AIMM

CELTIC-NEXT

Newsletter

December 2021
Issue 1

AIMM Testbeds and Platforms

Executive Summary

AIMM is a two-year European collaborative research and innovation project targeting radical performance improvements and efficiency dividends for beyond-5G/6G Radio Access Network (RAN), through novel multiple-input multiple-output (MIMO) enhancements and technologies, powered through and managed by the latest advancements in Artificial Intelligence (AI)/Machine Learning (ML).

This document provides an outline of the development of research testbeds within AIMM project Work-Package 6 (WP6). The developed testbeds, including both centralised and cell-free architectures, are utilized for proof-of-concept (PoC), evaluation, and demonstration of algorithms and technologies developed in the other work packages within AIMM. The overall objective is to verify and demonstrate the practicality of the concepts proposed in AIMM, as well as to capture and analyse performance data.

Testbeds and Platforms in AIMM

MIMO serves as a key air-interface technology in nearly all contemporary communications systems. MIMO, through utilization of multiple antennas at the radios, provides several benefits including enhancing spectral efficiency and quality-of-service. Despite significant performance improvements achieved through MIMO to date, there exists a significant gap between the theoretical limits versus practical performance of active antenna systems.

The above trend highlights the importance of further innovation on novel concepts and architectures for future cellular antenna systems. This is the main driver behind AIMM, a two year collaborative European research and development project defined in the area of end-to-end RAN delivery chain with six tightly coupled work packages, as illustrated in Figure 1.

The AIMM project WP6, titled “Testbed & Demonstration Development” focuses on design and build of testbeds for proof-of-concept, evaluation, and demonstration of algorithms and technologies developed in the other WPs. The overall objective is to verify and demonstrate the practicality of the AI/ML concepts proposed in the AIMM project as well as to capture and analyse performance data.

The total of five testbeds covers different use-cases (co-located versus distributed for antenna locations, centralized versus cell-less for operation) and technologies (real-time implementation using FPGAs versus non-real-time processing using CPUs and GPUs). The remainder of the document provides additional details and progress to date of testbeds and platforms being developing in AIMM WP6.

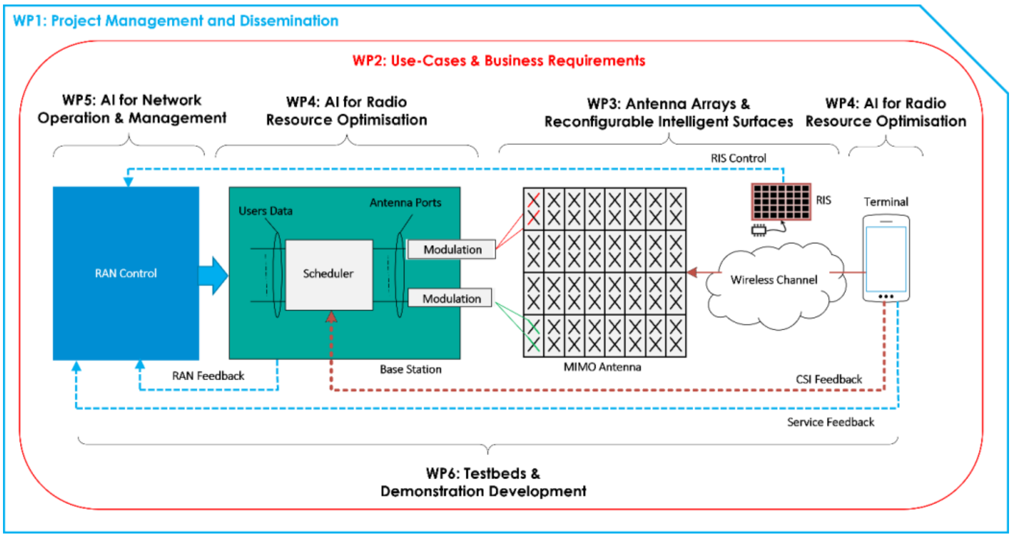

Figure 1. AIMM project work-package structure.

Massive MIMO Testbed

This task will focus on extending the existing University of Bristol (UoB) mMIMO SDR testbed achieved by connecting it to an external machine with AI capabilities to demonstrate the performance of the concepts proposed in WP4 and WP5.

