
Peer reviewed version

Link to published version (if available):
10.1109/GLOCOM.2016.7841965

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at http://ieeexplore.ieee.org/document/7841965/authors?ctx=authors. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/
LOS Throughput Measurements in Real-Time with a 128-Antenna Massive MIMO Testbed

Paul Harris*, Siming Zhang*, Mark Beach*, Evangelos Mellios*, Andrew Nix*, Simon Armour*, Angela Doufexi*
Karl Nieman†, Nikhil Kundargi†

*Communication Systems & Networks Group, University of Bristol, Bristol, UK
†Advanced Wireless Research Group, National Instruments, Austin, Texas, USA

Email: {paul.harris, sz1659, m.a.beach, evangelos.mellios, andy.nix, simon.armour, angela.doufexi}@bristol.ac.uk
Email: {karl.nieman, nikhil.kundargi}@ni.com

Abstract—This paper presents initial results for a novel 128-antenna massive Multiple-Input, Multiple-Output (MIMO) testbed developed through Bristol Is Open in collaboration with National Instruments and Lund University. We believe that the results presented here validate the adoption of massive MIMO as a key enabling technology for 5G and pave the way for further pragmatic research by the massive MIMO community. The testbed operates in real-time with a Long-Term Evolution (LTE)-like PHY in Time Division Duplex (TDD) mode and supports up to 12 spatial streams, providing an excellent basis for comparison with existing standards and complimentary testbeds. Through line-of-sight (LOS) measurements at 3.51 GHz in an indoor atrium environment with 12 user clients, an uncoded system sum-rate of 1.59 Gbps was achieved in real-time using a single 20 MHz LTE band, equating to 79.4 bits/s/Hz. To the best of the author’s knowledge, this is the highest spectral efficiency achieved for any wireless system to date.

Index Terms—Massive MIMO, Testbed, Field Trial, Indoor, 5G

I. INTRODUCTION

MIMO has become a mature communications technology in recent years, finding itself incorporated today within both Wi-Fi and fourth generation (4G) cellular standards. Current systems typically deploy between 2 to 4 antennas at the Access Point (AP) or Base Station (BS), and they can be used to either enhance the achievable throughput for a single device or allow 2 to 4 devices to be served simultaneously in the same frequency resource. Massive MIMO takes the latter Multi-user (MU) MIMO concept one step further by deploying hundreds of antennas at the BS, each with their own individual Radio Frequency (RF) chain. The result is greatly enhanced spatial multiplexing performance allowing many tens of User Equipments (UEs) to be served with greater reliability than in standard MU MIMO. It is well-recognised as one of the key enabling technologies for 5G that could provide superior spectral and energy efficiencies. The theoretical benefits of massive MIMO can be found discussed and well documented in [1], [2] and [3].

However, despite promising theoretical results, massive MIMO introduces many new and challenging problems when it comes to a real-time system implementation. At the air interface, it is crucially important to coherently measure and track the channel conditions for all active users, where hundreds of signal paths need to be suitably managed and processed at the BS to ensure an acceptable turnaround time (latency) from channel estimation to Downlink (DL) precoding. Prototyping work for various aspects of massive MIMO technology is thus ongoing in many different parts of the world and Table I provides a summary of known testbeds. The ArgosV2 testbed [4], [5] developed at Rice University is a TDD system based upon the Wireless Open-Access Research Platform (WARP) platform. Some basic cell capacity predictions have been made by measuring the received Signal to Interference plus Noise Ratio (SINR), but true real-time performance with a specified Bandwidth (BW) has yet to be reported. A field trial of a TDD massive MIMO prototype by ZTE was reported in [6] where 64 transceiver units served 8 TDD-LTE commercial handsets located at different floor levels in a high rise building, achieving a 300 Mbps sum rate in 20 MHz of BW. In Australia, the Frequency Division Duplex (FDD) Ngara demonstrator [7] uses practical low-cost hardware and was previously reported to have achieved a coded Uplink (UL) spectral efficiency of 67.26 bits/s/Hz in a lab environment at 638 MHz. The Titan massive MIMO platform [8] claims to provide a distributed solution for baseband processing with a large number of RF chains through a two-layer linear processing method, but an implementation using the platform has not yet been reported. And finally, Lund University, also in collaboration

