White Paper

Intel Agilex® FPGAs
Massive MIMO Solutions

Build Efficient and Cost-Effective mMIMO Solutions with Intel Agilex® 7 FPGAs

Leverage Intel's mMIMO ORAN Enablement Package on Intel Agilex® 7 FPGAs to Reduce TTM and CAPEX

Authors

Alberto Alonso
Sr. Wireless Business Developer

Ankur Vora
Wireless Business Developer

Bret Gustafson
Wireless Business Developer

Christian Lanzani
Director, Head of Wireless Business Unit

Kully Dhanoa
System Architect

Navid Lashkarian
Director, System Solutions Engineering

Richard Maiden
Sr. Director,
System Solutions Engineering
Intel Programmable Solutions Group

Table of Contents

Executive Summary

Prior to 1973, mobile telephony was limited to phones installed in cars and other vehicles. Several companies started work on handheld mobile phones, but Motorola was the first company to succeed. On 3 April 1973, a Motorola researcher and executive called Martin Cooper made the first mobile telephone call from a handheld device (albeit one that measured 9 x 5 x 2 inches and weighed in at 4.4 pounds). Standing on Sixth Avenue in New York City, Martin placed a call to Dr. Joel S. Engel, who was his rival at Bell Labs.

To make this call, Motorola had to deploy the equivalent of a cell phone tower to allow Martin to connect into the existing telephone infrastructure. Like all of the first-generation towers that were to follow, this tower was omnidirectional in nature, which means it broadcast its power equally (and wastefully) in all directions.

The capacity of a radio link can be increased by using multiple antennas to transmit data and multiple antennas to receive data. This is known as multiple-input, multiple-output (MIMO). Moreover, it's possible to control the timing and phase difference between the different antennas to implement beamforming, which allows the tower to direct focused beams at individual users. By means of sophisticated signal processing, it's possible for the multiple antennas at the transmitter and receiver to carry multiple data streams simultaneously at the same frequency band without interfering with each other.

Massive MIMO (mMIMO) is an extension of MIMO that involves the use of a much higher number of antennas on the radio tower, thereby allowing the system to support many more users simultaneously. Beamforming by means of mMIMO provides significant benefits with respect to increasing the bandwidth, quality, strength, and efficiency of signals.

Intel Agilex 7 FPGAs and SoC FPGAs have been designed from the ground up to service the extreme capacity and performance demanded by applications like mMIMO. Using Intel Agilex 7 FPGAs and SoC FPGAs, Intel has developed a state-of-the-art mMIMO beamforming solution for use with 5G open radio access networks (O-RANs). This paper introduces MIMO and discusses Intel Agilex 7 FPGA and Intel® Xeon® CPU-based 5G O-RAN offerings, including its mMIMO solution.

What is mMIMO?

The term "mobile communications" refers to the use of technology to communicate information like voice and data without any physical connection, such as wires or cables. This covers a wide range of technologies, including Bluetooth and Wi-Fi. For this paper, however, we will focus on mobile communications using cellular technologies, in which radio towers communicate with user equipment (UE) in the form of smartphones and IoT devices, for example. Depending on the local geography and other factors (e.g., environmental), each radio tower has an associated transmission and reception range allowing it to service a specific area. Collectively, the areas covered by radio towers are known as "cells." This is what led to the use of the term "cellular" in this context, and also to names like "cellular phone" and "cellphone."

The demand for mobile communications is growing exponentially with respect to the number of users and the amount of data each user is consuming and generating. Today's users demand low latency, high fidelity, and high data rates for applications such as online gaming, video streaming, virtual reality (VR), and so on. To meet these growing demands, mobile network operators (MNOs) are transitioning to 5th generation (5G) mobile networks and increasingly relying on the use of high-frequency (HF) radio frequency (RF) bands. MNOs are predominantly using the centimeter bands (100 MHz to 7.12 GHz), which are known as frequency range-1 (FR1), and they are moving into the millimeter band (24 to 100 GHz), which is known as frequency range-2 (FR2) and referred to as mmWave. The 3rd Generation Partnership Project (3GPP)[1] is an umbrella term for a group of standards organizations developing mobile telecommunications protocols. For example, 5G NR (New Radio) is a new radio access technology (RAT) developed by 3GPP for 5G that is designed to be the global standard for the air interface of 5G networks.

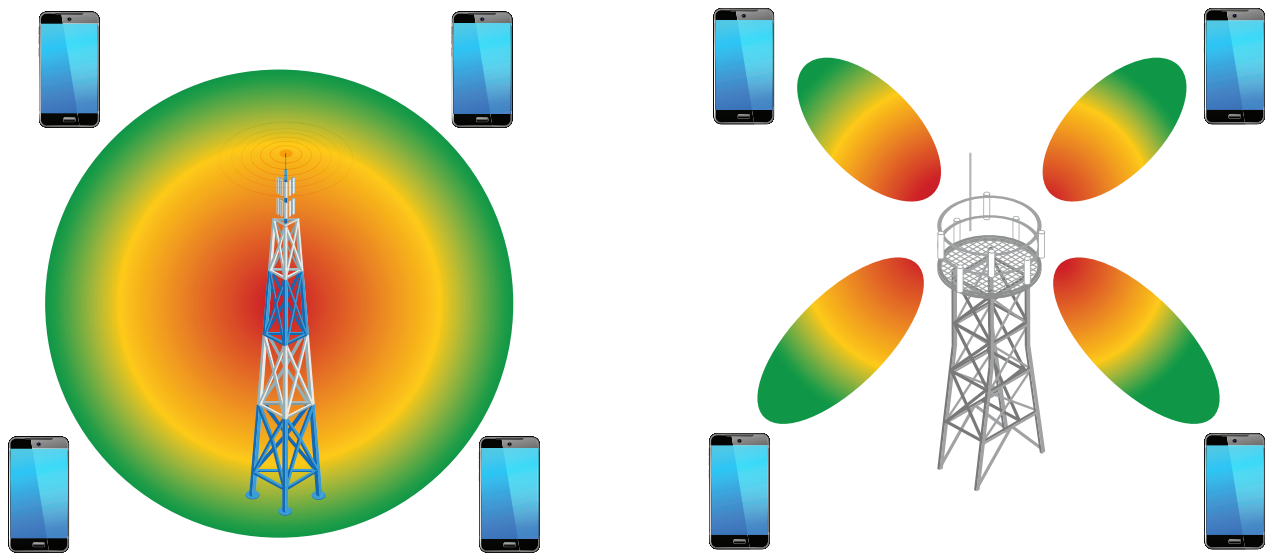
Traditional radio antennas transmit their power omnidirectionally (that is, equally in all directions) with the result that most of the power is wasted in locations where UEs are not present (Figure 1a). This is exacerbated by the fact that power drops off as a square of the distance. Also, assuming four users, for example, each user "sees" the same signals, which necessitates the use of a different frequency for each user or the time division multiplexing (TDM) of a single frequency between multiple users.