The mMIMO testbed, depicted in Figure 2, comprises of a BS and up to 12 users. The BS is divided into 4 racks, providing 32 RF ends each, i.e., 128 in total. The RF ends are connected to a patch panel antenna array in a 4x32 configuration with vertical and horizontal polarisations operating at 3.51 GHz. The BS can serve simultaneously: a) up to 12 users with single antenna from 6 USRPs, or b) up to 6 users with two antennas. The system was designed and build to align closely with the TDD LTE air interface with a scalable sub-6 GHz carrier frequency. The 128 antennas are connected to 64 dual-channel USRPs divided equally into 4 racks as shown in Figure 2.

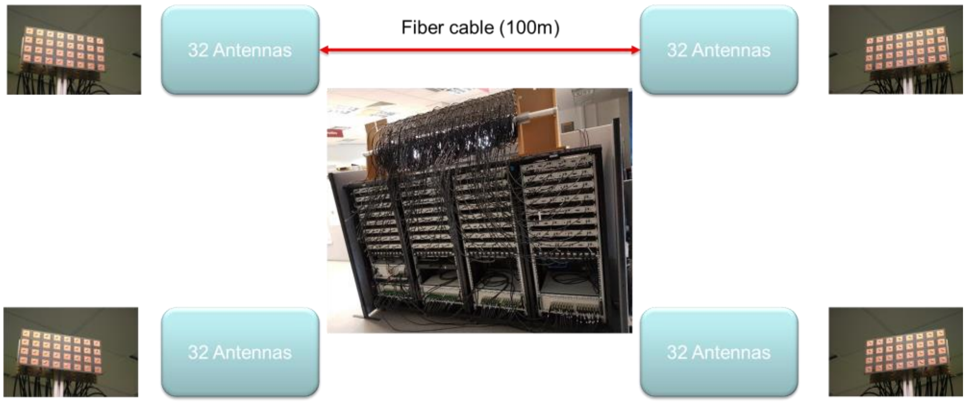

Figure 2. UoB 128-antenna massive MIMO testbed (BS side).

The mMIMO testbed was connected to an external machine controller as shown in Figure 3. This will enable extending the existing massive MIMO SDR testbed into AIMM SDR testbed. Data transfer connection was established between the mMIMO testbed and external machine controller by using Ethernet connection. The mMIMO testbed uses LabVIEW as a programming language, while the majority of ML and AI algorithms use Python-based frameworks. Therefore, a LabVIEW to/from Python interface was created between the mMIMO testbed and the external machine controller. This provides the external machine to control several of the mMIMO testbed functionalities. Real time data can be transferred from the mMIMO testbed to the external machine through an Ethernet cable. The external machine can then send commands to the massive MIMO testbed based on the results of ML&AI algorithms. This will allow for flexible coordination between WP4, WP5 and WP6. It will also provide WP4 and WP5 with the flexibility to modify their algorithms without impacting the work progress on WP6. Some AIMM algorithms that require a very low latency will be implemented on the FPGAs of mMIMO testbed.

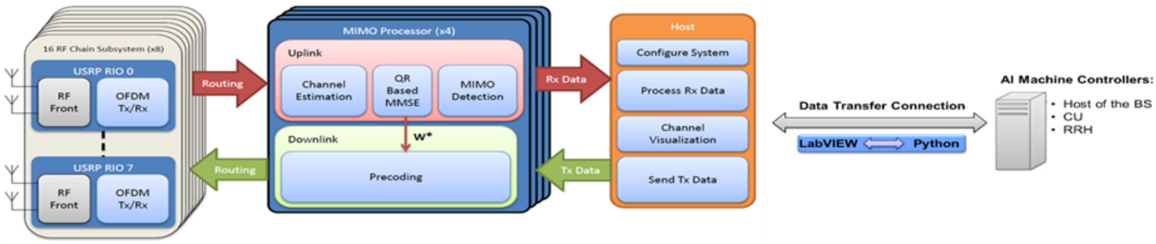

Figure 3. UoB mMIMO testbed controlled by external machine.

Digital Pre-Distortion Testbed

The core of the testbed is an RFSoC evaluation board as shown in Figure 4. The main board consists of the core component FPGA from Xilinx, together with peripheral devices interfaces. The RFSoC FPGA consists of a main part with the FPGA macro cell array, an arm-based processor and an RF section which consists of up/down conversion, resampling sections, and A/D and D/A conversion. The interfacing to the user is done via ethernet. The A/D converters support up to 8 12-bit 4.096GSPS ADCs, the D/A converter section consists of 8 14-bit 6.554GSPS DACs. With this hardware, a direct sampling of RF signals up to 4 GHz is possible. In the project, the D/A converters and the A/D converters on the chip are used to transmit and receive RF signals at 3.6 GHz center frequency with a bandwidth of 500MHz.