<table>
<thead>
<tr>
<th>Name</th>
<th># of Antennas</th>
<th># of Streams</th>
<th>BW (MHz)</th>
<th>Band (GHz)</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argos V2</td>
<td>96</td>
<td>16</td>
<td>n/a</td>
<td>2.6</td>
<td>TDD</td>
</tr>
<tr>
<td>Ngara</td>
<td>32</td>
<td>32</td>
<td>28</td>
<td>0.6</td>
<td>FDD</td>
</tr>
<tr>
<td>Lund</td>
<td>100</td>
<td>12</td>
<td>20</td>
<td>3.7</td>
<td>TDD</td>
</tr>
<tr>
<td>ZTE</td>
<td>64</td>
<td>8</td>
<td>20</td>
<td>2.6</td>
<td>TDD</td>
</tr>
</tbody>
</table>
with National Instruments (NI), have published numerous papers such as [9] describing their 100-antenna co-located massive MIMO testbed, which utilises the same underlying architecture presented here.

In this paper, we present initial field trial results for the Bristol Is Open (BIO) massive MIMO research platform, previously introduced in [10], which has been developed within the Communication Systems & Networks (CSN) Research Group at the University of Bristol in close collaboration with both National Instruments and Lund University. The BS prototype consists of 64 NI Universal Software Radio Peripheral (USRP) Reconfigurable Input/Output (RIO) [11] Software-Defined Radios (SDRs) providing 128 RF chains, with a further 6 USRP RIOs acting as 12 single-antenna UEs. A linear array of 128 dipole antennas was deployed at the BS for this trial and a reconfigurable patch-panel array is being produced for future work. Here, real-time channel measurements taken simultaneously at 112 active BS antenna ports at 3.51 GHz for 12 transmitting UEs over a 20 MHz bandwidth are reported. The more deterministic orthogonality between spatial streams, also referred to as the channel hardening effect, is presented graphically, and the Singular Value Spread (SVS) for each scenario is discussed. Furthermore, with an UL dominated frame schedule, we measured a real-time uncoded sum capacity of 1.59 Gbits/s for 12 UEs, which equates to a spectral efficiency (SE) of 79.4 bits/s/Hz.

The remainder of the paper is organised as follows: Section II presents an overview of the testbed architecture and the novel Field-Programmable Gate Array (FPGA) implementation of the QR Decomposition and Minimum Mean Square Error (MMSE) detection. The measurement environment, test scenarios and methodologies are then described in Section III, before Section IV presents the throughput results and channel observations. Finally, Section V summarises the work and outlines our future measurement plans.

II. SYSTEM OVERVIEW

The top-level system specifications are provided in Table II. The testbed design has been based on the LTE specifications to a substantial extent; however, there are some major aspects for implementing a TDD based massive MIMO system in which it diverges from the LTE numerology.

A TDD massive MIMO architecture exploits the reciprocity of the wireless propagation channel to avoid the large DL channel feedback overhead in an FDD implementation. Instead, orthogonal pilots are transmitted in the UL, and used to estimate the UL channel. This estimated channel is used to compute the DL precoder, which under perfect channel reciprocity is simply the conjugate of the UL decoder matrix.

A. HARDWARE AND SOFTWARE ARCHITECTURE

The hardware architecture of the system is shown in Fig. 1 and represented pictorially in Fig. 2. The architecture separates the various functional blocks of the massive MIMO base station into three units - the Remote Radio Head (RRH), the Base Band Unit (BBU) and the Software Controller.

RRH: The RRH is implemented on a USRP RIO, which is a commercial Software Defined Radio. Each USRP RIO has two RF transceiver chains, and an FPGA for implementing the per-antenna processing such as Orthogonal Frequency Division Multiplexing (OFDM) modulation, demodulation, sample re-timing, reciprocity calibration, and IQ impairment correction.