In the context of radio, multiple-input and multiple-output (MIMO) refers to a method for increasing the capacity of a radio link by using multiple antennas transmitting data and multiple antennas receiving data. Furthermore, in the case of signals being transmitted from the radio tower, by using channel state information (CSI) received from the UEs, it's possible to control the timing and phase difference of the signals from the antennas in such a way as to achieve constructive interference (boosting the signal) in a desired direction along with destructive interference (rendering the signal undetectable) in other directions. This process, known as beamforming, allows the tower to direct a beam directly at a UE.

Spatial multiplexing in wireless communications is based on MIMO technology, whereby multiple antennas at the transmitter and receiver carry multiple data streams simultaneously within the same frequency band.

There are two major types of MIMO for how users utilize a radio tower: single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO). In the case of SU-MIMO, multiple data streams can be transmitted and received between the cell tower and a single piece of UE. By comparison, in the case of MU-MIMO, multiple data streams can be transmitted and received between the cell tower and multiple pieces of UE using the same frequency bands.

MIMO provides the ability to focus a radio beam on an individual UE while nulling the signal to other UEs. This process is known as zero-forcing beamforming. In the case of MU-MIMO, the radio tower can communicate with multiple users simultaneously (Figure 1b). Beamforming techniques can also be employed for isolating and receiving signals from multiple users.



(a) Omnidirectional transmission

(b) MIMO beamforming

Figure 1. Omnidirectional versus traditional codebook-based MIMO beamforming.

Massive MIMO (mMIMO) is an extension of MIMO that involves using a much higher number of antennas on the radio tower, also known as the base station. This allows the radio to support many more users simultaneously (Figure 2).

mMIMO wireless access technology is currently being deployed in 5G base stations worldwide for use in both the sub-6 GHz and the mmWave bands.

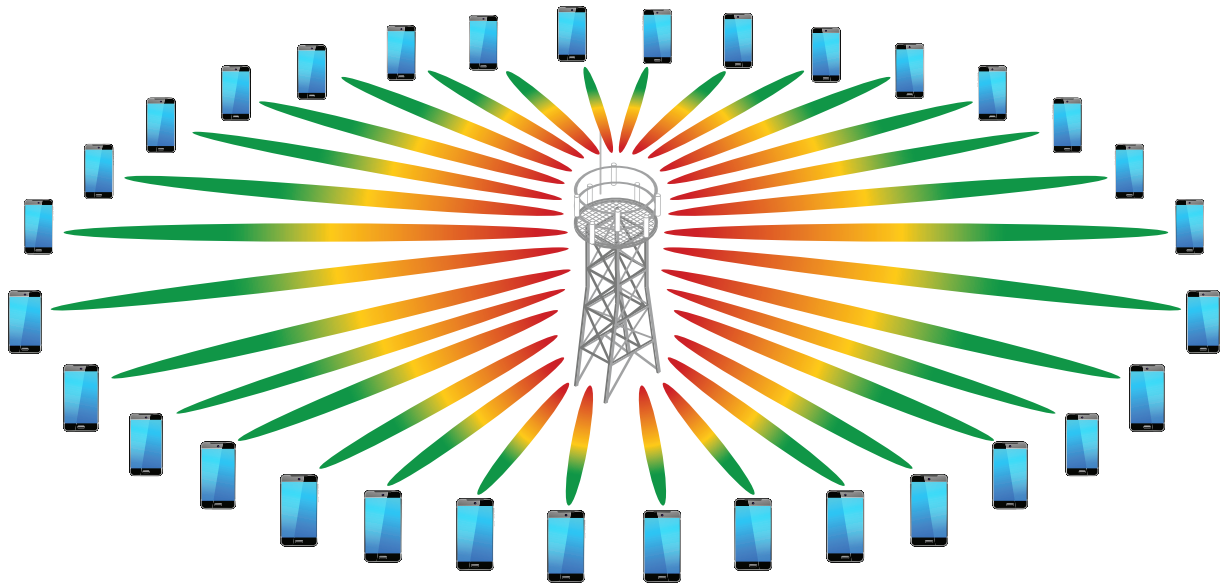


Figure 2. mMIMO allows a base station to communicate with many more users simultaneously.

A useful way to visualize all this is that an omnidirectional antenna is like dropping a boulder into the middle of a pond, where the energy (in the form of ripples) radiates out in all directions equally. If we break that boulder into four smaller rocks and drop these rocks into the pond, then the size of each rock will determine the amplitude of its associated ripples. Meanwhile, the times the rocks are dropped will determine the phase relationships between their ripples. By controlling the size of the rocks and the times they are dropped, we can get the four sets of ripples to combine coherently in certain places (where the UEs are) and cancel out in other places (where the UEs aren't). Alternatively, suppose we break our original boulder into 64 pebbles and control the size of the pebbles and the time each pebble is dropped. In this case, we can get the resulting ripples to combine and subtract in such a way as to create more "beams" (target more UEs).