The RF chain consists of modular components which can be assembled in several different configurations by plugging them sequentially together. The following components have been provided for the RF signal chain: Wide band bandpass filter for image rejection, narrow band bandpass filter for rejection of spurious components and image rejection, Balun component to change between symmetric/asymmetric RF signal flow, 2 pre-amplifiers to compensate for the several attenuations caused by the other passive components, and the main 5G amplifier from NXP with an output power of 5W.

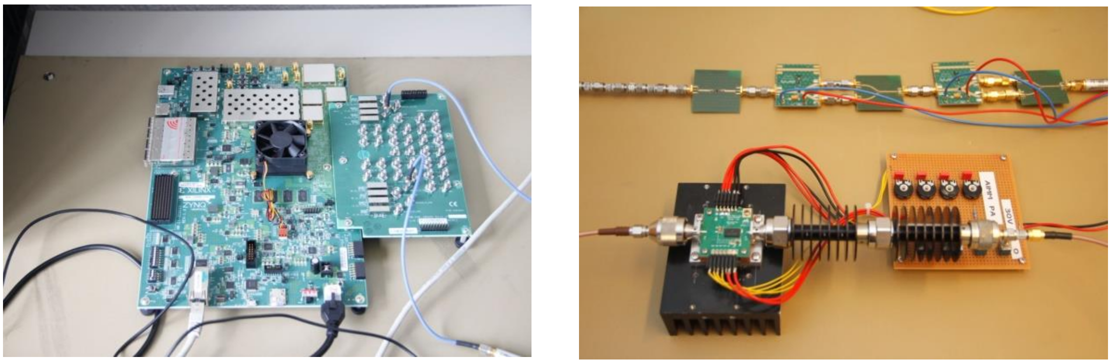

Figure 4. IMST RFSoC evaluation board and RF chain.

The gain curves at the left of Figure 5 shows a clear frequency selectivity of the amplifier gain, which results in a certain unwanted pre-distortion of a transmitted LTE signal shown on the right at high output levels. Within the project, these raw signals are used to be pre-distorted for the compensation of the non ideal transmission function to satisfy the 5G standard requirements.

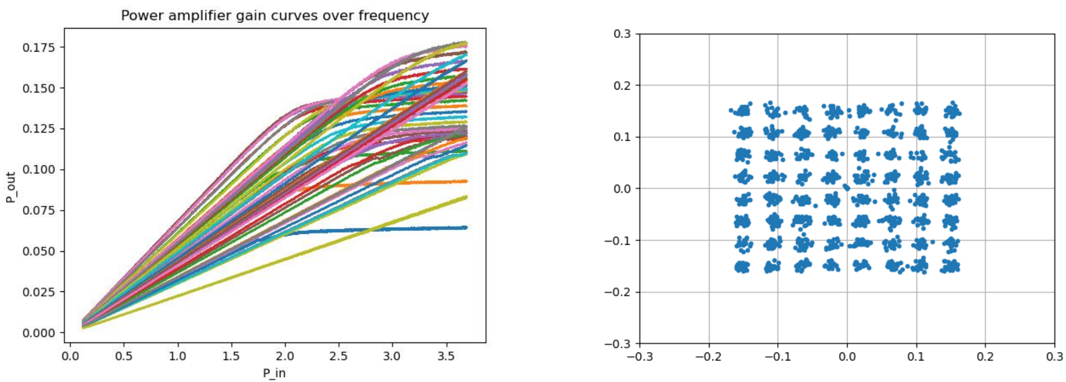

Figure 5. Gain curves over frequency and 256-QAM modulation symbol (IMST)l.

Distributed Cell-less MIMO Testbed

Comparing to the traditional centralized architecture, the distributed cell-less MIMO architecture has many more access points (APs), while each AP is equipped with much fewer antennas and transmit power. The advantage for such architecture is that all users could benefit from the distributed APs within the cell-less MIMO service area, where they could be jointly served by multiple APs nearby, instead of solely relying on the central AP. This will help address the problems such as the performance concerns for the cell-edge users, because with a scalable cell-less architecture, the service coverage can be easily expanded. The users at any location, no matter it is close to the edge or next to the centre, will receive the similar wireless access service.

