A flexible 100-antenna testbed for Massive MIMO

Vieira, Joao; Malkowsky, Steffen; Nieman, Karl; Miers, Zachary; Kundargi, Nikhil; Liu, Liang; Wong, Ian; Öwall, Viktor; Edfors, Ove; Tufvesson, Fredrik

DOI:
10.1109/GLOCOMW.2014.7063446

2014

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
A flexible 100-antenna testbed for Massive MIMO

Joao Vieira¹, Steffen Malkowsky¹, Karl Nieman², Zachary Miers¹, Nikhil Kundargi², Liang Liu¹, Ian Wong², Viktor Öwall¹, Ove Edfors¹, and Fredrik Tufvesson¹

¹ Dept. of Electrical and Information Technology, Lund University, Sweden
² National Instruments, Austin, Texas, USA

Abstract—Massive multiple-input multiple-output (MIMO) is one of the main candidates to be included in the fifth generation (5G) cellular systems. For further system development it is desirable to have real-time testbeds showing possibilities and limitations of the technology. In this paper we describe the Lund University Massive MIMO testbed – LuMaMi. It is a flexible testbed where the base station operates with up to 100 coherent radio-frequency transceiver chains based on software radio technology. Orthogonal Frequency Division Multiplex (OFDM) based signaling is used for each of the 10 simultaneous users served in the 20 MHz bandwidth. Real time MIMO precoding and decoding is distributed across 50 Xilinx Kintex-7 FPGAs with PCI-Express interconnects. The unique features of this system are: (i) high throughput processing of 384 Gbps of real time baseband data in both the transmit and receive directions, (ii) low-latency architecture with channel estimate to precoder turnaround of less than 500 micro seconds, and (iii) a flexible extension up to 128 antennas. We detail the design goals of the testbed, discuss the signaling and system architecture, and show initial measured results for a uplink Massive MIMO over-the-air transmission from four single-antenna UEs to 100 BS antennas.

Index Terms—Massive MIMO, testbed, system description, prototype, 5G, large array.

I. INTRODUCTION

MASSIVE MIMO is a promising technology and a strong candidate for future-generation wireless systems. Compared to conventional MIMO, potential benefits brought by the extra degrees-of-freedom due the excess number of BS antennas include [1] [2]: (i) both system capacity and radiated energy efficiency can be improved by several orders of magnitude; (ii) hardware requirements on the base station (BS) radio frequency (RF) chains can be greatly relaxed; (iii) simplification of the multiple-access layer; all of this with (iv) reduced complexity at the user equipment (UE). To take the next steps in the development and verification of the potential, it is necessary to have proof-of-concept platforms, i.e. testbeds, where Massive MIMO can operate under real-life conditions (e.g., with analog front-end impairments and real wave propagation conditions) to assist further algorithm development and circuit design. Testbeds can improve the overall understanding of, so far conceivable, issues and help maturing the technology for standardization.

Table I lists existing many-antenna testbeds as of today. The first system is a channel sounding system used at Lund University to measure the wireless channel with a large number of antennas to validate theoretical gains [3]. 50 MHz channel measurements were taken over slow continuous user movements and then processed offline. The results confirm favorable propagation for measured channels with low eigenvalue spread. Second, Rice University [5] constructed a testbed and evaluated practical performance gains of Massive MIMO in indoor environments. Channel measurements were collected over a 0.625 MHz bandwidth for both LOS and NLOS conditions, and promising capacity results based on SINR computations were presented. Third, researchers at Samsung [6] recently made their work in many-antenna MIMO systems public. This testbed is targeted at millimeter wave bands but can be applied to cellular band applications. The press release is not very detailed, though it is mentioned that a throughput of 1 Gbps is achieved at 2 km range.