BBU: Multiple Kintex 7 FPGAs manufactured by Xilinx

The remainder of the paper is organised as follows: Section II presents an overview of the testbed architecture and the novel Field-Programmable Gate Array (FPGA) implementation of the QR Decomposition and Minimum Mean Square Error (MMSE) detection. The measurement environment, test scenarios and methodologies are then described in Section III, before Section IV presents the throughput results and channel observations. Finally, Section V summarises the work and outlines our future measurement plans.

II. SYSTEM OVERVIEW

The top-level system specifications are provided in Table II. The testbed design has been based on the LTE specifications to a substantial extent; however, there are some major aspects for implementing a TDD based massive MIMO system in which it diverges from the LTE numerology.

A TDD massive MIMO architecture exploits the reciprocity of the wireless propagation channel to avoid the large DL channel feedback overhead in an FDD implementation. Instead, orthogonal pilots are transmitted in the UL, and used to estimate the UL channel. This estimated channel is used to compute the DL precoder, which under perfect channel reciprocity is simply the conjugate of the UL decoder matrix.

A. HARDWARE AND SOFTWARE ARCHITECTURE

The hardware architecture of the system is shown in Fig. 1 and represented pictorially in Fig. 2. The architecture separates the various functional blocks of the massive MIMO base station into three units - the Remote Radio Head (RRH), the Base Band Unit (BBU) and the Software Controller.

RRH: The RRH is implemented on a USRP RIO, which is a commercial Software Defined Radio. Each USRP RIO has two RF transceiver chains, and an FPGA for implementing the per-antenna processing such as Orthogonal Frequency Division Multiplexing (OFDM) modulation, demodulation, sample re-timing, reciprocity calibration, and IQ impairment correction.

BBU: Multiple Kintex 7 FPGAs manufactured by Xilinx

The remainder of the paper is organised as follows: Section II presents an overview of the testbed architecture and the novel Field-Programmable Gate Array (FPGA) implementation of the QR Decomposition and Minimum Mean Square Error (MMSE) detection. The measurement environment, test scenarios and methodologies are then described in Section III, before Section IV presents the throughput results and channel observations. Finally, Section V summarises the work and outlines our future measurement plans.
are used to implement the BBU unit on NI FlexRIOs [12], which are FPGA coprocessors. The FlexRIO devices perform centralized MIMO processing including channel estimation, MIMO detection, MIMO precoding, and Quadrature Amplitude Modulation (QAM) symbol mapping/demapping. Software Controller: A quad core Intel Xeon processor runs the Software Controller. The upper layer Medium Access Control (MAC) functions, system setup and configuration, FPGA bitfile management, bit source and sink for BBU and visualizations are implemented on the Software Controller. The software architecture of the system is shown in Fig. 3. A massive MIMO implementation imposes two major constraints on the system design - a) massive data transfer and b) massive and rapid matrix computations. This testbed has been designed to seamlessly tackle both of these constraints for a 128 antenna system with up to 20 MHz of channel bandwidth.

Massive Data Transfer: Efficient massive MIMO coding and decoding algorithms such as MMSE and Zero-Forcing (ZF) need availability of the per antenna samples at a single centralized location. For example, the 128 antenna system generates 51.6 Gb/s of baseband samples. Aggregation and disaggregation of this enormous amount of data into the FPGA coprocessors is implemented in a dense grid of high throughput Gen 3 Peripheral Component Interconnect Express (PCIe) switching fabric. Massive Matrix Computation Massive MIMO performs UL/DL switching within the channel coherence interval to be able to use the UL pilots to compute the Downlink Channel State Information (CSI) at the transmitter. This channel reciprocity based approach needs an ultra fast computation of the channel estimate and generation and multiplication with the precoder matrix on a per subcarrier basis.

B. FPGA Implementation of MIMO Processing

In this section we will present the design and implementation of a wide data path Linear MMSE decoder for a 128x12 Multi User MIMO system.

