast to DICHASUS, the WiFi antenna array is completely integrated onto a single four-layer circuit board shown in Fig. 4, with WiFi System-on-Chip receivers on the back and ceramic 2.4GHz antennas on the front. This makes the channel sounder easy to manufacture, which, in addition to the cost-aware choice of BOM components, contributes to its low procurement cost.

Early experimental results indicate that the channel sounder is working as intended, but may need some improvement to enable absolute phase calibration. It will be possible to combine multiple antenna arrays into a single phase-aligned receiver array, though completely distributed deployments as with DICHASUS are not possible, since the system relies on cable-based reference clock distribution.

2.1.2 Cell-less Massive MIMO for Fixed Wireless Access

Recently a new paradigm of wireless access, termed as cell-free massive MIMO, has drawn significant research interest. Its primary distinction from conventional massive MIMO aided cellular networks is the ability to eliminate the detrimental inter-cell interference (ICI) or to convert ICI into extra power for the wanted signal via a network MIMO treatment. However the information-theoretical limit of cell-free access is achieved at the expense of large network configuration overhead and high MIMO processing complexity. Under the constraint of wireless channel dynamics, the global channel state information (CSI) invoked for network MIMO quickly becomes outdated, leading to CSI errors. This study focuses on the cell-free implementation of fixed wireless access (FWA), which can be an attractive alternative to fibre-to-the-premise (FTTP) in exceptional circumstances. In particular, we discuss the centralisation architectures and channel characteristics of cell-free FWA, as well as their joint implications on imperfect CSI performances. Measurement-based offline simulations shows that the ‘quasi-static’ assumption of real-world FWA channels is only valid against a completely static background, and thus it should not be used in FWA system design and performance analysis.

Worldwide 5G rollout provides a robust hardware infrastructure for a new generation of fixed wireless access (FWA), complementing the fibre-to-the-premise (FTTP) wireline access paradigm. As a complement of massive multiple-input-multiple-output (MIMO) antenna arrays, cell-free cooperation between base stations (BSs) can further optimise the spectral efficiency (SE), coverage and power consumption of FWA systems. With the ambition of ubiquitous broadband connectivity, enhanced FWA solutions are crucial for use cases such as remote rural/mountain side areas and emergencies like pandemic hospitals, where fibre deployment is too expensive, intrusive, or inefficient.

Conventional cellular communication networks rely on a system of BSs with each one covering a certain geographic area termed a cell. At any given location, a user equipment (UE) is exclusively served by a single BS. If the signals from adjacent BSs are not perfectly orthogonalized, they will become interferences which have to be avoided or overcome, thereby reducing the achievable SE of the UE. Moreover, the presence of multiple active UEs belonging to different cells also causes mutual interference. In contrast, cell-free cooperation exploits diversity gain when concurrently-arriving signals of overlapping frequencies from all adjacent BSs are relatively strong at a given UE’s location, in which case the multi-user interference as well as the inter-cell interference (ICI) becomes beneficial. This occurs most strongly for small cells, at high user densities, and in a low-path-loss environment. Additionally, cell-free cooperation can also improve the reliability of FWA services, since each UE’s outage probability shrinks exponentially in the number of cooperating BSs.

Cell-free massive MIMO is perceived as the evolution and consolidation of multiple cooperative MIMO technologies from the past two decades. In Long Term Evolution (LTE), common radio resource management (CRRM) specified systematic management of interference and radio resources across multiple sectors and cells. CRRM aims to utilise the limited radio frequency (RF) spectrum resources and radio network infrastructure as efficiently as possible. Under relatively moderate traffic demands, CRRM can reduce interference levels, but under overwhelming traffic demands, the limited pool of radio resource becomes insufficient. However, the more powerful type of cell-free deployment known as coordinated multi-point (CoMP) joint-transmission (JT) is less susceptible to such limitations, because the size of the radio resource pool is increased per extra cooperating BS. In CoMP-JT, multiple BSs cooperatively perform downlink (DL) transmit precoding (TPC) by exploiting global channel state information (CSI). Hence, the signals from adjacent BSs will constructively combine at the UEs, leading to improved SE performances. See Fig. 5 for different cell-free FWA coordination mode schematics.

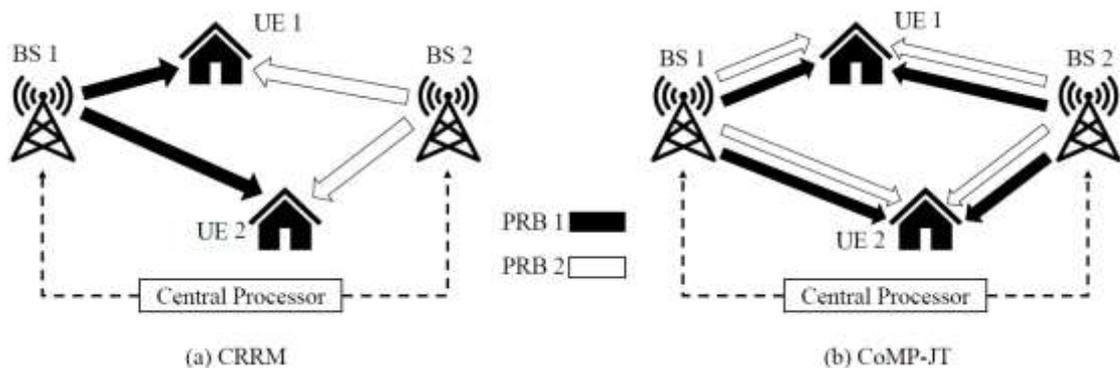


Fig. 5: Cell-free FWA coordination modes. (a) CRRM prohibits simultaneous use of the same physical resource block (PRB) by different BSs; (b) CoMP-JT allows PRB sharing across coordinated BSs.

Even though cell-free cooperation is a readily standardised concept, it has not been extensively used in commercial FWA or mobile networks. This is mainly because of the high processing delay incurred by essential MIMO TPC related operations and by BS-side information exchange via non-ideal fronthaul, as well as of the fact that the transfer characteristics of a wireless channel remain almost constant only for a short period, termed as coherence time. Practically, if the DL channel from a BS to a UE changes significantly between the point of CSI acquisition at BS and the point of DL data reception at UE, then a performance degradation is inevitable as a result of CSI estimation error. Therefore, the channel needs to be stable and the signal processing delay needs to be low for cell-free cooperation to be most beneficial. Due to the immobile UEs, FWA channels are widely assumed to change more slowly than mobile ones, even though the assertion that FWA exhibits quasi-static channel states remains as a myth in practice. In a sharp contrast to previous CoMP case studies such as [7] which mainly addressed the design trade-off between performance gain and network configuration overhead for multiple CoMP modes, this study focuses on the physical realities behind FWA and the optimality conditions for coherent JT.

Cell-free FWA: Coordination Strategies

In cell-free FWA, ICI exploitation can have a constructive impact on the target UE's received signal-to-noise ratio (SNR), and it is therefore more powerful than reduction or avoidance. However, optimal cell-free FWA incurs considerable network configuration overhead due to the need for global CSI aggregation and sometimes even user data sharing at the central processing unit (CPU) via fronthaul. In addition, the global CSI matrix naturally has a higher rank than those associated with a single BS or UE, which also increases the complexity of channel estimation and TPC-related matrix operations. To mitigate these predicaments, time division duplexing (TDD) has been widely accepted as the more realistic duplexing option over frequency division duplexing (FDD) since it facilitates channel reciprocity exploitation between UL and DL. UL/DL channel reciprocity in general frees up the radio resource budget reserved for CSI sample feedback as well as reducing the associated processing delay. In this section, we present an overview of the main architecture of the interference-avoiding and the interference-exploiting BS coordination strategies.

CRRM: Generally, CRRM does not permit the sharing of a single physical resource block (PRB) across multiple BS sites, because the CPU does not have access to the global CSI which is essential for ICI exploitation. A primitive attempt at CRRM is the classical frequency-reuse scheme, widely employed by existing 4G and 5G cellular BSs. However, frequency-reuse performs fixed spectrum allocation that does not adapt to the channel environment. As a result, frequency-reuse is suboptimal in a dynamically changing environment.

In TDD CRRM (Fig. 6(a)), the UL signal (including the CSI-bearing pilots) received at each BS undergoes the typical analogue-to-digital conversion (ADC), bit timing and multi-user synchronisation, as well as demodulation (e.g. a discrete Fourier transform (DFT) for the orthogonal frequency division multiplexing (OFDM) scheme). After demodulation, while the UL user data is passed to error correction decoders and backhaul transmission protocols, UL channel estimation is performed locally at the BSs. With local UL CSI available, each BS can perform independent CSI preprocessing, including UL-DL conversion as well as other mandatory TPC initialisation steps (e.g. matrix inversion). Given the availability of PRB instructions obtained via fronthaul and the locally available TPC weights, BS can precode the requested DL user data fetched via backhaul. The precoded DL data is modulated (e.g. by inverse DFT) and converted to analogue before finally being sent to the UEs via the DL FWA channel.

CRRM facilitates more relaxed BS coordination and generally incurs less configuration overhead than CoMP-JT. More specifically, since the role of CPU is to schedule PRB allocation to the BSs, it only needs to know qualitatively an abstraction of each BS' local CSI, namely a channel quality indicator (CQI). In comparison to sharing full size channel matrices, CQI sharing significantly reduces the amount of information that needs to be exchanged via fronthaul. An immediate benefit of it is the reduction of processing delay and quantisation loss happening in both directions of the fronthaul. Moreover, since the centrally managed PRB allocation information does not have to be updated in line with the instantaneous CSI, the role of CPU is not a part of the critical signal processing timeline (i.e. the red paths in Fig. 6).

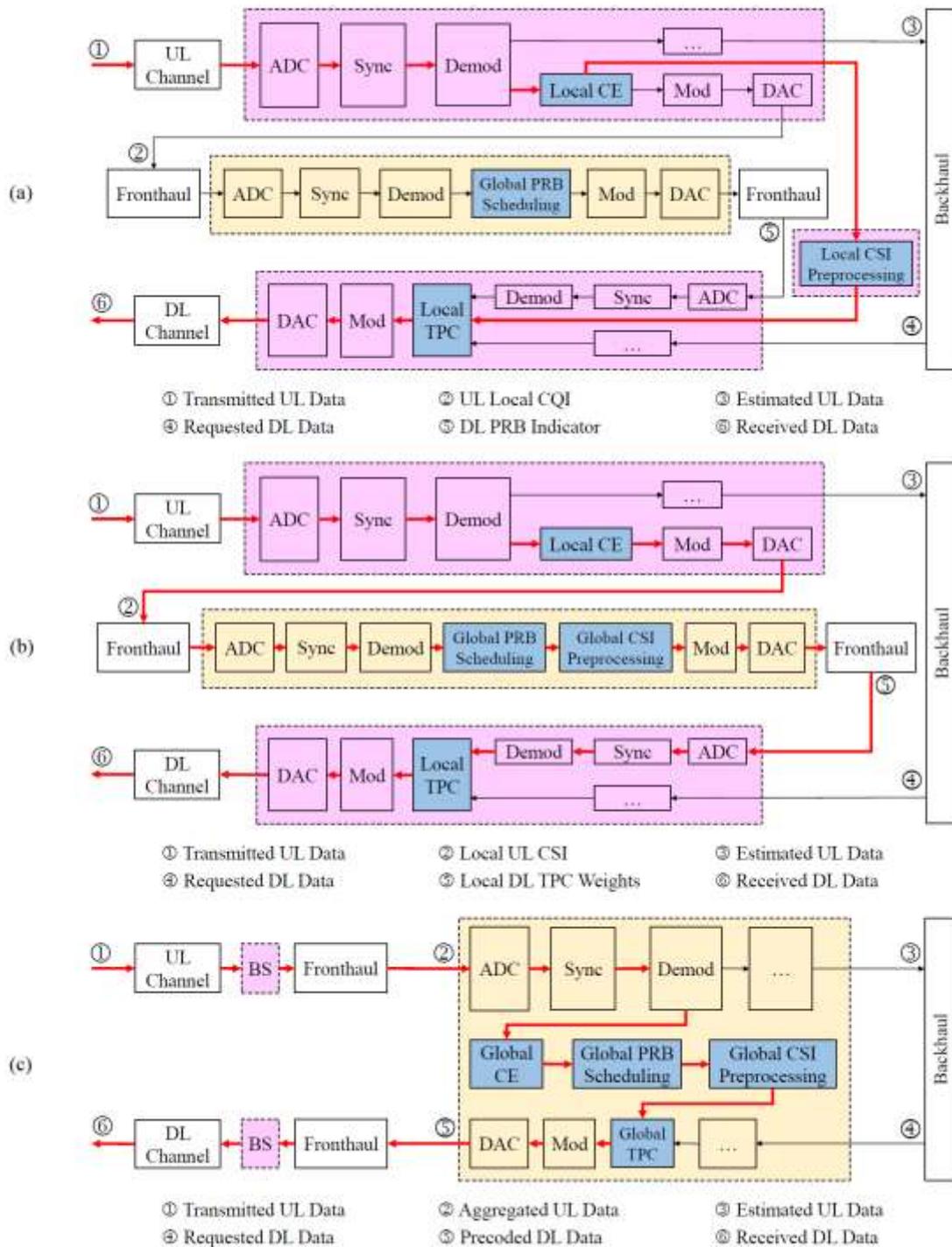


Fig. 6: TDD cell-free FWA architectures. The red chains signify critical CSI processing paths associated with channel stability requirement. Pink blocks are the BSs while the yellow blocks represent CPUs. A CPU can be deployed remotely or within a master BS. (a) Only PRB scheduling is performed at the CPU; (b) CSI processing is performed at the CPU but TPC is performed locally at each BS; (c) Both CSI processing and TPC are performed at the CPU.

CoMP-JT: In a nutshell, CoMP-JT based cell-free access may be perceived as the ultimate form of network MIMO. Similar to distributed antenna systems, CoMP-JT can utilise geographically separated transmitting antennas belonging to different cooperating BSs. Therefore, it is capable of achieving the optimal multiplexing or diversity gain. Despite the existence of other CoMP modes such as coordinated beamforming (CB) or scheduling (CS), the information-theoretical performance limit of network MIMO can only be truly achieved by coherent (i.e. centrally synchronised) JT.

Coherent CoMP-JT relies on the availability of instantaneous global CSI at the CPU. Depending on the level of centralisation, the CPU can operate in one of the following two ways. If the CPU has no access to DL user data (Fig. 5(b)), it will simply aggregate, synchronise the local UL CSI and perform global CSI preprocessing in order to derive the global DL TPC matrix, which will be distributed alongside the global PRB allocation instructions to the BSs for local TPC. On the other hand, if the CPU has direct access to all of the DL data fetched via backhaul in addition to global CSI (Fig. 5(c)), it will perform centralised global TPC before distributing the precoded DL data to the BSs. In both cases, the CSI has to be shared in full, which imposes much more stringent requirements upon the latency and capacity of fronthaul networks than CRRM does. Furthermore, in comparison to mode (c), the additional ADC, bit timing synchronisation and modems occurring in both directions of the fronthaul in mode (b) lead to more quantisation loss and processing delay. This is also different from CRRM in (a) because 7 fronthaul is used much more frequently in (b) due to mandatory CSI and TPC matrix updates. Therefore, fronthaul uses constitutes a part of the critical path of (b) but not for (a).

In a practical setting, only a limited number of BSs may be clustered to perform CoMP-JT due to MIMO processing complexity, synchronisation requirement and fronthaul constraints. It is therefore expected that inter-cluster interference will become the dominant problem in dense networks with low path loss. Fortunately, clustering is trivial for FWA services since it does not require dynamic optimisation. However, the coordination between mobile and FWA services in a cell-free scenario, both radio resource and interference wise, remains as an open challenge.

Cell-free FWA: Channel Characteristics

Cell-free FWA exhibits different channel characteristics than cellular massive MIMO or mobile services, due to the stationary and geographically distributed nature of UE terminals. In a nutshell, Cell-free FWA channels do not usually behave like independent and identically distributed (i.i.d.) random variables, because the path losses from different serving BSs to the same UE tend to be exceptionally different due to the varying distances. Moreover, cell-free FWA channels are also expected to have low spatial correlation but high temporal correlation, with the former being more likely in CoMP-JT types of implementations. We discuss the soundness of these common expectations in further details in the following sections.

Spatial Correlation: A simulated scenario is considered for quantifying the spatial correlation characteristics of cell-free FWA, with particular respect to the effect of UE distribution and of antenna correlation. Specifically, we consider a cluster of six cooperating BSs arranged in a hexagonal lattice pattern in the shape of an equilateral triangle, deployed in a sub-urban environment. The fixed UEs are assumed to be uniformly distributed (at fixed locations) within a circle centred at the joint of three hexagonal cells. Due to path loss, out-of-cluster signals, detrimental or beneficial, are heavily attenuated in this area. Hence, the generality of this assessment is not affected even if more BSs are introduced. The simulation considered a path loss exponent of 2.8 and an antenna correlation factor of 0.5 between adjacent elements of the same array. Spacing between the closest pair of BSs is 346 m. Since FWA UEs are typically in customer premises where line of sight (LOS) paths to the outdoor BSs are blocked, the Rayleigh model is chosen for multi-path fading.

Overall, it is shown in Fig. 7(a) that the SE performance of CoMP-JT scales better than CRRM with an increasing number of active UEs. This result aligns with the fact that CRRM cannot glean diversity gain from the additional UE-side channels while CoMP-JT can. On the other hand, the addition of transmit antenna elements in the array yields less gain for CoMP-JT. Due to proximity, conventional (massive MIMO) arrays deployed over a single BS have densely packed antenna elements which inevitably spawn spatially correlated channels. This is circumvented in CoMP-JT, which jointly utilises antenna elements from distributed BS locations, but not in CRRM. Hence, increasing the number of (correlated) antenna elements has more explicit benefits for CRRM. For the sake of exploiting spatial diversity/multiplexing gain, it may be more beneficial to deploy smaller arrays of fewer antenna elements over a larger number of cooperative BSs. Unfortunately, this is likely to destroy the 'channel hardening' property of massive MIMO.

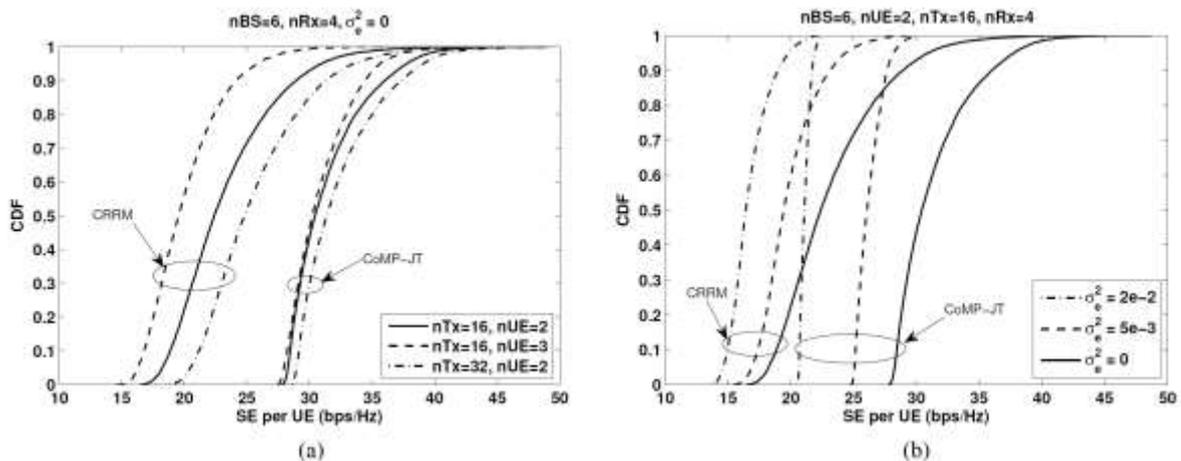


Fig. 7. Cell-free FWA performances and CSI error tolerance.

Temporal Correlation: One of the most prominent challenges associated with cell-free FWA (as well as mobile access) is the variability of wireless channels. In most of the information theoretical analysis of network MIMO systems, the CPU is required to have access to perfect and non-causal knowledge of the global DL CSI matrix. However, as a result of signal processing delay, such an assumption is never valid in practice. A typical wireless channel remains constant only for a limited duration as a result of Doppler shift, arising from the relative motion between transceiver terminals. This characteristic time interval is termed as coherence time, which shows the temporal correlation of the propagation environment.

Longer coherence time relative to computational time is key for BS coordination in FWA. The coherence time will determine how much cyclic prefix or suffix is needed for orthogonal frequency-division multiplexing (OFDM) based transmissions, as well as the pilot length requirement for channel estimation. Those cyclic extensions of message-carrying data consume extra radio resources. Consequently, their length has a pivotal effect on the SE capacity limit. It should be mentioned that the minimum coherence requirements for field trial and commercial deployment are hardware-specific variables which also depend on signal processing designs. Generally, the minimum coherence time needs to be longer than the total transmitter-side data processing delay.

Implications of Channel Coherence

CSI Error Considerations: Nearly four decades ago, the seminal work of Costa on TPC showed that, the optimal interference-free SE capacity of a interference-contaminated system can be fully recovered if the interference is perfectly and non-causally known at the transmitter. Unfortunately, it is impossible to fulfil such conditions in practice. Despite the performance promises, cell-free systems are in fact very vulnerable to CSI errors. Fig. 6(b) portrays the collective influence of additive CSI errors on the performance of cell-free FWA with varying CSI error variances σ_e^2 . While CoMP-JT always outperforms CRRM at the same level of σ_e^2 , its performance degrades much faster than CRRM when σ_e^2 increases. More importantly, given the fact that CoMPJT needs to share critical information (CSI and/or user data) via unreliable fronthaul whilst CRRM does not, the actual value of σ_e^2 may appear larger for CoMP-JT even when their overall signal processing frameworks, channel estimation procedures and TPC designs are equivalent (particularly Fig. 8(a) and 8(b)).

Aside from the theoretical limitations of random variable estimators, imperfect channel estimation is also the collective result of short coherence time, imperfect BS coordination and hardware impairment. Among these factors, coherence time is regarded as the least influential for FWA. In fact, the so-called 'quasi-static' assumption is frequently employed for characterising FWA channels in the literature. The following sections will be dedicated to a series of channel measurements that identifies the extent to which the 'quasi-static' assumption is valid.

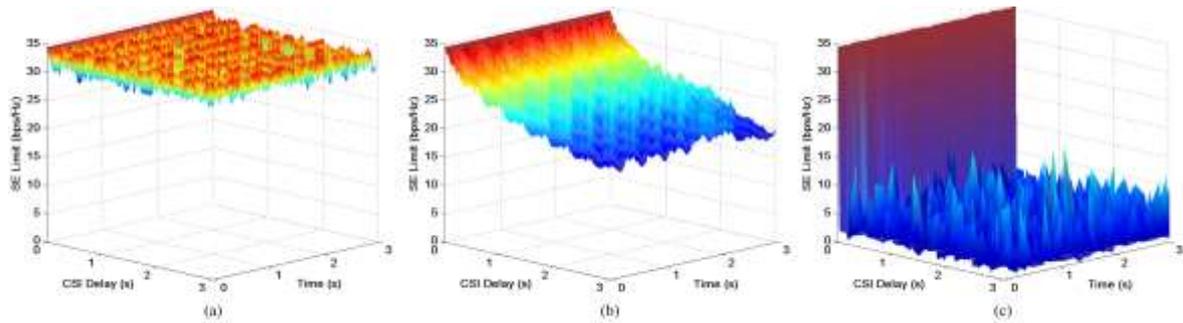


Fig. 8. Channel coherence time measurements for fixed 2 x 2 MIMO terminals in (a) a static indoor environment, (b) a dynamic indoor environment, and (c) an outdoor environment.

Fixed Terminal Coherence Time: Preliminary analysis of coherence time in the literature mostly relies on an autocorrelation-based statistical definition. However, statistical modelling of wireless channels is mostly for characterising an ensemble of channels rather than a particular deployment environment. Therefore channel coherence evaluation based on statistical models, such as a Rayleigh distribution, does not necessarily match the real-world observations. Such discrepancies in channel characterisation are likely to result in unrealistic performance expectations. Even though FWA does not experience Doppler shifts, nevertheless the environment between the transmitter and the receiver is not necessarily static, and this may have a significant role in determining the multipath profile. The coherence time of fixed terminal 2 x 2 MIMO channels is characterised for indoor and outdoor scenarios, respectively.

Indoor Measurement: Two indoor measurements were taken at the Communications Lab, Loughborough University. The first test scenario involves a completely static room environment without any moving scatters or reflectors, while the second scenario includes a person moving near the LOS path in the static background. Both sets of CSI are measured with a 0.8264 ms time resolution at a carrier frequency of 900 MHz, while the terminal spacing is 1 m between each other. The coherence time of indoor environment is portrayed in Fig. 7(a) and (b) by means of simulated SE limit. More particularly, the simulation depicts the effect of outdated CSI via an offline time-sliding methodology using a total transmit SNR of 60 dB, where we gradually increase the delay between two hypothetical points in time, namely the point of CSI acquisition and the point of data transmission. The 'time' axis in Fig. 7 represents the hypothetical point when precoded data is transmitted, while the difference between the two channel states is treated as CSI error. The immediate observation is that FWA channel is only 'quasi-static' when there is no moving scatters or reflectors in the background. In the presence of only one moving background obstacle, the quasi-static property of FWA channel is lost despite the preserved LOS path. Moreover, the 'block-fading' assumption, where the CSI remains constant for a time block and changes in an i.i.d. manner between blocks, is not true for FWA, either. Instead, the dynamic environment CSI shows a continuous, gradual channel variation.

Outdoor Measurement: In order to evaluate the coherence time of a more realistic FWA scenario, an outdoor channel measurement was taken at Adastral Park, Ipswich. The fixed 2_2 MIMO channel was measured from a building-top transmitting array to a ground-level receiving terminal with a LOS distance of approximately 60 m. The full measurement was taken with an average time resolution of 15 ms at a carrier frequency of 3.56 GHz. Given that the time-sliding SE calculation relies on a transmit SNR of 120 dB, we may observe from Fig. 8(c) that the SE variation over time is negligible for the error-free case (where CSI delay = 0) despite the non-constant CSI, which was also true for indoor scenarios. The consistent 'SE capacity stability' of both indoor and outdoor scenarios implies that, at reasonably high SNRs, the SE capacity of FWA becomes insensitive to the relative channel variations, thus exhibiting a 'quasi-hardening' effect at the performance level. In comparison to prior outdoor massive MIMO mobile channel measurements reported in [8], an outdoor FWA channel exhibits strong channel hardening even for a minimalist 2_2 MIMO scale. However, we do expect that the SNR requirement for the 'SE capacity stability' will be lowered with an increasing amount of antenna elements.

CoMP-JT based cell-free massive MIMO constitutes a promising solution for next gen FWA that can be readily deployed over the existing 5G network with low capital expenditure, at the expense of

some radio resources intended for mobile access. The performance of CoMP-JT based BS coordination is shown for a theoretical network topology to be considerably improved compared with the state-of-the-art BS coordination strategy CRRM, as long as the severity of CSI error is comparable in both cases. It is also shown by measurements that the ‘quasi-static’ assumption for FWA holds valid only for a completely static environment, while the coherence of a typical outdoor environment is expected to be much weaker.

Cell-free channel state stabilization

As portrayed in Fig. 9, the stability of wireless channels can be characterized both globally and locally. A globally stable channel varies within a small dynamic range, while a locally stable channel varies slowly within a given time/frequency interval. Global channel stability is commonly observed in MIMO systems exhibiting the channel hardening phenomenon, which is one of the defining characteristics of massive MIMO in an i.i.d. Rayleigh fading environment. This has been shown in Fig. 9, where a 2 x 2 MIMO FWA setup exhibits channel hardening at high SNR. Under general SNR conditions, channel hardening is achieved by having an asymptotically large number of transmitting and/or receiving antennas. The array-wise global stability achieved by channel hardening can simplify network-level power budget scheduling for the cell-free downlink, but it is less useful for reducing channel estimation errors. Hence the main challenge for channel state stabilization is the local stability of individual channels.

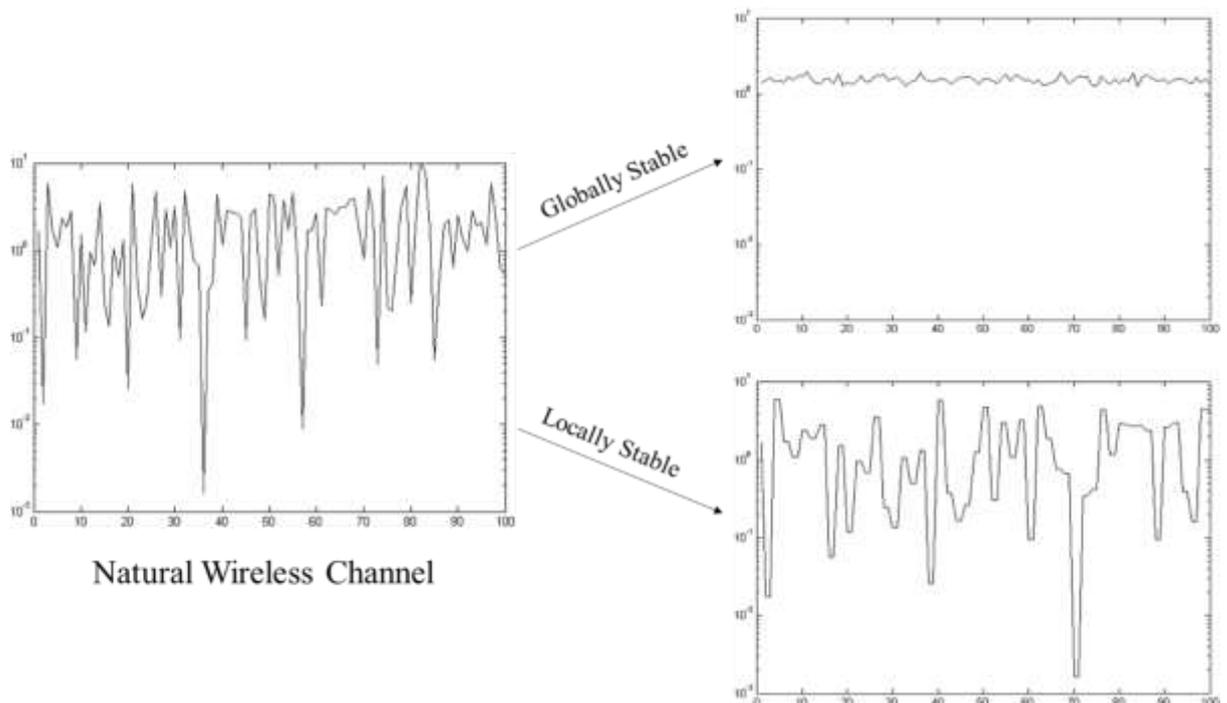


Figure 9. Global and local channel stability.

Recently, reconfigurable intelligent surface (RIS) was discussed as an option for channel state stabilization. This serves as a complementary CSI error correction technique to the popular channel coding strategy such as the low-density parity check (LDPC) codes adopted by 3GPP New Radio. Current research activities within the general RIS community focus on the low power consumption (‘passiveness’), inherent full-duplexing, and noiseless aspect of smart surfaces. However, metamaterial has a wide range of radio wave manipulation capabilities, which need further attention in terms of both practical functionality and regulation compliance.

In the context of wireless channel stabilization, the key to improve the local stability of cell-free FWA channels lies in the extension of coherence time and bandwidth, which is based on the mitigation of Doppler effect and multipath spread. The following sections present an overview of technical considerations for RIS aided cell-free channel coherence extension.

Time reversal: Active multipath cancellation by spatio-temporal symmetry

Time reversal is a well-known strategy for cancelling the multipath effect. It exploits the spatial reciprocity of radio propagation in a static environment. A RIS that is specialized in performing time reversal of (radio) waves is called a time reversal mirror (TRM). More explicitly, within a TRM-enclosed chamber, the propagation of radio waves during a particular time interval can be recorded at distributed locations of the TRMs. Implicit information related to the scatterers and reflectors within the chamber is embedded in such recordings. A time-reversed playback of such recordings will then be able to refocus the scattered radio waves, thus removing the multipath spread. As a side effect, the scattered signal power is also recollected, leading to improved network energy efficiency. An application of time reversal to stabilized wireless communications was demonstrated in [9]. Moreover, with a static internal environment, it is shown that a TRM chamber can precisely construct an electromagnetic field of any prescribed shape [10]. Additionally, the recoding of radio waves does not necessarily require external storage devices since metamaterial can store radio waves locally via an electromagnetically induced transparency approach [11]. However, when the propagation environment is time-varying, the channel's spatial reciprocity is destroyed. In this case, TRM will not be able to fully mitigate multipath scattering, unless the environmental changes can be accurately predicted.