Despite prior work in large scale MIMO systems, many shortcomings are evident. Existing testbeds are either proprietary, non-real-time, or both. These limitations hinder researchers from developing algorithms tied to real wireless channels. To address this, we have developed an extensible platform, the LuMaMi testbed, to realize up to 20 MHz bandwidth 100-antenna MIMO. It is built up of commercial off-the-shelf hardware, making it accessible and modifiable. The main objectives for this testbed are:

- implementing BS architectures to meet high-throughput/low-latency processing requirements;
- evaluating practical performance of different baseband processing algorithms;
- implementing time and frequency synchronization solutions between BS RF chains;
- identifying scenarios where favorable propagation conditions for Massive MIMO exist (or do not exist);
- demonstrating a Massive MIMO proof-of-concept by concurrent high-speed data streaming to and from multiple users, via high-density spatial multiplexing within the same time-frequency resource. The link quality can

### Table I

<table>
<thead>
<tr>
<th>Institution</th>
<th>Band (GHz)</th>
<th>Hardware</th>
<th># of BS antennas</th>
<th># of users</th>
<th>(&lt; 1 ms) turnaround?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice [5]</td>
<td>2.4</td>
<td>WARP, powerPC Proprietary</td>
<td>64 (planar)</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>Samsung [6]</td>
<td>1-28</td>
<td>Proprietary</td>
<td>64 (planar)</td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>


be accessed either by: (i) evaluating performance metrics by streaming pseudo-noise (PN) sequences, such as bit-error-rate (BER), error vector magnitude (EVM), etc; (ii) visualizing streamed high-definition (HD) videos;

The remainder of this paper is structured as follows: Sec. II details the system architecture and hardware components implementing the BS; Sec. III addresses different aspects of the communication protocol; Sec. IV presents the initial testbed results in terms of RF-chain synchronization and illustrations of received signal constellations under maximum-ratio combining (MRC) and zero-forcing (ZF) uplink spatial multiplex; and Sec. V presents conclusions drawn from the work.

II. TEST BED DESIGN

A. Problem formulation

In a massive MIMO context, a potential BS architecture designed to yield low processing latency, transport latency and high transport reliability would:

- use an all-mighty central controller (CC) aggregating and processing data from/to all (100) antennas;
- be architected in a star-like fashion yielding hundreds of input/output ports;
- shuffle large amounts of baseband data between the CC and RF front ends through high bandwidth/low latency interconnects;
- operate with hundreds of perfectly synchronized RF chains with low RF impairments;

While the second point imposes a tight hardware constraint, potentially preventing flexibility and scalability of the system, the first is the toughest to meet with today’s off-the-shelf solutions since 100 antennas of baseband data far exceeds the input/output (IO) capabilities of most practical hardware. Flexible implementations of massive MIMO BSs with real-time processing requirements are thus non-trivial.

B. Hierarchical overview

Fig. 1 shows the hierarchical overview of our system, whose main blocks are detailed as follows:

1) Central controller (CC): A master chassis embeds a x64 controller (NI PXIe-8135) which runs LabVIEW on a Windows 7 64-bit OS and serves three primary functions: (i) it provides a user interface for radio configuration, deployment of FPGA bitfiles, system control, and visualization of the system, (ii) it acts as source and sink for the user data—e.g. HD video streams—sent across the links, and (iii) the CC measures link quality with metrics such as BER, EVM, and packet-error rate (PER). It connects to three switches through cabled Gen 2 x8 PCI Express (MXIe) in a star fashion.

2) Switches: The switches consist of three (NI 1085 PXIe) 18-slot chassis. The first slot is reserved for the modules (NI PXIe-8381) that connect to the master chassis, and the remaining slots hold MXIe interface cards (NI PXIe-8374) to link with the SDRs. The MXIe interface between the Gen 2 x8 PCIe backplane and the SDRs is Gen 1 x4. Switches yield no processing but allow data to be transferred between SDRs using peer-to-peer direct memory access (DMA) streaming and between SDRs and the CC using target-to-host and host-to-target DMA transfers.

3) Software defined radios: The SDRs (NI 2943R/USRP-RIO) each contain a reconfigurable (Xilinx Kintex-7) FPGA and two full-duplex 40 MHz RF bandwidth transceivers that can be configured for center frequencies 1.2-6 GHz, and can transmit with up to 15 dBm. Baseband processing is partitioned and distributed across the fifty FGPAs, as detailed in Sec. II-D, and the RF transceivers connect to the antenna array.

Please check [7] for further hardware specifications.