The Benefits of mMIMO

Beamforming provides significant benefits with respect to increasing the bandwidth, quality, strength, and efficiency of signals. Beamforming addresses a wide variety of communication challenges, such as improving the accuracy of the wireless links between the base station and UEs, increasing throughput, increasing the number of connections (i.e., the cell capacity) that can be supported within a given cell area, and saving energy consumption during transmission. Beamforming is particularly beneficial for improving the signal-to-noise ratio (SNR) of a link by directly targeting the UEs as illustrated in Figure 2.

MNOs spend billions of dollars in frequency spectrum acquisitions. For example, the auction by the US Federal Communication Commission (FCC) of 280 megahertz of C-band has raised nearly $81 billion.[2] Thus, operators want to use their portions of the frequency spectrum as efficiently as they can while still offering the best user experience possible. mMIMO provides higher spectral efficiency by allowing its antenna array to focus narrower beams on users.

With a higher SNR, the operator can attain a more efficient modulation and coding scheme. Also, an operator can exploit time (through spatial separation) and frequency to reuse the spectrum. These result in higher spectral efficiency, which means the operator can pull more revenue from the same spectrum allocation. MIMO, especially mMIMO with appropriate beamforming algorithms, allows operators to communicate with multiple users simultaneously using the same frequency. It manages to minimize interference by using the spatial diversity of individual users. Furthermore, the combination of antenna array gain, narrower beams, and more precise spatial multiplexing results in higher throughputs leading to better user experiences.

Mobile networks have various use cases, ranging from sparse rural environments to dense urban settings. In dense urban cases, operators cannot simply keep adding base stations to cater to the increasingly high throughputs required. The network would become interference limited as different base stations would start to interfere with each other if they use the same frequency simultaneously. There are some coordinated multi-point (CoMP) techniques, but these reduce spectral efficiency. mMIMO allows for higher throughput by providing more beams that do not interfere.

By comparison, rural settings tend not to have interference or throughput challenges. Instead, coverage or cell footprint (cell size) is key. Having said this, mMIMO also has advantages in these settings. If we take the omni antenna in Figure 1a, for example, all of the RF energy is transmitted in every direction equally, regardless of where the users are situated. With mMIMO's narrower beams, the RF energy is focused only on the users. In turn, this means a mMIMO system can have a longer reach resulting in a larger footprint that covers more users while requiring fewer cell sites (reducing the number of fronthaul optical fibers and saving fronthaul costs), thereby reducing capital expenses (CapEx) and operational expenses (OpEx).

Narrow Beams and Mobility

While narrow beams have several advantages as outlined above, narrow beams, compounded with mobility, introduce an additional layer of complexity. If a user is moving, there is a risk that the beam is sent to where the user was rather than where the user is now. The narrower the beam, the more troublesome this can be. New beam-forming weights need to be recalculated quickly enough to track the user. Depending on the amount of mobility supported (it can be more than 100 mph in the case of high-speed trains, for example), the channel estimation and new weights need to be recalculated with low latency. For a mMIMO system, this typically requires tens or hundreds of billions of complex multiply and accumulate (MAC) operations per second.

Different Beamforming Approaches

As illustrated in Figure 3, three different hardware approaches may be used to implement beamforming: analog, digital, and hybrid (this paper is focused on digital beamforming).

Analog Beamforming

Analog beamforming—which should more properly be thought of as analog beam steering—commences with generating a single signal. This signal is fed into multiple phase shifters, each of which employs coefficients to modify the phase and amplitude of the signal. The output from each phase shifter is fed to its antenna. Collectively, all the antennas focus their energy in a particular direction, resulting in a single directed beam. By varying the coefficients, the direction of the beam can change over time.

Digital Beamforming

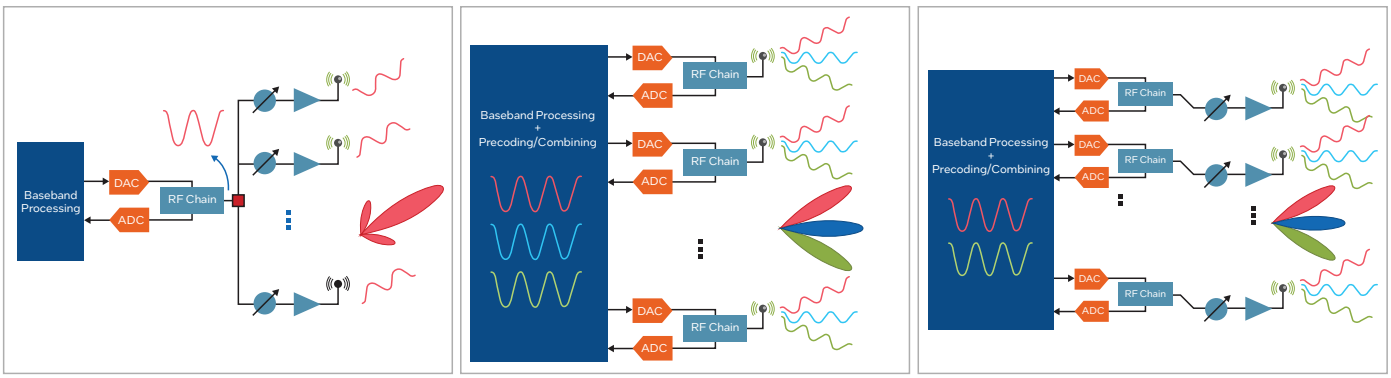
In the case of digital beamforming, each antenna element will have its own dedicated RF chain, digital-to-analog converter (DAC), and analog-to-digital converter (ADC). Each spatial sample's gain and phase are adjusted individually along with baseband processing before up-conversion at the transmitter or after down-conversion at the receiver. This enables the underlying mathematical algorithms to be implemented with maximum flexibility.

Digital beamforming and its dynamic finetuning capabilities of multiple weights allow for the antennas to be employed as multiple sub-arrays producing sets of uniquely distributed beams. These multiple beams can be emitted simultaneously, maximizing the signal strengths in desired directions. As a result, nulls are created in undesired directions, thereby minimizing and suppressing interference. By leveraging spatial multiplexing, multiple spatial streams can be simultaneously created and transmitted. More complicated precoders can be implemented to generate multiple beams and enable multiuser communications. A digital architecture allows for catering to large bandwidths by selecting the weights for a frequency-selective scenario. This is one of the reasons digital beamforming is widely popular for 5G NR radios.