Motion compensation: Doppler cloaking vs Motion prediction

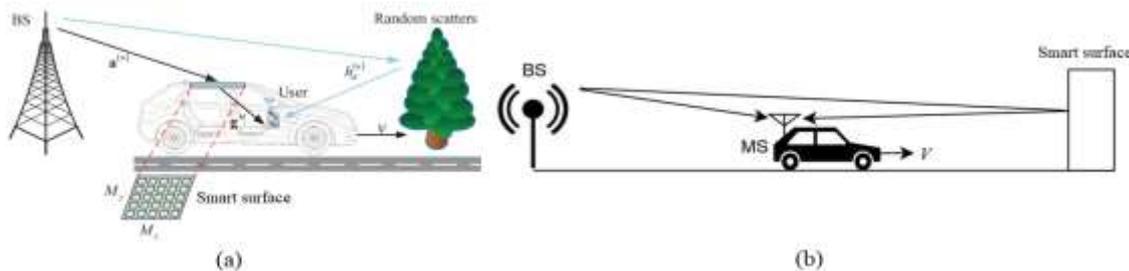


Figure 10. Motion compensation via smart surfaces. (a) The moving terminal is equipped with a smart surface [12], (b) A stationary smart surface is placed near the moving terminal's trajectory [13].

As demonstrated by the fixed terminal CSI measurements, a time-varying channel is due to not only the relative motion between transmitting and receiving antennas, but also the motion of any scatterers or reflectors. The resulting Doppler shift and its associated frequency dispersion have to be removed in order to achieve a stable, time-invariant channel state. In the ideal case, RIS-based motion compensation can transform a time-varying channel into a time-invariant one given the availability of precise non-causal knowledge of all environmental motions (the 'predictable mobility' scenario).

More generally, motion compensation can be achieved by either coating the target moving terminals with RISs (Fig. 10a [12]), or by placing stationary RISs near the trajectory of the moving terminals (Fig. 10b [13]). In the former case, perfect knowledge of the moving trajectory is readily available to the controller of the RIS, which makes motion compensation trivial. This approach is sometimes referred to as 'Doppler cloaking', where a moving transceiver terminal appears as electromagnetically stationary. However, the second approach will instead rely on accurate motion prediction, which can be trivial for scenarios such as railway or satellite, but difficult for other use cases such as an urban macro cell or a crowded exhibition hall.

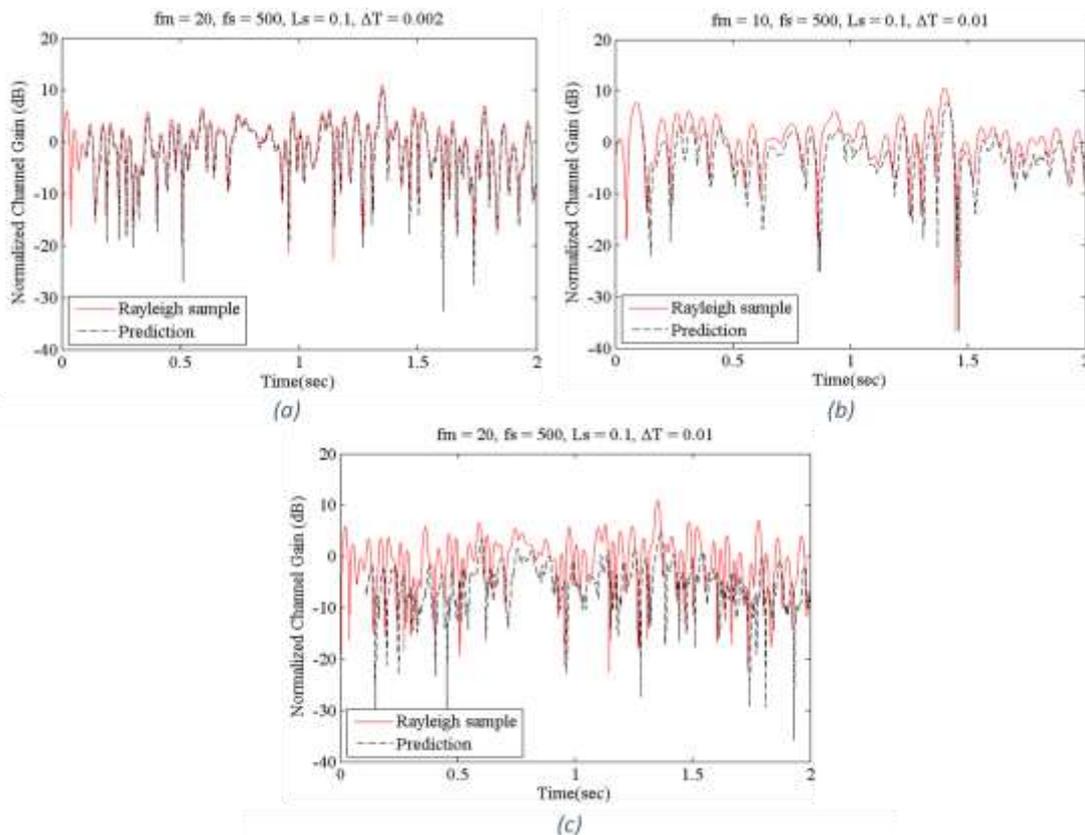


Figure 11. Yule Walker AR [14] prediction vs correlated Rayleigh channel samples. f_m and f_s are the maximum Doppler frequency and sampling frequency measured in Hertz, while L_s and ΔT are the training sequence duration and the prediction timing offset measured in seconds.

Channel prediction: A toy example

It is evident that powerful yet efficient prediction techniques have a critical role in channel state stabilization with respect to the extension of both coherence time and bandwidth. In the simplest case, wide-sense stationary (WSS) channels can be somewhat accurately predicted via simple autoregressive (AR) techniques. AR represents a class of filtering techniques that make predictions by extracting the autocorrelation from a consecutive sequence of known samples from the immediate past. For the sake of stabilization, a timing offset needs to be considered, which includes the processing delay required for performing prediction itself in addition to the other steps of downlink preprocessing.

To demonstrate the applicability of AR to simple channel prediction, we portray a group of comparisons between temporal-correlated Rayleigh sequences and their corresponding predictions made via the Yule Walker method [14] in Fig. 11. Here we may observe that an environment under the influence of weak Doppler effect can be predicted accurately within a short time span into the future, since the inherent temporal correlation is sufficiently strong under such conditions. Unfortunately, real-world cell-free channel states rarely satisfy the WSS criterion due to the change of large-scale parameters and long-term statistical CSI.

Recent activities in channel prediction have identified more powerful variants of AR such as the spatio-temporal AR (STAR) of [15] for massive MIMO-OFDM, which exploits channel sparsity in the delay-angle domain. Other channel prediction methods have also been studied in the literature, with neural network (such as [16]) and sinusoidal modelling (such as [17]) being the most popular machine learning and physical approaches, respectively.

Delay-Doppler mapping could make prediction better and simpler

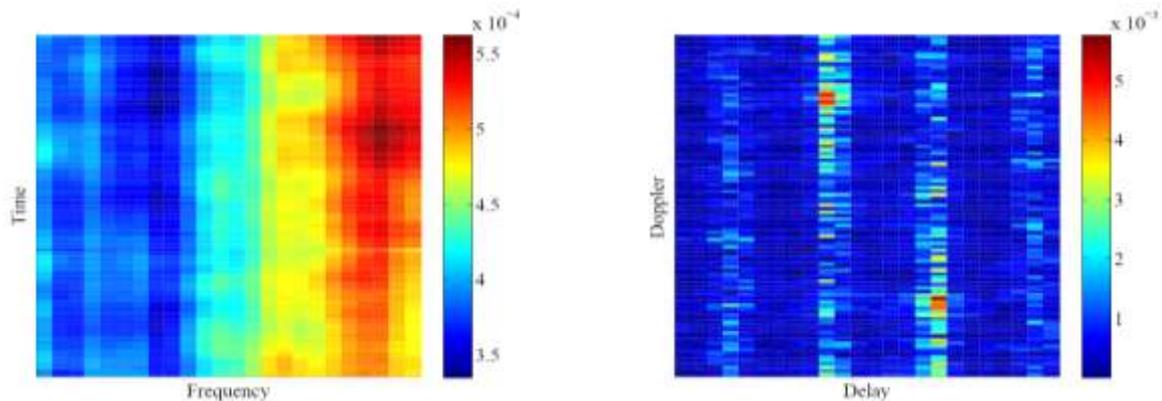


Figure 12. Time-frequency channel response (left) vs DDM (right). Converted from the CSI in Fig. 8c.

Predicting the time and frequency responses of real-world cell-free channels is challenging because of the dispersions caused by a multitude of uncoordinated, moving scatterers and reflectors in a complex background. However, the superposition of these dispersions can be decoupled to some extent if the channel states are measured in the delay-Doppler domain instead of the time-frequency domain. This approach is expected to reveal a sparse and more predictable channel representation, which we elaborate as follows.

In radar communications and Global Navigation Satellite System (GNSS), delay-Doppler map (DDM) has been conceived for estimating geophysical parameters via remote sensing [18]. DDM conventionally extracts the dominant Doppler and delay components via a spread-spectrum matched-filter approach [19]. More specifically, a spreading sequence such as a pseudo random or Walsh-Hadamard sequence that undergoes Doppler shift and multipath scattering can only be recovered by a synchronized copy of itself having the same ('matched') Doppler and delay values. We note that a rough DDM can also be derived from 2D time-frequency domain CSI directly via the symplectic Fourier transform as performed by orthogonal time frequency space (OTFS) demodulation [20]. These matching Doppler and delay values appear as peaks in the resulting DDM (e.g., Fig. 12), which will reveal the behaviours of major scatterers and reflectors in the environment. Higher resolution DDMs are generally more powerful for environmental characterization but they consume more computational resources. Regardless, DDM should be able to lower the requirement for channel prediction if it can provide a sparse representation of the radio environment. The computational resources expended in generating a DDM will need to be considered in future assessments.

Building on the cell-free channel stabilization concept and Doppler compensation presented in this section for WP 3.1, further work has been performed to link with the RIS studies in WP 3.2 and this is described in the following section 2.2.

2.1.3 Virtual Prototyping of Large-Scale Antenna Arrays

Even though electromagnetic simulation of antennas and antenna arrays is a relatively well-established field, fast and accurate electromagnetic characterization of antenna arrays with large number of antennas still remains a very challenging task. The ability to use accurate electromagnetic simulations within antenna array design flow is crucial for antenna designers to find a cost-effective optimal antenna design that satisfies given specifications. This is particularly problematic for antenna arrays working at mmWave frequencies, as the size of the computational problem grows dramatically making.

One of the objectives of Task 3.1 is to develop an accelerated electromagnetic simulation platform for large-scale antenna array virtual prototyping that would include a fast full-wave electromagnetic simulator that is able to accurately predict an electromagnetic response of antenna arrays consisting of a large number of elements. To achieve that, a novel full-wave electromagnetic solver based on method of moments and layered-medium formulation [21] was tailored for the antenna array modelling. In particular, the solver has been enabled to tackle planar and non-planar antenna array structures, to handle typical electromagnetic excitations required for antenna characterization, and to produce the visualizations typically needed for antenna designers (far field, gain, network parameters).

Fig. 13 depicts an 8x8 patch antenna array design for 28 GHz 5G mobile base station [22] which was used to benchmark the developed capabilities of virtual prototyping platform. The discretized model of the design in Fig. 13 required 58 904 method-of-moments unknowns situated in layered medium, and the entire structure was characterized in 2.5 min using 1.12 GB of memory on a single r5.24xlarge Amazon EC2 node. The frequency response from 27.25 GHz to 28.75 GHz was obtained (20 s per frequency, 9 frequency samples in total) using a mesh consisting of 58 904 method-of-moments unknowns situated in layered medium. The return loss (S_{11}) and gain are depicted in Fig. 14 and matches well the response obtained via state-of-the-art commercial solvers [22].

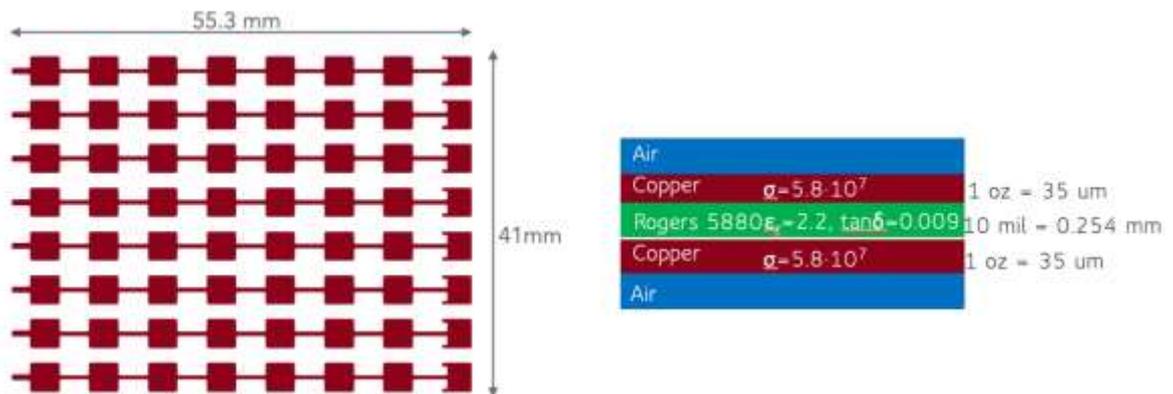


Fig. 13. 8x8 patch antenna array design for 28 GHz 5G mobile base stations [18].

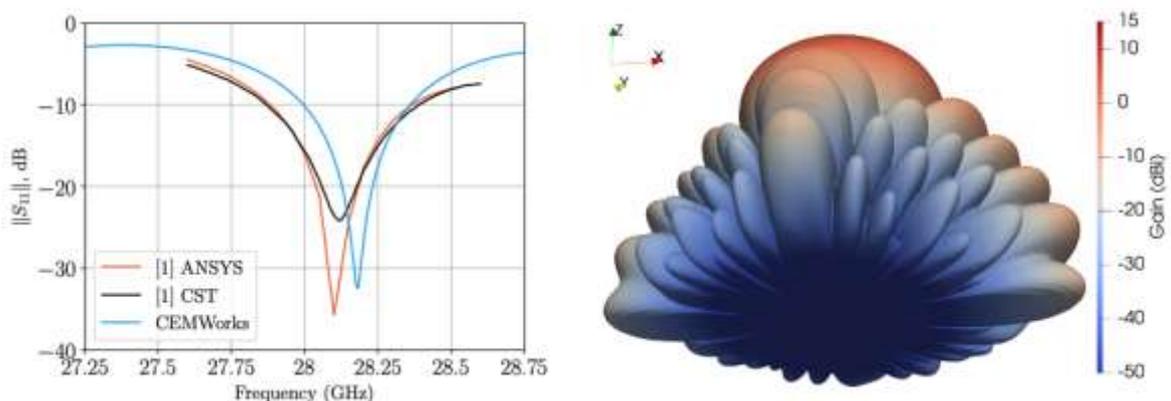


Fig. 14. (left) Simulated return loss and (right) Gain of the 8x8 patch antenna array.

The demonstrated ability to quickly simulate antenna arrays with large number of elements (64) allows to include such full-wave simulation into the most complicated design flows, including the

generation of the simulated antenna parameters for machine-learning-based optimization techniques.

2.1.4 Centralized and Scalable Cell-Free Massive MIMO with Realistic Fronthaul

Cell-Free massive MIMO is considered as a promising network architecture for beyond-5G & 6G communications with its potential of supporting the exponentially increasing mobile connections and resolving the inherent disadvantage of canonical cellular networks. The fundamental physical infrastructure of the cell-free massive MIMO network is employing a large number of distributed access points (APs) which can potentially serve the user equipment (UE) in the service area all together. The APs are connected to the central processing unit (CPU) or cloud radio access network (C-RAN) [23] through the wireless or rate-limited wired fronthaul links. Different network implementations for cell-free massive MIMO have been studied in the previous work. For the distributed implementation [24], APs perform local channel estimation and combining, and CPU performs final decoding using the combined signals from all the APs. In the centralized implementation, one possible operation is that APs perform local channel estimation then forward the channel estimates and received data signals to CPU for combining [25]. Another operation is the fully centralized implementation [26], where APs forward both received pilots and data signals to CPU for combining. No matter which operation is employed, the resulting fronthaul signalling load should be considered to meet the requirement of practical cell-free networks with rate-constrained fronthaul. In previous research [27], an optimal uniform scalar quantizer was designed for the cell-free uplink forwarding and Bussgang decomposition [28] was employed to model the nonlinear quantization process. Transform coding was employed in [29] to decorrelate the signals first, then the transformed signals were quantized using scalar quantization. Gaussian test channel was employed to model the quantization noise. The optimal vector quantization based on Linde-Buzo-Gray (LBG) algorithm was proposed in [25] for a centralized cell-free massive MIMO network with correlated Rayleigh fading channel. To obtain the optimal vector codebooks, significant training is required especially under the high-mobility scenarios, which is prohibitive for a large distributed system. To balance the performance and the computational complexity, a coarse vector quantization based on the lattice codebook was proposed in [30]. It has been shown in [30] that, under a correlated channel model, a close-to-optimal performance can be achieved by the near-ellipsoidal lattice codebook without requiring much training.

To implement a relatively realistic cell-free Massive MIMO network, a lattice quantization is studied under a large-scale cell-free Massive MIMO regime. For the signal processing, a fully centralized operation is considered. Specifically, in the uplink transmission, after receiving data from UEs, each AP quantizes the locally received data using the local codebook, then forwards the quantized data to CPU for further channel estimation and combining. Lattice quantization and Bussgang decomposition are briefly explained in the next session. Then the network architecture as well as the channel model between the AP and UE are introduced. At the end, two lattice-based quantization schemes used in the centralized and scalable cell-free massive MIMO network are shown.

Lattice Quantization and Bussgang Decomposition

In a n -dimensional real Euclidean space \mathbb{R}^n , a lattice Λ is defined as the integer span of the basis vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. For each lattice point $\mathbf{m} = (m_1, \dots, m_n) \in \Lambda$, the Voronoi region $V(\mathbf{m})$ is defined as the region consisting of all points at least as close to \mathbf{m} as to any other lattice points in \mathbb{R}^n [31]. The Voronoi region of the lattice $V(\Lambda)$ is specially defined as the Voronoi region of the origin. Voronoi codes have been used for vector quantization and Gaussian channel coding. Fast quantizing/decoding [32] and encoding [31] algorithms are available for the well-known A_n, D_n, E_n lattices. When lattice is used for quantization, with the finite quantization bits, only a subset of the lattice points should be included in the codebook. The region consisting of all the codebook points is defined as the support region [33]. With a rate of b bits/dimension, the ideal support region for uncorrelated Gaussian source should be $\Pi^c = V(\Lambda^c) + \mathbf{a}$, where $\Lambda^c = \{(2^b m_1, \dots, 2^b m_n) | \mathbf{m} \in \Lambda\}$ and $\mathbf{a} \in \mathbb{R}^n$ is the offset vector to ensure the boundary of Π^c contains no lattice points. $V(\Lambda^c)$ is essentially the Voronoi region of Λ shaped by a factor 2^b equally at each dimension and has a near-circular shape. Π^c is also a sublattice of Λ consisting of all lattice points (slightly translated by \mathbf{a}) in the codebook. A rule of thumb while designing Voronoi codes is to match the shape of the support region with the distribution of the input signals of the quantizer. Hence, to quantize the correlated Gaussian vector

sources, near-ellipsoidal Voronoi codes were proposed in [34]. The support region of a near-ellipsoidal codebook is defined as $\Pi^e = V(\Lambda^e) + \mathbf{a}$, where $\Lambda^e = \{(f_1 m_1, \dots, f_n m_n) | \mathbf{m} \in \Lambda\}$. Different to the near-circular Voronoi region, the shaping factor $\mathbf{f} = (f_1, \dots, f_n)$ will shape the Voronoi region of the lattice Λ differently at each dimension.

Bussgang theorem provides a linear model for the non-linear process. By employing Bussgang decomposition, the output of the quantizer can be modelled by a linear transformation applied to the input first, then plus a non-linear distortion. Leveraging the property of Bussgang theorem that the distortion is uncorrelated to the input of the quantizer, both the linear transformation matrix and non-linear distortion can be calculated through the Monte Carlo method.

Network Architecture and Channel Model

A classic architecture of the cell-free network is shown in Figure 15 with K single-antenna UEs and L N -antenna APs randomly distributed in the service area. The channel between the AP and the UE is the traditional Rayleigh channel. Specifically, the channel between AP l and UE k is denoted by \mathbf{h}_{kl} . When the correlation matrix of \mathbf{h}_{kl} is the identity matrix, the channel becomes the classic i.i.d channel, otherwise the channel is the correlated. Both uncorrelated and correlated Rayleigh channel models are considered to evaluate the performance of different quantization schemes. To make sure the network can still function when the number of UEs in the service area increases to infinity, instead of assuming that all APs can potentially serve all UEs, a serving relation matrix \mathbf{S}_{kl} is defined. If AP l is assigned to serve UE k , the relation matrix \mathbf{S}_{kl} is set as identity matrix, otherwise this is set as a zero matrix. During the simulation process, the service area is set to 1 km by 1km, the number of UEs are 40, the number of 2-antenna APs are 200.

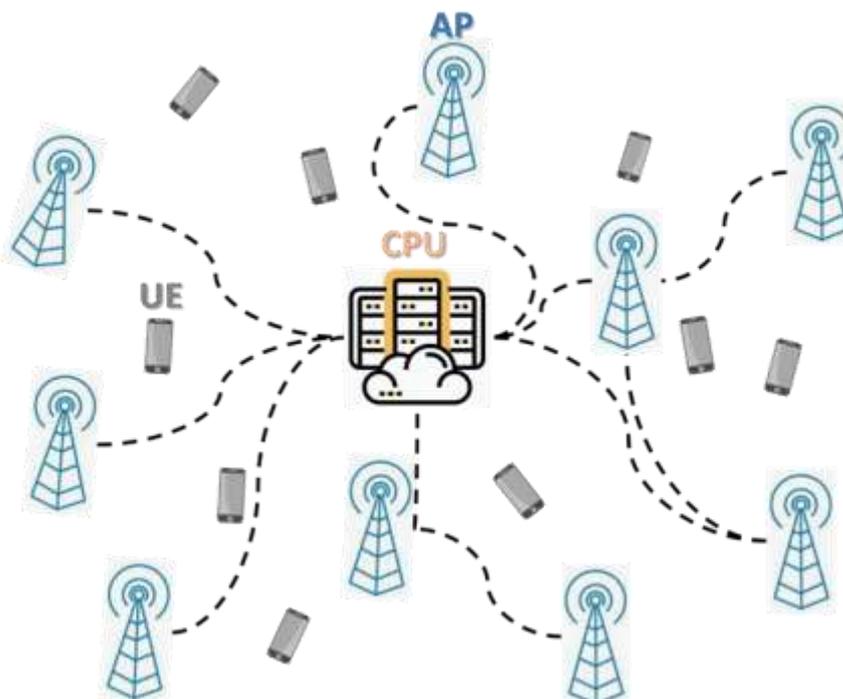


Fig. 15. cell-free massive MIMO network with K single-antenna UEs, L N -antenna APs connected to CPU through rate-constrained fronthaul links.

Two Lattice-based Quantization Schemes

After receiving data signals from UEs, each AP performs local quantization independently using the given lattice codebook. The real parts and imaginary parts of the signals are quantized separately using the same codebook since they follow the same distribution. For the first quantization scheme (Scheme 1), each AP quantizes the received signals directly using a near-circular codebook. The quantized signals after near-circular quantizer are then modelled by Bussgang decomposition. The codebook has a near-circular shape since at each dimension of the codebook, the same rate is adopted.

For the second scheme (Scheme 2), the near-ellipsoidal codebook is applied to the transformed signals at each AP locally. Firstly, at AP, the eigendecomposition of the local correlation matrix is performed. Secondly, a Karhunen-Loeve transformation based on the result of the eigendecomposition is applied to the received signals at the AP to get the transformed signals. By doing this, the transformed signals will have the similar distribution of the eigenvalues. In Scheme 2, instead of the received signals, the transformed signals are quantized using a near-ellipsoidal codebook. The shape of the codebook is adjusted to the distribution of the eigenvalue in order to match the distribution of the transformed signals. For instance, Figure 16 has shown two 2-dimensional codebooks truncated from the hexagonal lattice A_2 . These two codes are used by two 2-antenna APs ($N = 2$) independently. For each codebook, the horizontal axis represents the real parts of the transformed signals at the first dimension and the vertical axis represents the real parts of the transformed signals at the second dimension. Since the local correlation matrix at each AP can be very different, after decomposition, the eigenvalue distribution is also different for different APs. This is why in UoB1-Fig2 the shapes of the two local codebooks are different, but both are similar to the distribution of the locally transformed signals.

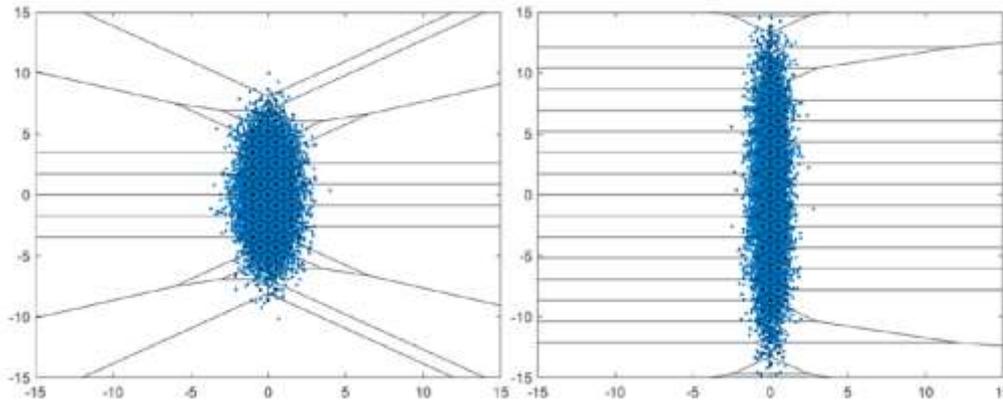


Fig. 16. Near-ellipsoidal codebooks at two APs. Blue dots represent the transformed signals (real parts). Black dots are the codebook points.

Under uncorrelated channel model, Scheme 1 and Scheme 2 are essentially the same, so only Scheme 1 is considered during the simulation. Under correlated channel model, both schemes are considered. For channel estimation, the minimum mean-square error (MMSE) estimator is employed. For centralized combining, both Maximum Ratio (MR) and MMSE combining methods are employed at CPU. Furthermore, to provide scalability to the network, the scalable alternative of MMSE proposed in [35] is also considered. The processing band is 2 GHz band (bandwidth 20 MHz). To evaluate the performance of two quantization schemes, the achievable spectral efficiency (SE) is adopted as the metric. In Figure 17 and Figure 18, it is shown that only 3 bits/dimension for MR and 6 bits/dimension for MMSE can achieve close-to-perfect fronthaul performance. In general, MR requires less fronthaul capacity than MMSE. Under MR combining, the dominant part in the Signal-to-Distortion plus Interference and Noise Ratio (SDINR) is the inter-user interference, so introducing extra quantization distortion does not decrease the SE significantly. With the same reason, at the lower end of the curves in Figure 17, the performance gap is negligible between 3 bits/dimension and the perfect fronthaul case, but the gap is bigger at the upper end indicating the UEs experiencing relatively good channel conditions. MMSE is an interference-compressing scheme, so UEs are more sensitive to the distortion compared to MR and more bits are needed to achieve close-to-perfect fronthaul performance. Moreover, the quantization distortion influences all UEs almost equally under both optimal and scalable MMSE combining methods.

It is shown in Figure 19 and Figure 20 that near-ellipsoidal codebook always outperforms the case of directly applying near-circular codebook no matter under MR or MMSE combining. This indicates that under correlated channel model, by exploiting the eigenvalue distribution of the channel correlation matrix, the near-ellipsoidal codebook has essentially improved the quantization efficiency. Furthermore, similar to the uncorrelated channel model, MR requires lower rates than MMSE generally, and UEs with relatively bad channel conditions are less sensitive to the quantization distortion under MR combining. Specifically, with near-ellipsoidal codebook, 3 bits/dimension and 6 bits/dimension are required for MR and MMSE respectively to achieve close-to-perfect fronthaul performance. Practically, one of the two quantization schemes can be specifically selected based on the real propagation terrain which introduces different level of angular deviation among multipath components.

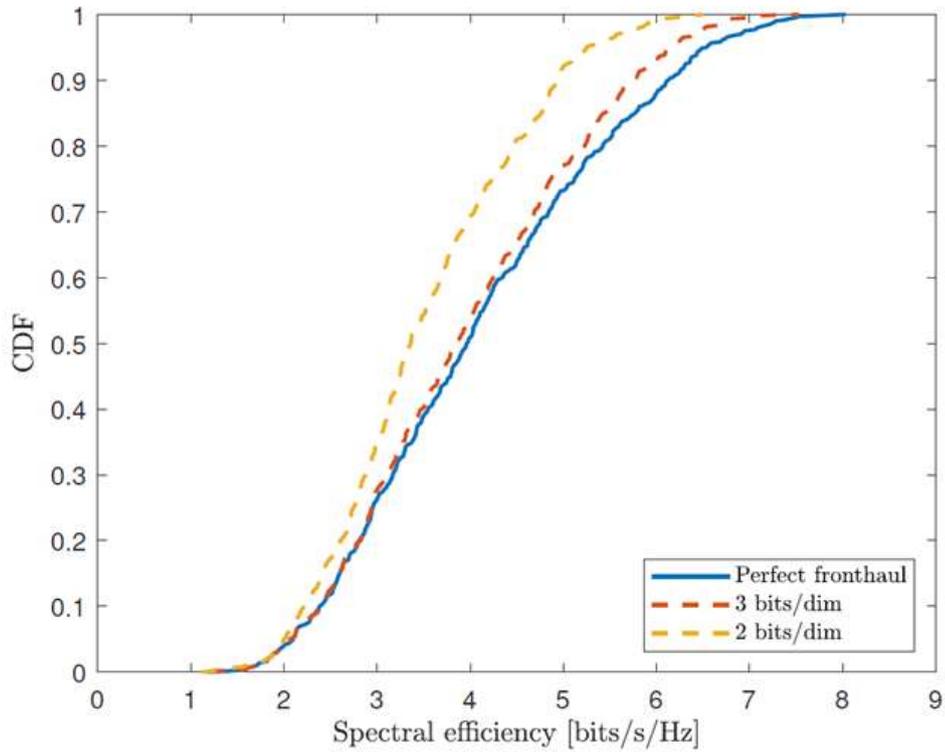


Fig. 17. CDF of the uplink SE per UE under uncorrelated channel. The MR combining scheme is used at the CPU.