C. Streaming IO rates

For proper baseband processing partition, the limitations of the hardware components implementing the system in Fig. 1 are:

- Each Gen 2 x8 PCI Express interface linking the three chassis handles up to 3.2 GBps bidirectional traffic.
- Two Gen 2 x8 switches link the interface cards through the backplane of the chassis. Their streaming rate is bounded to 3.2 GBps of bidirectional traffic in each slot with an aggregate total of 32 GBps inter-switch traffic.
- Each SDR has 13 available DMA channels (three are used for the radio configuration) that share the total IO rate for Gen 1 x4 PCIe of 800 MBps bidirectional.

D. Sub-system partitioning

Below we detail how the baseband processing is partitioned across FPGAs. The functional representation of an OFDM Massive MIMO system is shown at the left part of Fig. 2. To
map this to hardware, the occupied bandwidth is divided into eight OFDM sub-bands which are processed independently to relax the IO requirements of a single FPGA. One subsystem of eight FPGAs, shown in the right part of Fig. 2, operate per sub-band. Additional folding of MIMO detectors and precoders (since we do not have eight subsystems) is performed and end nodes are inserted to achieve a full 100-antenna platform. For each RX chain, the received RF signals are digitized, followed by analog front-end calibration and time/frequency synchronization. From the synchronized data, the cyclic prefix (CP) is removed, followed by FFT OFDM demodulation and guard-band removal. Note that the OFDM symbols contain the superposition of the transmitted signals by all users. In each sub-system, consisting of 16 receive antennas, the yet unequalized OFDM symbols are streamed into an FPGA with an “Antenna Combiner” function. This combines all the uplink streams from the 16 antennas and passes the result to another FPGA in the sub-system with a “Bandwidth splitter” function, which splits the signals into eight bandwidth chunks. In each sub-system, we have one FPGA with a “MIMO Detector” function collecting data of a given bandwidth chunk from the other seven sub-systems. Using the channel matrix estimated from uplink pilots, the “MIMO Detector” cancels interference and detects the frequency-domain symbols from each user equipment. The detected symbols are then sent to the CC for further processing, such as link quality evaluation.

At the downlink, the estimates of both channel and reciprocity calibration weights are passed to the “MIMO Precoder”, and the reverse processing performed, e.g., modulation instead of demodulation.

It can be noted that each subcarrier data sample is quantized with 12 bits for each in-phase and quadrature component. This allows meeting the SDRs IO rate limitations listed in II-C.

### E. Latency analysis

To support fast precoder turnaround time, the system has been architected to provide low latency in the signal path from channel estimation to MIMO precoding, shown in Fig. 2. The turnaround time must meet the frame structure shown in Fig. 4. This structure leaves 214 µs for total latency

\[
\Delta = \Delta_f + \Delta_o + \Delta_p + \Delta_r + \Delta_s + N_b \Delta_b + \phi \quad (1)
\]

in the critical path, including RX front-end delay \(\Delta_f\), OFDM RX (CP removal, FFT, guard subcarrier removal) \(\Delta_o\), channel estimate calculation \(\Delta_r\), precoder calculation \(\Delta_p\), OFDM TX (guard subcarrier interleave, IFFT, CP addition) \(\Delta_s\), and TX front-end delay \(\Delta_p\). Additional sources of latency include overheads in data routing, packing, and unpacking \(\phi\) as well as latency for each hop across the PCIe backplane \(N_h \Delta_h\). The worst-case latency of each hop is \(\Delta_h \approx 5 \mu s\) for the seven-hop path \((N_h = 7)\), resulting in a worst-case total PCIe latency of \(N_h \Delta_h = 35 \mu s\) in the critical signal path. \(\Delta_f + \Delta_s\) was measured to be \(\approx 2.25 \mu s\), \(\phi \approx 0.1 \mu s\), \(\Delta_r \approx \Delta_s \approx 27.5 \mu s\). \(\Delta_p\) depends on the type of precoder and respective implementation type. The MRT precoder can be processed point-by-point, allowing for a high degree of pipelining. Similarly, channel estimation can be performed point-by-point. The highest latency configuration will be that for the ZF precoder due to the matrix inverse, matrix-matrix multiplications, and its serial-to-parallel conversions.