The fundamental advantage of digital beamforming is that it allows multiple different directed beams to be generated at the same time.

Hybrid Beamforming

As its name suggests, hybrid beamforming employs a combination of analog and digital techniques. While powerful, hybrid beamforming leads to more complexity and less flexibility, thereby making it unsuitable with respect to adopting and accommodating rapidly changing wireless standards.



1) Analog Beamforming 2) Digital Beamforming 3) Hybrid Beamforming

Figure 3. Different hardware approaches for beamforming.

O-RAN Supported Beamforming Techniques

The Open RAN Alliance or O-RAN Alliance[3] is a worldwide community of mobile network operators, vendors, and research and academic institutions operating in the RAN industry that is dedicated to evolving radio access networks (RANs) into being smart and open. This goal moves previously proprietary RANs towards an ecosystem of innovative, multi-vendor, interoperable autonomy, thereby reducing costs while increasing performance and agility.

As connectivity and connected devices are growing, the demand for a flexible RAN architecture also arises. An Open RAN (O-RAN) is a non-proprietary implementation of a RAN that allows interoperability between cellular network equipment provided by different vendors. As the O-RAN Alliance defines the typical 5G RAN, the base station is split into three logical nodes: the radio unit (RU), the distributed unit (DU), and the central unit (CU).

Intel provides silicon technologies that address every aspect of the 5G O-RAN architecture, including O-RAN RUs (O-RUs), O-RAN DUs (O-DUs), and O-RAN CUs (O-CUs). These technologies include Intel Agilex FPGAs, Intel Xeon CPUs, and Ethernet solutions (network adapters, cards, controllers, and accessories). Intel augments and enhances these hardware technologies with software counterparts, including the FlexRAN™ software stack and the Virtual RAN (vRAN) Enablement Package (Figure 4).

Of particular interest in the context of this paper is the fact that Intel Agilex 7 FPGAs offer the ability to execute 5G RAN algorithms in a massively parallel fashion, thereby providing extreme performance while consuming relatively low power as compared to a CPU implementation of the same algorithms. Also, the programmable fabric in Intel Agilex 7 FPGAs allows developers to respond to rapidly changing standards and evolving protocols, including updates after systems have been deployed into the field.

The 5G NR O-RAN architecture addresses the fact that different network configurations may better serve different use cases. The phrase “3GPP Functional Splits” refers to the fact that 3GPP specifies a variety of options for distributing (“splitting”) the functionality of the 5G NR RAN stack across the fronthaul network. Split 7.2 is a commonly used functional split in which the Lower PHY functions remain in the radio and the Higher PHY and above are split between the O-DU and O-CU.

In addition, O-RAN has selected a specific point of separation known as "7-2x" but provides some flexibility in the placement of the precoding function above or below the fronthaul interface. This decision can have a huge impact on fronthaul traffic. O-RUs that perform precoding are classified as "Category B" (in which case it is essential to provide the O-RU with the necessary information to perform the precoding operation), while those that do not are labeled as "Category A."

Various methods have been proposed by the O-RAN Alliance to implement beamforming solutions: predefined beamforming, weight-based dynamic beamforming, attribute-based beamforming, and channel information-based beamforming. All of these methods can be used with each of the beamforming techniques introduced above.

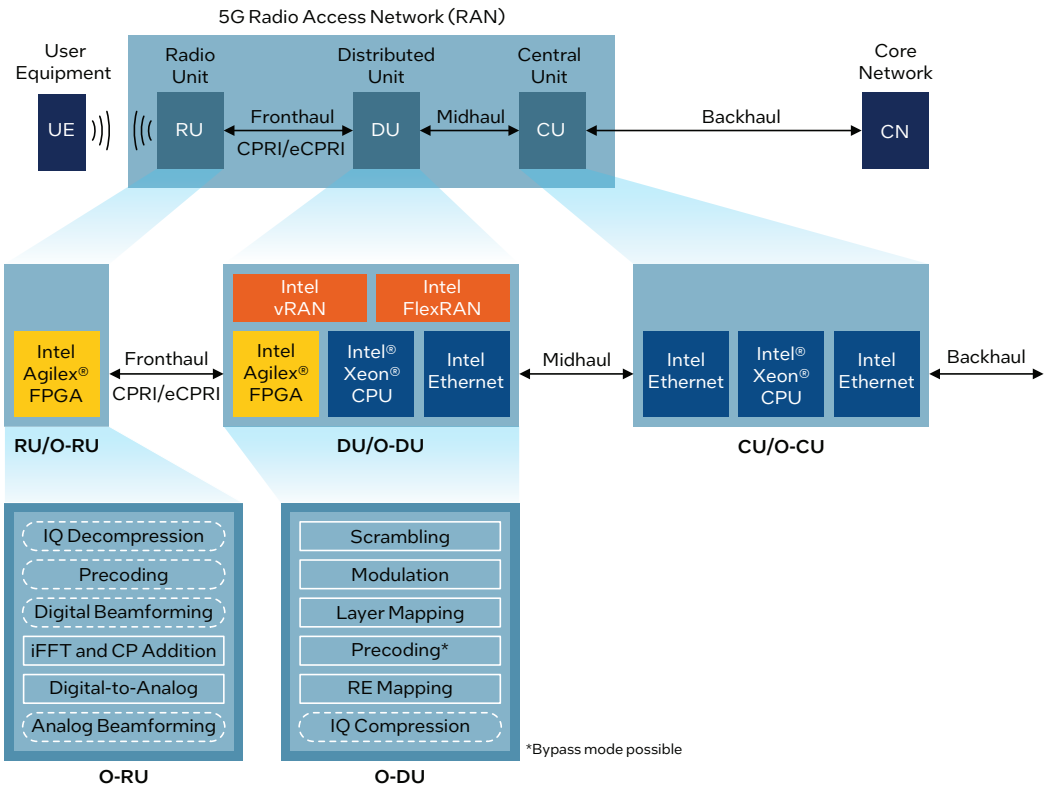


Figure 4. Intel offers end-to-end silicon solutions for 5G O-RANs.