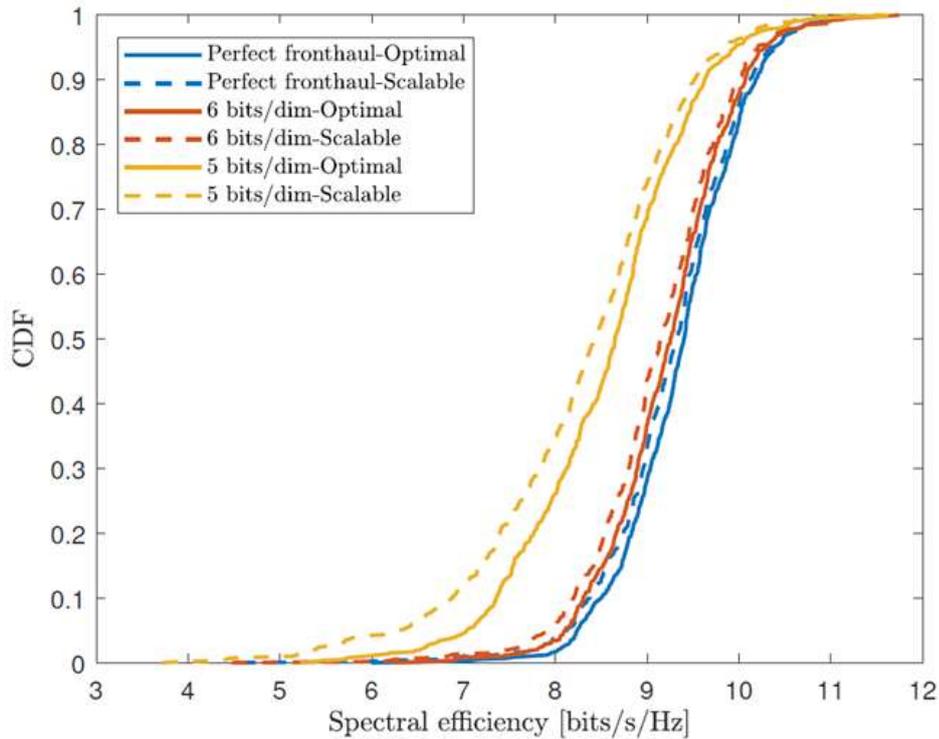


Fig. 18. CDF of the uplink SE per UE under uncorrelated channel. The MMSE combining scheme is used at the CPU.

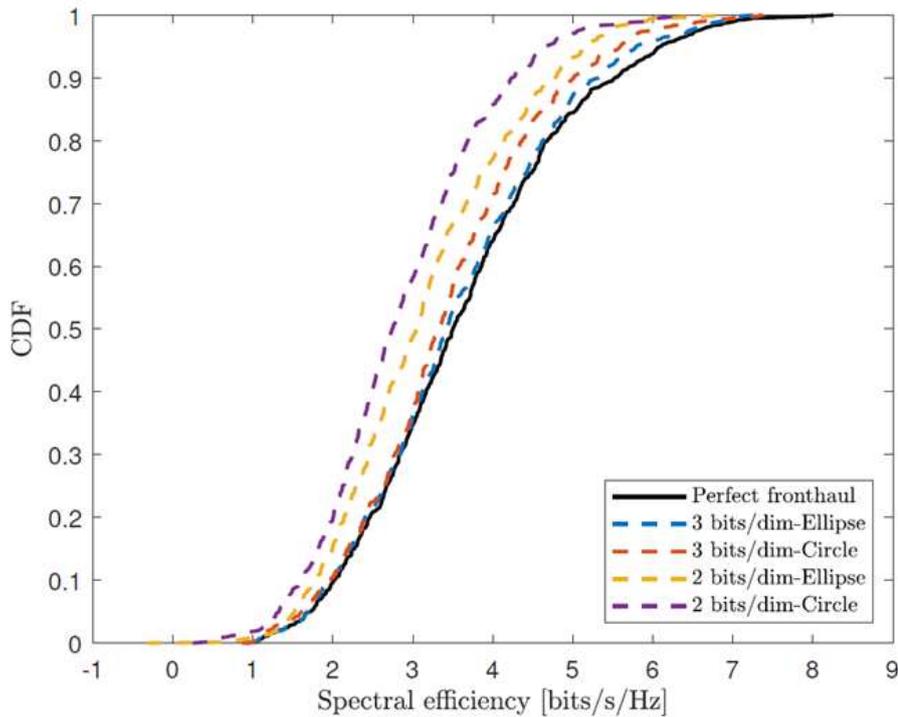


Fig. 19. CDF of the uplink SE per UE under correlated channel. The MR combining scheme is used at the CPU.

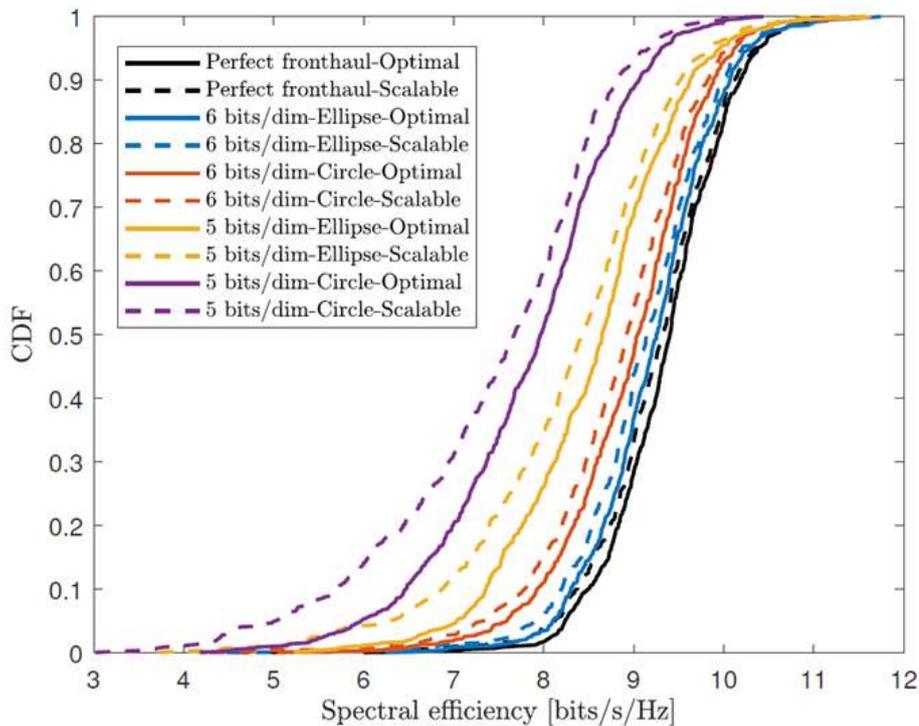


Fig. 20. CDF of the uplink SE per UE under correlated channel. The MMSE combining scheme is used at the CPU.

2.2 Reconfigurable Intelligent Surfaces

RISs can be used in multiple ways, e.g., as passive and controllable mirrors for enhancing the coverage and in low-complexity massive multiple-antenna transmitters. The RISs are a unique technology since they are aimed to intentionally and deterministically control the propagation environment in order to boost the signal quality at the receiver. In this task, we are interested in identifying the potential benefits from the incorporation of RIS technology in 5G NR and beyond networks by using system-level analysis and experiments, and in designing and optimizing them with the aid of machine learning tools. Another research and development effort will target

characterization of intelligent metasurfaces using 3D full-wave or hybrid methods which will allow for better accuracy with particular focus on scaling to large sizes and co-simulation with antenna arrays.

2.2.1 RIS – Physical-layer Options & Trade-offs

The benefits of advanced 5G networks with gigabit throughput and low-latency will be demonstrated in the next decade. Nevertheless, there are already predictions that capacities 10x greater may be required for next-generation networks beyond 2030, supporting both fixed wireless access and mobile services. A conventional sub-6 GHz Massive MIMO (“Massive Multiple-Input Multiple-Output”) wireless system uses phase array technology, where capacity can be improved indefinitely by simply increasing the number of antenna elements without limit. However, in practice the weight, size and power consumption of such a system will limit the phase array to at most 200 elements. This performance roadblock can potentially be overcome by using a reconfigurable intelligent surface (RIS) and a local transmitter.

Such reflective metasurfaces can offer a number of benefits including easy deployment (on any surface), reduced power usage (thanks to their “nearly passive” nature¹), robustness to noise (not affected by receiver noise) and multi-band operation (due to reconfigurability). All components are high-volume commercial off-the-shelf parts, leading to significant economies of scale. Some early concept demonstrators have offered encouraging results. Device performance may improve at higher mm-Wave and THz frequencies due to the changed statistics of electromagnetic propagation. However, there are many open questions around their use, including channel estimation and feedback mechanisms (how to acquire and exchange channel information to achieve beamforming i.e. beam focusing) and actual performance benefits in practice (e.g., how many reflective surfaces are needed per unit area to give better performance at reduced cost).

The surface itself will be very large and control complexity will be enormous, but nevertheless these may be more tractable engineering challenges than building ever-larger phase array systems. At much higher frequencies such as millimetre wave (“mm-Wave”, 30 GHz to 300 GHz, and beyond) a reconfigurable reflective surface may prove to be the only practical Massive MIMO solution.

This assessment considers the performance comparison of reconfigurable reflective surfaces compared to conventional phased array systems. There is indeed an opportunity to exceed the performance roadblock of conventional systems, the “price” for this being the physical size of the surface and the computing resources needed to control it. This therefore means that the suitability of such surfaces will depend on the specific use-case, so an enclosed shopping mall will be a much more suitable location for smart surfaces than a rural macro-site.

The following section discusses RIS technology as a candidate for overcoming the current phase-array limitations discussed above. This new technology might supplement or even replace conventional approaches, to an extent depending on use-case envisaged. The treatment is largely descriptive, with numerical performance data presented later. The subsequent section compares the performance of the new RIS technology against the baseline, providing some quantification of the discussion in previous sections. To do so requires consideration of both phased array limitations and possible use-cases in a working network.

Reconfigurable Intelligent Surfaces

This section discusses a promising novel technology as a candidate for overcoming the current limitations. This might supplement or even replace conventional approaches, to an extent depending on use-case envisaged. The treatment is largely descriptive, with numerical performance data presented later.

Novel metasurface based technologies can offer alternative mechanisms for implementing beamforming which could be more responsive and provide more highly focused beams for communicating between the network and customer terminals. Meta-surfaces are layered dielectric-metal structures where electromagnetic properties have been engineered to depend entirely on details of the surface. Passive frequency selective surface devices (typically employed as filters)

¹Emphasising the need for ultra-low (but non-zero) control power.

have been heavily investigated historically due to their low loss, high degree of integration, and ease of fabrication. They usually consist of periodic patches or apertures in a metallic or dielectric sheet using a variety of geometries and dielectric substrates. Such devices can be carefully designed to control incident electromagnetic radiation. New meta-surfaces are being produced which can be configured using a wide range of physical mechanisms.

With regard to available spectrum and directly analogous to more conventional mm-Wave technologies mentioned above, the terahertz frequency range (usually defined as 0.1 to 10 THz) is also expected to be particularly useful for wireless communication systems, due to the much-reduced communication wavelength (hence, possibilities for device miniaturisation etc.) and even greater bandwidth available. However, this will require technological improvements in many areas. Furthermore, it is generally accepted that THz cannot be used for highly mobile applications but might be appropriate for indoor scenarios where user movement is constrained. THz might also support mobile wireless communication indirectly by offloading static users from a lower band (e.g. microwave) to THz, making more room for mobile users in the lower bands.

Devices

Electronically tuneable arrays of meta-surface devices have attracted much attention recently. This has led to the idea of the reconfigurable intelligent surface (RIS) as a revolutionary technology for smart (cost-effective) and energy-efficient (green) communication. Specifically, a RIS is a planar array consisting of a large number of “nearly passive” elements (e.g. low-cost printed dipoles with ultra-low power consumption which controls their switching between on/off states). In principle, by integrating active elements (e.g. positive intrinsic-negative PIN diodes, varactor diodes and microelectromechanical switches) a reconfigurable RIS device can be fabricated.

Each element is then capable of electronically controlling the phase of incident electromagnetic (plane wave) radiation to collaboratively achieve beamforming. The RIS typically functions as a “smart” reflector. This is conceptually similar to conventional array systems but the radiation from a RIS comes from an external source, incurring no additional power consumption. The key difference between a RIS and a reflectarray² is that a RIS is located with neither transmitter nor receiver but can be arbitrarily deployed in the propagation environment to aid communication.

A RIS functions as a full-duplex transparent relay that does not amplify signals but can synthesise the scattering behaviour of an arbitrarily shaped object. In principle, a RIS can carry out most things normally implemented using traditional beamforming. Crucially, a RIS cannot beat an active array of the same size (due to propagation loss, there being no RIS amplification) but if a larger surface is used than in conventional relays or multi-antenna transceivers, it can deliver comparable performance.

A RIS device can be configured to operate as an anomalous reflector and realised at low cost, size, weight and power. Compared with phased arrays or multiple-antenna systems, each RIS may require a relatively large number of antenna elements but each of these only needs to be backed by fewer, less costly components. In addition, RISs can be nearly passive as active elements are usually not necessary. While some energy is needed to control RIS switches and receive one-off control signals for configuration, no dedicated power supply is needed for signal transmission.

Conversely, in a traditional network, any relay must be equipped with a dedicated power source and the necessary front-end circuitry. This results in increased network power consumption and may require significant capital expenditure for deployment.

To summarise, the potential RIS advantage over traditional relays is that large surfaces can be implemented with lower power consumption and cost since there are no amplifiers, but only power dissipation in the hardware controlling the configurability. The main drawbacks are the reduced signal range due to lack of amplification, together with any delay in reconfiguring the surface.

A major issue for enabling wireless communications in the mm-Wave and THz bands is the large path loss, especially when the transmission path is blocked by obstacles. Common solutions (e.g. relays) usually imply increased network power consumption, complexity and cost. Therefore, it is

²A reflectarray consists of an array of unit cells illuminated by a feeding antenna, focusing a beam in a similar way to a parabolic reflector.

necessary to explore new hardware architectures to enhance network coverage for transmission in high frequency bands, without such overheads. In this context, RISs emerge as a promising enabling technology for 6G networks in the mm-Wave and THz frequency bands.

For realistic network applications (e.g. Massive MIMO), establishing wireless communication links will first require gathering channel state information (CSI) to determine link reliability and aid connection setup etc. Equipping a RIS with low-power sensors to collect CSI would increase cost and power consumption of the entire surface. Accordingly, the vast majority of current research activities rely on the assumption that RISs are not equipped with sensing elements, so any CSI must be gathered by a separate, purpose-built network.

Some early protocols and algorithms to estimate channels and control a RIS in real time have been explored, though much research remains to be done:

- One approach is to transmit a test sequence repeatedly and measure the received signal when using different RIS configurations. Even when CSI has been acquired, it is computationally complex to select appropriate time-delays, particularly in wideband channels. To reduce complexity, it may be necessary to group adjacent RIS elements and accept a compromise on performance.
- Another approach is to alter the “nearly passive” nature of the RIS and allowing a few elements to do sensing and channel estimation directly at the RIS. Theoretically, this might work better at higher (mm-Wave/THz) frequencies due to changed statistics of the propagation channels.

Appropriately configuring on/off state of the PIN diodes or bias voltage of the varactors controls the (macroscopic) beamforming. RIS configurability is then limited to a finite set of states per element with given delays and amplitudes (mutual coupling between adjacent elements is a possible limitation). For example, in a RIS made of two PIN diodes the unit cells may be designed to simply rotate the phase of an incident radio wave of 0, 90, 180, and 270 degrees; i.e., four discrete states needing only two bits configured.

It is also worth noting that deployment strategy for RISs is generally different than for active communication nodes (such as base stations i.e. BSs or relays) for several reasons:

- Since RIS-reflected channels suffer severe path-loss due to passive reflection without signal amplification/regeneration, RISs need to be placed close to the transmitter or receiver. This is in sharp contrast to active BSs that need to be geographically well separated for coverage maximisation.
- Thanks to the much lower cost of RISs compared to active BSs, they can be more densely deployed in the network. This may prove a significant advantage in Cell-Less networks, decreasing latency for improved performance.
- One challenging issue when using active communication nodes arises from their mutual interference, which greatly complicates deployment. In contrast, since RISs are passive their reflected signals decay in power rapidly over distance. As long as RISs are deployed sufficiently far apart, their mutual interference is practically negligible.

Performance Comparison

This section compares the performance of new RIS device technology against the baseline, providing some quantification of the discussion in previous sections. To do so requires consideration of both phased array limitations and possible use-cases in a working network. While initial progress has been made studying sub-6 GHz RIS applications, the more complicated mm Wave systems remain harder to analyse in detail.

Baseline: Phased Array Limitations

As outlined above, conventional Massive MIMO antenna arrays will face significant challenges when increasing beamforming capability to support future customer demand, due to technical limitations on all of the following:

- hardware cost (e.g. phase shifters)
- power consumption
- complexity (feeder network)
- physical size (accommodation)

Current commercial systems employ approximately 64 antenna elements, with 100-200 a likely upper limit in any practical implementation. With regard to size, fundamental physics means individual antenna elements have dimensions in the order of the radiation wavelength (e.g., 10cm at 3GHz).

Accordingly, Massive MIMO network designers must trade off channel capacity improvements against size and power requirements. Using representative network component and propagation parameters, Figure 21 shows how a limiting 200-antenna Massive MIMO system can support a cell rate (aggregated over all 10 users) in the order of 1 Gbps but only if relatively complex M-MMSE (multi-cell minimum mean-squared error) signal processing is employed. M-MMSE complexity grows quadratically with number of antennas, so a 200-antenna system has ten times the compute complexity of a 64-antenna system. As computational cost also translates into consumed power, a simpler (less power-hungry) baseline MR (maximum ratio combining) approach may be more appropriate for most users, albeit at reduced communication rate³.

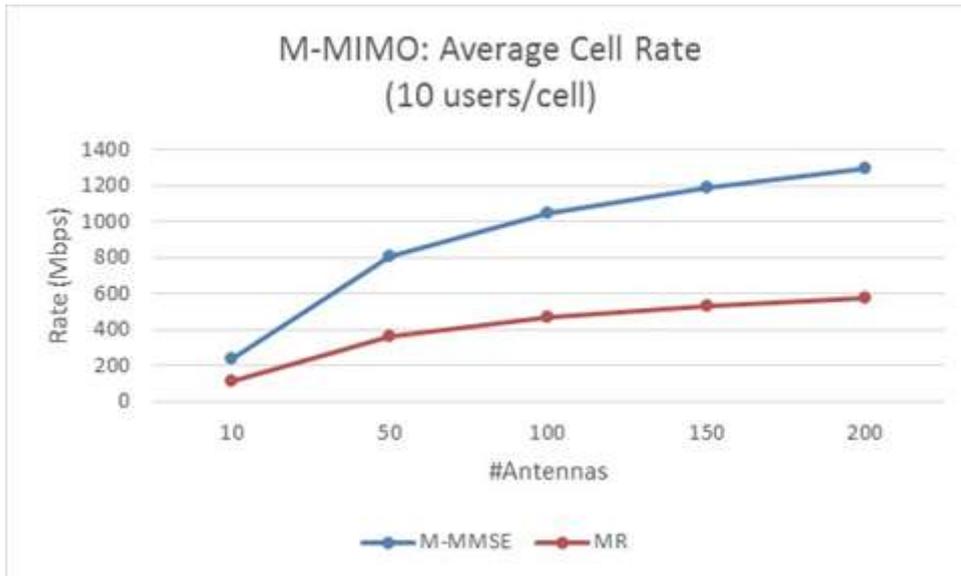


Fig. 21. Massive MIMO Cell Throughput from [35].

Network Use-cases

In view of the above limitations, we consider how new RIS technology might supplement or even replace conventional approaches, to an extent depending on use-case envisaged.

Specifically, two use-cases are considered here:

- Our primary interest is in a purely reflective RIS panel at some position between Tx and Rx, taking on either no active beamforming function or all of the beamforming function, or maybe something shared. As a (trivial) special case, the simplest comparison is then a beamforming RIS reflector co-located with a non-beamforming Tx, as compared to a straightforward beamforming Tx. The RIS device supplements conventional technology (which is limited to about 200 antenna elements).
- We also consider situations where a combination of transmitter (i.e. power source) and RIS with appropriate DSP on a physical link might replace a conventional solution. This could offer a viable alternative when a phased array approach beyond ~200 antenna elements is not feasible. The RIS device (plus transmitter) replaces conventional technology.

For the sake of comparison, our baseline scenario in Figure 22 considers a rooftop-mounted base station transmitting to a static indoor user. A thick wall between them causes significant propagation losses but the signal can pass through a nearby window with only minor losses. A RIS (for example) deployed inside the window could capture signal energy and re-radiate it to the receiver, bypassing the wall. To ensure a beam focused towards the user, the RIS would need to be reconfigurable.

³In theory, a 64 antenna Massive MIMO MMSE calculation might need approximately 64x64 more computation than the MR equivalent; efficient software coding techniques would reduce this estimate

- In use-case 1, the RIS aids direct Massive MIMO transmission from base station to user via a phased array (in the limit of an impenetrable wall, the RIS provides the only route for Massive MIMO communication to the receiver).
- In use-case 2, the RIS replaces direct Massive MIMO transmission from base station to user, powered by a suitable transmit antenna with appropriate DSP.

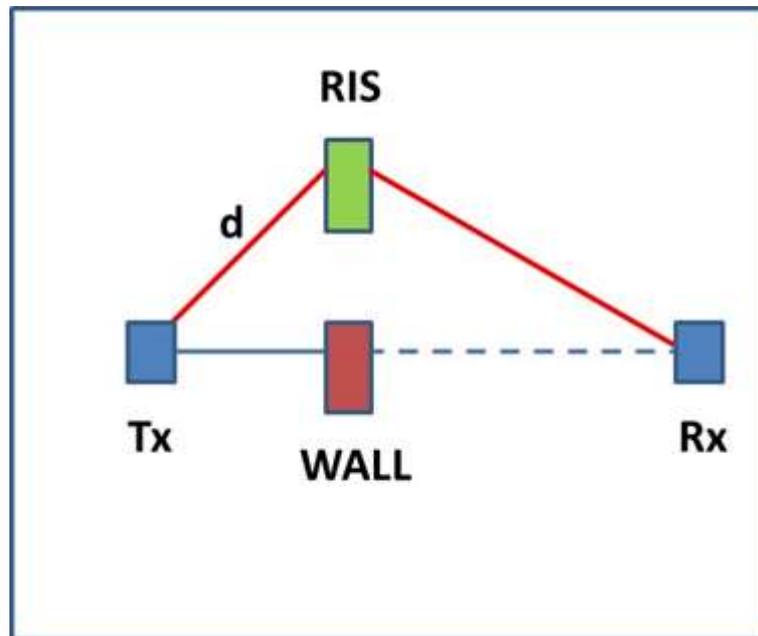


Fig. 22. Network Use-case.

Obviously, there are many variations depending on factors such as

- whether base station and user are situated indoors or outdoors
- balance between LOS (line-of-sight) and NLOS (non-line-of-sight) propagation
- precipitation/foilage losses en-route
- user mobility

Some particularly interesting scenarios might utilise higher frequency bands (mm-Wave and THz) for communication. Since the corresponding wavelengths shrink by orders of magnitude (compared to current microwave spectrum) in these newer bands, electromagnetic radiation suffers greater attenuation. This in turn increases the importance of NLOS propagation, reflection and scattering. Consequently, RISs may become an enabling technology for controlling radiation and enhancing connectivity in such higher frequency networks whenever LOS communication becomes problematic.

However, at higher mm-Wave and THz frequencies, statistics of the electromagnetic propagation channel are significantly different to behaviour in the conventional regime. Further research on relevant channel modelling and estimation is therefore needed to better understand and quantify RIS performance at such frequencies.

RIS-Massive MIMO Comparison (sub-6GHz)

RIS technology employs a frequency selective meta-surface to implement beamforming. This potentially offers advantages of low loss, high degree of integration and ease of fabrication. In this sub-section, we consider the comparison between RIS and Massive MIMO technologies.

Limits

In general, we can explore possible maximum size RIS systems to understand what performance might be obtainable in an ideal scenario. There is an upper limit on number of elements possible in a real RIS (imposed largely by the fabrication process), together with constraints on transmission power (e.g. due to health and safety precautions). The user-rate available may then be calculated with suitable assumptions about the noise environment. Reasonable parameter estimates (e.g. maximum and minimum RIS sizes) are shown in Table 1 below. The corresponding user-rate is shown in Figure 16 suggesting that maximum spectral efficiency achievable is likely to be in the neighbourhood of 20bps/Hz, representing a significant performance improvement of ~150% over

current Massive MIMO. We conclude that below 6Gbps, RIS devices may support a welcome (but not excessive) capacity upgrade, so the main advantage of RIS technology in this frequency range lies in providing more sustainable (“greener”) network solutions.

PARAMETER	VALUE
Minimum RIS Size	256 elements
Maximum RIS Size	5000 elements
Maximum Transmit Power	50dBm
Noise	-85 ± 10 dBm
Maximum Massive MIMO Size	200 elements
Bandwidth (sub-6GHz) per user	10MHz
Bandwidth (mm-Wave) per user	100MHz [36]

Table 1. Parameter Limits.

In a RIS array, it is commonly assumed that there is negligible EM mutual coupling interaction between elements, which is strictly only valid if RIS antennas are well separated from one another. When the overall size of the array is constrained (as here), increasing the number of antenna elements naturally implies that antenna separation decreases, which can alter the radiation pattern and have a substantial impact on capacity [40].

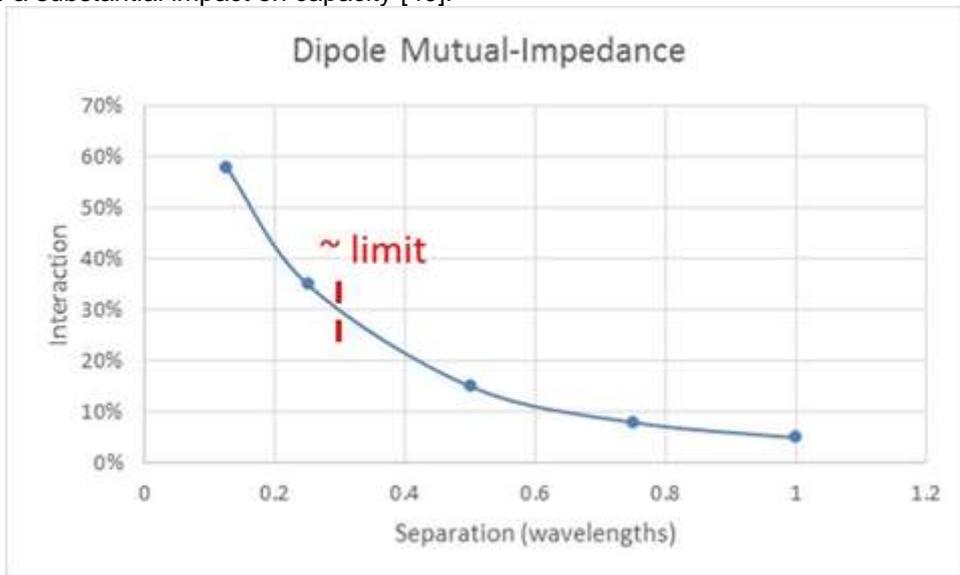


Figure 23. RIS Mutual Coupling Interaction.

Furthermore, the effect will be greater in rectangular (rather than linear) arrays due to the increased number of adjacent elements. As a simple illustration, for two typical side-by-side dipoles the effect may be well-quantified by the ratio of mutual coupling interaction to direct antenna impedance [41] relative to some acceptable limit, as shown schematically in Figure 23. This demonstrates that mutual coupling between scattering elements of a RIS can significantly affect capacity at separations below half a wavelength [42]. We neglect such effects in our simplified treatment here, so any RIS capacity predictions will be correspondingly optimistic.

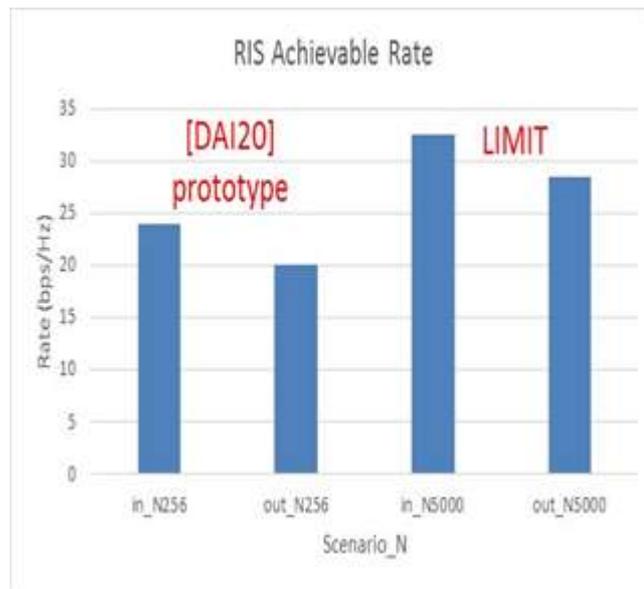


Fig. 24. Maximum Achievable Rate Varying Noise.

The user-rate corresponding to maximum size RIS systems is shown in Figure 24 suggesting that maximum spectral efficiency (SE) achievable is likely to be in the neighbourhood of 20bps/Hz, representing a significant performance improvement of ~150% over current Massive MIMO. We conclude that below 6Gbps, RIS devices may support a welcome (but not excessive) capacity upgrade, so the main advantage of RIS technology in this frequency range lies in providing more sustainable (greener) network solutions.

Similarly, we may compare RIS and Massive MIMO technologies in terms of performance per given number of antenna elements. Figure 25 shows that below 200 elements there is a choice between Massive MIMO and RIS, the latter at significantly reduced rate. Above 200 elements, only RIS solutions are feasible at some 50-80% of the (hypothetical) Massive MIMO rate. Beyond 5000 elements, no technology solution is available. Results for the equivalent outdoor use-case are broadly similar.

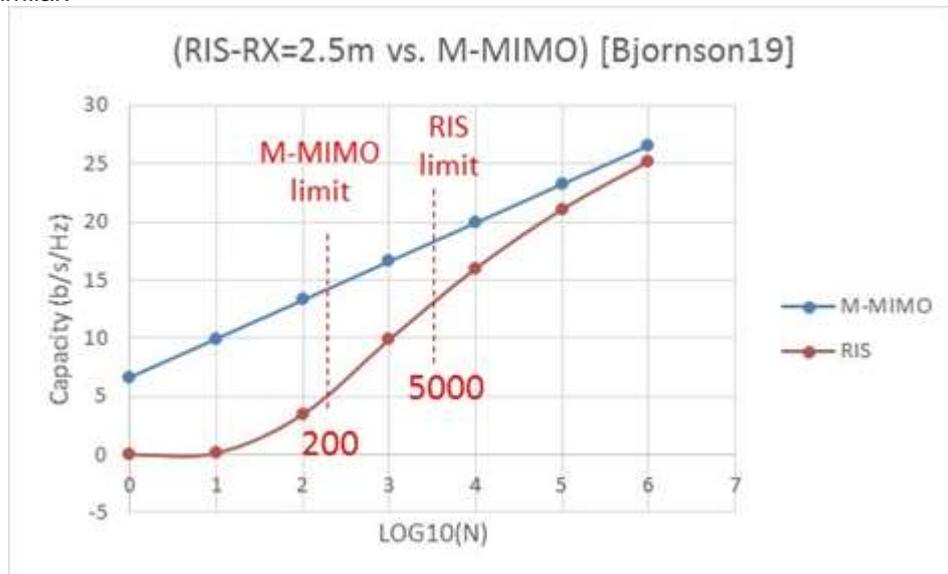


Fig. 25. Indoor Scenario: $D(\text{TX-RIS}) = 25\text{m}$.

Performance may be further quantified in terms of the network use-cases identified above.

Use-case 1

Recently, a prototype high-gain, yet low-cost, RIS employing 256 two-bit elements has been developed [43]. The device combines functions of phase shift and radiation together on an electromagnetic surface, using PIN diodes to realise phase shifting for beamforming. PIN diode

switching frequency can be up to 5 megahertz, corresponding to a switching time of 0.2 microseconds. This is much smaller than typical channel coherence times (of order milliseconds) so well suited for mobile applications with time-varying channels.

The size of the surface is 800 mm x 800 mm and the distance between the primary feed and the surface is 720 mm. To achieve a similar gain, a conventional phased array might require 64 elements for a total array size of 1m x 1m. This is an example of use-case 1 above, with source and RIS almost co-located.

For the same total radiated power (64 W), power consumption of the RIS is about 153 W, while that of the conventional phased array is about 370 W. Hence, the proposed RIS can reduce power consumption by some 60%, while radiating power comparable to a conventional phased array for static users. The prototype uses a relatively small number of beamforming solutions previously calculated off-line by a purpose-built controller. These are then pre-loaded into memory, avoiding real-time beamforming computations.