There have been several challenges to implement such architectures, where traditional analytical based methods cannot deal with the time-critical signal processing tasks. One of the solutions is to push the AI algorithms to the network edges. This is because both real-time baseband signal processing and AI algorithms require high computing resources, while the distributed APs have less power and resources for computationally intensive jobs.

This task in AIMM exploits a hybrid solution of CPU/GPU and FPGA for the distributed APs, to accelerate the baseband signal processing and support the real-time computing requirements. Specifically, the general-purpose CPU/GPU architecture implements the baseband functionalities of the radio system exploiting CPUs for serial functionalities in bit-level processing, while the GPUs are used to accelerate computationally demanding functionalities with a high level of parallelism. The FPGA is used for low-level signal processing jobs at the physical layer, such as synchronization and channel status indicator (CSI) estimations, which could demand large computing time due to its high sample rate in the baseband.

So far, the key FPGA-based baseband modules, as well as the corresponding drivers for CPU cooperation have been completed. The modules provide the key functionalities to enable the embedded signal processing on the distributed APs, including synchronization, CSI estimation and BPSK demonstration. These FPGA modules have been developed and implemented on USRP X310 and N321, which have been tested in the lab environment with real-time experiments as shown in Figure 6.

Figure 6. Lab experiments for the CSI estimations with the developed baseband signal processing modules (Loughborough University).

Machine Learning for Layer 1 in GPU-enabled gNB

The goal of this task is to investigate the performance gains of AI/ML-enhanced PUSCH channel estimation on real-time GPU-based gNB architecture. This should provide useful insights towards future developments of native AI/ML air interface.

The hybrid architecture of general-purpose gNB architecture is shown in Figure 7. It implements the baseband (L1) functionalities of the DU of the radio system. It exploits CPUs for bit-level processing, while the GPUs are used to accelerate computationally exhaustive functionalities with a high level of parallelism. The RU is based on an USRP X310 SDR board, interfaced over split 8 (time-domain IQ samples) with the DU.

The processing architecture follows 5G NR release 15 SA and will be extended according to the needs of the project, e.g., by upscaling to a higher number of antennas. The functionalities will be enhanced to support the AI/ML algorithms for radio resource optimisation, developed in WP4, i.e., AI-enhanced PUSCH channel estimation.