Predefined Beamforming

In the case of predefined beamforming, the O-DU instructs the O-RU as to which weights to use as a function of the BeamID. The O-RU already has the weights stored in a table indexed by the BeamID.

This table contains beam weights, which can be frequency or time domain based. The BeamID is used as a pointer to a vector of predefined beamforming weights. Thus, the BeamID is exchanged between the O-DU and the O-RU to indicate which beam weight should be applied.

Attribute-Based Beamforming

In this case, the O-RU is responsible for generating beamforming weights using a set of beam properties signaled by the O-DU. These properties can reflect different parameters such as angle-of-arrival, angle-of-departure, azimuth and elevation, beamwidth, and sidelobe suppression.

Channel Information-Based Beamforming

With respect the channel information-based beamforming, each antenna has a dedicated RF signal and path and provides many beams that can be dynamically switched. With this method, the O-RU generates beam weights based on the channel information supplied by the O-DU.

Weight-Based Dynamic Beamforming

With this method, the O-DU sends real-time generated beamforming weights to the O-RU, which associates them with user-specific data. Beam indices can be assigned and used in subsequent messages. In addition, the beamforming gets updated in real-time using the BeamID value associated with the weights coming from the O-DU.

All mMIMO beamforming techniques involve massive amounts of computation. The challenges are to achieve this computation within the required cost, bandwidth, latency, and power consumption envelopes. This whitepaper focuses on a digital weight-based dynamic beamforming implementation using Intel Agilex 7 FPGAs. The implementation of other beamforming techniques will be discussed in future whitepapers.

Intel's mMIMO Solution for Digital Beamforming

Intel has developed a Split 7.2B (ORAN Category B O-RU) mMIMO optimized workload solution supporting 32 and 64 antennas with up to two component carriers (CCs) using 100 MHz of bandwidth. This solution implements O-RAN-compliant, weight-based dynamic beamforming on a single Intel Agilex 7 SoC FPGA. The implementation is extremely resource-efficient and it supports a full datapath, including 2x25GbE interfaces with IEEE 1588v2 PTP on F-tile, eCPRI/O-RAN IP, along with an extremely compact and optimized beamforming engine with centralized layer-1 processing connecting to external digital front ends (DFEs) using JESD204 protocol.

The Intel Agilex 7 FPGAs and SoC FPGAs enable the extreme capacity and performance requirements demanded by applications like mMIMO.

In the downlink (DL) path, the O-RU receives frequency domain layer data from the O-DU and sends this data into the beamforming precoder, which converts this into frequency domain antenna data, thereby creating multiple beams. This is then converted into time-domain antenna data by the iFFT, and cyclic prefix data is added before being interleaved and sent to the RF chain.

In the uplink (UL) path, the O-RU receives time domain down converted antenna data from the chain and sends this to the PRACH branch and de-interleaver. The de-interleaved output has the cyclic prefix removed before being fed to the fast Fourier transform (FFT) and Shock Response Spectrum (SRS) modules.

The FFT converts the data into frequency domain antenna data and this output is fed into the beamformer combiner which combines the antenna data to create layer data. Lastly, this is sent to the O-DU.

The SRS signals are stored in external DDR and then read and sent to O-DU over time. SRS is a non-delay-managed signal, so there is a greater time budget for transferring these signals to the O-DU.

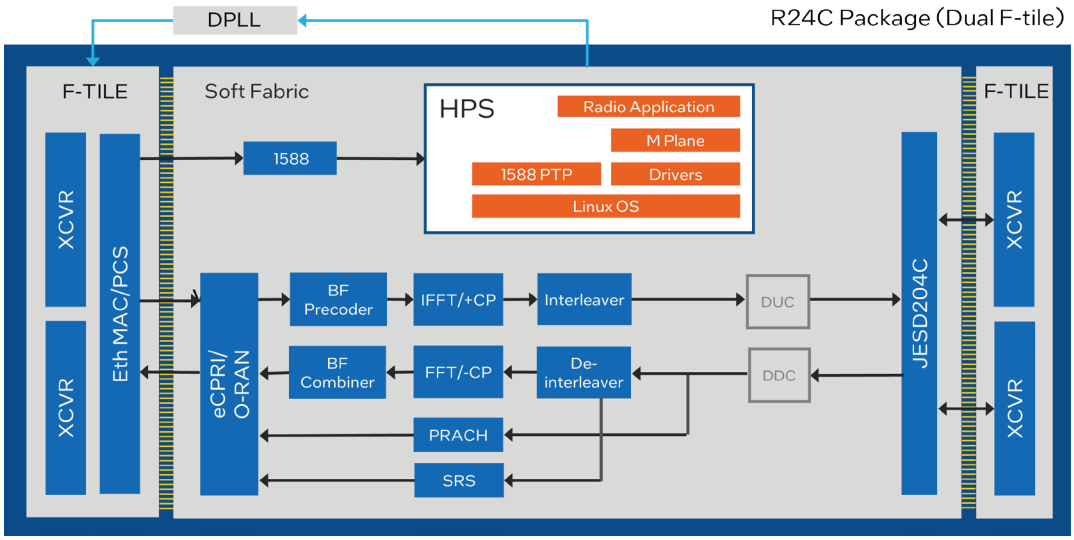


Figure 5. Intel Split 7.2B Beamforming Solution

Intel Solution Specification

The following table summarizes the overall specification for this solution.