To achieve high system throughput at the BS of 30.72 MS/s \times 128 channels in UL and DL directions, baseband signal processing is performed on FPGAs. The MIMO processing architecture in the system has been designed to use a common MIMO core to produce a linear 12 \times 128 decoder and 128 \times 12 encoder for the UL and DL channels respectively. In this implementation, the linear encoder is the Hermitian transpose of the linear decoder. This MIMO core is capable of producing encoders and decoders based on several popular linear MIMO algorithms including MMSE, ZF, and Maximal Ratio Combining (MRC).

In [13], [14], the authors presented an FPGA based architecture to solve for \( W_{MMSE} \), the 12x128 detection matrix, using QR decomposition. This testbed extends that algorithm to scale up significantly in throughput for a massive MIMO system. The QR Decomposition based MMSE is computed as shown in Fig. 4. ZF and MRC MIMO decoders are reformulated as special cases of the MMSE and implemented using additional logic that reconfigures the MIMO core to bypass orthogonalization in the MRC case and use a small \( \sigma \) parameter in the ZF case.

Due to the high system throughputs and the relatively low FPGA clock rates, much of the processing must be performed in parallel by processing multiple input and/or producing multiple output samples every clock cycle. This type of processing is also known as wide data path processing. Depending on the clock rate and throughput, individual signal processing functions can be implemented in fully parallel (most resources, highest throughput), partial parallel (medium resources, medium throughput), or fully sequential (least resources, least throughput) forms.
To relax throughput, the same linear decoder and encoder is used for each resource block (12 subcarriers). In this formulation, the $128 \times 12$ matrix throughput is $16.8 \text{ MS/s}$ (LTE subcarrier rate) divided by $12 = 1.4 \text{ Mmatrices/s}$. To meet this throughput, the matrices are processed at a $200 \text{ MHz}$ clock rate using four FlexRIO FPGAs which each perform a partial-parallel (4 path) QR decomposition. An example of a $N \times 12$ systolic array QR decomposition based on modified Gram Schmidt for full-parallel, partial-parallel, and full-sequential is shown in Fig. 6 and described in [13], [14].

Fig. 6 shows three different systolic array formulations of the $N \times 12$ QR decomposition. Three different arrays are shown for full-parallel (12 path), partial-parallel (4 path), and full-sequential (1 path). Each formulation reuses three common nodes “orthogonalize”, “normalize”, and “orthogonalize and normalize”. To balance throughput requirements and resource utilization, the implementation for the $128 \times 12$ QR decomposition is chosen to be the four path partial parallel formulation.

After the $W_{MMSE}$ and $W_{MMSE}^H$ decoder and encoder matrices are computed for the full bandwidth (100 resource blocks), UL subcarriers are decoded and DL symbols are encoded. Fig. 5 shows a partial parallel (4 path) implementation of the linear decoder, where a $12 \times 128$ by $128 \times 1$ matrix vector multiplication is performed in thirty two $200 \text{ MHz}$ clock cycles by decomposing the operations into four parallel submatrices and subvectors, and summing the result. The implementation of this function in LabVIEW FPGA is shown in Fig. 7.

C. Frame Schedule

The PHY frame schedule used for the system is shown in Fig. 8. In its default configuration, it is based closely upon the TDD LTE standard, but it can be completely customised at the OFDM symbol level to allow a range of configurations to be applied for different applications. For the throughput results shown in this paper, the schedule was configured to use UL only transmission with two pilots per radio frame.

D. Synchronisation

Cabled synchronisation was used for this trial with a network of Octoclock distribution modules [15]. The BS utilises 8 of these to phase align the USRP Local Oscillators (LOs) by redistributing a $10 \text{ MHz}$ clock with $50 \text{ parts-per-billion}$ (ppb) accuracy provided by a PXIe-6674T timing card [16].

To achieve sample alignment, a start trigger that signals the start of the first radio frame is fanned out to all radios by equal length cables. The user clients were connected to a 9th Octoclock by extended cable and the timing offset corrected for, but future deployments will eventually use Over-the-air (OTA) synchronisation.