### F. Synchronization

A massive MIMO basestation requires time synchronization and phase coherency between the RF chains. This is achieved using a reference clock and timing/trigger distribution network. This synchronization network consists of eight OctoClock modules in a tree structure with a master OctoClock.
feeding seven secondary OctoClocks. Low skew buffering circuits and matched-length transmission cables ensure that there is low skew between the reference clock input at each SDR. The source clock for the system is an oven-controlled crystal oscillator within an NI PXIe-6674T timing module. Triggering is achieved by generating a start pulse within the Master SDR via a software trigger. This trigger is then fed from an output port on the master to the NI PXIe-6674T timing module, which conditions and amplifies the trigger. The trigger is propagated to the master OctoClock and distributed down the tree to each SDR in the system (including the master itself). This signal sets the reference clock edge to use for start of acquisition for the transmitter (TX) and receiver (RX) within each channel. Initial results show that reference clock skew is within 100 ps and trigger skew is within 1.5 ns, which is well below the sampling period of 33 ns.

G. Antenna Array

The three different stages of the array building process are described below.

1) Material and characterization: We choose Diclad 880 with thickness of 3.2 mm as the printed circuit board substrate. The dielectric constant and dissipation factor were confirmed using a trapped waveguide characterization method [8]. To verify the substrate characterization, a six element patch array with slightly different element sizes was built, measured, and compared with the simulated data. To fit the final results, a final re-characterization of the substrate was performed, and the simulated and measured bandwidth matched within 1 MHz.

2) Design: A planar "T"-shaped antenna array was built with 160 dual polarized λ/2 shorted patch elements. The "T" upper horizontal rectangle has 4 × 25 elements and the central square 10 × 10 elements, (see Fig. 3). This yields 320 possible antenna ports that can be used to explore different antenna array arrangements. All antenna elements are center shorted which improves isolation, bandwidth, and reduces risk of static shock traveling into the active components if the elements encounter a static electric discharge. The feed placement shifts by 0.52 mm from the center of the array elements to the outer edge elements in order to maintain match with changing array effects which impact individual elements differently. The size of the element changes by 0.28 mm from the center of the array to the outer elements this maintains constant center frequency of 3.7 GHz throughout the entire array.

3) Measurements: The final 160 element array was simulated at 3.7 GHz. Results showed an average match of -51 dB, and an average 10dB-bandwidth of 185 MHz. Similar tests were done to the manufactured array which yielded an average 10dB-bandwidth of 183 MHz centered at 3.696 GHz and the average antenna match was found to be -28 dB.

H. Mechanical structure and electrical characteristics of BS

Two rack mounts assemble all BS components with combined measures of 0.8 × 1.2 × 1 m shown in Fig. 3. They were attached on top of a four-wheel trolley not to compromise its mobility when testing different scenarios. Approximate combined weight and average power consumption are 300 kg and 2.5 kW, respectively.

I. User Equipment

Five SDRs (NI 2953Rs/USRP-RIOs) are used at the terminal ends to emulate the UEs. They yield similar properties as the ones at the BS with the additional feature of their internal clocks can be locked to a GPS reference signal. This provides a reliable timing reference for sampling purposes, and a frequency offset of less than 1 ppb.

III. SYSTEM SPECIFICATIONS

A. General parameters

In the current setting, the testbed operates with many parameters similar to LTE-like cellular systems, as shown in Table II.

| Table II
| HIGH-LEVEL SYSTEM PARAMETERS |
| Parameter | Variable | Value |
| Bandwidth | W | 20 MHz |
| Carrier frequency | f_c | 3.7 GHz |
| Sampling Rate | F_s | 30.72 MS/s |
| FFT Size | N_{FFT} | 2048 |
| # Used subcarriers | N_{used} | 1200 |
| Slot time | T_S | 0.5 ms |
| Sub-Frame time | T_{sf} | 1 ms |
| Frame time | T_f | 10 ms |
| # UEs | K | 10 |
| # BS antennas | M | 100 |

B. Supported precoders

The heavy real-time processing requirements for massive MIMO have, in general, been restricting the attention mostly to linear precoders/equalizers. For a proof-of-concept of massive MIMO, we focus on the implementation of two standard linear precoders:

1) Maximum Ratio Transmission (MRT): The MRT precoder maximizes the signal-to-noise ratio (SNR) at the terminal side and precoding-weights simply consist of the complex conjugate of the estimated channels. Thus, precoding is both of low complexity and can, in principle, be performed independently close to each antenna, i.e., in a non-centralized fashion.