As an example of use-case 1, this is an encouraging demonstration of the RIS concept, offering a reduction in both physical size and power requirements. Naturally, many additional questions remain over performance and coverage. Crucially, the user is relatively close to the RIS (i.e., user and RIS effectively co-located) so the communication channel experiences almost no free-space attenuation.

Use-case 2

In terms of the use-case 2 outlined above, performance of a RIS system will clearly depend on the distances between base station, user and RIS. For comparison purposes, consider a base station communicating with a fixed user via a RIS positioned at a variable intermediate distance from the user. Figure 25 shows the number of RIS elements required to match (at a 90% or 50% level) performance (i.e. capacity b/s/Hz) of a current Massive MIMO (64-antenna) system. For reference, the red asterisk indicates the (use-case 1) prototype mentioned above.

The blue 90% curve emphasises that to fully match performance of conventional phased arrays, it will normally be necessary to install a very large number of RIS elements⁴. However, early theoretical studies [44] do suggest it may be possible for a RIS of perhaps 500 elements (consistent with current laboratory prototypes) to contribute successfully to 5G Massive MIMO, as shown by the brown 50% curve for very short User-RIS distances (e.g. < 5m). This could indeed be helpful in the short-term, perhaps using a RIS instead of full Massive MIMO to bypass obstructions when lower customer performance is acceptable.

If the RIS-user distance is small, the RIS is relatively close to the user (effectively co-located); a relatively small number of RIS elements is then needed to match Massive MIMO (as for the Figure 26 prototype). If the RIS-user distance is of the order of kilometres, then only a small proportion of the transmitted power will reach the RIS to be beamformed to the receiver. This low signal strength may nevertheless be sufficient to support communications if there is no other available path (direct or indirect) e.g. from a surface macro-cell to a remote reflector illuminating an underground space.

Future work will be required to explore wider scenarios where the distance between user and RIS might be increased, representing more complex environments.

In general, there will also be a trade-off between computational complexity of RIS control and the equivalent Massive MIMO digital signal processing overhead. Simple heuristics scaling linearly with size have already been shown to work well in practice [45]. However, RIS control power cannot be fully predicted before channel estimation and configurability algorithms have been developed, validated and standardised.

⁴For example, with 5m User-RIS distance, matching Massive MIMO at 90% requires 3000 RIS elements while matching at 50% requires only 500 elements.

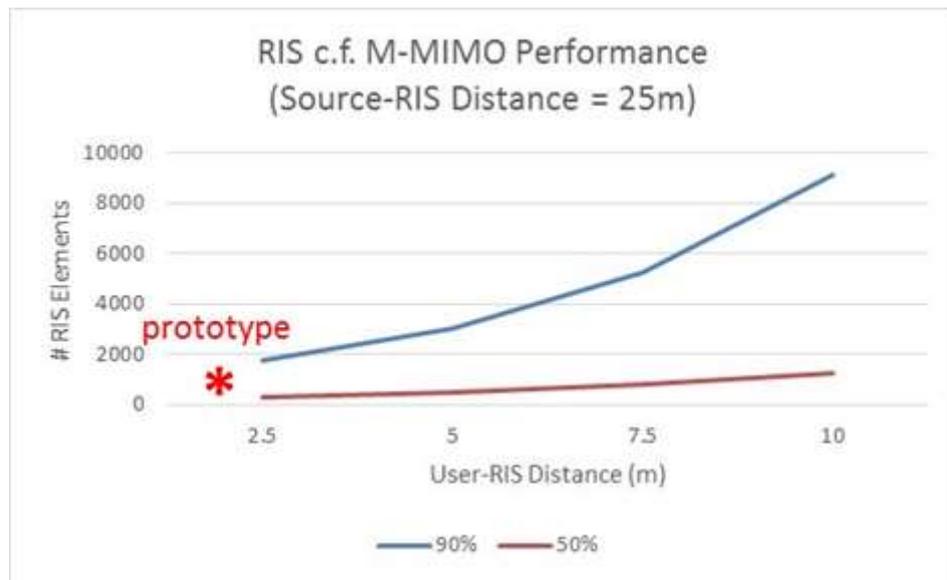


Fig. 26. RIS and Massive MIMO Performance Comparison.

RIS-Relay Comparison (sub-6GHz)

In this sub-section, we seek to make a fair comparison between RIS-supported transmission and conventional decode and forward (DF) relaying, with the aim of determining how large a RIS needs to be to outperform relaying. To this end, we consider both technologies by computing optimal transmit powers and optimal number of elements in a RIS to allow a fair comparison. This is an example of Use-case 2 with source-RIS/relay distance 25m or 100m for indoor or outdoor scenarios (respectively).

Broadly speaking, a RIS is a software-defined surface which can be realised at low cost, size, weight and power (C-SWaP). Configurability is ensured through low-power and low-complexity electronic circuits (switches or varactors). Alternatively, network relays must each be equipped with a dedicated power source and the necessary front-end circuitry for reception, processing and re-transmission. Hence, the use of relays may result in an increase of network power consumption and require a larger capital expenditure for deployment.

The idea of configuring the RIS to beamform its received signal towards the destination is the same scenario as for relays, the key difference being that a relay actively processes the received signal before retransmitting an amplified signal, while a RIS passively reflects the signal without amplification but with beamforming. The relay achieves a higher signal-to-noise ratio (SNR) at the cost of a penalty due to two-hop transmission.

We consider communication from a single-antenna source to a single-antenna destination i.e. a single-input single-output (SISO) channel. The capacity can potentially be increased by involving additional equipment (RIS or relay, deployed at the same location) in the communication.

If the destination requires a particular data rate, this can be used to identify the required transmit power for each of the communication setups. We consider data rates (spectral efficiency) of 5, 15 and 40bps/Hz representing LTE, Massive MIMO and some likely “future” scenarios (respectively). Channel gains are evaluated at a carrier frequency of 3 GHz.

The total power consumption of the system consists of both transmit power and dissipation in hardware components. Network energy efficiency (EE) may then be defined as “the number of bits that can be reliably transmitted per unit of energy”, a high EE thus implying a suitably “green” network.

Figure 27 shows the transmit power required to support such indoor data rates using SISO, DF and RIS solutions. Only the RIS network remains feasible (within the allowed maximum 50dBm from Table 3.1) across the full range, the others requiring too much power at higher future rates. RIS technology out-performs DF relaying above Massive MIMO rates.



Fig. 27. Indoor Transmit Power.

With regard to EE shown in Figure 28, while none of the systems can adequately support future rates, RIS offers the most efficient solution for LTE, Massive MIMO and beyond. This indicates strong potential for the technology wherever energy consumption is a high priority and “green” solutions are preferred or mandatory.

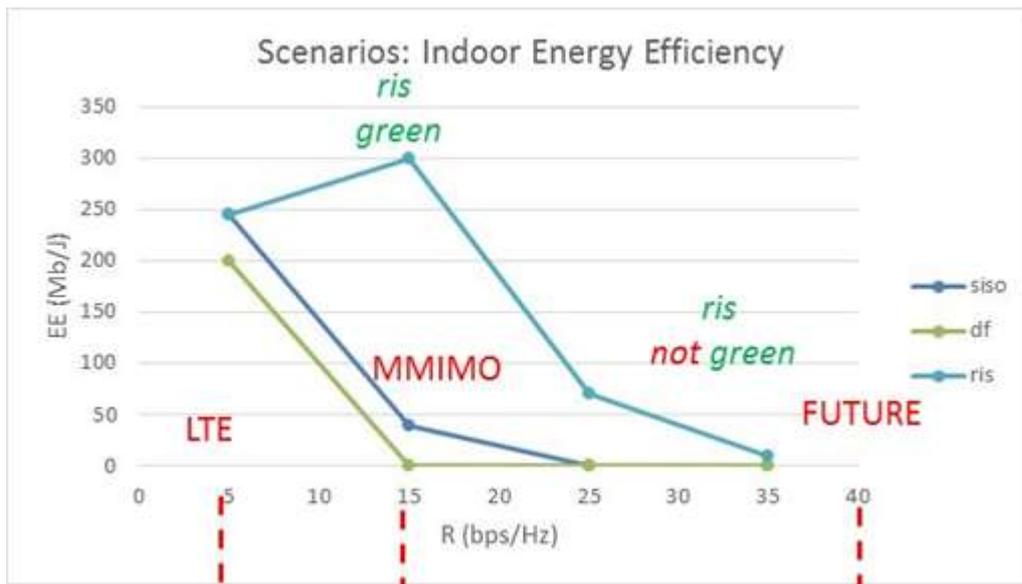


Fig. 28. Indoor Energy Efficiency.

The final choice between a RIS and a relay depends on the SNR and number of elements. The key observation is that a RIS needs many (hundreds/thousands) of reconfigurable elements (each the size of an antenna) to be competitive. Table 2 shows how rapidly RIS size increases, particularly in the outdoor scenarios (which are typically only both feasible and green at lower LTE rates). The reason is that source transmit power must travel over two channels to reach the destination, leading to a very small channel gain per element in the RIS, which needs many elements to compensate for the low gain.

While the large number of elements is a weakness for a RIS solution, the advantage is that a RIS requires no power amplifiers in its ideal form. However, in practice, active components are needed for adaptive phase-shifting. Even if the power dissipation per element is low, the total power is non-negligible.

R (bps/Hz)	N_indoor	N_outdoor
5	1	30
15	40	640
25	450	5000
35	5000	5000

Table 2. RIS Indoor/Outdoor Size.

mm-Wave (10GHz, 38GHz and 60GHz)

In this sub-section we consider some particularly interesting scenarios utilising higher frequency bands (mm-Wave 10-60GHz) for communication. The bandwidth available in such systems is expected to be at least an order of magnitude greater than at sub-6GHz frequencies, which should itself be sufficient to generate the desired customer bit-rate increase. This is, of course, a major motivation for developing mm-Wave networks. From a modelling perspective, we may study RIS assuming it will remain a viable technology if the underlying SE is similar to sub-6GHz values, relying on the increased bandwidth to support greater customer demands.

Since the corresponding wavelengths shrink by orders of magnitude (compared to current microwave spectrum) in these newer bands, EM radiation suffers greater attenuation. This in turn increases the importance of NLOS propagation, reflection and scattering. Consequently, RISs may become an enabling technology for controlling radiation and enhancing connectivity in such higher frequency networks whenever LOS communication becomes problematic. Of course, at these reduced length scales the engineering fabrication, EM coupling and control challenges are correspondingly increased.

Applying the same methodology as developed above for sub-6GHz, we may explore:

- Performance achievable using limiting system parameter values
- Extent to which mm-Wave offers a green solution (EE)
- Sensitivity of conclusions to modelling assumptions

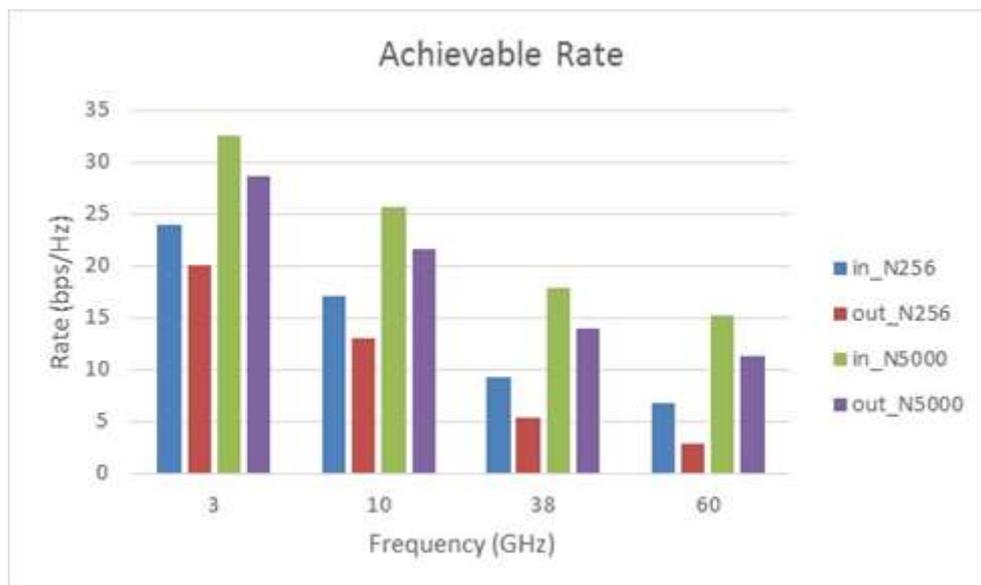


Fig. 29. Maximum Achievable SE @ mm-Wave 10GHz.

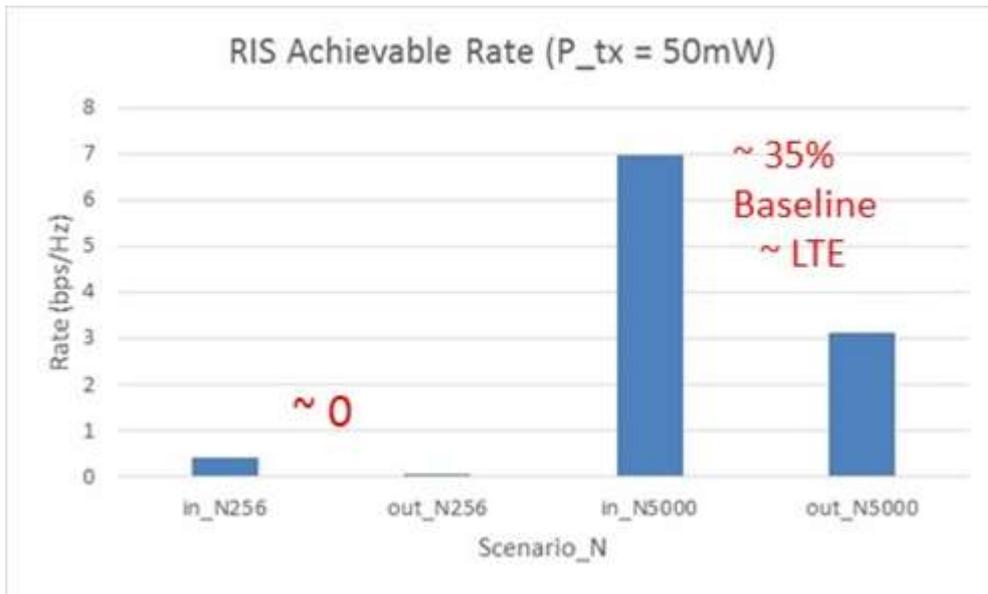


Fig. 30. Maximum Achievable SE @ mm-Wave 38GHz.

Accordingly, Figures 29 and 30 show the trend in limiting performance as mm-Wave frequency is increased and RIS size varied (from a minimum of 256 to maximum 5000 antenna elements, in a manner directly comparable to baseline). There is clearly a heavy bias towards indoor scenarios and large arrays. At 10GHz results suggest that both LTE and Massive MIMO services might be supported, albeit at lower rates than sub-6GHz. At higher frequencies (38 and 60GHz) only lower LTE rates appear viable assuming suitably large RIS arrays can be made.

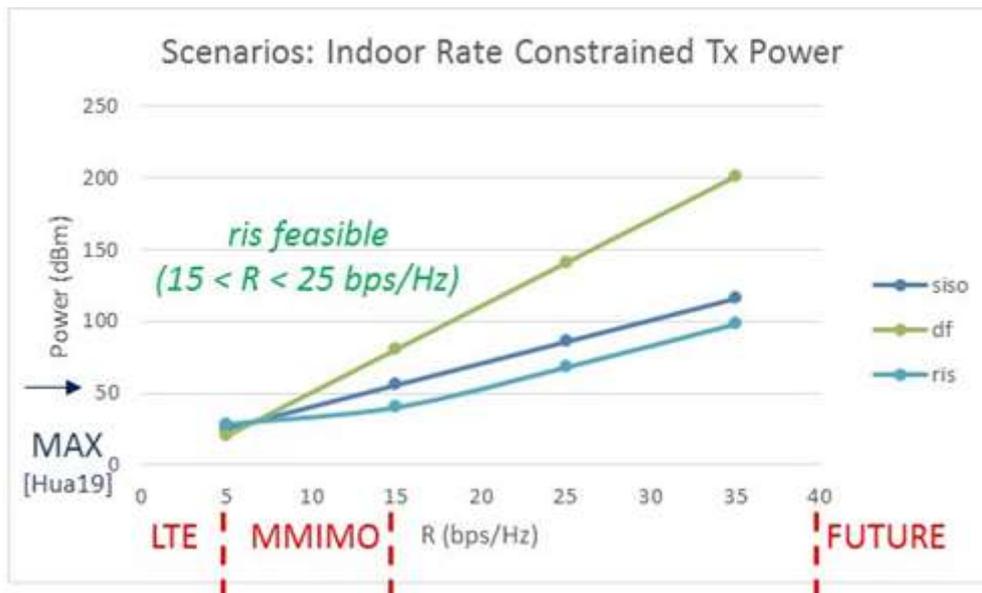


Fig. 31: Indoor Transmit Power @ mm-Wave 60GHz.

As was the case for sub-6GHz, none of the scenarios can support our hypothetical FUTURE ~ 40Mbps/Hz case but RIS offers the most efficient solution for LTE and Massive MIMO SE. Regarding EE, by definition this will increase linearly with bandwidth (by a factor 100/10 = 10 as indicated in Table 2.1), so tending to favour these higher frequencies over sub-6GHz operation. By analogy with sub-6GHz, we may seek an EE value of approximately 100 Mb/J as constituting a green solution. With this definition, all indoor LTE and Massive MIMO mm-Wave scenarios are both feasible (operable within the Table 2.1 50dBm maximum power) and green, as shown in Figures 31 and 32 at 60GHz.

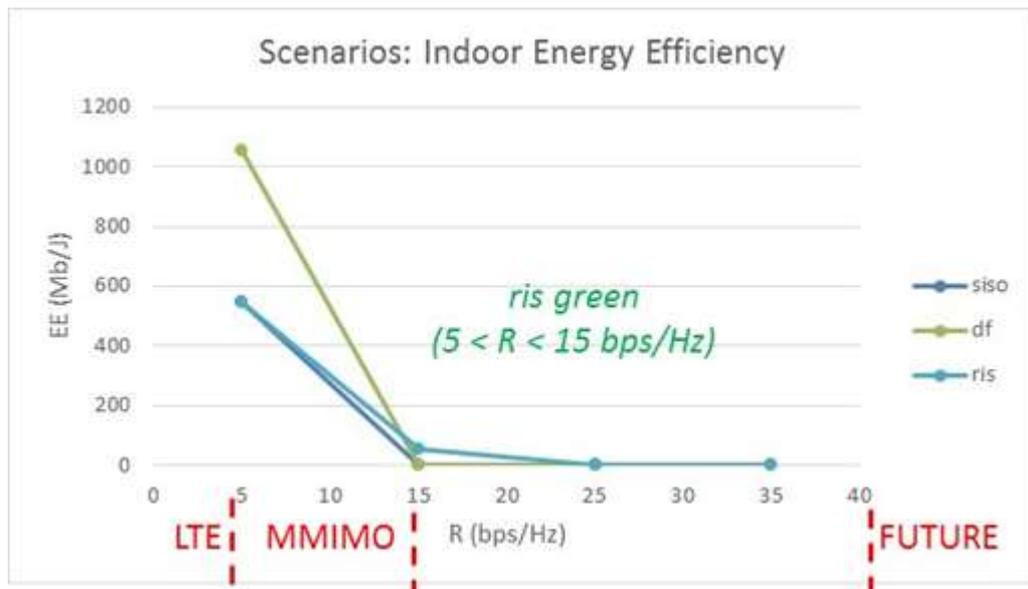


Fig. 32. Indoor Energy Efficiency @ mm-Wave 60GHz.

Of the outdoor scenarios, only the 10GHz case is operable within the 50dBm maximum power limit at lower LTE rates ($SE < 15$ bps/Hz), as shown in Figures 33 and 34. In this case, the increased bandwidth overcomes the effect of greater attenuation. At higher frequencies the attenuation is too great to provide a viable network solution.

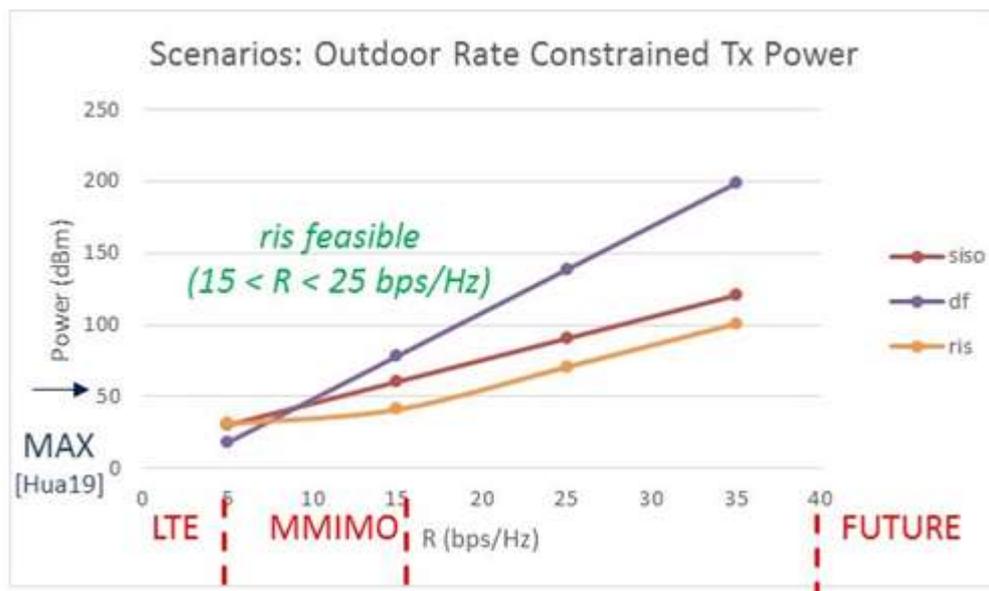


Fig. 33. Outdoor Transmit Power @ mm-Wave 10GHz.

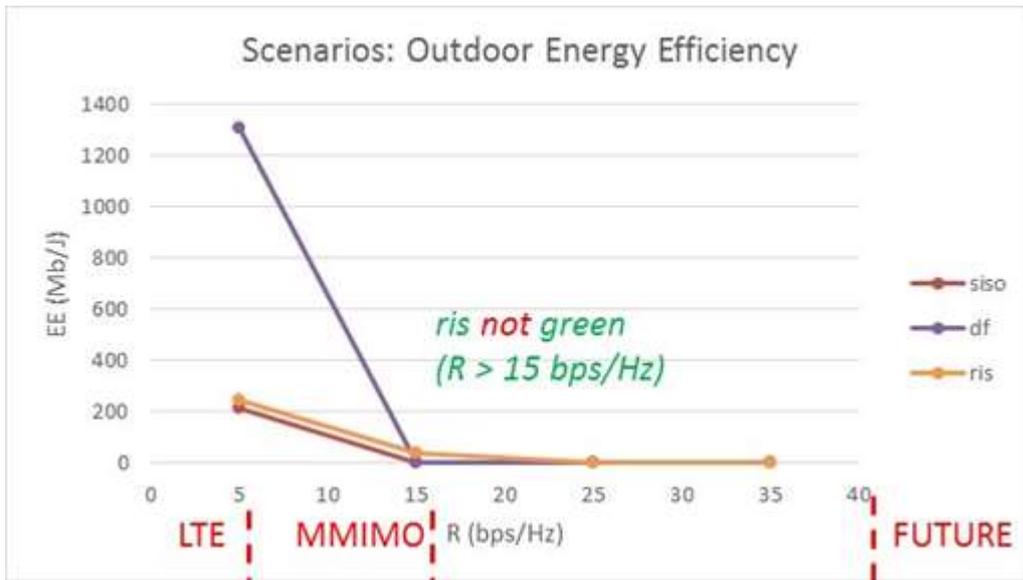


Fig. 34. Outdoor Energy Efficiency @ mm-Wave 10GHz.

Sensitivity Analysis

The discussion so far has used a deterministic model. Here, we study uncertainty in the various parameters and any likely effect on conclusions. Table 3 identifies the relevant quantities and best-estimates of modified values which might arise. Item 1 represents a more optimistic five-fold increase in bandwidth directly affecting EE, while remaining items 2-7 primarily affect SNR and hence SE. In particular, item 2 depends on the square of RIS size, item 3 represents likely device insertion loss, while item 6 represents the possible effect of variations in the environment (e.g. urban micro versus urban macro). Items 2-4 are more pessimistic than original values implicit in Table 1.

We identify items 1 and 2 as most affecting performance:

- In principle, the item 1 increase in bandwidth may be sufficient to make more scenarios at 38GHz viable. However, this must be weighed against the likelihood of any extra bandwidth being immediately utilised in serving increased customer demands.
- The greatest remaining uncertainty lies around item 2 maximum RIS size (contributing as a square to SNR) which can be fabricated and controlled, particularly as current prototypes are very much smaller than any values here [45].
- For remaining items 3-7, simple estimates suggest each individual 10dB uncertainty in SNR might correspond to approximately 3bps/Hz SE variation, so arguably unlikely to change our general conclusions.
- As a first approximation, items 3-7 might combine in quadrature as an overall ± 20dB uncertainty, comparable to item 2 in a pessimistic scenario. This emphasises once again that our overall RIS capacity predictions will be correspondingly optimistic.

ID	PARAMETER	ALTERNATIVE VALUE
1	Bandwidth (mm-Wave) per user	500MHz
2	Maximum RIS Size	1000 elements (-20dB)
3	Insertion Loss	-15dB
4	Maximum Transmit Power	-10dB
5	Noise	± 10dB
6	Scenario	± 5dB 3GPP Urban Macro
7	Shadow Fading	± 5dB

Table 3. Model Parameter Uncertainties.

Summary

As a broad average, outdoor scenarios require 25dBm more power than the corresponding indoor case. In general, scenarios may be assessed regarding both feasibility (remaining within the maximum allowed transmit power) and green operation (suitably high EE). In this sense, no scenario appears fully viable at our hypothetical 40bps/Hz FUTURE SE rates.

Practical limitations on antenna array size suggest that maximum SE achievable below 10GHz is probably in the neighbourhood of 20bps/Hz, representing a significant (but not excessive) performance improvement of ~150% over current Massive MIMO. The main advantage of RIS technology in this frequency range lies in providing more sustainable (greener) network solutions.

At this time, prospects for the Massive MIMO use-case appear promising (though probably entailing significant reduction in link bandwidth) while the relay use-case seems more speculative (except for relatively short distances). Comparing SISO, RIS and DF relaying technologies, there is a natural bias towards indoor scenarios at lower rates:

- Only an indoor RIS sub-6GHz network remains feasible (within the maximum allowed transmit power) across the full range of scenarios considered
- Below 6GHz, indoor RIS technology out-performs DF relaying up to Massive MIMO rates
- Only 3GHz and 10GHz outdoor RIS scenarios offer a small EE advantage up to Massive MIMO rates

- In all other mm-Wave scenarios, RIS offers no significant EE advantage

Overall, with regard to EE, RIS shows strong potential wherever energy consumption is a high priority and green solutions are necessary.

However, RIS technology needs many (hundreds/thousands) of reconfigurable elements (each the size of an antenna) to be competitive though it can be significantly lower power. Control complexity becomes a huge challenge because of the large number of RIS elements, particularly the need to reconfigure quickly when switching between multiple mobile devices. As the main design consideration, software rapidly controls beam direction primarily through adjusting phase of the propagating waves, possibly augmented by advanced Artificial Intelligence (AI) techniques. For easy reference, Tables 4 and 5 compare the different technologies in terms of both trade-offs available and attributes of working systems, for broadly similar performance targets.

SYSTEM	COST	SIZE	COMPLEXITY	COMMENT
Massive MIMO	HIGH: Distributed phase shifters and amplifiers	MEDIUM: Trade cost for thickness; thin is very expensive (accommodation)	HIGH: Antenna feeder chain drives price and power consumption	~ 200 Antennas maximum
RIS	LOW	HIGH Require large array	MEDIUM	Same-size performance never better than conventional M-MIMO equivalent

Table 4. Technology Trade-offs.

SYSTEM	POWER	SIZE	COMPLEXITY	COMMENT
Massive MIMO	50W	64 elements 1m x 1m area	64 feeder chains	~ 10x performance improvement if 64->200 elements
RIS	60% x Massive MIMO	256 elements 0.8m x 0.8m area	single feeder chain	size limits, prototype only

Table 5. Technology Systems (broadly similar performance targets).

2.2.1.2 RIS-assisted Channel Hardening

In the limit of many array elements and large element spacing, a feature of conventional multiple-input-multiple-output (MIMO) Rayleigh fading channels is channel hardening, which stabilizes the overall MIMO channel gain by virtue of central limit theorem given the high-rank channel matrix. Hence it facilitates a method for global channel stabilization. Channel hardening is widely exploited in massive MIMO for improving the design of channel estimation, transmit precoding and receiver detection. In the context of RIS-aided smart radio environment, the channel hardening phenomenon has also been demonstrated for channel models incorporating the spatial correlation between RIS elements. However, no prior evaluations of RIS-aided channel hardening have systematically studied the design of phase optimization for achieving the hardening effect.

To facilitate global channel stabilization with the aid of RIS, here we consider a cross-entropy optimization (CEO) method for SNR optimization as the first step, where each RIS element has a binary phase response. In a nutshell, CEO is a type of iterative Monte Carlo algorithm that relies on minimizing the cross entropy between a starting probability distribution and a target distribution, which will gradually improve the fitness of the initial sample drawn from the starting distribution. The operations of the proposed CEO algorithm are described in Alg. 1 below.

Algorithm 1: Cross Entropy Optimisation Algorithm for RIS Aided Communication

- 1: Set RIS element number as N , iteration index $t = 1$, sampling parameter $\beta = 0.1$, and initialize Bernoulli Probability vector $\mathbb{P}^0 = \{P_1^0, \dots, P_N^0\}$ with $P_i^0 = 0.5, i = 1, \dots, N$.
 - 2: Generate K RIS configurations as $\{\mathbf{X}_1, \dots, \mathbf{X}_K\}$, where $\mathbf{X}_k = \{x_{k,1}, \dots, x_{k,N}\}$ according to Bernoulli distribution with probability \mathbb{P}^t .
 - 3: Commit the RIS experiment and collect the SNRs.
 - for** $k = 0 : K - 1$ **do**
 - i: Commit \mathbf{X}_k as RIS configuration for the k th experiment.
 - ii: Record the SNR at the receiver side as e_k .
 - end for**
 - 4: Sort the SNRs $\{e_1, \dots, e_K\}$ in the descending order, denote the sorting index as $\{\delta_1, \dots, \delta_K\}$ and sort $\{\mathbf{X}_1, \dots, \mathbf{X}_K\}$ accordingly as $\{\mathbf{X}_{\delta_1}, \dots, \mathbf{X}_{\delta_K}\}$.
 - 5: Set $J = \lceil \beta K \rceil$ and update \mathbb{P}^t with $P_i^t = \frac{\sum_{j=\delta_1}^{\delta_J} \sum_{n=1}^N x_{j,n}}{JN}, i = 1, \dots, N$.
 - 6: Update the Bernoulli probability vector.
 - if** \mathbb{P}^t not converges to binary vectors **then**
 - Set $t = t + 1$ and go to Step 2.
 - else**
 - Return the optimized RIS configuration as $\mathbf{X}^* = \mathbf{X}_{\delta_1}$.
 - end if**
-

In [46], it was specifically shown that RIS-aided channel hardening is achieved when:

1. the number of RIS elements becomes large, while
2. the RIS has an SNR-optimal phase configuration.

However, it is trivial to see that the second condition requires the phase response of each RIS element to be continuous. Considering the control overhead for operating a RIS and practical circuitry, continuous phase RIS remains largely unrealistic for real-world scenarios. As a result, it is necessary to consider a new SNR bound encompassing the influence of quantization loss given the presence of binary state RIS elements.

To facilitate the implementation of RIS-aided channel hardening, the proposed Alg. 1 is benchmarked against the genetic algorithm (Ga) framework for characterizing the quantization loss of achievable SNR due to the binary phase constraint. In the simulated scenario, the RIS has $8 \times 16 = 128$ binary-phase elements, while the BS and UE are both equipped with a single antenna. The statistical channel follows a block fading Rician model, where the channel coefficient stays constant for a single run of CEO or Ga and changes in an i.i.d. fashion between different runs.