Requirement	Specification	Description
Antennas	16, 32 and 64	
Component carriers	Up to two	
Layers	Up to 16 layers in downlink Up to 16 streams in uplink	
Resources blocks	1-273	Programmable supporting 100 MHz carrier bandwidth
Subcarrier spacing	30 kHz	
Beamforming method	Weight-based dynamic beamforming	Can support predefined beamforming with minor design changes
O-RAN compliance	U-plane compression, BFP compression, fixed-timing delay/advance, Weight-based dynamic beamforming	BFP compression with 8, 9, 10, 11, 12, 13, 14, 15, and16 bits. Fixed-timing meets O-RAN version IoT profile. Ethernet transport encapsulation, eCPRI as transport method, jumbo frames, application layer fragmentation
UL/DL scheduling	Supports flexible time division duplex (TDD) slot format	Independent engines for DL beamforming and UL combing.
Fronthaul interface	2x25 GbE	eCPRI / ORAN with advanced-PTP 1588 v2

Intel Solution Advantages

One of the challenges of mMIMO is the huge amount of processing required. The beamformer for the solution described requires a staggering 96 Giga Complex MAC/second in the beamformer precoder alone in the downlink, plus the same amount for the beamformer combiner in the uplink. Intel has managed to implement this processing (along with the other RU processing shown in the previous illustration on an Intel Agilex FPGA with speed grade 3. Note that speed grade 1 is the fastest FPGA and hence the most expensive.

Another key highlight of this solution is the ability of the beamformer (precoder and combiner) to be able to process an entire OFDMA symbol in less than 27us. This latency is hidden in the overall system latency by the iFFT/FFT processing time: the beamformer precoder generates data in the order required by the iFFTs; the beamformer combiner processes data in the order generated by the FFTs. In addition to minimizing the system latency, this reduces the storage required between these modules.

The beamformer operates at 491.52 MHz with plenty of timing margin, which ensures the DSP Block usage is efficient due to the time sharing of multipliers. Power is optimized by ensuring that dynamically fewer layers in the beamformer are being used, power consumption is reduced by the toggling of signals on unwanted layers.

The beamformer portion of the solution is designed in the MATLAB Simulink graphical user interface (GUI) using Intel's DSP Builder for Intel® FPGAs advanced library. This approach has the following advantages:

- DSPBA sources delivery allows you to quickly and easily modify the design to meet your exact requirements
- Pushbutton synthesis for different devices/speed grades with the tool adding the required pipeline stages to meet timing
- Automatic RTL testbench generation using test vectors from Simulink simulation to confirm the RTL matches the DSP Builder for Intel FPGAs model

Digital Beamforming Validation Platform Powered by Intel

Intel, along with its partner ADI, is developing a mMIMO O-RAN radio whitebox that complies with O-RAN and 3GPP standards (Figure 6). This whitebox will serve as a comprehensive development platform for mMIMO applications and be the hardware component of Intel's mMIMO enablement package.

Powered by an Intel Agilex 7 FPGA Series with two or four F-Tiles, the whitebox comes equipped with all the necessary interfaces, ports, converters, and RF front-end (RFFE) to serve as a complete development platform for mMIMO radio applications.

As an open platform, the whitebox can be used in conjunction with Intel's O-RAN IP library, which includes mMIMO beamforming IP, or combined with other third-party IP to develop end-to-end mMIMO O-RAN radio products.

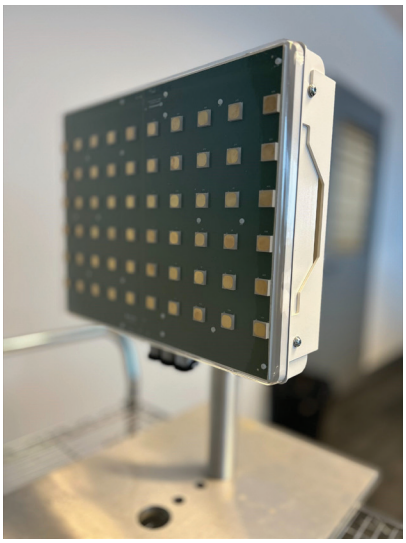


Figure 6. mMIMO O-RAN radio whitebox

The mMIMO whitebox can be used for various purposes, such as proof-of-concept, lab validation, indoor field testing, limited field trials, or as the foundation for the development of production hardware.

The mMIMO whitebox supports all four O-RAN synchronization topologies, including IEEE1588v2 and SyncE, a PTP software stack, and servo to comply with O-RAN enhanced O-RU specifications as well as full and partial timing support for telecom profiles.

Intel's End-to-End O-RAN Silicon and IP Solutions

As illustrated in Figures 4 and 7, Intel provides end-to-end (E2E) silicon technologies that address every aspect of the 5G O-RAN architecture, including O-Rus, O-Dus, and O-Cus. These technologies include Intel Agilex FPGAs, Intel Xeon CPUs, and a wide variety of Ethernet solutions (network adapters, cards, controllers, and accessories). Intel augments and enhances these hardware technologies with software counterparts, including a library of intellectual property (IP) blocks, the FlexRAN software stack, and the Virtual RAN (vRAN) Enablement Package.

This open architecture offers advantages to multiple vendors when developing specific implementations of O-Cus, O-DU, and O-Cus. Intel® FPGAs are used for accelerating O-RU, fronthaul gateway (FHGW), and O-DU applications, while Intel Xeon CPUs and Intel Ethernet solutions are used to power O-DU and O-CU functionality.

Intel Agilex 7 FPGAs for O-RUs and O-DUs

Intel Agilex 7 FPGAs and SoC FPGAs are designed from the ground up to service the extreme capacity and performance requirements demanded by applications like O-RANs. Intel Agilex 7 FPGA F-Series and I-Series are built using Intel 10 nm SuperFin process technology, which is Intel's third generation FinFET technology.