III. EXPERIMENT OVERVIEW

A. Measurement Environment

The length of the lower atrium in the University of Bristol’s Merchant Venturers Building was used for three different LOS measurements between the BS and 12 UEs. UEs were grouped both in a straight line parallel to the BS and at a slant, with a distance of $3.3m$, $12.5m$ or $18.1m$ to the nearest client in each scenario. At the BS side, a $128$ element linear array of length $5.44m$ was used as shown in Fig. 9, providing half-wavelength spacing at $3.5 \text{ GHz}$. A floor plan of the experiment is shown in Fig. 10 with the UE locations for each of the three scenarios. Measurements were performed outside of university hours to ensure the environment remained as static as possible.

B. System Configuration

The objectives for this initial trial were to evaluate the achievable uncoded UL throughput and establish the degree of spatial orthogonality that could be achieved for the three scenarios described above. As the environment was relatively static, the frame schedule was configured such that channel estimation was performed in $5\text{ms}$ intervals, and all remaining slots were assigned for UL data with a 256-QAM Modulation.
and Coding Scheme (MCS). ZF detection was used for all throughput measurements and equal transmit levels were fixed for all UEs. 100 channel data captures were recorded to disk for each scenario at an interval of approximately 200ms, resulting in a total measurement period of approximately 3 minutes for each run.

C. Channel Processing

For each scenario, the capture files contain the raw 12x128 channel matrix \(H\) for all 100 resource blocks of 180 kHz (12 subcarriers). To quantify the joint orthogonality between users, we obtain the SVS. This is found by first performing the Singular Value Decomposition (SVD)

\[
H_{n,t} = U_{n,t} S_{n,t} V_{n,t}^H
\]

where for resource block \(n\) and acquisition time \(t\), \(U_{n,t}\) and \(V_{n,t}\) represent the left and right unitary matrices, and \(S_{n,t}\) is a diagonal matrix containing the singular values. The SVS is then defined as the ratio of the largest to the smallest singular values of \(H\):

\[
\gamma_{n,t} = \frac{\max_i \lambda_1, \lambda_2, ... \lambda_i}{\min_i \lambda_1, \lambda_2, ... \lambda_i}
\]

and a single SVS value for each capture is obtained by averaging across resource blocks

\[
\tilde{\gamma}_t = \frac{\sum_n \gamma_{n,t}}{N}
\]

where \(N = 100\) is the total number of resource blocks. Cumulative Distribution Function (CDF) plots are then produced using the resource block averaged SVS values for all captures. This is performed in each scenario for 32, 64 and 112 elements by truncating the antenna dimension centrally. In addition to the SVS, the 12x12 \(HH^H\) symmetric matrix was calculated and averaged across all subcarriers and captures, providing an alternative way of visualising the UE pairwise orthogonality.

IV. INITIAL RESULTS

CDF plots for the resource block averaged SVS values obtained for each scenario with 32, 64 and 112 antenna elements are shown in Fig. 11. From 2.6 GHz measurements with a 128-element virtual linear array, [17] provides SVS statistics for closely spaced users in outdoor LOS conditions and compares these to the independent and identically distributed (iid) Rayleigh channel ideal of approximately 2.5 decibels (dB). For 32 antennas, the median SVS from these measurements with users spaced 13-17 wavelengths apart was approximately 10 dB. This most closely matches with scenario 2 where our UE separation was only 6 wavelengths. In the 128-antenna case, the median SVS in [17] only drops to 9 dB, whereas all three of our measurement scenarios show lower values with only 112 antennas. The above two observations imply that we are experiencing a richer scattering environment indoors, even under LOS conditions. Furthermore, despite LOS conditions in the 3.3m scenario, the angular arrival spread between elements of the array for each UE will be far higher than the 18.1m scenario due to the extreme length of the linear array, allowing us to achieve a very low median SVS of 3.3 dB. Our client separation was only 2.5-6 wavelengths and 16 fewer antennas were used, indicating that extremely good spatial separation can be achieved using massive MIMO in indoor LOS settings, even when users are closely clustered. However, whilst both the 112 and 64 elements cases perform best in Scenario 1, the 32 antenna case appears in reverse, with Scenario 1 providing the worst median SVS by nearly 2 dB. This likely indicates that close range LOS separation is more challenging with the 32 element array as it cannot exploit the angular arrival spread to the same extent. In addition, the difference between median SVS for Scenario 2 and 3 in the
112 element case is 1.3 dB, whereas these are closely aligned in the 32 and 64 element cases. This shows that even at range, the 112 element array was able to exploit the greater spread in angular LOS arrival expected for Scenario 2 over Scenario 3.