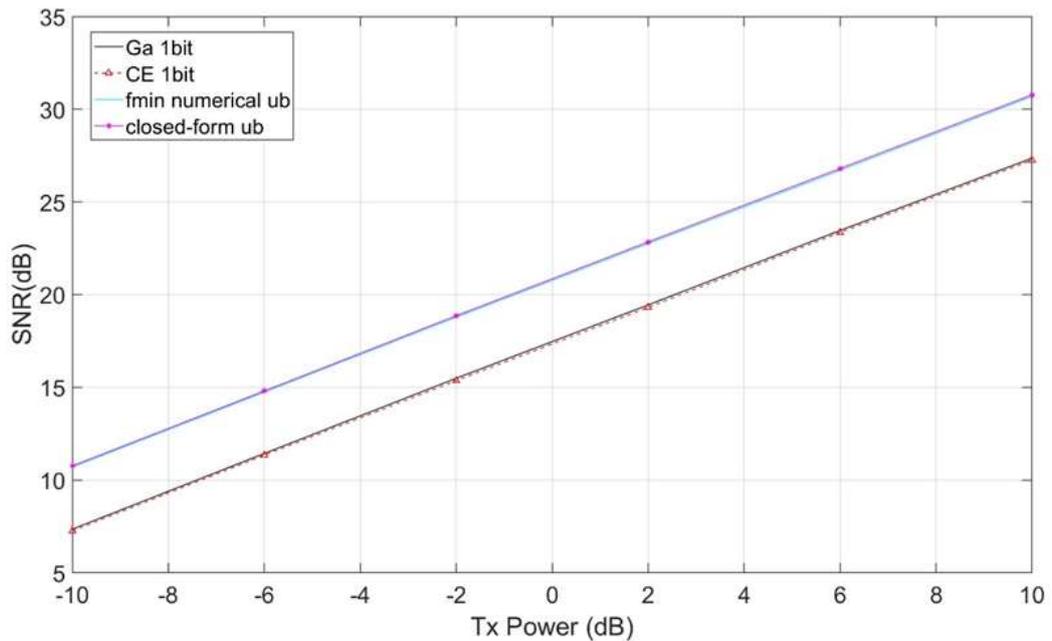


Fig. 35. Rx SNR vs Tx power.

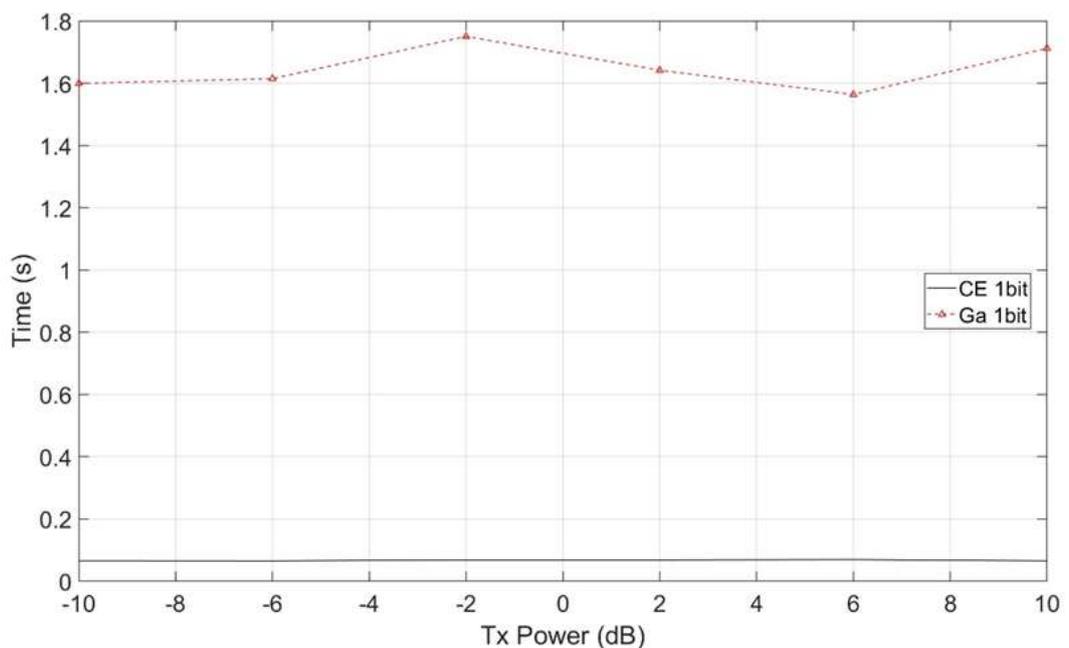


Fig. 36. Average time complexity vs Tx power.

The SNR performance and time complexity of Alg. 1 are compared to those of the Ga in Figure 35 and 36. Moreover, Fig. 1 also portrays the numerically evaluated SNR upper bound as well as the closed-form theoretical upper bound. The comparisons in Figure 35 shows that both CEO and Ga have an ~ 3 dB quantization loss towards the optimal Rx SNR limit under the assumption that the RIS consists of binary-state elements. On the other hand, Fig. BT_2 demonstrates that the complexity of CEO is less than 10% of Ga, making Alg. 1 a much more realistic technique in this context when the natural channel environment exhibits short coherence time.

Furthermore, preliminary demonstration of the channel hardening effect is portrayed in Figure 36. Here the variation of SNR between fading blocks (samples) is shown. In comparison to the continuous phase upper bound, it may be observed that the SNR in the binary-phase scenario varies over a wider range, leading to weaker channel hardening effect. Additionally, CEO and Ga are also shown to have very similar SNR variation ranges, making them equally effective for achieving RIS-aided channel hardening.

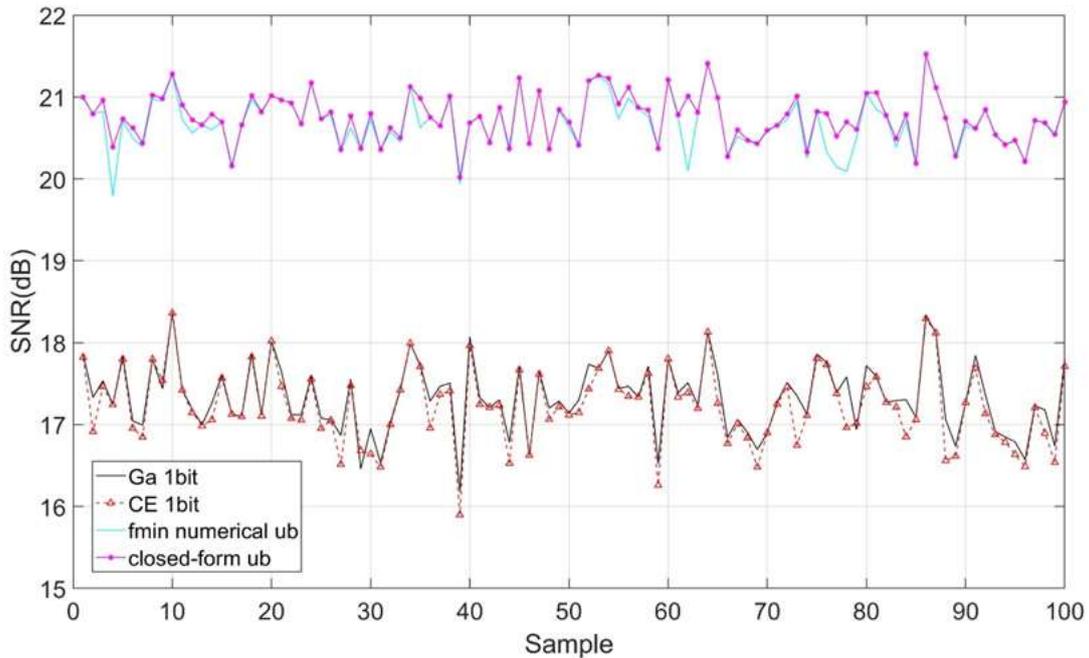


Fig. 36. Initial demonstration of RIS-aided channel hardening.

2.2.2 RIS-Integrated NR with AIML

Reconfigurable intelligent surfaces (RIS) have fast emerged as a prominent beyond 5G (B5G) candidate technology [48]. RIS broadly correspond to smart radio surfaces with many small antennas or reflecting metamaterial elements that receive and reflect electromagnetic (EM) waves with the capability of controlling the phase-shifts, thus offering tunable reflections. The distinctive characteristics of RIS in terms of reconfiguration of the propagation environment has motivated a host of potential new use cases. These include (i) utilising RIS as intelligent reflectors to enhance the coverage performance of existing multiple-input multiple-output (MIMO) systems, and (ii) to implement low-complexity MIMO radios particularly in the form of distributed antenna arrays.

A key motivation behind the existing works on RIS has been based on the notion that the operators would use these materials to control the propagation environment. This means the RIS would be part of the network operator infrastructure, which can be particularly suitable for improving coverage performance outdoors; this is especially relevant to operation in higher frequency bands (FR2) [49]. Another key area that has certainly received less attention in the literature is around the use of RIS for improving outdoor-indoor and indoor-indoor link-budget performance in sub-6 GHz bands. In these cases, RIS would more likely be controlled by the consumer or enterprise; some example indoor RIS implementations have been reported in the literature such as intelligent walls [50] and 3D reflectors [51].

The use of RIS as intelligent reflectors to improve coverage performance indoors, particularly, from the perspective of the terminals is the focus of the work. The reference architecture is 3GPP 5G NR networks and terminals considering mainly sub-6 GHz frequency bands. The key technical aspects to be covered are around developing standards-compliant air-interface and Radio Access Network (RAN) control solutions using tools from AI/ML for facilitating RIS-aided MIMO indoor communications. To train, develop, and validate the schemes for baseband and control functionalities, a 5G NR-compliant link-level simulator (LLS) which can capture the RIS-based end-to-end transceiver chain for the various data and control physical channels has been developed in MATLAB (see Figure 37).

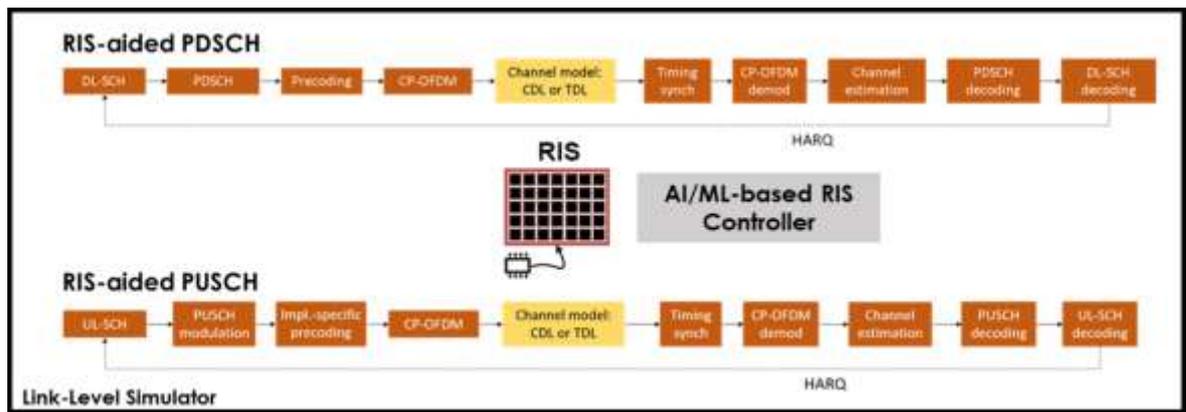


Fig. 37. RIS-aided 5G NR transceiver chains for downlink and uplink data channels with RIS-based RAN controller (with AI/ML-based xApp).

It is well known that the uplink channels in NR are considered bottlenecks for the overall coverage performance. Here, we provide some preliminary results for RIS-aided communications for the uplink data channel in NR (PUSCH), considering a transmitter-receiver (UE-gNB) pair, with a RIS component in between, as shown in Figure 38. The focus here is on enhancing coverage performance for indoor scenarios, using standards-compliant signals and waveforms. We consider two schemes, a baseline NR PUSCH system, and a RIS-aided NR PUSCH system. The RIS here is modelled as a planar array with isotropic elements. The throughput performance of PUSCH as a function of gNB-RIS distance is captured in Figure 39. It can be observed that the RIS component can significantly improve the link budget performance for PUSCH, highlighting the potential of using RIS as an intelligent reflector.

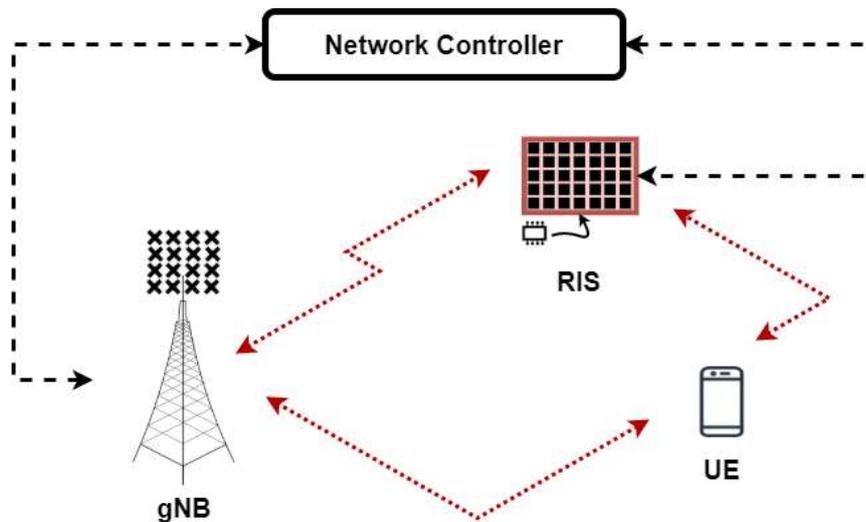


Fig. 38. RIS-aided communications for NR.

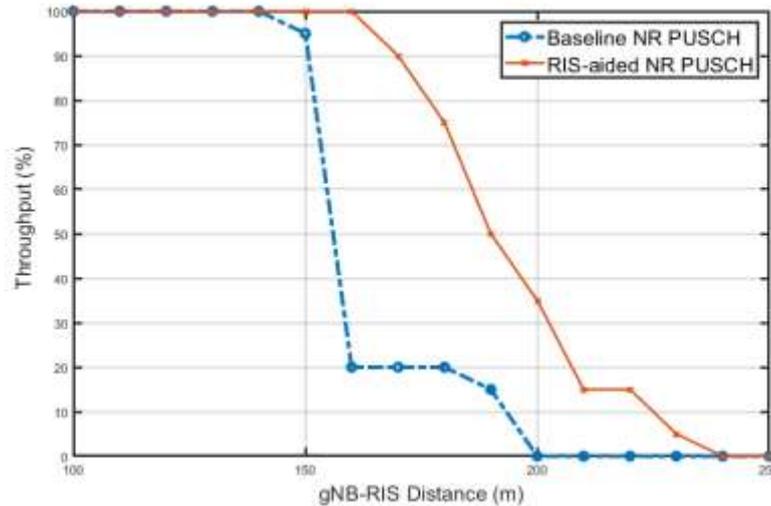


Fig. 39. Performance of NR PUSCH for RIS versus non-RIS aided systems. Parameters: SISO gNB/UE, 8x8 RIS, FDD@2.0GHz., CP-OFDM, NRB=52, SCS=15kHz, QPSK, 193/1024, UE-h=1.5m, gNB-h=10m, RIS-h=5m, UE-RIS Distance=5m, SNR=10dB, gNB-UE link UMi CDL-A, gNB-RIS link UMi custom-CDL, UE-RIS custom-CDL link InH-Office.

Importantly, achieving infinite phase resolution is infeasible and needs to be quantized for practical implementation. The resolution of the phase shifters that defines the number of possible phase shifts is another factor which widely impacts the system performance. Practically utilized discrete phase shifters (with limited number of possible phase shifts) may introduce errors if infinite resolution optimal configurations are obtained and then quantized. In a RIS-aided communication environment, this effect of the phase resolution on the performance is widely studied in [52] where it is shown that the overall performance of the system is proportionally impacted with the reducing number of phase shifts. In literature, most of the works focus on obtaining the continuous phase configurations for RIS elements which will inherently result in sub-optimal performance in practical implementation. Here we consider unsupervised-learning based method that predict the optimal discrete RIS phase configurations while maximizing the overall (signal-to-interference-plus-noise ratio) SINR performance.

We focus on an unsupervised learning-based method to estimate the optimal discrete RIS phase configuration by utilizing the available channel state information (CSI) as the input feature. It is assumed that the direct link between the base-station (BS) and a user equipment (UE) is unavailable and only the CSI of the RIS-aided link is utilized in the proposed method. Further, the proposed method is designed to be independent of the number of transmit and receive antennas at the BS and UE respectively, making the solution generalized and adaptable towards different BSs and UEs in the communication. Due to the unsupervised nature of the proposed method, the solution does not rely on any labelled data, obtaining which is an exhaustive process. The performance of the proposed solution is analysed through rigorous simulations with the developed LLS and shows the effectiveness of directly estimating the discrete RIS configurations as compared to obtaining the infinite resolution phases and then quantizing them. The performance analysis further demonstrates the generalized property of the method over different number of BS and UE antennas.

We will provide the specific details of the system and AI models in the subsequent deliverable for WP3. For performance analysis, we evaluate the performance of the proposed system in terms of achievable sum-rate, phase resolution of the discrete phase shifters associated with RIS elements and the bit-error-rate. Throughout the simulation, we have utilized the benchmark open source DeepMIMO dataset for network training and performance evaluation. We consider an outdoor scenario operating at 3.5 GHz frequency. From the dataset, we consider the defined BSs as RIS units and the operating BS and UE user pair is selected from the user distribution. The various parameters considered in the dataset are summarized in Table 6. To generate the training and testing data samples, we randomly select RIS from the list and then select the BS and UE randomly from the serving area of the selected RIS. The selected BS and UE are allocated random number of antennas from a set 1,4,16 and the respective channel between BS-RIS and RIS-UE are generated.

Table 6. RIS-aided Simulation Environment Parameters.

Operating Frequency	3.5 GHz
Active BSs and users rows	1-3852
Number of BS Antennas	1, 4, 16
Number of user Antennas	1, 4, 16
Active RISs	1, 4, 5, 6, 8, 10, 16, 17
Number of RIS Antennas	4, 8, 16
Antenna spacing (\times wavelength)	0.5
Bandwidth (GHz)	0.05
Number of OFDM subcarriers	512
OFDM sampling factor	1
OFDM limit	64
Number of paths	1

The convergence of the proposed network in terms of the sum-rate versus varying number of epochs is depicted in Figure 40. The proposed method is evaluated with the output phase resolution of 6-bit and 8-bits. We compare the performance of the proposed discrete phase resolution estimator network with the genie-aided system which is while considering the infinite phase resolution. It can be seen that the proposed network with both the 8-bit and 6-bit resolution.

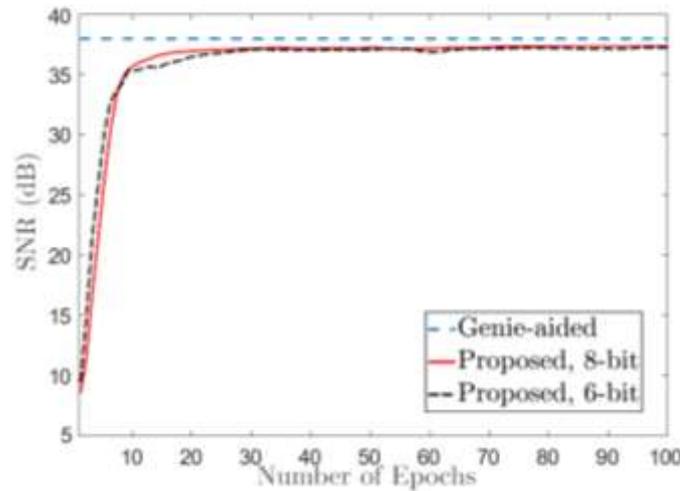


Fig. 40. SNR performance with varying number of epoch for the proposed method with 6-bits and 8-bit resolution comparing with the infinite resolution genie-aided system with infinite resolution.

In Figure 41, we illustrate the performance of the proposed method in terms of the sum-rate with varying transmit power. The performance is evaluated for 2-bits, 4-bits and 6-bits discrete phase resolution. It is observed that the sum-rate increases with increasing transmit power for all the considered resolutions. As expected higher sum-rate is achieved with higher bit-resolution of the phase shifter. This is due to the low quantization errors introduced due to the increasing phase step. Nevertheless, it can be seen that the performance of the methods with lower resolution phase shifts is also close-by and can be considered for the applications requiring low-power.

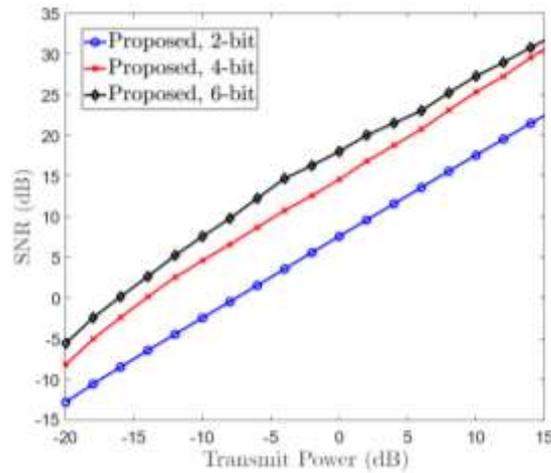


Figure 41. SNR performance with varying transmit power for the proposed method with 2-bit, 4-bits and 6-bit resolution.

2.2.3 Virtual Prototyping and Simulation of Reconfigurable Intelligent Surfaces

While electromagnetic characterization of large-scale antenna arrays (as described in Subsection 2.1.3) is challenging from accuracy and speed point of view, accurate simulation of reconfigurable intelligent surfaces poses unique challenges to the developers of EM CAD/CAE systems. The RIS are usually simulated using infinite flat periodic array assumption [36] in order to save computational time and avoid numerical issues coming from the resonant nature of RIS. These resonances render many common numerical methods (and their implementations) for solving Maxwell’s equations not feasible for the simulation of RIS.

However, fast simulation of RIS that take into account the finite size, curvature, and feature different unit elements with low degree of periodicity is quite important [19], particularly for RIS that will later enhance the performance of the antenna arrays or become their integral part.

In Fig. 42, we depict a metasurface (RIS) that is intended to block 5.8 GHz signals, which was designed using traditional equivalent circuit model method [40]. The RIS is based on the square loop unit element and based on the design process is supposed to totally block the 5.8 GHz field under the planar infinite periodic array assumption. In Fig. 43 we depict the simulation setup (YZ and XY planes), where two halfwave dipole antennas (Rx and Tx) are placed on the different sides of the RIS.

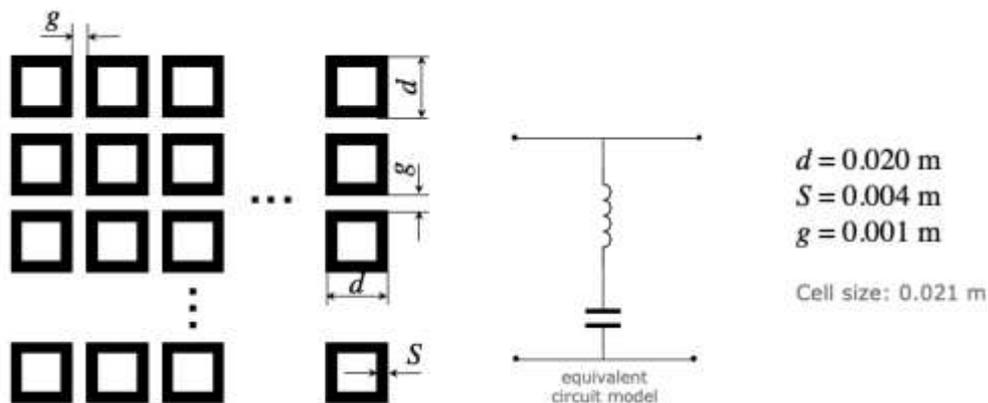


Fig. 42. Square-loop RIS designed to block 28 GHz signal [20].

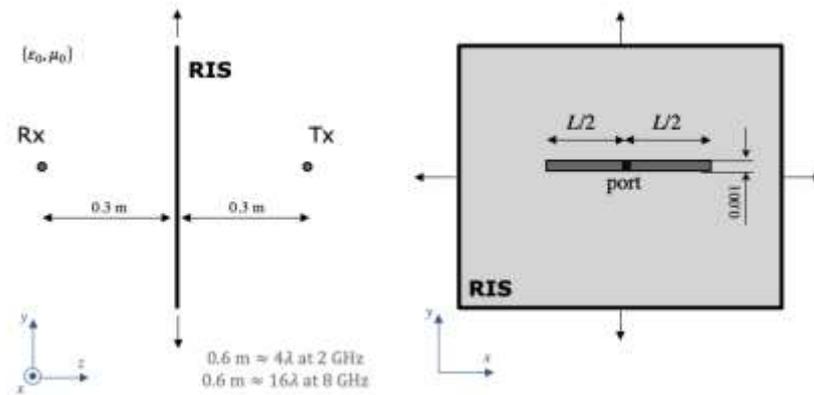


Fig. 43. Simulation setup for testing the effectiveness of the square-loop RIS.

By simulating RIS of different finite size (10×10 , 20×20 , ..., 100×100), we determined that 40×40 size closely converges to the infinite periodic array behaviour. In Fig. 44, we depict the normalized received power (relation between the received power when RIS is present to the received power when there is no RIS in dB) for RIS of different size. Both 40×40 and 100×100 RIS provide around 40 dB isolation; however, if a smaller RIS is used, the blocking of EM field degrades rapidly.

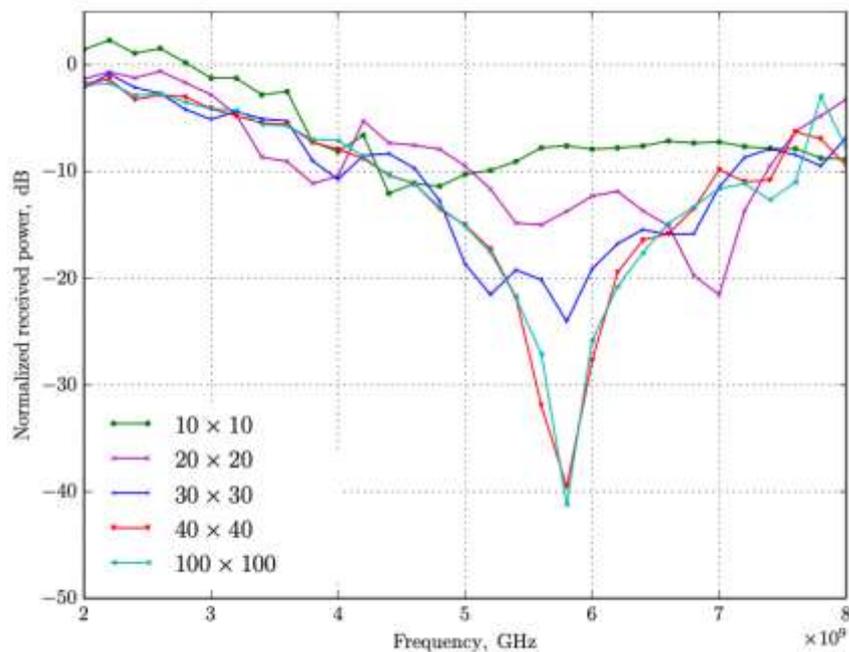


Fig. 44. Normalized received power for RIS of different sizes.

Next, we investigated the effect of the curvature to the performance of the RIS. In Fig. 36, we depict the electric field magnitude for four scenarios. Top-left: no RSS; bottom-left: 40×40 planar RSS (corresponding to the red curve in Fig. 45); top-right: slightly curved 42×42 RIS; bottom-right: medium-curved 45×45 RIS. The number of elements for curved RIS has been slightly increased to keep the corner points of the RIS at the same positions.

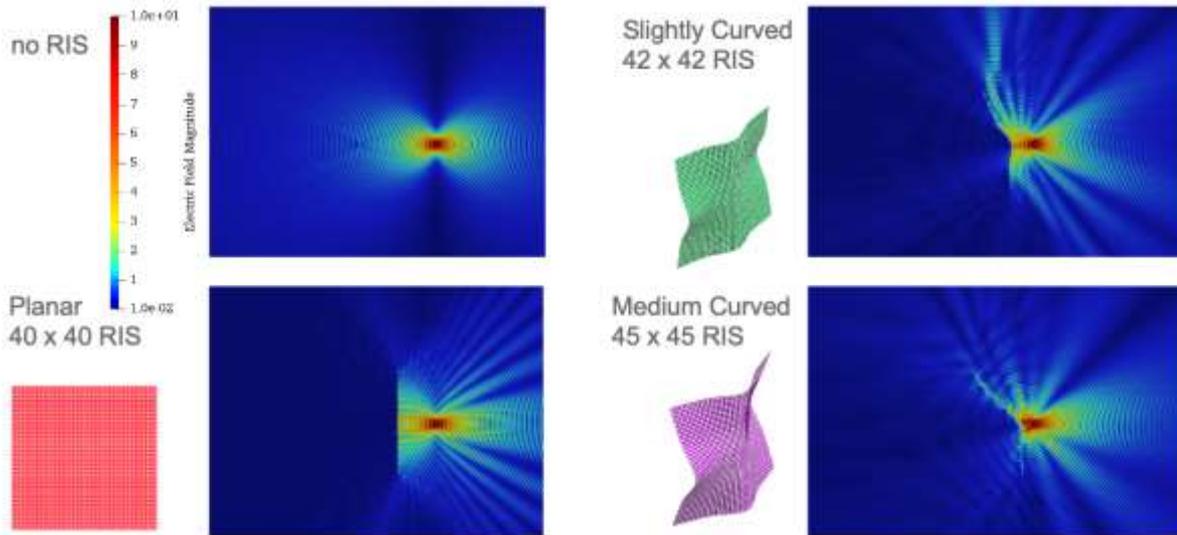


Fig. 45. Simulated electric field magnitude for RIS exhibiting different curvature.

It is clearly seen, that the EM fields quickly starts to penetrate beyond the RIS as soon as some curvature is introduced, and the performance of the RIS deteriorates quickly. In Fig. 46, it is shown that the RIS no longer provides 40 dB isolation for the normalized received power in the scenarios featuring curved RIS.

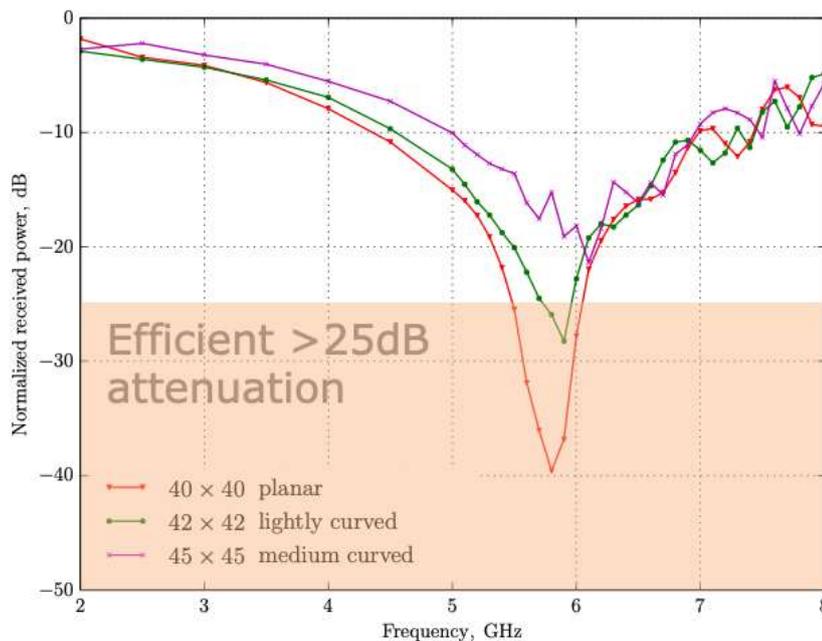


Fig. 46. Normalized received power for planar (red), lightly curved (green), and medium-curved RIS.

The simulated 45x45 medium curved RIS resulted in 240 924 unknowns and required 5 min simulation and 15 GB of memory on a single c5.24 Amazon EC2 node.