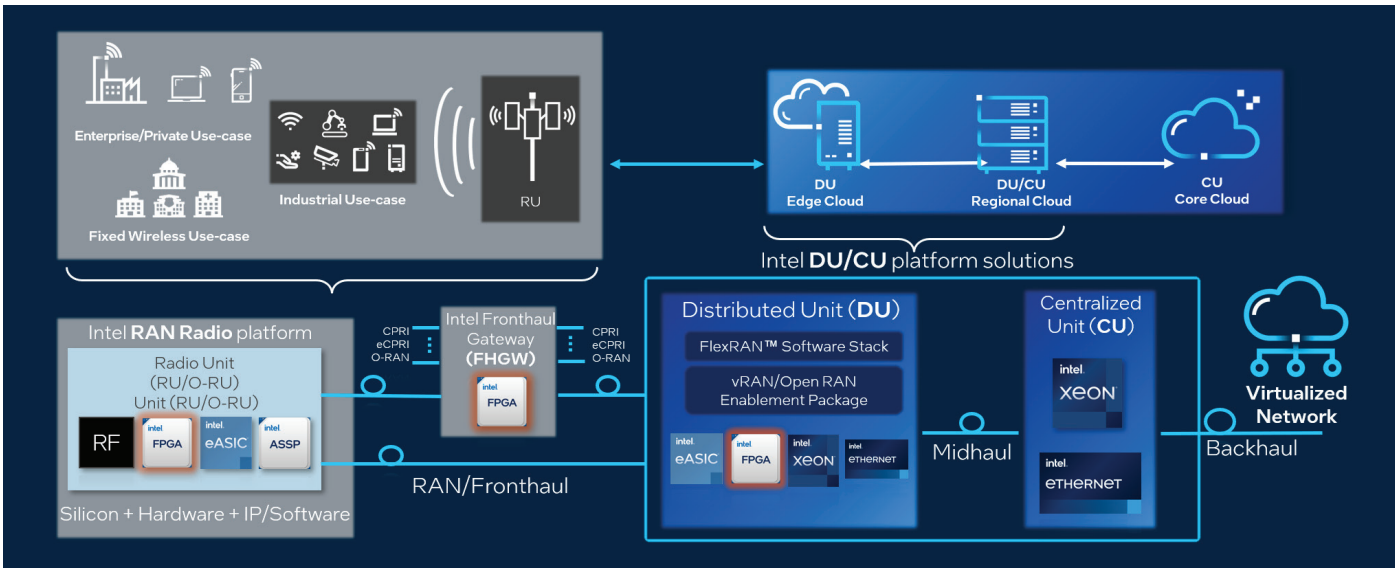


Figure 7. Intel's E2E O-RAN silicon solutions

These devices also use the second-generation Intel® Hyperflex™ FPGA fabric, which provides on average 50% higher performance or up to 40% lower power consumption for data center, networking, and edge computing applications compared to previous-generation Intel® Stratix® 10 FPGAs. Compared to our competitor's 7 nm FPGA portfolio, F-Series and I-Series FPGAs deliver approximately 2X fabric performance per watt, and are 15% to 20% faster on average, than competing 7 nm FPGAs for 5G O-RU functions4. Intel Agilex 7 SoC FPGAs also integrate a 64-bit quad-core Arm Cortex-A53 processor—also known as a hard processor system (HPS)—for a high degree of system integration and dynamic configuration.

Intel® Quartus® Prime Software is Intel's suite of FPGA design software. The Intel Quartus Prime Software supports the analysis and synthesis of hardware description language (HDL) designs, allowing designers to compile their designs, perform timing analysis, examine register transfer level (RTL) language diagrams, simulate design responses to various stimuli, and configure the target device. Intel Quartus Prime Software supports Verilog and VHDL, visual editing of logic circuits, and vector waveform simulation.

5G RANs involve a huge amount of digital signal processing (DSP). The DSP Builder for Intel FPGAs is a DSP design tool that generates DSP algorithms for HDL directly from the MathWorks Simulink environment. This tool allows developers to describe designs at a high level of abstraction and then generate efficient structures from MATLAB functions and Simulink models as high-quality, automatically pipelined, synthesizable VHDL/Verilog code. This RTL layer allows rapid exploration of the design space and faster time to market, and it can be combined with other RTL and intellectual property (IP) blocks before being imported into the Intel Quartus Prime Software.

Intel Xeon CPUs O-DUs and O-CUs

Intel offers a fully virtualized RAN infrastructure for both 4G and 5G as a software-based RAN stack named FlexRAN running on an Intel Xeon processor for the high-PHY layer. This provides operators with a platform to realize numerous operational and business benefits, including total cost of ownership reduction, cost, and power with continuous innovation in RAN.

4th Generation Intel Xeon Scalable processors with Intel® vRAN Boost have integrated accelerators. The Intel FlexRAN L2+ reference library uses data plane development kit (DPDK) memory pools and mbuf as the basis for buffer management, where HugePages also applies. This DPDK also offers the zero-copy-based 5GNR MAC/RLC packet processing application for O-DU. Intel C++ Compiler (ICC) is a group of C and C++ compilers that allows compilation optimization for core or targeting to a specific function(s). Intel Vtune™ amplifier is a performance analysis tool for Intel Xeon processor targeting specific cores or functionality. It optimizes performance while avoiding power and thermal-related throttling.

The O-RU implementation illustrated in Figure 4 can be connected to an Intel FlexRAN-based O-DU server by means of an O-RAN-compliant fronthaul interface using CPRI or eCPRI links in accordance with 3GPP Split 7.2X.

The O-DU and O-CU are both powered by the Intel Xeon family, which is a scalable server processor that complements the Intel Ethernet solution and vRAN enablement package, including FlexRAN, Intel C++ Compiler (ICC), Intel vTune Amplifier, and DPDK.

The O-DU and O-CU is powered by a multi-core processor loaded with native instructions to handle certain L1 and L2 functions, including efficient vector processing. The Intel Xeon processor family includes high-performance cores and instructions, such as the Intel AVX-512 and the upcoming enhanced AVX512-FP16 signal processing instructions, available on 4th Generation Intel Xeon Scalable processors, which are ideal for software-based high layer-1 functionality. It also offers true inline acceleration with forward error correction (FEC) and discrete Fourier transform (DFT) performed in a fixed-function accelerator block.