To visualise the pair-wise orthogonality, Fig. 12 shows intensity plots of $HH^H$ in Scenario 2 for 32, 64 and 112 antennas, with a dB scale relative to the maximum value. Each plot is averaged across all subcarriers and acquisitions. On inspection of the plots, a strengthening diagonal can certainly be seen as the antenna numbers increase. Interestingly, each pair of streams appear to interfere with one another slightly, even when the antenna numbers are increased. This can be explained by the fact that those two streams correspond to the two antennas on each USRP that have a fixed separation distance of 13cm. Thus, even with 50cm separation between USRPs, we should expect these to be more spatially correlated. As the number of antenna elements are increased, the cluster of user correlation seen in the top right for 32 elements begins to flatten out, and by 112 elements all the off-diagonal values excluding the user pairs are at least 10 dB below the diagonal.

For some streams, the level of correlation is far below this, with user 3 experiencing a rejection of 25 dB with respect to users 1 and 5.

Fig. 13 shows the channel visualisations that were displayed in real-time for scenario 2 during the experiment. In the top left we can see the real-time $HH^H$ intensity, below the PDPs for each UE for a selected BS antenna, and top right the received level for each spatial stream across the antenna array. It can be seen that the level is slightly lower for the extreme left elements, which corresponds to the portion of the array that is partially obscured by the staircase. The received fixed-point (FXP) constellations for each of the twelve streams in scenario 2 are also shown in Fig. 14. These are each extremely clean, further illustrating the high level of orthogonality we observed when analysing the channel. Whilst not tested in this initial trial, it shows hope for the effective use of simple MRC detection and Maximal Ratio Transmission (MRT) precoding with more BS antennas as we leverage the natural pair-wise orthogonality in $H$.

For each scenario, the uncoded UL throughput was also measured in real-time by continually transmitting a known pseudo-random byte sequence from each UE. The BS calculates the sum rate for a given radio frame using

$$Sum \ rate = 100 \times \left( Max - \sum_{i=0}^{11} E_i \right) \text{bits/s}$$

where $E_i$ is the number of bit errors for stream $i$ and $Max$ is the aggregate bit total for all streams in an ideal error-free frame. For our frame schedule, $Max = 1200 \text{ subcarriers} \times$
8 bits $\times$ 138 symbols $\times$ 12 users = 15897600 bits. All 112-elements were used in each case and the results are summarised in Table III. With the frame schedule applied, we could expect a maximum UL throughput of 1.589 Gbps, corresponding to 132.4 Mbps per stream. Scenario 1, where we have the lowest SVS, comfortably achieves this, and only drops slightly for the more distant scenarios. It is evident that the large degree of freedom not only results in high diversity gain, increasing the Signal to Noise Ratio (SNR) of each stream, but also improves user orthogonality for linear capacity growth.

V. CONCLUSIONS AND FUTURE WORK

We have introduced a flexible 128-antenna massive MIMO testbed that can perform 128x12 MIMO processing in real-time on a 20 MHz LTE band. As an initial trial, the performance in an indoor LOS environment was assessed using a physically large linear array, and a spectral efficiency of 79.4 bits/s/Hz was achieved using uncoded transmissions. To the best of the author’s knowledge, this is the highest spectral efficiency reported thus far for any system in the world [18]. Future work will include real-time DL performance evaluation, implementation of massive MIMO optimised power control and OTA synchronisation, rooftop deployment and node distribution on the BIO citywide fibre network.

ACKNOWLEDGMENT

The authors wish to thank Bristol Is Open for access to the hardware facility. They also acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training (CDT) in Communications (EP/I028153/1), NEC and NI. Contributions from the University of Lund (Professors Edfors, Tufvesson and their research team) are also most gratefully acknowledged.

REFERENCES