While it is possible to simulated RIS rigorously taking into account all the needed effects using full-wave methods, simulation of the antenna array designs that feature RIS is extremely challenging. It will likely require hybridization between full-wave methods and high-frequency asymptotics.

While electromagnetic characterization of large-scale antenna arrays (as described in Subsection 2.1.3) is challenging from accuracy and speed point of view, accurate simulation of reconfigurable intelligent surfaces poses unique challenges to the developers of EM CAD/CAE systems. The RIS are usually simulated using infinite flat periodic array assumption [36] in order to save computational time and avoid numerical issues coming from the resonant nature of RIS. These resonances render many common numerical methods (and their implementations) for solving Maxwell’s equations not feasible for the simulation of RIS.

2.2.4 Simulation and Optimization on Reconfigurable Intelligent Surface assisted Multiple-receiver Communication

For multiple-receiver communication based on RIS, many methods have been used to achieve reflecting a single incident wave to multiple directions at the same time. For example, the phase-only reconfiguration, the amplitude and phase reconfiguration, the space division multiplexing, and the time division multiplexing. A comparison of the performance between these different methods has been made in [37] showing that the phase-only reconfiguration is able to achieve a moderate side lobe level with the lowest complexity, power, and latency. To achieve the multi-reflection based on phase-only reconfiguration, several single-reflection reconfigurable profiles should be overlapped into one profile which can reflect the wave to these several reflect directions. But few articles considered the existence of periodicity distribution in these reflection sub-profiles, which makes the best superposition cannot always be achieved at the original position. Better performance may be achieved with a misplaced superposition that should be investigated with a good approach to find. In this work, some basic simulations of the RIS hardware design have been made, and based on the experience from the hardware simulation, a Deep Reinforcement Learning (DRL) based optimization method has been proposed with a verification systematically.

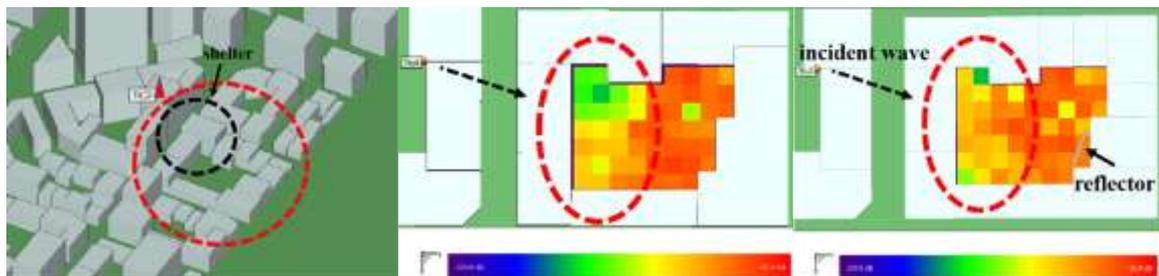


Fig. 47. The simulation of the RIS-assisted outdoor communication (a) the outdoor scenario with shelters, (b) the coverage without RIS, and (c) the coverage with RIS

Firstly, a basic simulation based on the ray tracing software Wireless Insite has been made. In this outdoor scenario, the users are in the area surrounded by many buildings and are blocked by the total buildings between the Base Station (BS) and them. The simulation results from Figure 47 shows that with the deployment of the RIS reflector, the coverage in the blocked area has been improved a lot.

Then, a RIS reflect element has been designed in the electromagnetic simulation software CST Studio with inspiration from [38] and [39]. As shown in Figure 48, two substrates and two varactor diodes have been utilized in each element. The varactor diodes based on the radiation pattern of the RIS element are used to achieve multiple reflect phases controlled by the Digital to Analog Converter (DAC). In the simulation, a group of six reflect phases have been selected and the amplitude and phase responses are shown in Figure 49. It can be seen from the figure that by selecting a moderate voltage, the different capacities and hence the accurately reflect phases can be achieved. By managing the arrangement of these RIS elements with six different states, a reflect profile can be calculated with basing on the optical path difference. With the joint simulation of MATLAB and CST, the reconfigurable profile reflecting the incident wave to any expected directions can be achieved and the far-field radiation pattern of a 14x14 array is shown in Figure 50.

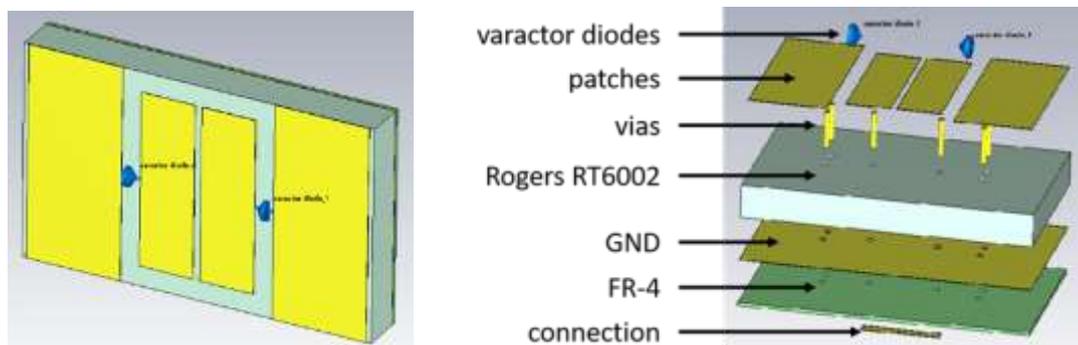


Fig. 48. The structure of the designed RIS element

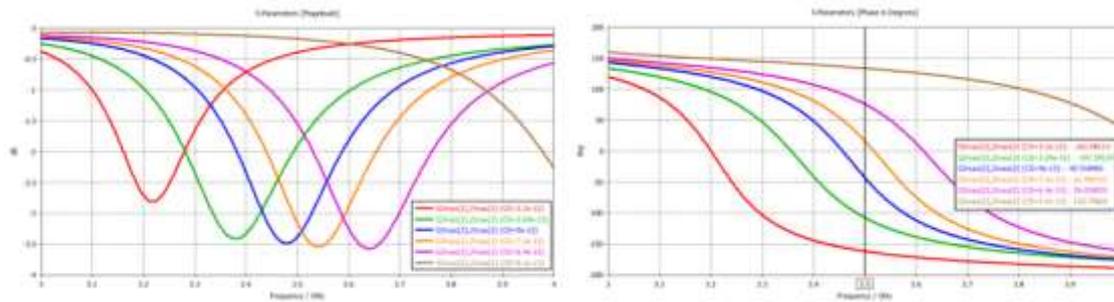


Fig. 49. The (a) amplitude and (b) phase response of the designed RIS element

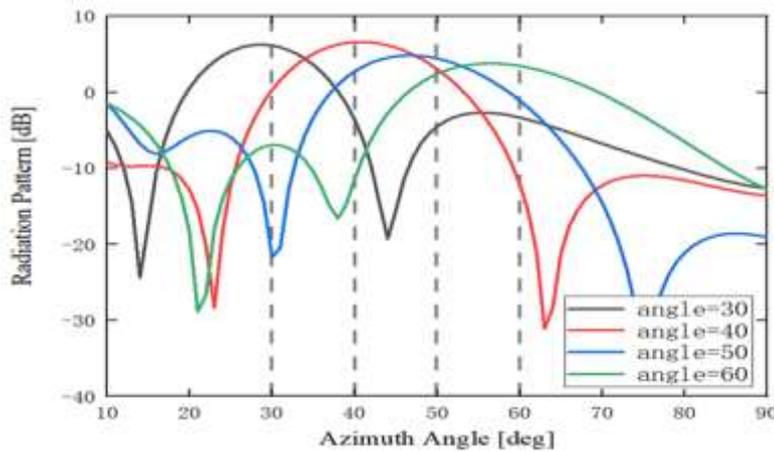


Fig. 50. The radiation pattern with different reflected angles on the reflected direction

After achieving the reflect profile for a single reflect direction, then the profile reflecting the incident wave to multiple directions at the same time can be designed by the superposition of these several single-reflection profiles. As Figure 51 shows, the dual-reflection profile is overlapped by the single-reflection profiles of [30°, 40°] and [50°, 60°], and the far-field pattern shows two accurate beams pointing at these two directions with a few of high side lobes. But the reflective profile of each reflective direction shows a strong periodicity so different overlapping ways will achieve different performances. Usually, the simple superposition of multiple reflections is hard to achieve the best performance, so it is important to find the optimal superposition method which can improve the RIS's performance. Because the optimal superposition is merely relying on an appropriate movement of the original reflective profiles, both the hardware design and the algorithms do not need any change, which ensures such improvement without increasing the complexity of the system.

A series of superpositions of multiple reflective profiles without DRL has been done to verify the feasibility and the predictable performance of this improvement. For two different reflect directions, calculating the reflective profile of every single direction that the profile of direction 2 has a size four times bigger than normal. As shown in Figure 52(a), the red and blue rectangles are the reflective profiles of direction 1 and 2 respectively, which are denoted by the term profile 1 and profile 2 in the following sections. Then, these single reflective profiles will have an overlapped profile lying in the same location as profile 1 and with the same size. The area enclosed by the yellow dotted line is the moving/searching window which is the range that the central point of profile 2 can move in. The length of this window is determined by the maximum period length of the single reflective sub-profiles. When the central point of profile 2 moves to a different position, a new superposition profile will be achieved which is denoted as the different overlapped positions (or the superposition on different positions) in this work. In the verification stage, move profile 2 with its central point has an exhaustive search on each cell of profile 1, calculating the far-field radiation pattern on each position. Set the peak value of the original far-field pattern as the baseline $E_{baseline}$, then calculate the size of area having a higher value over the $E_{baseline}$ within the circle centering at the given reflect directions with 20 degrees radius R for each movement as shown in Figure 51(b). The schematic diagram and results are recorded in Figure 52, from Figure 52(b) and Figure 52(c), it can be seen that:

- The superposition of different positions has an impact on the far-field pattern of RIS. Moving to adjust the position of each sub-reflective profile can improve the far-field performance.

- The distribution of the performance with superposition on different positions has the characteristic of periodicity, which means the exhaustive searching is not needed and the maximum performance can be found within a small scope.
- The variation from the poor position to the best position has a gradual gradient. So, the problem can be transformed into finding the shortest path moving from any random initial position to the best position which can be formalized as a typical RL problem.

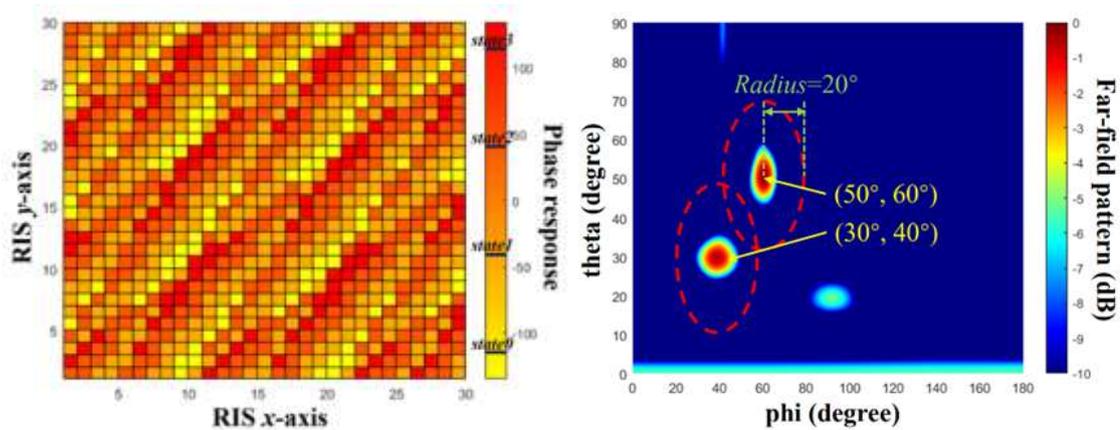


Fig. 51. The (a) overlapped reconfigurable profile and (b) its far-field performance at the direction [30°, 40°] and [50°, 60°]

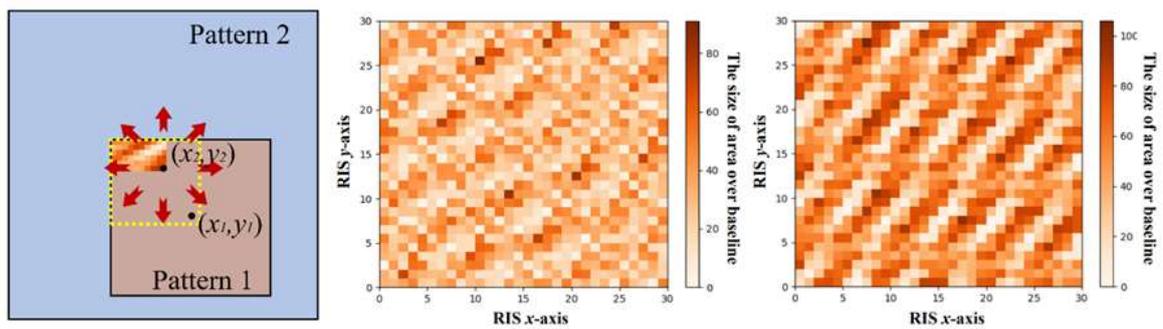


Fig. 52. The superposition (a) schematic diagram, (b) on [45°, 30°] and [45°, 60°] directions, and (c) on [30°, 40°] and [50°, 60°] directions.

By setting the Deep Q-Networks (DQN) with an exploration rate of 0.9, reward discount factor of 0.98, and learning rate of 0.001, the DQN training curve of the proposed scenario with two reflect directions are shown in Figure 53, where the y-axis is the mean reward with upper and lower bounds that each episode can achieve and the x-axis is the running steps. Then, the trained policy function can be exported to validate the performance of the DRL model compared with the original profile without any move, random search, and exhaustive search drawn in Figure 54. For preciseness, the maximum time-steps of random search is also set to 11 steps which is equal to DRL method. The random search was run three times from the same starting point while the DRL validation was also run three times from different but adjacent starting points. The result shows after several steps, the DRL method can make the superposition reach the position with the optimal performance quickly. In contrast, the random search needs a long period running and then may go through a good position but cannot stand there. Also, it cannot achieve the best position in most runs especially when the best position is sparse enough in the entire searching area.

The comparison of far-field patterns before and after DQN adoption is shown in Figure 55, where 1.2 dB improvement of the peak value and a wider beam has been achieved. Table 7 lists the accuracy, max A_{total} achieved, and time cost of each method. Compared with the exhaustive search, the DRL method can reach the position with the max A_{total} within the shortest time in the same size as the searching window. Though exhaustive search can always find the best value, the DRL method can also find it if the learning process is completed.

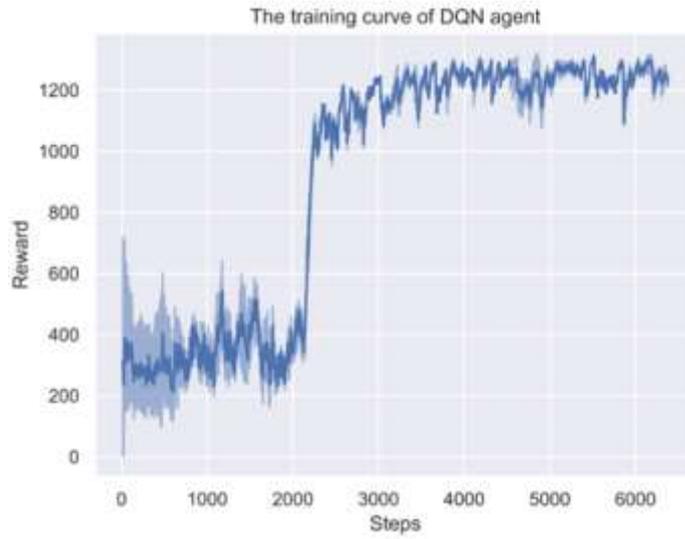


Fig. 53. The training curve of DQN.

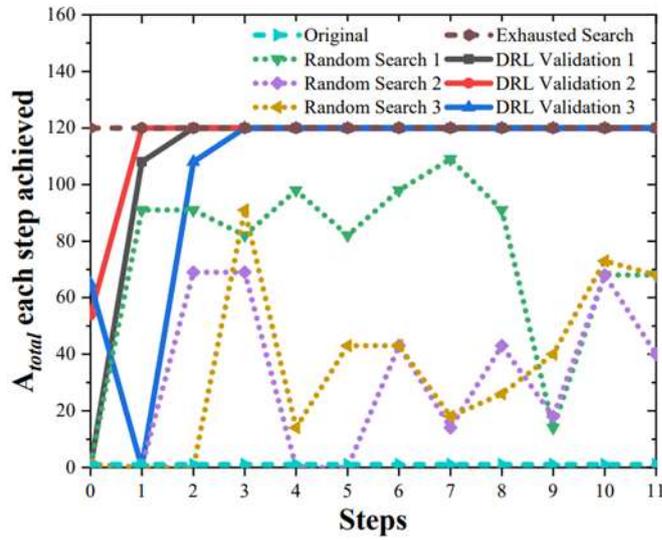


Fig. 54. The performance of each step in different methods.

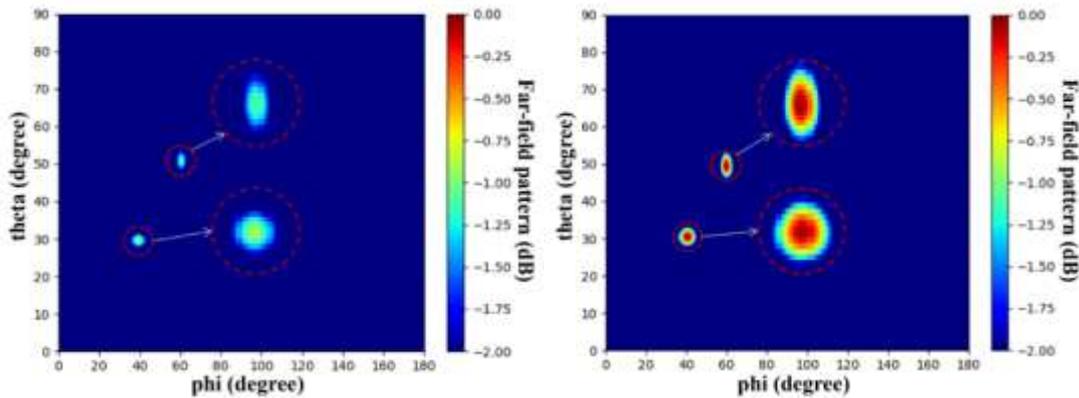


Fig. 55. The comparison of far-field patterns of (a) before and (b) after DQN adopting.

TABLE 7

COMPARISON OF DIFFERENT METHODS

Name	Random	Exhausted	DRL(proposed)
Accuracy	low	highest	high
A_{total} achieved	/	120	120
Time cost (s)	36.16	504.01	11.70

2.3 Power Amplifier Enhancements

Hardware complexity must be minimised as well in the power amplifier system implementation of large antenna arrays. The efficiency of the power amplifiers must be further increased, and the overall energy needed for digital signal processing has to be minimised in power amplifier array implementations. In this task, a DPD algorithm will be designed which exploits the similarities of the different power amplifiers of a MIMO antenna array. Using the a-priori information of statistical correlations between the properties of the power amplifiers, the problem of N independent DPD estimation loops can be reduced to a coupled system where each DPD loop benefits from calculations and results of the other calculations. Thus, computational complexity in terms of the number of needed iterations and matrix sizes will be significantly reduced. It may be even possible to reduce the overall amount of running DPD loops. DPD with nonlinear recursive-least-squares estimation is a well-established technology for broadband PA optimisation. The two most important optimisation requirements are the EVM of the modulated signal and the out-of-band spectral energy imposed by the nonlinearities.

Objective

In a 5G MIMO array, both hardware complexity and energy consumption is a crucial factor for implementation. A major part of the signal stream has to be implemented for each path of the MIMO array. This includes the analogue frontend as well as many digital algorithms for signal processing of the array elements.

The power amplifier is only one component within the complex framework, but it is surrounded by a bunch of supporting measures to minimize the power losses. Optimizations can be done with two different approaches:

- Optimization of the efficiency of every single amplifier
- Exploiting a-priori knowledge of the amplifier array to save computing power

Both approaches are studied within this work package, while starting with the first one. The study will include classical methods for optimization as well as AI support.

Therefore, a digital pre-distortion (DPD) algorithm will be designed which is first applied and optimized to a single power amplifier and later on exploits the similarities of the different power amplifiers of a MIMO antenna array.

Popular DPD algorithms are based on behavioural modelling. The behaviour of broadband non-linear power amplifiers is very often modelled using Volterra series and their derivatives. The models are implemented with a certain complexity, depending on the quality of the used power amplifier and the utilised bandwidth. The PA behaviour will be characterized while finding a Volterra model which is close to the PA regarding its input-output relations. Only in recent times it became possible to sample the PA output signal coherently fulfilling the Nyquist criterion and to use this information for the full characterisation of the PA behaviour. Model parameter estimation is done with Least Mean Square (LMS) or Recursive Least Squares (RLS) algorithms in most cases. State-of-the-art solutions will find an optimum solution for a single PA.

The stimulus for the classical solution as explained above will be extended to serve as a training model for the AI based support as well. Therefore, an essential part of this WP is the generation of the training model and the validation by a reference hardware and, later on, by a real power amplifier.

The final outcome of this WP shall be the basis for the test bed implementation, which is done in WP6. For this reason, the WP shall establish a bit-true model for the finally selected algorithms for implementation. The bit true model shall be verified against the simulation work.

The following sub tasks will be executed in this WP:

1. Algorithmic design of classical DPD algorithms
2. Algorithmic design of DPD AI support
3. Algorithmic design of AI-enabled array analysis
4. Simulation of DPD algorithms
5. Simulation of array analysis
6. Complete system simulation
7. Implementation of a bit true model
8. Use of testbenches for generation of training data

Progress

The algorithmic design of the classical DPD algorithms has been completed in Q1/2021. This includes the programming of the Volterra model and the RLS algorithms to solve the Volterra kernel based optimization algorithm for the model. Afterwards, the synthetic model for non-matching stimulus templates (with Gaussian basis wave forms) has been programmed and tested.

Both matching and non-matching synthetic models have been used as stimulus for linear and non-linear model estimation first. Afterwards, they have been used for estimation of the inverse model and final simulation of the DPD algorithm with error compensation instead of error estimation only.

First samples of the test bed of WP6 have been taken with a non-linear reference chain and the model estimation and compensation has been tested with real measurement data.

Although it was originally intended to use MATLAB for the simulation work, first experiments with the Tensorflow framework together with Keras showed that it will be easier in terms of simulation and validation if the whole framework including the traditional algorithms would run in python instead of MATLAB. Therefore, the simulation framework has been shifted to a python-based platform. The core algorithms (which are programmed as very fast C++ models) have been integrated to the python framework to enable high speed simulation even if not yet on the final FPGA target.

The RLS algorithm

The RLS algorithm is capable of solving non-linear optimization problems as long as the output is a linear function of the input variables. Although, on the first glance, this is not true for the non-linear Volterra model, the problem can be separated in a non-linear part (mainly multiplications of different tap weights of the FIR filter) and a linear part (located in the RLS algorithm), where the RLS algorithm is fed with the already multiplied tap weights. This combination of Volterra model and RLS algorithm has been implemented and tested and shows a very good convergence even for hundreds or thousands of parameters, up to order 5 at least.

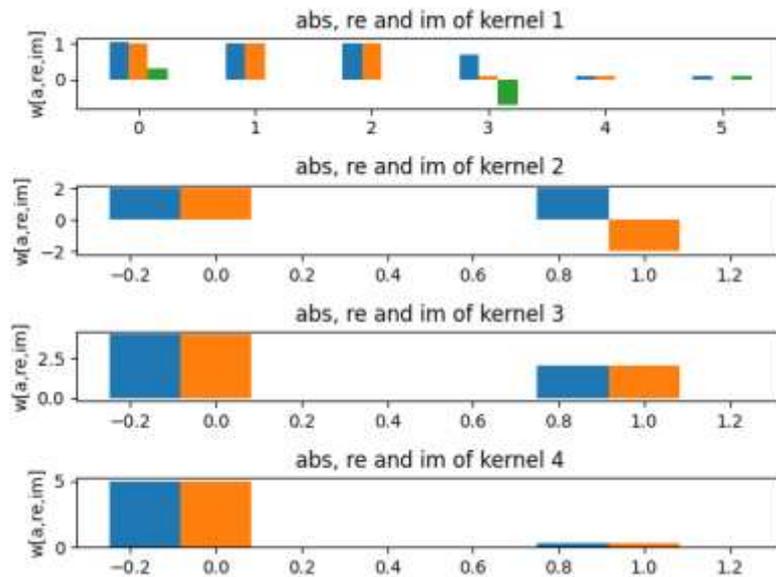


Fig. 56. Estimated kernels ($M=6$, $O=4$).

Figure 56 shows an example of a complex Volterra kernel with a linear part of memory $M=6$ and two additional polynomials of order 4 at delay 1 and 3 (no cross multiplications between the different delays). Only non-zero kernel elements are plotted. This kernel has been estimated via RLS estimation (without additional noise for illustration reasons), the resulting error is shown in Figure 57.

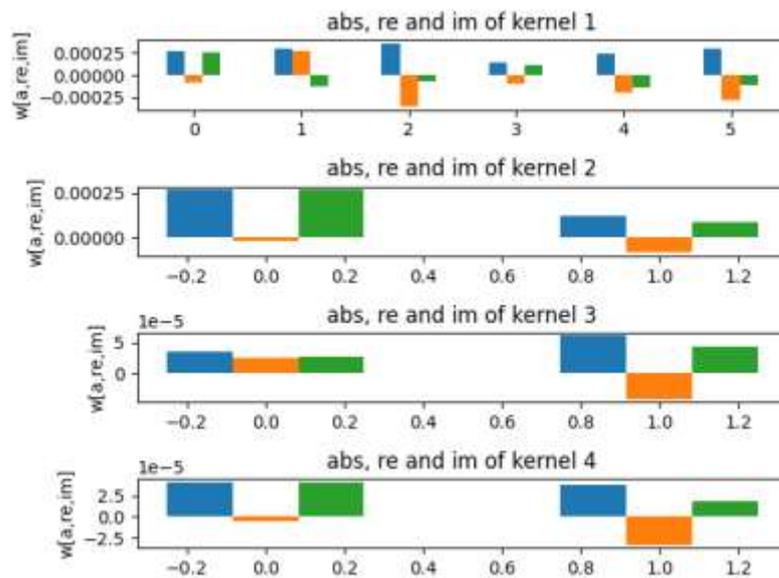


Fig. 57. Final kernel estimation error per kernel component.

It can be seen that over all kernels the convergence of the algorithm is very good ($<10^{-3}$), even at higher order. The RLS algorithm is known to be very fast, Figure 58 shows a convergence in less than 100 iterations of the RLS filter.

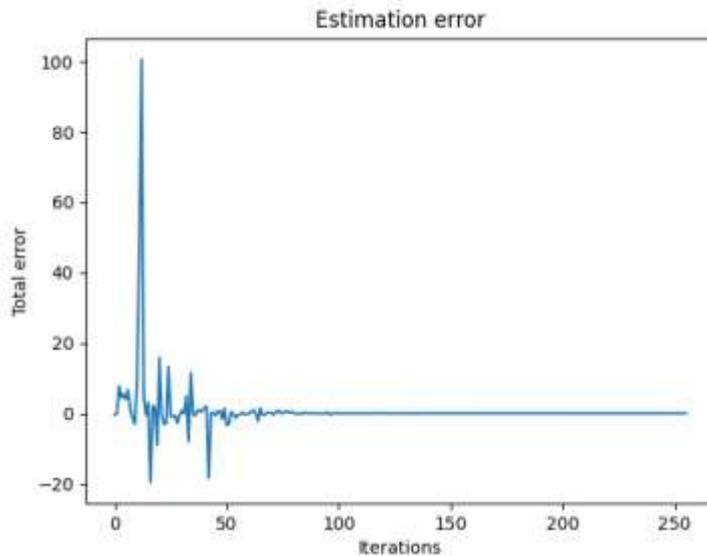


Fig. 58. Kernel estimation error over time.

The good and fast convergence can be observed while comparing the absolute values of the distorted signal (blue) with estimation of the distorted signal (green). This is illustrated in Figure 59 for the first 50 iterations.

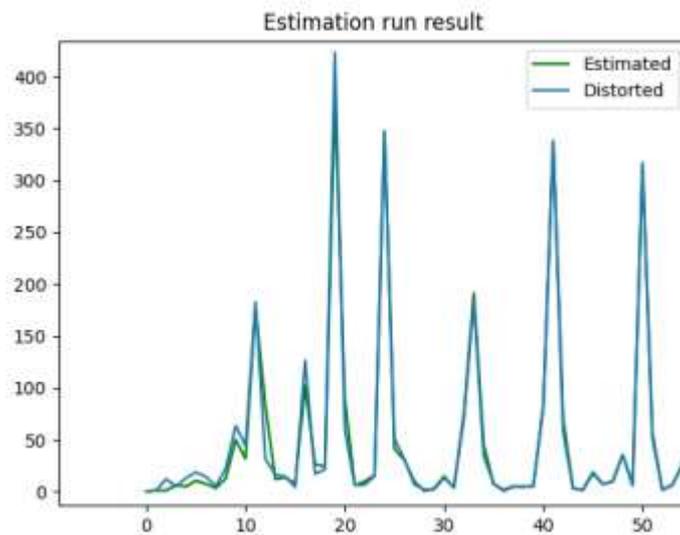


Fig. 59. Estimation run over time (magnified).

The algorithm has also been tested for noisy signals and with a bigger number of kernel coefficients and shows a similar good behaviour in this case.

Generation of synthetic training models

The training model generation has been done, programming has been concluded and first tests have been started as well. The training model strives to simulate the behaviour of a 5G Doherty power amplifier, which is one of the most used implementations for broad band amplifiers today.

A Doherty amplifier consists of two different amplifiers, where the first one is used for low power outputs and the second one is added when transmitting high power. The challenges in the design of a Doherty amplifier (beside the proper design of each separate amplifier chain) lies mainly in the

seamless combination of the two amplifier chains, especially in the switching area where the second amplifier is starting. Three aspects are important in the modelling of the amplifier:

- the non-linear characteristics of each amplifier
- the memory / spectral properties of each amplifier
- the behaviour of the combination network for the two amplifiers in terms of the final linearity of the characteristics, the phase stability, and the combined spectrum mask

All three aspects have been considered in the simulation model.

The following illustrations show a Doherty model implemented with a Gaussian base function set.

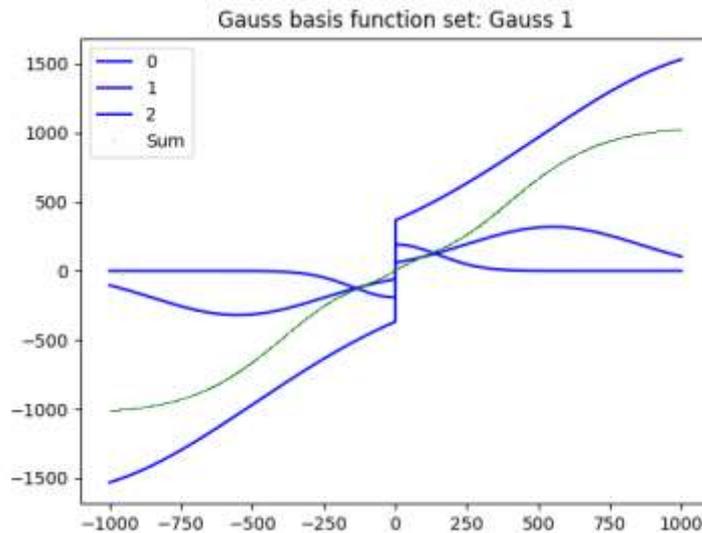


Fig. 60. Sample Gaussian base set for training model (Doherty amplifier).

Figure 60 shows an illustration of the composition of the non-linear characteristics out of the base function set. Three base functions with different normal distributions in terms of mean value, variance, amplitude, and DC offset have been used to compose a Doherty-like non-linear function. The function is normalized for negative input values so that it is smooth at point zero.