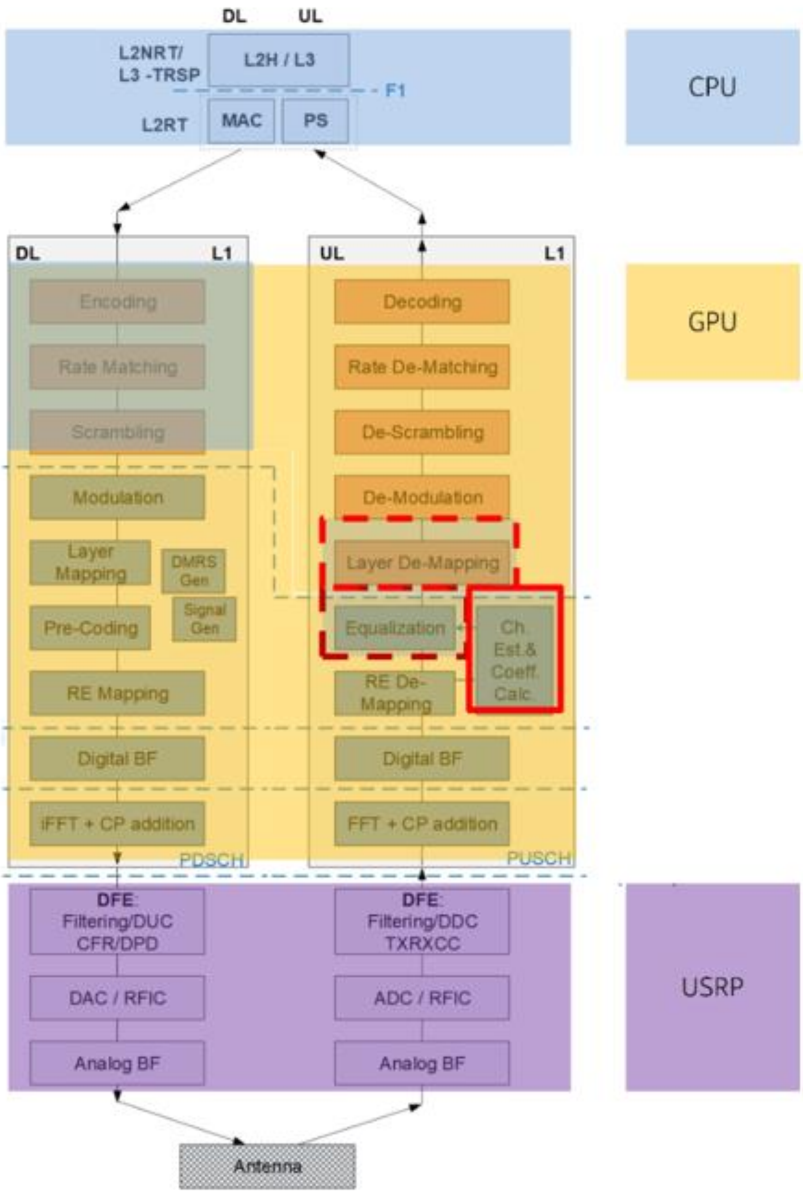

Figure 7. The overall architecture of GPU-enabled gNB (Nokia Bell Labs Stuttgart).

Since the original focus has been on optimizing the UL receiver chain for the shared data channel, i.e., PUSCH channel, of the gNB, the existing platform is extended such that it supports necessary tools for performance evaluation and data generation. This includes the following efforts:

- The implementation of a "mirrored" gNB, i.e., the existing PDSCH chain is modified in a way to serve as PUSCH transmitter, i.e., emulating the respective UL functionalities of a device. By doing so, the PUSCH receiver performance can be evaluated for various system parameters and different use cases.

- Building a framework for performance testing and visualisation of either individual blocks or grouped conventional signal processing units and ML modules. The KPIs to be evaluated within the project are link performance (in terms of improved NMSE – normalized mean squared error - and BLER), and the latency of the RX chain.

- Defining and implementing set of data collection points for the training of AI models.

The training of ML modules/models initially starts by using synthetic data (i.e., data artificially generated by simulated channel models and hardware impairments). However, to enable expected KPI improvements in the considered testbed (with real RF chain in an OTA environment), it is of crucial interest to collect real data within the TRX chain. This real data is used to further train the ML models, bringing them to a more realistic operational environment. This data collection points, and corresponding acquisition methods are implemented on the CPU/GPU platform.

Finally, once the the ML models are trained using the collected data sets, they are integrated into the Rx chain. Using the implemented framework, their performance and latency will be evaluated, both in simulation mode and within the end-to-end setup shown in Figure 8.

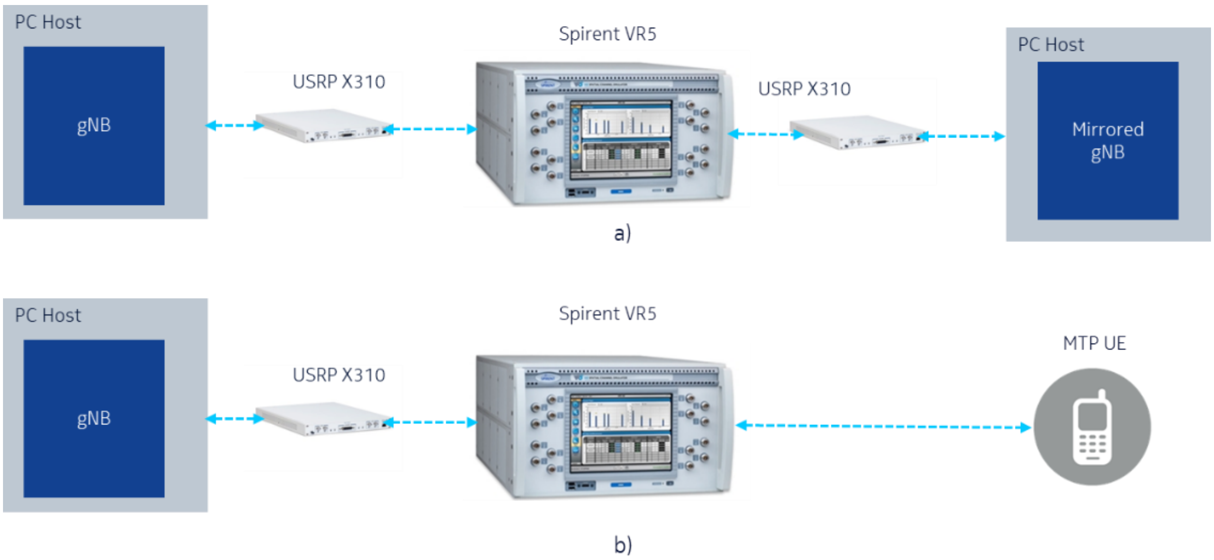

Figure 8. Nokia Bell Labs Stuttgart setup for data generation, testing and demonstration: a) RF loopback mode; b) E2E connection with MTP UE.