Intel IP Solutions that Support O-RAN Implementations

Beamforming is one piece of an mMIMO O-RAN solution. Key IP components of an O-RU digital front end (DFE) include fronthaul interface processing (CPRI, eCPRI, O-RAN), low layer 1 (LL1) These simplified, production-ready features include key processing IP elements and software frameworks bundled as an integrated radio solution. The solution is a resource-efficient implementation of the O-RAN standard optimized for Intel Agilex 7 FPGAs and SoC FPGAs. It allows easy reconfiguration of functionality for different applications while maintaining an ultra-small footprint.

The IP elements shown here include finite impulse response (FIR), fast Fourier transform (FFT), inverse FFT (iFFT), cyclic prefix (CP) add/remove (+/-), radio timing, physical random-access channel (PRACH), layer mapping, IQ compression/decompression, user plane framer/deframer, control plane multiplexer/demultiplexer, and TDD switching. The software environment runs on the Intel Agilex 7 FPGA hard processor system. This 5G O-RAN solution integrates JESD204B/C, IEEE1588, Ethernet solutions, and other modules to form a complete radio sub-system. The solution is also scalable to support different antenna configurations, bandwidths, RATs, and carrier configurations. Intel Agilex 7 devices are supported by a wide range of hardware and software IP blocks and functions. In the case of 5G RANs, some key IP products are as follows:

- Ethernet: Ethernet is a family of wired computer networking technologies commonly used in Local Area Networks (LANs), Metropolitan Areas (MANs), and Wide Area Networks (WANs). Intel Agilex 7 devices use heterogeneous 3D SiP technology. In addition to the main Intel Agilex 7 FPGA die, these SiPs also contain many smaller chips, also known as chiplets or tiles. The Intel Agilex 7 FPGA F-Tile is an evolution of E-Tile, providing hardened Ethernet IP, including a split, configurable, hardened Ethernet protocol stack that supports rates from 10G to 400G and that is compliant with the IEEE 802.3 specification and other related specification from the Ethernet Alliance. There are several variants of IP cores that provide different combinations of Ethernet channels and functions. These include optional Reed-Solomon Forward Error Correction (RSFEC) and optional IEEE 1588v2 Precision Time Protocol (PTP). Users can choose between Media Access Control (MAC) variants and Physical Coding Sublayer (PCS) variants, PCS only variants, FlexE variants (FlexE) or Optical Transport Network (OTN) variants.

- JESD204C: JESD204C is a Joint Electronic Devices Engineering Council (JEDEC) standard. It is a high-speed interface designed to connect high-speed analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) to high-speed processors, FPGAs, and ASICs. JESD204C Intel® FPGA IP address synchronization of multiple devices using subclass 1 for deterministic latency. It also supports TX-only, RX-only, and duplex (TX and RX) modes.
- eCPRI: Common Public Radio Interface (CPRI) is a legacy interface used to enable fronthaul connectivity between RRH and BBU in 4G RAN. Enhanced or Evolved CPRI (eCPRI) provides a way to divide baseband functions to reduce system traffic load and enable more flexible and efficient communication between O-RU and O-DU in 5G RAN. Intel eCPRI IP implements version 2.0 of the eCPRI specification. It can support 10G and 25G Ethernet ports while supporting one-way latency measurements similar to IEEE Standard 1588 PTP hardware-based timestamping
- O-RAN Fronthaul Interface: The O-RAN WG4 fronthaul interface defines the interface between the O-RAN radio unit (O-RU) and the shared O-RAN distributed unit lower layer (O-RAN). Interface). The current Intel O-RAN Fronthaul Interface IP supports compression, decompression, and both CAT-A and CAT-B radios.

Conclusion

Intel understands the challenges associated with implementing 5G O-RANs and has the expertise to address these challenges. Intel provides end-to-end (E2E) silicon technologies that address every aspect of the 5G O-RAN architecture, including O-RUs, O-DUs, O-CUs. These technologies include Intel Agilex 7 FPGAs, Intel Xeon CPUs, and a wide variety of Ethernet solutions (network adapters, cards, controllers, and accessories).

With respect to mMIMO, Intel has a compelling solution that includes silicon (Intel Agilex 7 SoC FPGAs deliver approximately 2X fabric performance per watt, and are 15% to 20% faster on average, than competing 7 nm FPGAs for 5G O-RU functions[4]), a workload enablement package, and a hardware whitebox that is flexible, programmable, consumes less power, and is more affordable than competing solutions.

The Intel Agilex 7 SoC FPGA meets the extreme capacity and performance requirements demanded by applications like mMIMO. Intel has implemented a weight-based dynamic beamforming solution on a single Intel Agilex 7 SoC FPGA that is O-RAN compliant and supports 32 and 64 antennas for up to two component carriers using 100 MHz of bandwidth. This beamforming solution is designed in MATLAB Simulink using Intel's DSP Builder for Intel FPGAs Advanced Blockset, which offers designers the flexibility and logic capabilities to modify the design to meet their exact requirements and feature sets.

One of the key challenges of mMIMO is the huge amount of processing required. Intel has managed to implement this processing on an Intel Agilex FPGA with speed grade 3, which is the most cost-effective SKU. The beamformer precoder and combiner can process an entire OFDMA symbol in less than 27 us, this minimizes system latency and reduces the storage required between modules. Power consumption is also optimized by reducing the number of layers in use.

Intel's solution offers several advantages over competing solutions. The beamformer operates at 491.52 MHz with plenty of timing margin, ensuring efficient DSP Block usage. The solution supports flexible TDD slot format and has independent engines for DL beamforming and UL combing. The 2x25 GbE fronthaul interface supports eCPRI with advanced-PTP 1588 v2, which is O-RAN compliant and supports U-plane compression, BFW compression, fixed-timing delay/advance, weight-based dynamic beamforming, and more.

If you wish to build your own efficient and cost-effective mMIMO solutions with Intel Agilex 7 FPGAs, reach out to your local Intel representative today to learn more.

Learn More

- Intel RAN FPGA Solutions
- Intel Agilex 7 FPGAs

References

1 https://www.3gpp.org/

2 https://spacenews.com/c-band-raises-81-billion/

3 https://www.o-ran.org/

4 https://edc.intel.com/content/www/us/en/products/performance/benchmarks/agilex-fpga/