Power scaling of precoding weights is still needed if an average transmit power level is to be met. This requires a centralized control structure with relatively low signaling overhead.

2) Zero Forcing (ZF): The ZF precoder forces interference among users to zero and precoding weights are obtained from inverting the inner Gram matrix of the full channel matrix, which contains all estimated channels. This implies a more complex precoder calculation and leads to a centralized architecture where all processing typically happens at a central controller.

The MRT precoding process is also known as conjugate beamforming.
Fig. 3. Left: Side view of the mechanical assembly the BS. The two racks sit side-by-side (not as shown) with the SDRs facing the same direction (towards the antenna array). Two columns of URSPs are mounted in each rack, totaling 50 of them. Right: Picture of the assembled BS, with mounted antenna array.

C. Frame structure

The transmission of massive MIMO data is divided into 10 ms radio frames as shown in Fig. 4. The frame consists of 10 subframes, each containing two 0.5 ms slots. The radio frame starts with a special down-link broadcasting subframe (may consist of PN sequences) to setup the initial synchronization of the network, e.g., UEs can synchronize their frequencies (both carrier frequency and sampling frequency) and align the time offset due to their variable distance to the BS. The remaining 9 subframes are used for UL and DL data transmission.

As also demonstrated in Fig. 4, one slot consist of 7 OFDM symbols, where the 1st is used entirely for UL pilots, followed by 2 UL data symbols, a guard period for UL→DL switching, and 2 DL data symbols, followed by a guard period for DL→UL switching.

D. Pilot allocation

The frequency domain uplink pilots are sequentially interleaved to each of the 10 users in the system, as shown in Fig. 5, where $P_{i,j}$ is the pilot for user $i$ and subcarrier chunk $j$, where each subcarrier chunk consist of 10 subcarriers. For a particular user, non-trained subcarrier channels can be estimated through an interpolation/extrapolation scheme using the trained ones. At the downlink, since users are spatially multiplexed, pre-coded pilots are inserted every $10^{th}$ subcarrier in the first DL OFDM symbol to allow compensation for the RF chain responses of the terminals.

E. Throughput

The total amount of aggregated baseband traffic that can be handled by the testbed both in uplink and downlink directions is given by

$$R_{Bs} = \#BS\ antennas \times 2 I/Q_{bits} \times ADC_{SR} = 384 \text{ Gbps}$$

where $I/Q_{bits} = 16$ is the maximum number of quantization bits per I/Q sample and $ADC_{SR} = 120 \text{ MS/s}$ is the ADC sampling rate.

An example of the data rate per user per direction is given by

$$R_{ue,ul/dl} = \frac{N_{used} \times N_{ul/dl}}{0.5 \text{ ms}} \times \frac{N_{bf} - N_{bf}}{N_{sf}} \times R_c \times N_{mod} ,$$

Fig. 4. Frame structure.

Fig. 5. Frequency-domain pilot-symbol allocation.
where $N_{ul/dl}$ is the number of UL or DL OFDM symbols within one slot, $N_{sf}$ is the number of broadcasting subframes within one radio frame, $R_c$ is the coding rate, and $N_{mod}$ is the number of bit per modulated symbol. In case of 16-QAM modulation with rate $2/3$ channel coding, the system provides $11.52$ Mbps data rate per user per direction, which can be enhanced to $17.28$ Mbps if 64-QAM is used.

IV. INITIAL RESULTS

In this section, the synchronization capabilities of the BS RF front ends are verified, and as a proof-of-concept, we realize an indoor uplink massive MIMO transmission with 100 BS antennas and four single antenna users and show equalized signal constellation points.