Advanced AI/ML Algorithms for RF Detection and Location

Using AI to detect and locate adversarial devices and attacks is a relatively new field in the Cyber Security and Defence domains with several Governments supporting local companies to develop a much needed capability.

This task in AIMM will use state of the art Anomaly Detection algorithms to detect RF signals based on monitoring and analysing RF spectrum in real-time to gain awareness of the surrounding wireless environment in a fashion that was previously prohibitive. The proposed AI technique will detect RF Spectrum anomalies such as jamming, Spectrum Hijacking, as well as malicious cyber threat transmissions from secure facilities. The system will also collect and analyse Spectrum signatures to categories malicious activity for future forensic assessment and Cyber Security based remediation steps that localize, isolate, and mitigate the threat. AI Anomaly Detection Algorithms, such as Auto & Variational Encoders as well as other Probabilistic density models will be Trained and tested against real-world scenarios.

So far work on several steps of the project including modelling, training and inference detection of single & multi-tone interference (jamming sources) as well as Co-Channel and Adjacent channel interference typical in many 4G/5G wireless networks have been completed. This represents one of the more challenging aspects of this type of application, since this form of interference can vary across customer scenarios. In addition, this class of interference is somewhat stochastic in nature and varies across frequency spectrum, amplitude and time.

The Interference Detection testing and performance set-up consists of an ThinkRF Realtime Spectrum analyser (RTSA) and wideband omni-directional antenna, embedded processor (Nvidea Jetson Nano) running software to capture real-time RF spectral data, Feature engineering processes and AI bases anomaly detection algorithms.

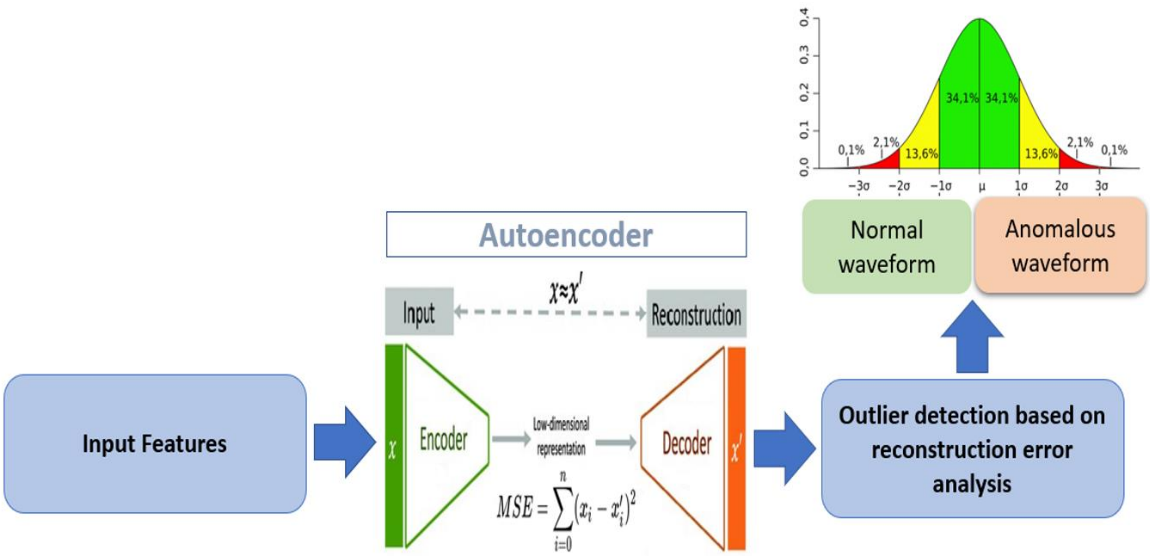

Figure 9. Illustration of AI-based RF anomaly detection in AIMM (ThinkRF).

BT

CEMWORKS

IMST

interdigital

Loughborough University

NOKIA

thinkRF

University of BRISTOL

Universität Stuttgart

Vilicom
signalling the future

Bundesministerium für Wirtschaft und Energie

UKRI Innovate UK

Canada
NRC·CNRC