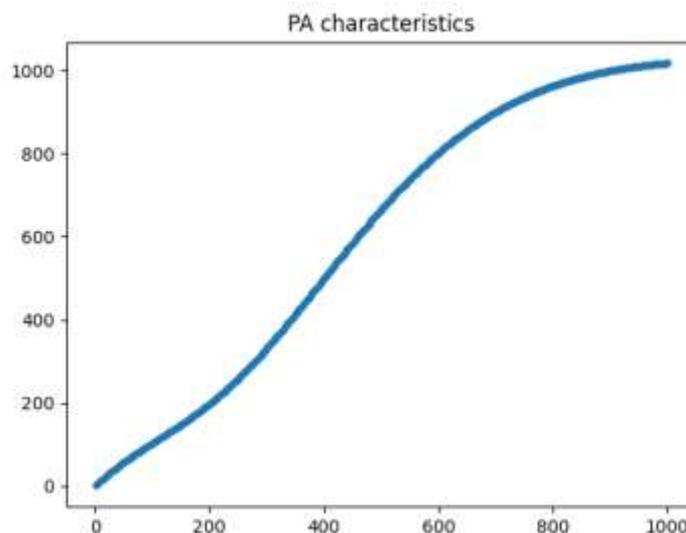


Fig. 61. PA characteristics without memory.

Figure 61 shows the output signal of the amplifier model versus the input signal based on normally distributed samples fed into the input and taken at the output of the complete PA

model when no memory is included in the model. As expected, the resulting function resembles the characteristics of the Gaussian base model.

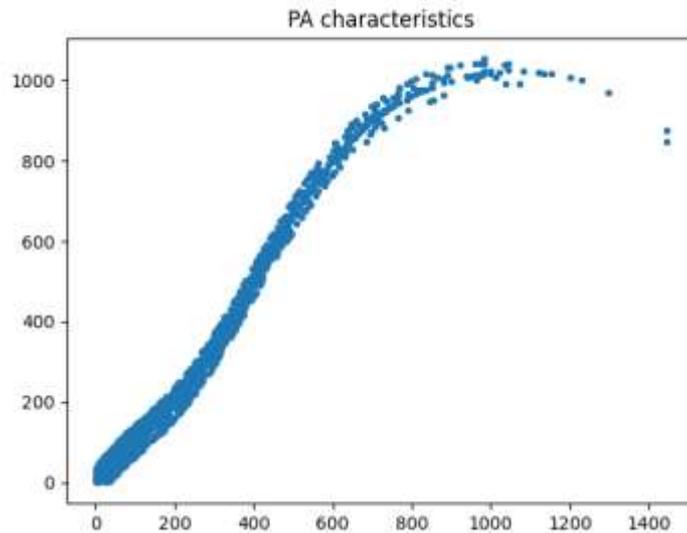


Fig. 62. PA characteristics with output memory.

When memory is introduced into the PA model, the characteristics of the amplifier significantly changes. Figure 62 shows the change when with a simple FIR filter at the output of the PA model. Due to the memory of the model, the input-output relation changes towards a power distribution since not only the actual input power has an effect on the output, but the most recent input value(s) as well. It has to be noted that, in this figure, the memory is located after the non-linearity.

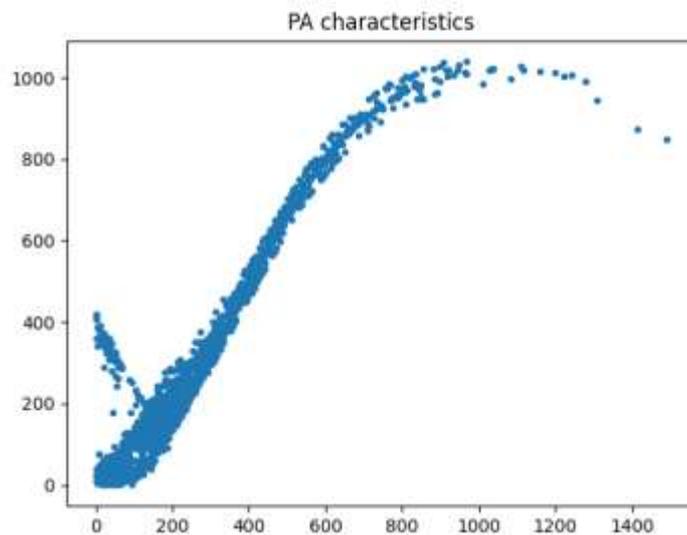


Fig. 63. PA characteristics with input memory.

Figure 63 illustrates the behaviour with a memory before the non-linearity. The resulting power distribution is even more complex since the non-linearity has an effect on every single memory tap.

Non-linear model estimation with synthetic models

Non-linear model estimation with zero-error models

The first test to check the non-linear RLS algorithm has been designed with Volterra models as the stimulus for the estimation. This results in stimulus signals which match the solution space, so that a near-zero error shall be achieved when estimating non-linear systems with memory. Several tests

have been made with zero-error models to validate the function of the RLS algorithm. Results have been presented in the chapter "The RLS algorithm" above.

Non-linear model estimation with non-matching models After check with zero-error models, the synthetic models have been taken to feed the estimation algorithm. Due to the nature of the now non-matching model template, a remaining residual error will be present which cannot be overcome by the optimizer. Besides, the number of iterations will increase due to this residual error.

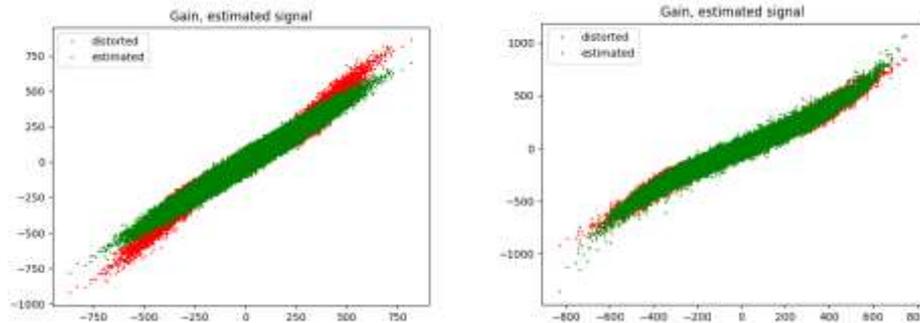


Fig. 64. Non-matching model estimation, linear (left) and order=5 (right).

Figure 64 shows the distorted gain curve using the non-matching Doherty model with a linear inverse estimation on the left and a non-linear inverse estimation with order=5 on the right. At low amplitudes, a good match of the nonlinearity can be observed. At higher amplitudes, the polynomial of order 5 is no longer able to follow the characteristics of the Doherty model. This behaviour is expected to be observed when using measurement data later on as well.

Non-linear model estimation with reference chain measurements

Although synthetic models are an excellent tool to figure out the behaviour of algorithms at well-defined preconditions, the real-world behaviour depends on additional, mostly unwanted effects as well. Some important items to be considered are:

- Volterra model is applied to band pass signal instead of low pass band
- influence of noise and other non-idealities like spurious emissions
- A/D conversion and D/A conversion quantization and sampling effects
- additional tasks to solve like resampling different sampling rates, fractional sampling, etc.

These effects have influence on both function and performance of the tested algorithms. For a step-wise approach, a hardware reference model is used before doing measurements with the unknown power amplifier. The reference chain contains both memory and non-linearity and is implemented by a simple limiter together with a matching network.

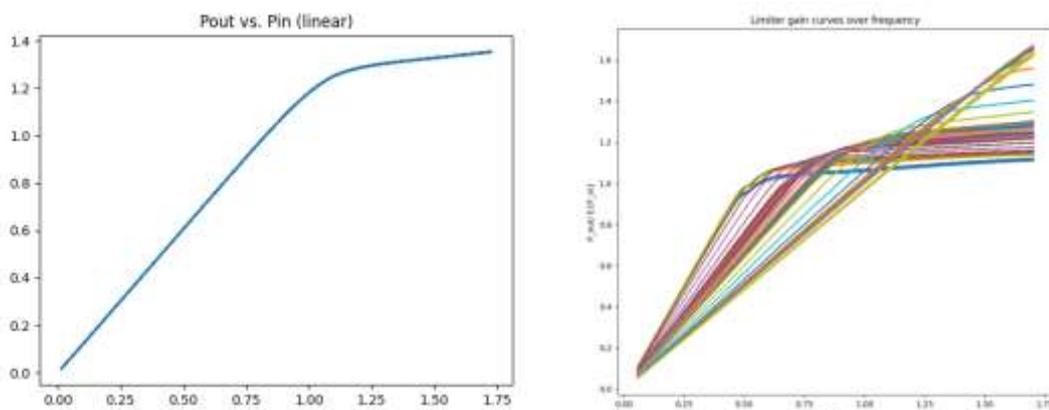


Fig. 65. Gain of reference chain at center frequency (left) and parameterized over 500MHz bandwidth.

The gain curve is caused by the pure limiter characteristics, whereas the gain curve differences in frequency are mainly originated by the matching network. As it can be seen, there is a very significant

difference in gain at different frequencies, which makes the model estimation a complex task, which actually is more challenging than the power amplifier to be analysed later on.

Model estimation has been performed at different orders of the Volterra model, with several different bandwidths. The following figures show an analysis with a bandwidth of 200MHz and a band-limited white noise stimulus.

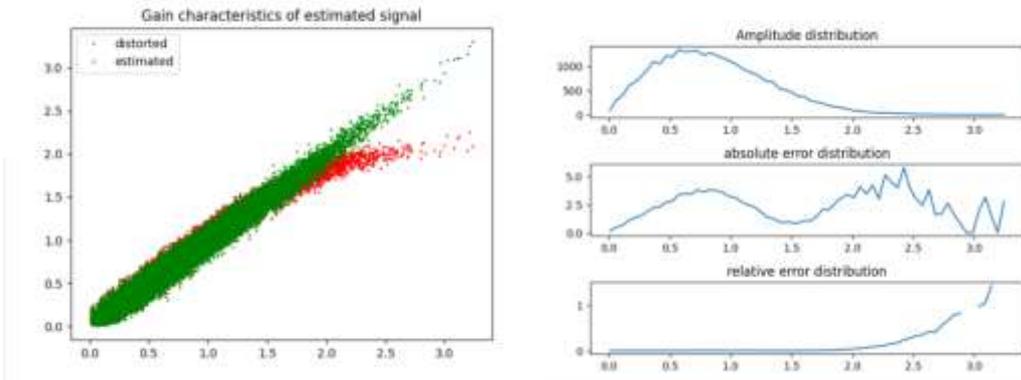


Fig. 66. Pout vs. Pin and amplitude distribution with linear estimation;

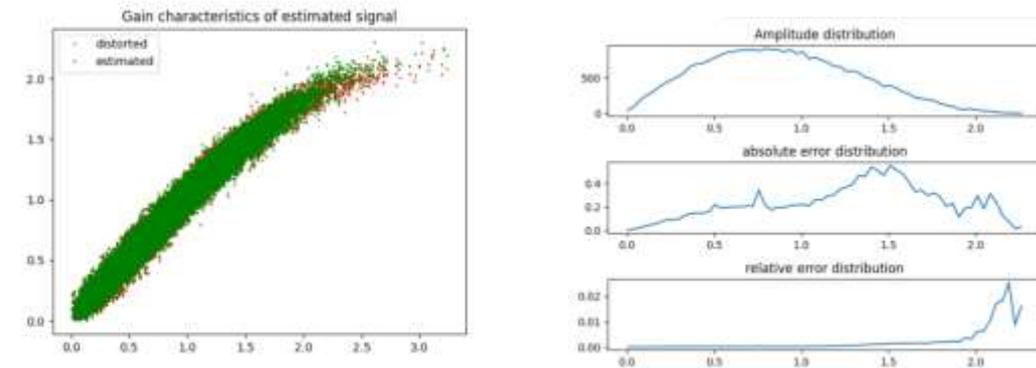


Fig. 67. Pout vs. Pin and amplitude distribution with O=3;

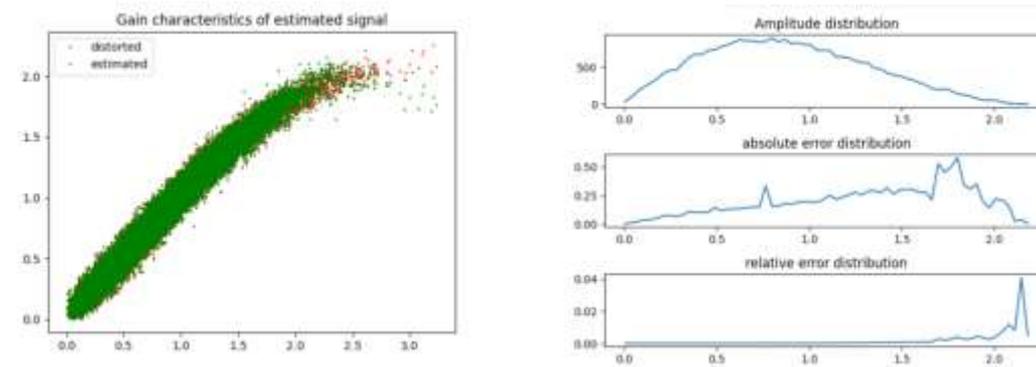


Fig. 68. Pout vs. Pin and amplitude distribution with O=5;

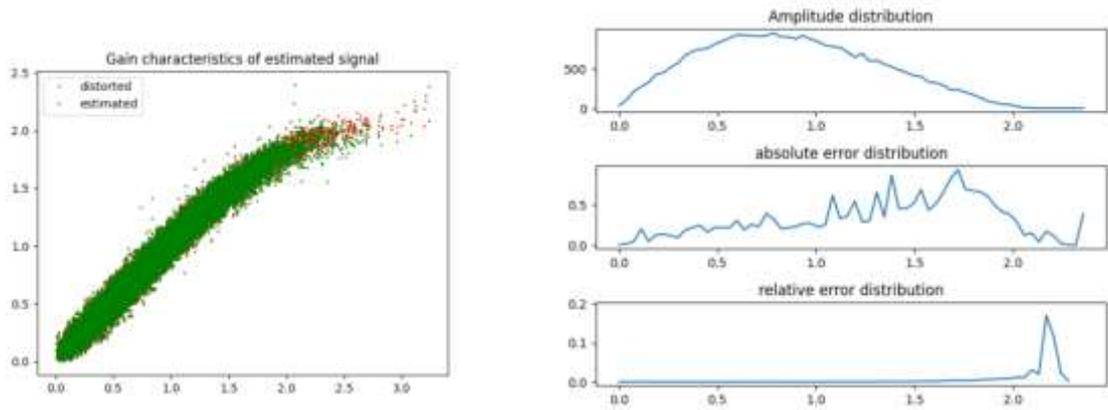


Fig. 69. Pout vs. Pin and amplitude distribution with O=7.

Order	Error Power
1	-23 dB
3	-33 dB
5	-33.7 dB
7	-31 dB

Table 8. Error power at different orders of estimation.

Observing the amplitude statistics of the error power, it can be observed that the error is typically greater at high signal levels, where the polynomial fitting is no longer accurate and there are less samples for estimation, too. It will be interesting to see if other estimation types will have a better fitting of the high amplitude samples.

DPD with synthetic models

The first step towards a DPD is the estimation of the inverse model. This has been performed for the same synthetic model as described above. The performance of the inverse estimation is of similar performance (error power -29 dB, order 3) as with forward estimation. The resulting gain curve can be seen in the next figure.

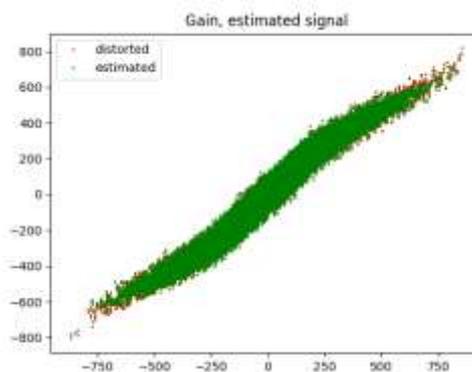


Fig. 70. Inverse model estimation as part of complete DPD.

A complete DPD is constructed through the combination of the estimated inverse model and the synthetic training model, i.e. the transmit signal is first sent through the inverse model to become pre-distorted and then sent through the training model. The result is a compensated transmit signal of the following form in Figure 71:

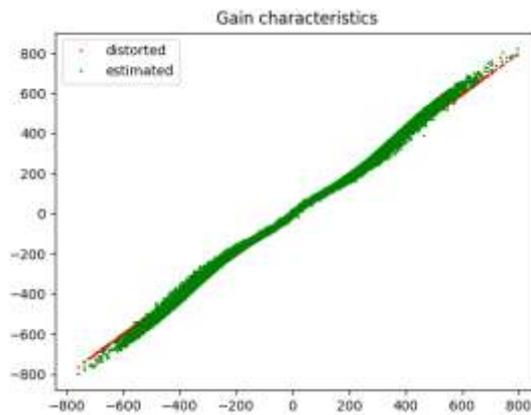


Fig. 71. Training signal (red) and compensated tx signal (green), complete synthetic DPD.

It can be seen that the linearity of the compensated signal (green) is close to that of the training sequence (red) and the point cloud is somewhat "narrower" than in the pure model estimation, which indicates the compensation of the memory effect. The overall error floor power which can be achieved with this pre-distortion is -24.5 dB.

Optimization of the signal processing algorithms

Nonlinear optimization is a complex task and the best parameters for the optimization algorithms and the Volterra model structure are not known a priori, nor can they be calculated in a deterministic way. Numerical simulation therefore is mandatory to optimize the RLS parameter configuration and the Volterra structure. It is important to know that the two domains depend on each other, so that these two domains are not orthogonal to each other, but may need some kind of iteration on the way toward a good overall optimization constellation.

The primary goal of the optimization is the minimization of the model error power, which is elaborated in this report. There are important secondary optimization goals like the overall structural complexity, but this goal will not be optimized in detail. However, too complex structures will be avoided by good engineering practice.

The optimization is done with real measurement data, exhibiting a strong non-linearity of the used power amplifier. This ensures the consideration of spurious effects which are present in the hardware like noise, band-limiting effects of the signal chain, quantization effects, and temperature drift.

Parameter Optimization for the RLS algorithm

The most important parameters for the RLS optimization is the memory of the RLS algorithm, which typically is denoted by the variable "lambda". Good choice of lambda will maintain the balance between robustness against non stationary effects in time like temperature drift and quality of long term estimation. Depending on lambda, the number of samples which are needed to let the algorithm converge will vary and thus must be studied. The number of needed samples, however, changes with the complexity of the Volterra structure as well. To summarize, in the simulations the following RLS parameters are used:

- lambda
- number of samples

Volterra Structure Optimization

The Volterra structure which is chosen for simulation is the memory-polynomial model. This model has been selected to reduce the needed number of Volterra coefficients in a first step, since it does not contain any cross-memory terms at higher order. The number of polynomials in the simulation can be arbitrarily selected (polylength). The order of the polynomials can be set in addition to the length of the polynomials (order). The polynomials are positioned at the center of the linear filter, with an adjustable offset (polypos). The linear part of the Volterra model can be adjusted independently from the higher order terms, so that the memory of the structure is independent from the number of

polynomials (M). To summarize, in the simulations the following Volterra structural parameters are used:

- polylength
- polypos
- order
- M

Simulation results

First, RLS algorithm optimization has been performed:

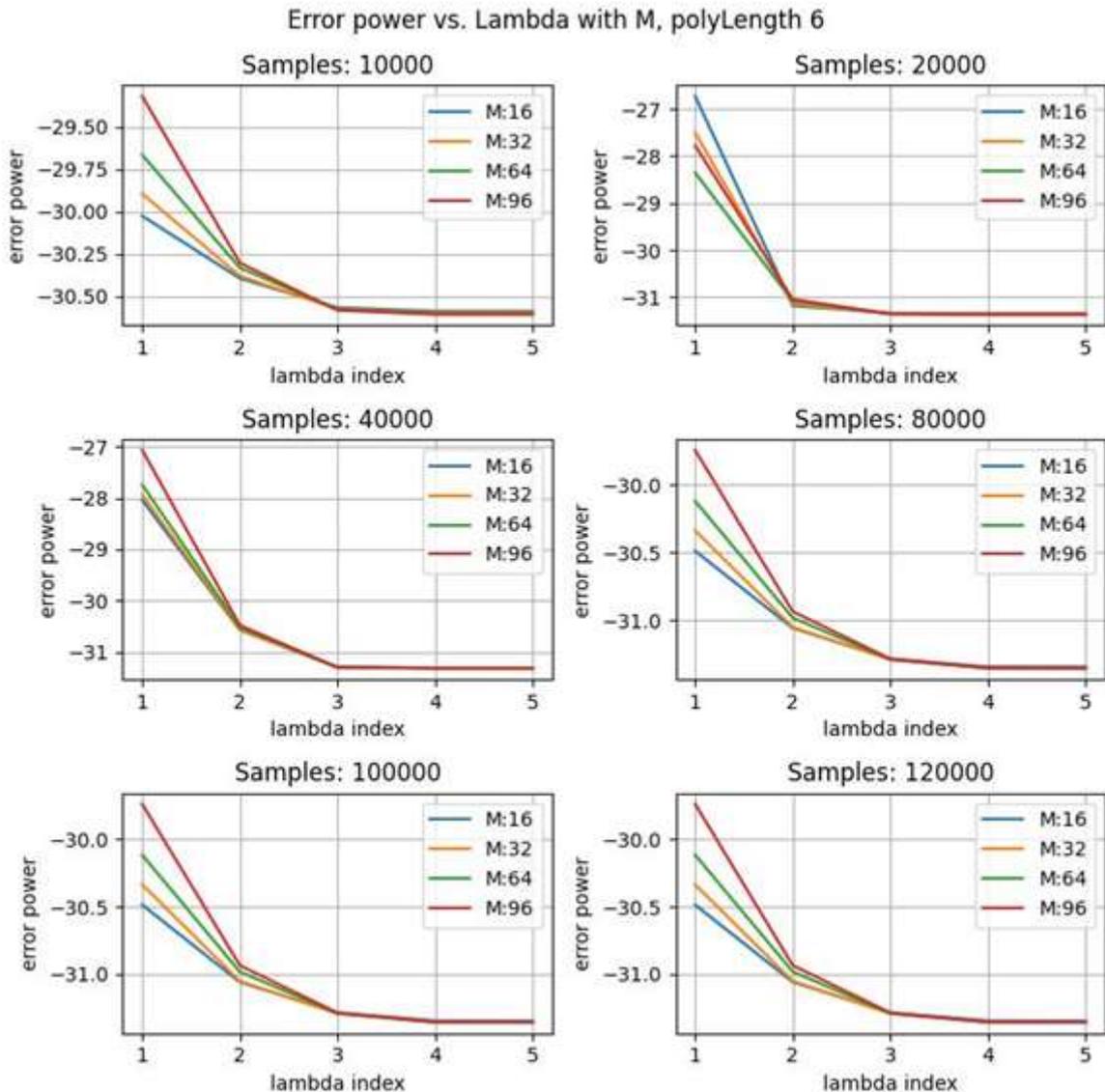


Figure 72 shows the error power for different lambda values, parameterized with the number of samples and for different linear filter lengths. The “lambda index” denotes the lambda values 0.99, 0.999, 0.9999, 0.99999, and 0.999999. It can be seen that a minimum lambda value of 0.999 must be chosen to achieve good convergence of the RLS algorithm. A number of 10000 samples is sufficient to achieve an error power better than -30dB, but error power will be slightly lower when increasing the number of samples up to 120000. However, regarding the complexity of the final algorithm running in hardware, the additional gain is less than 1 dB, so that a number of samples

~20000 may be a good choice. A linear filter length of 16 taps is sufficient for good results. All these simulations have been performed at a polynomial order of 5.

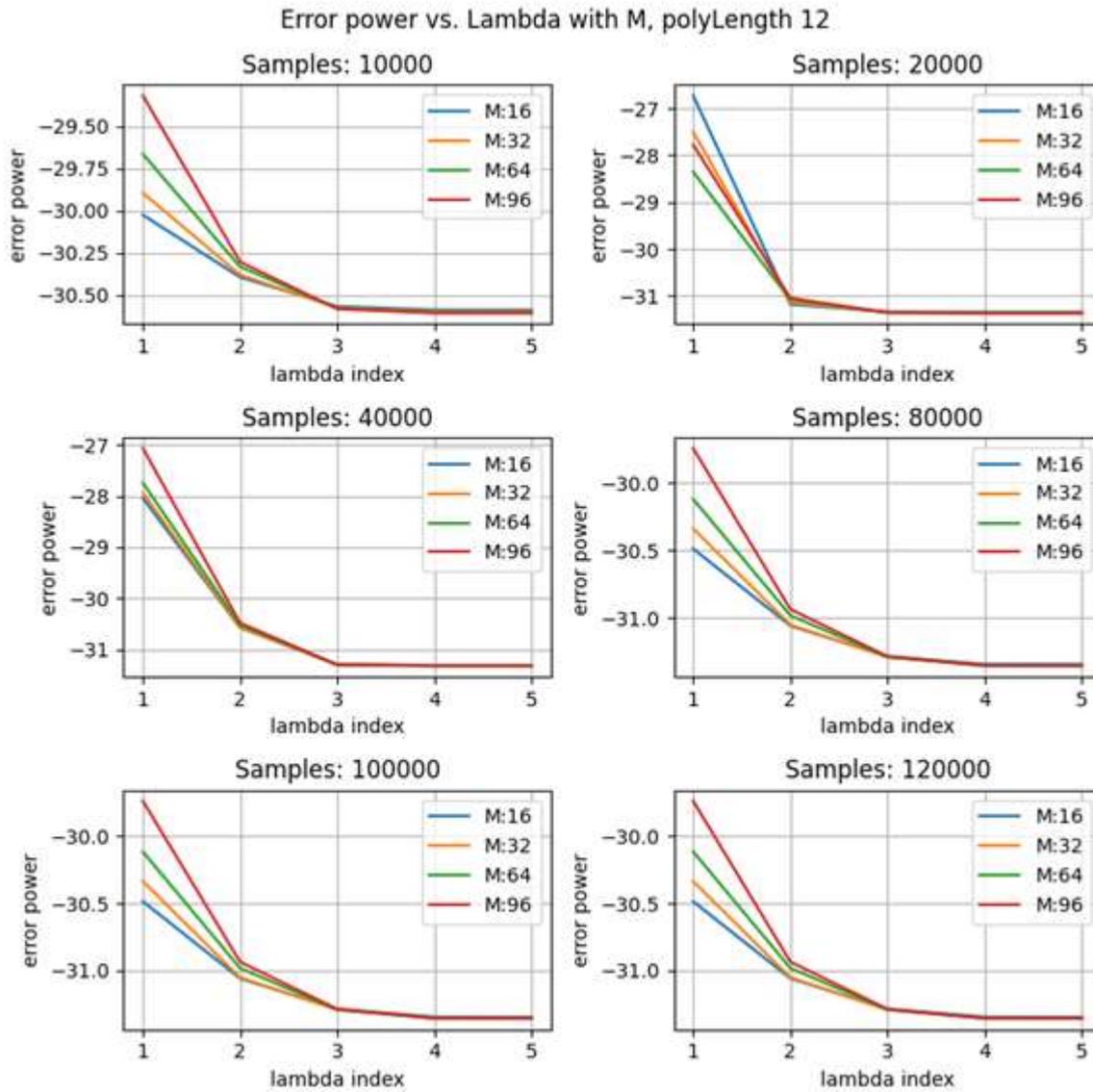


Fig. 73. Error power in dB at different lambda values, polynomial length=12.

Figure 73 shows the same as Figure 74, but for a polynomial length of 12. At order 5, no additional gain is achieved when using longer polynomials, nor must the RLS parameters be changed.

The next step is the Volterra structure analysis:

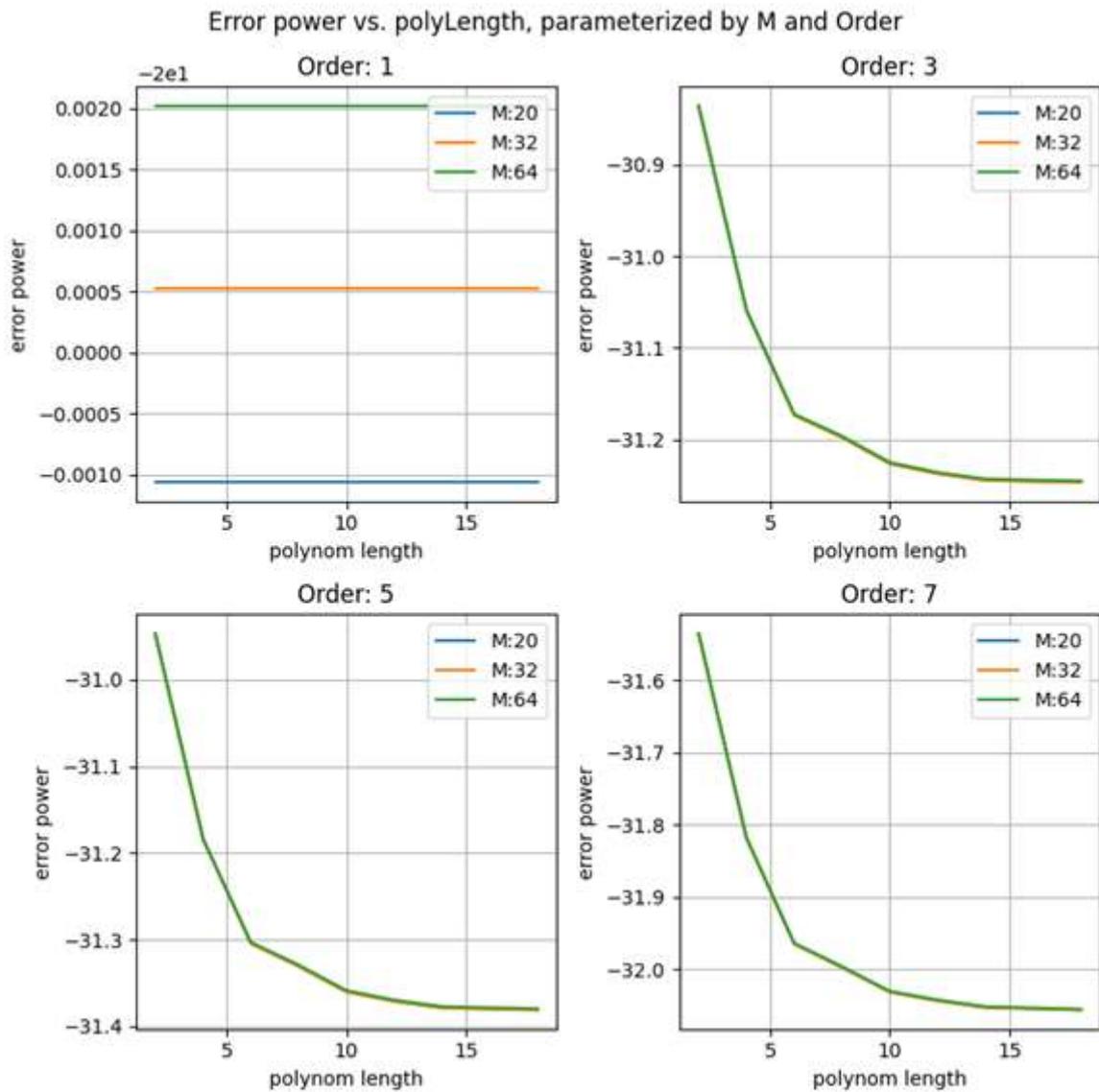


Fig. 74. Error power in dB at different length of polynomial, for M and order.

Figure 75 shows the error power at different lengths of the Volterra polynomial, parameterized by order and memory length M at sufficient number of samples. The linear optimization fails, what is obvious since the DUT has a strong nonlinear behaviour. For all simulations, results are better with higher number of polynomials, approaching a limit with ~14 polynomials. With order=3, a good error power of below -31dB can be achieved. Order=5 and order=7 both give slightly better results, an additional ~1dB can be achieved. A polynomial length of 12, however, seems sufficient for implementation. The memory length M of the linear part has no effect on the quality of the results.