A. Phase coherence

We measured the phase drift of different RX RF-chains. A tone transmitted by one SDR is split into four signals, and input to four SDRs spanning four different OctoClocks and two switches. SMA cables and RF splitters were used as the channel for this experiment. Since all four channels share the same TX RF chain, and the cables/RF splitter have static responses, the phase drift is solely due to the RF chains of the receivers. Fig. 6 shows the phases of the measured signal phases which remain within 5 degrees across 1 hour of measurements. The largest change in phase is observed within the first 10 minutes, as the devices are coming up to temperature. After that warm-up period, phases are stable to within a few degrees over a one-hour period. The results suggest that reciprocity calibration can be performed on an hourly basis, without severe performance degradation [9].

B. Time Synchronization

An 800-sample 30.72 MHz Gaussian PN sequence is repeatedly transmitted by a single antenna. The transmitter is positioned about a meter in front of a $4 \times 25$ antenna array arrangement. All 100 receiving antennas are roughly at the same distance from the transmitter and their respective RF-chains share the same reference clock signals. This setup yields a strong LOS channel that can be used to verify the sampling synchronization capabilities of the RF chains. For each channel, the impulse response (IR) is obtained by performing a circular cross-correlation of the received signals with the original PN sequence. Fig. 6 shows that the measured channels yield a distinctive planar wavefront with a small delay spread. These results indicate that the received samples are well time aligned within one 30.72 MS/s sample, i.e. within 33 ns.

C. Uplink Massive MIMO transmission test

As proof-of-concept, we performed an uplink Massive MIMO transmission from four single-antenna UEs to 100 BS-antennas in our lab. Each UE is equipped with SkyCross UWB antennas (SMT-3TO10M-A) and radiate 0 dBm of power. The uplink transmission mode was chosen since it can be realized without performing reciprocity calibration and implementation of the uplink/downlink frame structure, i.e., the base station simply equalizes the data symbols using their respective channel estimates. This also allows all baseband processing to be implemented solely at the CC if no real-time constraints are to be met. We took this provisional approach to be able to showcase a massive MIMO transmission. All baseband processing will, however, be moved to the FPGAs in subsequent work to meet the testbed description given in Sec. II-B. Note that the parameters specified in Table II are still valid for this experiment, but slots are transmitted at a rate which can be handled by the CC. We used the same slot structure as Fig. 4 but no downlink data symbols were transmitted during this test.

Fig. 7 and Fig. 8 show the equalized signal constellation points of an received OFDM data symbol under different channel conditions, users separations and MIMO decoders. For a given user, we used zero-order hold to interpolate between trained subchannels. This explains the small rotation that can be observed for the measured signal constellation points.

Overall, ZF outperformed MRC in all experiments, and showed to be possible to separate both: (i) closely spaced users, and (ii) users at different distances to the BS, if enough
power is transmitted. For the MRC case, its interference limited performance constrains user scenarios yielding acceptable performance to those where users are being spaced rather far apart with some sort of power control.

V. CONCLUSIONS AND FUTURE WORK

In this paper we detail our solution for realizing massive MIMO in a practical testbed. The testbed is operating with a 20MHz bandwidth and 100 antennas at the BS, entirely made of off-the-shelf hardware. To tackle the main hardware bottlenecks, we propose a hierarchical hardware architecture, baseband processing partitioning and a communication protocol that allows the processing to meet real-time requirements. To unveil key performance trade-offs for different system settings, it is of particular interest to be able to operate with flexible communication parameters and antenna array configurations. Synchronization tests between the BS RF chains show small and slow relative phase drifts and tight time alignment of received samples. As proof-of-concept, an over-the-air uplink massive MIMO transmission with spatial multiplexing of four users was performed with all baseband processing being conducted at the CC.

In future work, we intend to move the baseband processing to the SDRs FPGAs, such that both uplink and downlink transmissions can be realized under full real-time requirements. In addition to the distributed processing architecture presented in Fig. 2, we also intent to investigate alternative architectures based on a more centralized processing scheme.

ACKNOWLEDGMENTS

This project has partially been funded by grants from the Swedish Foundation for Strategic Research, the Swedish Research Council and the Strategic Research Area ELLIIT. The authors would also like to thank Xilinx for donating LabVIEW compatible IP blocks.

REFERENCES