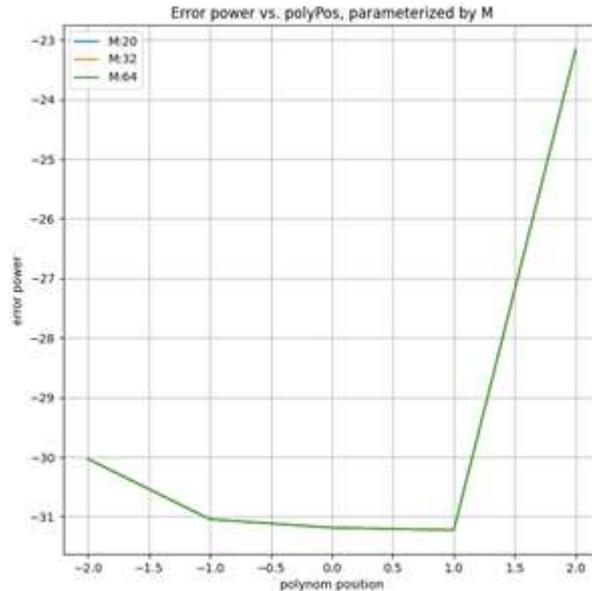


Fig. 75. Error power in dB at different positions of the polynomial

To validate the correct choice of the position of the polynomial structure with relation to the linear filter, the position has been changed around the center. Figure 76 shows the result of this simulation. It can be seen that at an offset of zero the error power is minimal.

KI System Model

Starting from the basic knowledge of the Volterra Approach IMST also investigated the use of artificial intelligence (AI) using TensorFlow Keras. The idea would be to use the many measured points as input data to train a neural network and thus enable to generate a model of the Tx-System with many non-linearities and at the end having a neural network with direct output rather than non-linear calculations.

The neural networks inner function is chosen to be “tanh” type of function since tanh- is pretty close to the saturation behaviour of power amplifiers in their non-linear region. In the next figure a fixed frequency measurement is depicted (red) with the “modelled” representation using a simple neural network (NN). As can be seen the non-linear behaviour and the gain change of this Doherty amplifier can be modelled very accurate with such a neural network using just enough measurement data and no knowledge about filtering and function.

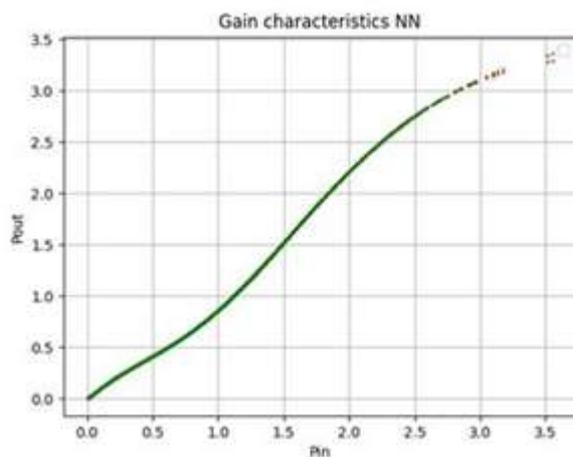


Fig. 76. Training signal (red) and modelled (green) using a neural network.

From this experiment neural networks seem to be a perfect match to model the system behaviour and thus to enable a very good correction. This is true for narrow bandwidths, measuring the very same PA with a broadband noise signal thus having a lot of frequencies and power levels the PA response is not only non-linear behaviour. The answer to such input signals is a time-dependent output signal and thus the neural network is reacting to the training data with a model of a mean value system behaviour. This can be seen in the next plot, where the red measured dots are modelled by the neural network with a mean gain curve (green) right in the median of the data points.

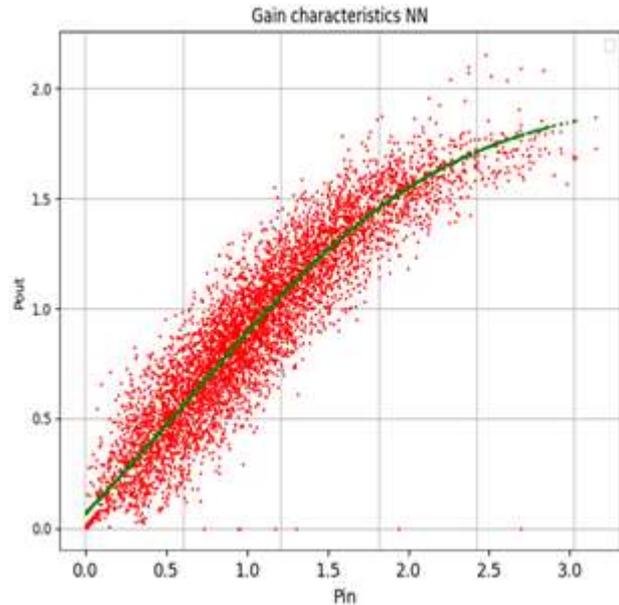


Fig. 77. Broadband training signal (red) and modelled (green) using a neural network.

To encounter the time dependent behaviour of the system we change the use of a simple neural network to a more modern KI-algorithm such as long-term short-term memory (LSTM) networks. With such networks the neural network is given the ability to use past data and thus enables modelling of time dependent training data. Due to the nature of TensorFlow no complex values can be used. Therefore, the algorithm is using a vector of real and imaginary part data instead of a direct complex value.

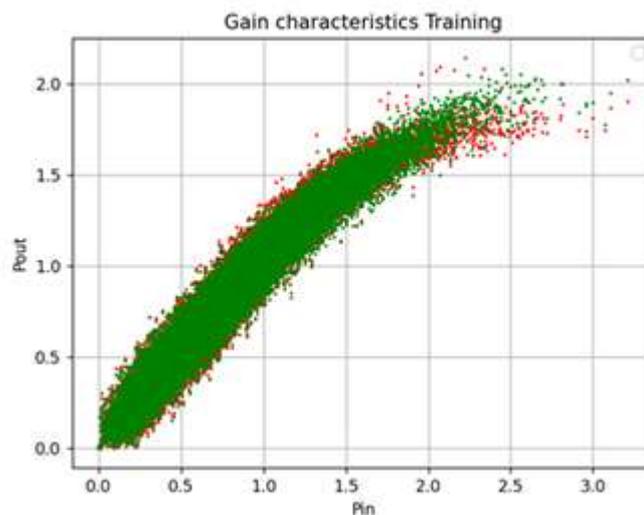


Fig. 78. Broadband training signal (red) and modelled (green) using a LSTM neural network.

Using the LSTM based neural network is resulting in a very good representation of the training data. The calculated error vector is in the range of -40 dB, which is a more or less **5-6 dB improvement** of the model accuracy compared to the Volterra series model.

KI Digital Predistortion

The KI System model is the most important input for the digital predistortion. With the system model the predistortion model can be trained to find an input based on the time variant behaviour of the transmitting system. To account for aging and device spread also KI models can be further trained during the life. This then would be a machine learning approach. This is not covered in the section. We concentrated to find a digital predistortion model trained using distorted LTE data.

The figure below is describing the training and function of the DPD implemented using Keras LSTM in python. To train the digital predistortion first a PA LSTM model is generated using measured input data (Tx data) and the distorted tx data (Tx* data). This is the first Training step (Training 1). The result of this training is a model of the system performance as it was described in the section above.

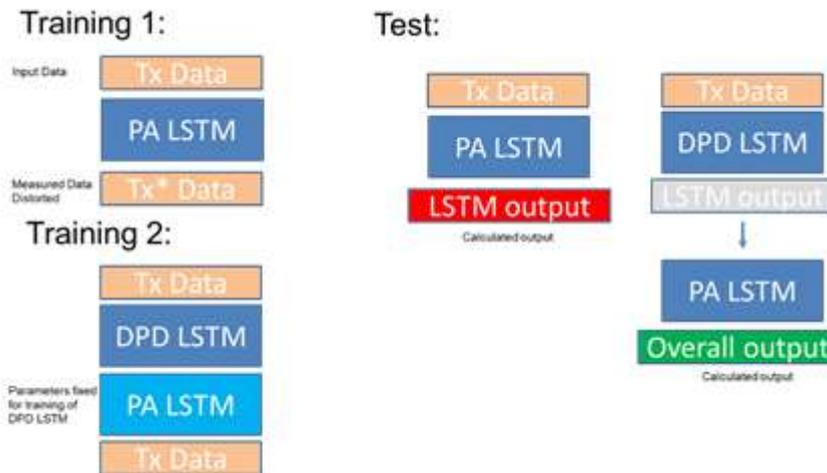


Fig. 79. Training and Test steps for LSTM based AI-DPD-algorithm.

In a next step a new LSTM neural network is built using the PA model as hidden layer and its output as output layer. The input to the PA-LSTM model is a neural LSTM network consisting of two hidden LSTM layers and a hidden linear layer with the same output dimensions as the input needed for the PA model. In our case it's a 2-dimensional array with real and imaginary amplitude samples.

This new model is then trained using the Tx input as input values and the very same values again as output. Ideally the DPD-LSTM model then is trained in a way that the input gets distorted and together with the PA output the results are linear and have the same quality as the Tx data before send through the amplifiers and filters. To enable proper training, the values of the PA-LSTM model are fixed during training.

For the later test the PA-model is removed from the DPD-LSTM model. As a result, we get two AI-models for the pre-distortion and the power amplifier, respectively.

DPD – simulation results

At IMST a power amplifier is used to get real measured data of a broadband LTE signal. This signal is then transferred through the PA and measured afterwards. The input is very clean and features a measured average EVM of -34.5 dB. Due to time relevant effects the EVM is dropping from below -40 dB to -34 dB. 10 different power levels are measured. Both at lower and higher power-levels the EVM is reduced due to quantisation noise of the output D/A converter, because the output power is reduced by reducing the amplitude of the D/A converter output. The EVM of the 10 power levels are given in the following picture:

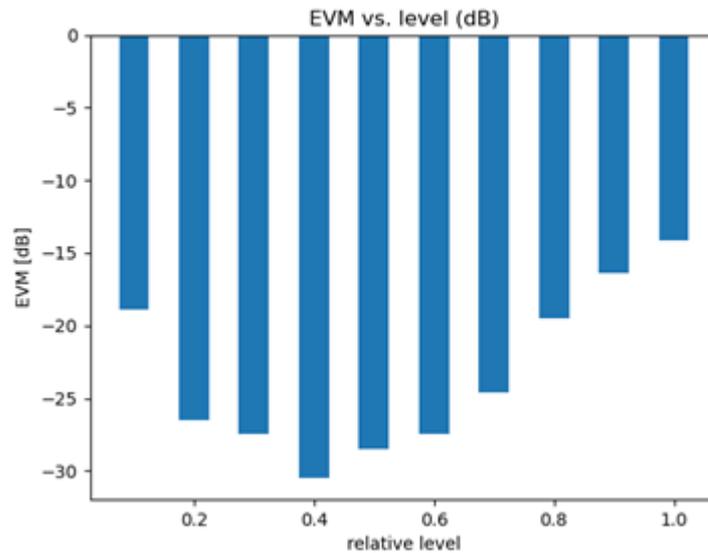


Fig. 80. EVM of 10 input power levels of the LTE signal used for measurement.

In the following set of figures the results of the PA modelling and the pre-distortion are shown. The results are presented for 3 different power levels ranging from slight compression to heavy compression, where both modelling and pre-distortion are more challenging for the algorithm. All data is normalized to enable easier comparison.

The left graph is showing the measured system output power versus input power. Red is the actual measured data while green is showing the PA model output for the same input points.

The middle graph is showing in red the output if no PDP would have been used and the blue dots are representing the output of the pre-distortion algorithm thus the input data into the PA. As a result the green data is representing the overall output of the PA with DPD input, thus ideally the corrected output.

To enable a better understanding the error vector magnitude is calculated over symbols for training data (blue), distorted output (red) and the compensated output using the DPD (green).

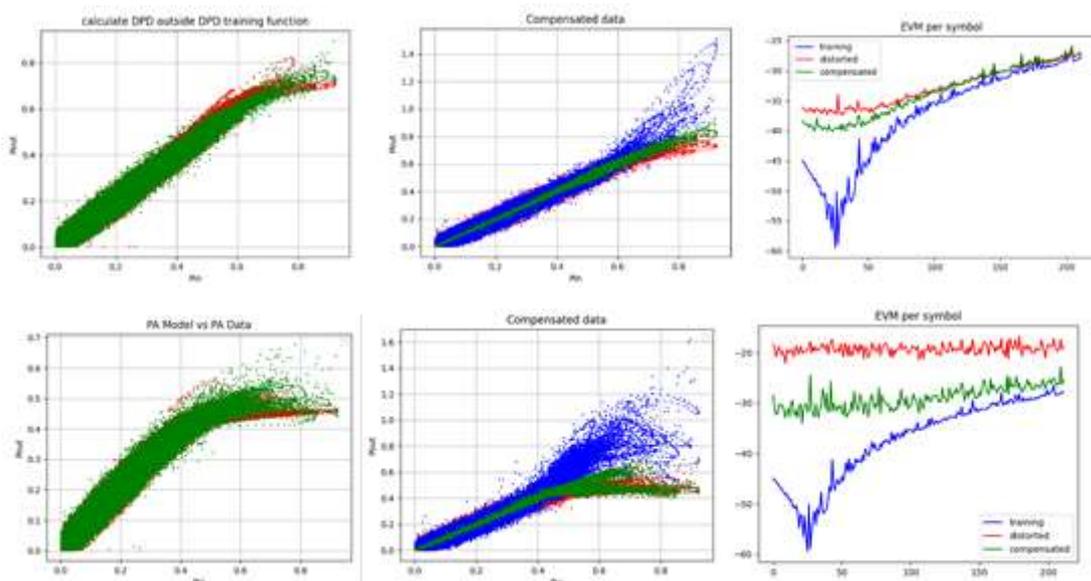


Fig. 81. Simulation results of PA model and DPD correction using LTE signal measurements.

As can be seen from the data plots the PA model is fitting very well to the measured data. This is very good, the DPD is relying on a good PA model. For simulation purpose the PA model is used for both DPD training and as PA system model to prove DPD functionality. A hardware implementation of the DPD must be made to demonstrate the function in a real system.

Nevertheless, the simulated data is clearly showing a very good performance of the DPD also when the output signals are facing heavy compression. At power levels with low compression the compensated output is as good as the input data and almost all non-linearities can be compensated even for high bandwidth systems. At deeper compression the DPD enables almost 10 dB improvement of the EVM compared to a non-compensated output.

This can be used as additional feature to the DPD. We can use the measurement data to safe the highest available output power provided by the output amplifier, then we can directly restrict the input signal to this value, thus disabling many dBs of compression. With such a limitation the EVM can be corrected to the input signal quality while still maintaining compression of the output and thus using the effective region of the output amplifier.

The results of this investigation are given in the following figure. While the un-compensated output is having EVM values worse than -20 dB, the corrected, but still in compression output EVM is reaching <-30 dB values and at later symbols even reaching the same levels as the input signal.

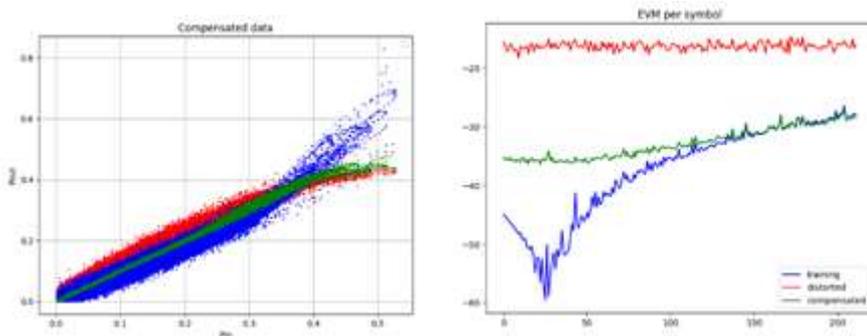


Fig. 82. Simulation of DPD performance using output limitation to saturated output power.

The LTE data measurements have been performed at 10 different output power levels. For all levels the DPD algorithm output performance is plotted together with the non-compensated direct output EVM and the ideal EVM of the input signal. Obviously at low power levels the amplifiers introduced non-ideal behaviour and is leading to large error vectors. Then a sweet spot is reached where the PA is showing very good error vector and thus almost no distortion.

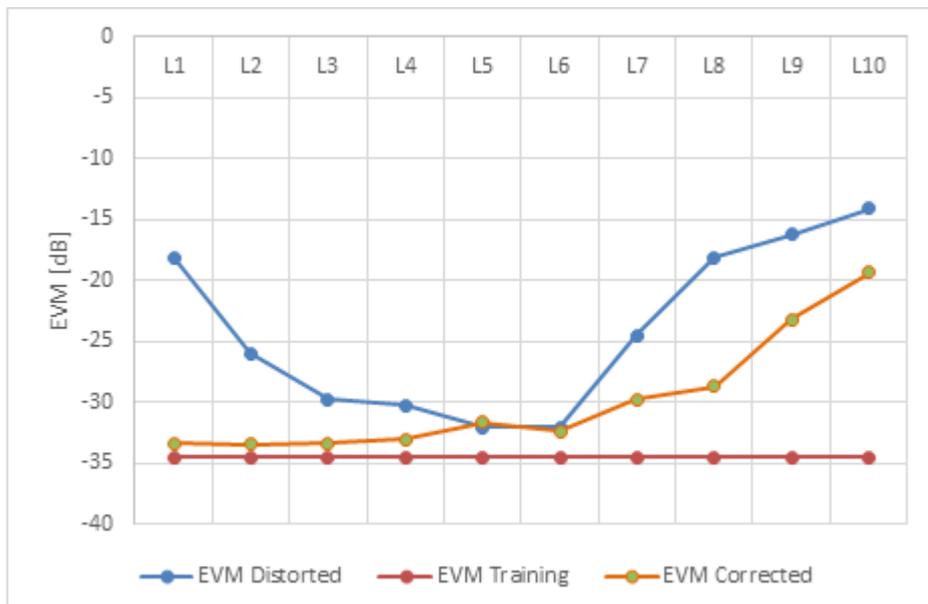


Fig. 83. Simulation of EVM performance for different power levels.

At increasing power levels, the output signal is starting to get distorted by the time variant and non-linear behaviour of the used power amplifier. This leads to an increase of the EVM from below -30 dB to values reaching -15 dB, where the complex data points are fading into each other. Although the DPD is not perfect it enables very good correction at low input power levels. Beyond the PA linearity sweet spot, the DPD is capable of correction of the EVM by up to 10 dB. If using the output level limitation feature even better results are possible.

From the current design, it can be concluded that the classical algorithms are validated per simulation and work as expected. The training model shows a great flexibility in terms of model non-linearities and model memory, especially when configured with many coefficients and different used bandwidths. The reference chain works as expected and the used algorithms work with the taken training data from the measurement chain as well. However, the non-linearity of the limiter component is extremely frequency dependent, so that inverse model estimation and pre-distortion do not work very well with the classical signal processing algorithms (which is known from theory). Furthermore, it could be shown that LSTM AI-models are very suitable for predistortion of PAs. In the analyses carried out so far, it was even possible to achieve a slightly better EVM performance with the AI approach. However, the two approaches are not directly comparable. There are different configuration parameters for both models that have a direct impact on the performance. Further investigations are required to carry out a meaningful comparison.

IMST has received an approval for the extension of the project by 6 months: Therefore future work will be dedicated to the remaining sub tasks of WP T3.3:

- The reference loop will be used to generate real training data to validate the algorithms which have been tested with synthetic training data so far
- With support from WP6 (test bed), real data from a 5G power amplifier will be taken as soon as the implementation is implemented and ready to use. With this data, the first complete system run will be enabled to get even more realistic model data

For this 5G amplifier, instead of using noise-like signals, signals with 5G physical layer modulation schemes will be used as well to check the performance of the algorithm for various modulation formats

3 Summary

In terms of dissemination, at the starting point of the AIMM project, a press release in the 6GWorld website around the RIS aspects of work within AIMM WP3 was produced.

A closed workshop on RIS was organised later in Jan. 2021, where key stakeholders within the AIMM project, as well as several partners from outside of the AIMM consortium working on this topic, were invited to present their latest research findings. The event was well received and facilitated closer interworkings with some key external partners.

Further, an open RIS workshop, jointly organized by AIMM, RISE-6G, and ARIADNE project consortiums working on radio intelligent surfaces, was held at the IEEE ICT'21 conference. This AIMM WP3 dissemination event, which took place in June 2021, included original paper presentations and invited talks from experts on the topic of RIS.

An extended version of the RIS workshop described above was also held at IEEE PIMRC'21 conference, September 2021. The workshop provided an excellent opportunity to discuss the latest research progress on the technology as well as the road ahead for standardisation, and eventually deployment of RIS. This event attracted the most attendance within the IEEE PIMRC'21 workshop program – highlighting the importance of RIS to the community as a key candidate technology trend for 5G-Adv/6G wireless systems.

A public newsletter was also issued based on the work within AIMM WP3 in August 2021. This document provided an outline of the transmission technologies and architectures being developed within the work-package, namely, Antenna Arrays, Reconfigurable Intelligent Surfaces, and Power Amplifier Enhancements, with AI/ML serving as a key tool for systems operation and optimisation.

Furthermore, AIMM WP3 participants were instrumental in establishing a new Industry Specification Group (ISG) on RIS at ETSI, the very first pre-standards group on a 5G-Adv/6G topic aiming to bridge the gap between research and standards. This group will provide the means for disseminating the results of the AIMM project, and indeed paving the way for all other collaborative projects working on these topics, onto future standards. Results from the AIMM WP3 have already been disseminated into the work of ISG RIS in the form of ETSI technical contributions made into the work programme.

WP3 partners have further disseminated their results in the form of several patent applications and paper publications. A full list of all dissemination activities of WP3 is available in the AIMM dissemination report (D1.6).

In summary, AIMM WP3 has delivered significant results in accordance with the project plan, and the partners involved will accordingly utilize the findings through exploitation activities.

References

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, p. 3590–3600, 2010.
- [2] P. v. Butovitsch, D. Astely, C. Friberg, A. Furuskär, B. Göransson, B. Hogan and J. L. E. Karlsson, "Advanced antenna systems for 5G networks," Ericsson White paper, 2019.
- [3] A. Shojaeifard (InterDigital), "[AI for Massive MIMO](#)", Cambridge Wireless Event, 2020.
- [4] M. Arnold and S. t. Brink, "Properties of Measured Massive MIMO Channels using Different Antenna Geometries," WSA 2019; 23rd International ITG Workshop on Smart Antennas, 2019.
- [5] M. Arnold, J. Hoydis and S. t. Brink, "Novel massive MIMO channel sounding data applied to deep learning-based indoor positioning," *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, 2019.
- [6] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis and T. L. Marzetta, "Massive MIMO is a reality—What is next?," Arxiv 1902.07678, 2019.
- [7] M. S. J. Solaija et al., "Generalized coordinated multipoint framework for 5G and beyond," *IEEE Access*, vol. 9, pp. 72 499–72 515, 2021.
- [8] P. Harris et al., "Performance characterization of a real-time massive MIMO system with LOS mobile channels," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1244–1253, 2017.
- [9] G. Lerosey et al., "Time reversal of electromagnetic waves and telecommunication", *Radio Science*, vol. 40, no. 6, 2005.
- [10] B. Li et al., "Precise transient electric field shaping with prescribed amplitude pattern by discrete time reversal", *IEEE Access*, vol. 7, pp. 84558–84564, 2019.
- [11] T. Nakanishi et al., "Storage of electromagnetic waves in a metamaterial that mimics electromagnetically induced transparency", *Phys. Rev. B*, vol. 87, no. 16, p. 161110, April 2013.
- [12] Z. Huang et al., "Transforming fading channel from fast to slow: Intelligent refracting surface aided high-mobility communication", in *IEEE Transactions on Wireless Communications*, 2021.
- [13] E. Basar, "Reconfigurable intelligent surfaces for Doppler effect and multipath fading mitigation", *Front. Comms. Net.*, vol. 2, p. 672857, May 2021.
- [14] M. Hayes. *Statistical Digital Signal Processing and Modeling*. New York: John Wiley & Sons, 1996.
- [15] C. Wu et al., "Channel prediction in high-mobility massive MIMO: From spatio-temporal autoregression to deep learning", *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 1915–1930, July 2021.
- [16] Y. Zhu et al., "An adaptive and parameter-free recurrent neural structure for wireless channel prediction", *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 8086–8096, Nov. 2019.
- [17] I. C. Wong and B. L. Evans, "Sinusoidal Modeling and Adaptive Channel Prediction in Mobile OFDM Systems", *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1601–1615, Apr. 2008.
- [18] V. U. Zavorotny et al., "Tutorial on Remote Sensing Using GNSS Bistatic Radar of Opportunity", *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 4, pp. 8-45, Dec. 2014.
- [19] A. Fish et al., "Delay-Doppler channel estimation in almost linear complexity", *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7632–7644, Nov. 2013.
- [20] R. Hadani et al., 'Orthogonal time frequency space modulation', *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2017, pp. 1–6.
- [21] J. Aronsson, F. Ling, A. Menshov, S. Zheng, and V. Okhmatovski, "New trends in analysis of electromagnetic fields in multilayered media," in *New Trends in Computational Electromagnetics*, Ö. Ergül, Ed., Scitech, pp. 401–474, 2020.
- [22] M. K. K. Ishfaq, T. A. Rahman, Y. Yamada, and K. Sakakibara, "8x8 Phased series fed patch antenna array at 28 GHz for 5G mobile base station antennas," in *Proc. IEEE-APS Top. Conf. Antennas Propag. Wireless Comm. (APWC)*, Verona, Italy, pp. 160–162, 2017.

- [23] “Link budget analysis for Reconfigurable Smart Surfaces in aerial platforms”, Alfattani, August 2021.
- [24] A. Burr, M. Bashar, and D. Maryopi, “Ultra-dense radio access networks for smart cities: Cloud-ran, fog-ran and “cell-free” massive mimo,” arXiv preprint arXiv:1811.11077, 2018.
- [25] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, “Local partial zero forcing precoding for cell-free massive mimo,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4758–4774, 2020.
- [26] D. Maryopi and A. Burr, “Few-bit csi acquisition for centralized cell-free massive mimo with spatial correlation,” in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2019, pp. 1–6.
- [27] E. Björnson and L. Sanguinetti, “Making cell-free massive mimo competitive with mmse processing and centralized implementation,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2019.
- [28] M. Bashar, H. Q. Ngo, K. Cumanan, et al., “Uplink spectral and energy efficiency of cell-free massive mimo with optimal uniform quantization,” *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 223–245, 2020.
- [29] J. J. Bussgang, “Crosscorrelation functions of amplitude-distorted gaussian signals,” 1952.
- [30] F. Wiffen, M. Z. Bocus, A. Doufexi, and A. Nix, “Distributed mimo uplink capacity under transform coding fronthaul compression,” in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, IEEE, 2019, pp. 1–6.
- [31] A. Burr and D. Maryopi, “On the modelling of coarse vector quantization in distributed massive MIMO,” in *2021 IEEE Statistical Signal Processing Workshop (SSP)*, IEEE, 2021, pp. 346–350.
- [32] J. Conway and N. Sloane, “A fast encoding method for lattice codes and quantizers,” *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 820–824, 1983.
- [33] J. Convay and N. Sloane, “Fast quantizing and decoding algorithms for lattice quantizers,” *IEEE Trans Inform Theory*, vol. 28, no. 2, pp. 227–232, 1982.
- [34] M. V. Eyuboglu and G. D. Forney, “Lattice and trellis quantization with lattice-and trellis-bounded codebooks high-rate theory for memoryless sources,” *IEEE Transactions on Information theory*, vol. 39, no. 1, pp. 46–59, 1993.
- [35] S. Ragot, M. Xie, and R. Lefebvre, “Near-ellipsoidal voronoi coding,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1815–1820, 2003.
- [36] E. Björnson and L. Sanguinetti, “Scalable cell-free massive mimo systems,” *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.
- [37] H. Taghvaei, A. Jain, S. Abadal, G. Gradoni, E. Alarcón, and A. Cabellos-Aparicio, “On the enabling of multi-receiver communications with reconfigurable intelligent surfaces,” *IEEE transactions on Nanotechnology*, 2022.
- [38] X. L. Pei, H. F. Yin, L. Tan, and et. al., “RIS-Aided Wireless Communications: Prototyping, Adaptive Beamforming, and Indoor/Outdoor Field Trials,” *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627-8640, Sep. 2021.
- [39] D. Rotshild, E. Rahamim, and A. Abramovich, “Innovative Reconfigurable Metasurface 2-D Beam-Steerable Reflector for 5G Wireless Communication,” *Electronics*, vol. 9, no. 8, pp. 1191, Jul. 2020.
- [40] “A review of mutual coupling in MIMO systems”, Chen, 2018.
- [41] “5G Study on channel model for frequencies from 0.5 to 100 GHz”, ETSI, 2018.
- [42] “On the rate and energy efficiency comparison of RISs with relays”, Ntontin, 2020.
- [43] “Electromagnetic Waves and Antennas”, Orfanidis, 2016.
- [44] “Mutual coupling and unit cell aware optimization for Reconfigurable Intelligent Surfaces”, Quian, 2020.

- [45] L. Dai et al., "Reconfigurable Intelligent Surface-Based Wireless Communications: Antenna Design, Prototyping, and Experimental Results," in *IEEE Access*, vol. 8, pp. 45913-45923, 2020.
- [46] E. Björnson and L. Sanguinetti, 'Rayleigh Fading Modeling and Channel Hardening for Reconfigurable Intelligent Surfaces', *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, Apr. 2021.
- [47] E. Björnson and L. Sanguinetti, "Demystifying the Power Scaling Law of Intelligent Reflecting Surfaces and Metasurfaces," 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 549-553, 2019.
- [48] M. D. Renzo, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Net.*, 2019.
- [49] P. Wang, J. Fang, X. Yuan, Z. Chen and H. Li, "Intelligent Reflecting Surface-Assisted Millimeter Wave Communications: Joint Active and Passive Precoding Design," in *IEEE Transactions on Vehicular Technology*, doi: 10.1109/TVT.2020.3031657.
- [50] L. Subrt and P. Pechac, "Controlling propagation environments using intelligent walls," in *Proc. 2012 6th Eur. Conf. Antennas Propag. (EUCAP)*, Prague, Czech Republic, 2012.
- [51] X. Xiong, J. Chan, E. Yu, N. Kumari, A. A. Sani, C. Zheng, and X. Zhou, "Customizing indoor wireless coverage via 3D-fabricated reflectors," in *Proc. 4th ACM Int. Conf. Syst. for Energy-Efficient Built Environ.*, 2017.
- [52] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?" *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4498–4502, 2020.
- [53] L. E. R. Petersson and J.-M. Jin, "A three-dimensional time-domain finite-element formulation for periodic structures," *IEEE Trans. Antennas Propag.*, vol. 54, no. 1, pp. 12–19, Jan. 2006. doi: 10.1109/TAP.2005.861547.
- [54] H. Zahra et al., "Bending analysis of switchable frequency selective surface based on flexible composite substrate," in *Proc. 2019 IEEE Int. Symp. Antennas Propag. USNC-URSI Rad. Sci. Meet.*, Atlanta, GA, Jul. 2019. doi: 10.1109/APUSNCURSINRSM.2019.8889266.
- [55] G. H. H. Sung, K. W. Sowerby, M. J. Neve, and A. G. Williamson, "A frequency-selective wall for interference reduction in wireless indoor environments," *IEEE Antennas Propag. Mag.*, vol. 48, no. 5, pp. 29–37, Oct. 2006. doi: 10.1109/MAP.2006.277152.
- [56] [AIMM Leverages Reconfigurable Intelligent Surfaces | 6GWorld](#), Nov. 2020.
- [57] Workshop on Reconfigurable Intelligent Surfaces for B5G/6G, ICT'21 conference: <https://ict-21.org/workshop-3-reconfigurable-intelligent-surfaces-for-b5g-6g/>, June 2021.
- [58] Reconfigurable Intelligent Surfaces for B5G/6G, PIMRC'21 conference: <https://pimrc2021.ieee-pimrc.org/reconfigurable-intelligent-surfaces-for-b5g-6g/>, Sep. 2021.
- [59] AIMM WP3 Newsletter: Antenna Arrays and Reconfigurable Intelligent Surfaces <https://aimm.celticnext.eu/2021/08/31/newsletter-antenna-arrays-reconfigurable-intelligent-surfaces/>, Aug. 2021.

