MegaMIMO and Full Duplex Radios

Neha Sharma

Both of the papers talk about approaches to improve wireless throughput at the physical. The physical layer which is the lowest layer of network stack is the most fundamental and any improvement in the physical layer ripples up all the way. Range of frequencies available at the physical layer determines capacity. Therefore, with improvements in physical layer, capacity increases leading to an increase in throughput.

Both of these papers have led to a great academic research as well as had real impact in early enforcing some trends that are actually showing up in emerging commercial wireless technologies.

Brief definitions:

MegaMIMO is a joint multi-user beamforming system that enables independent access points (APs) to beamform their signals, and communicate with their clients on the same channel as if they were one large MIMO transmitter. Full Duplex Radios are in-band, single antenna full duplex WiFi radios that can simultaneously transmit and receive on the same channel using standard WiFi 802.11ac PHYs and achieves close to the theoretical doubling of throughput in all practical deployment scenarios.

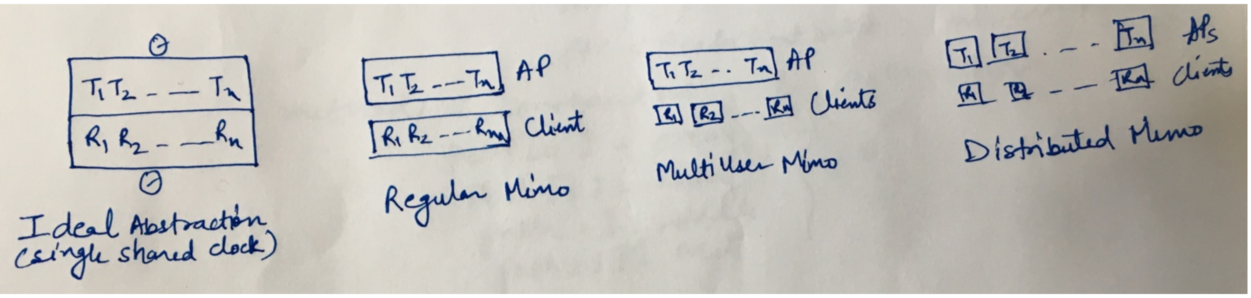
MegaMIMO

MIMO stands for multiple input and multiple output, basically referring to multiple antennas. The goal of systems relying on MIMO device (eg- N transmit antennas and N receiver antennas on a standard MIMO device) is to increase throughput and support more users essentially by two methods-

1. Multiplexity Gains – gains in form of increased performance by sending independent streams of information in parallel along the different spatial paths between transmit and receive antennas. This improves performance because, if we take care in how we construct and decode signals, adding an antenna and independent stream of information need not slow down the streams that are already being sent.
2. Diversity Gains – gains in form of increased reliability and range by sending or receiving redundant streams of information in parallel along the different spatial paths between transmit and receive antennas. The use of extra paths improves reliability because it is unlikely that all of the paths will be degraded at the same time. Improved range, and some performance increase too, comes from the use of multiple antennas to gather a larger amount of signal at the receiver.

Difference between Standard MIMO and MegaMIMO

Standard MIMO can have only one router send signal at a single channel (not more than 1 APs can send signal at the same channel at same time) whereas MegaMIMO can have multiple APs sending signal at the same channel at the same time. Standard/Regular MIMO and Mega MIMO represent two ends of spectrum as can be seen below:



Ideally, we want to have multiple transmit antenna and multiple receive antennas all operating as if they have a single shared clock. There are 3 possible implementations of this abstraction:

1. Regular MIMO- AP has multiple antenna and client also have multiple antenna (ex- with user having different Wi-Fi cards and each Wi-Fi card having multiple antenna)
2. Multi User MIMO- AP have multiple antenna but clients have single antenna (available in some version of LTE, 802.11n); clients have single antenna because of size constraints.
3. Distributed MIMO- MegaMIMO is an implementation of distributed MIMO. A collection of independent APs having 1-2 antennas each and arranged in a distributed fashion to provide an impression of one large MIMO. Biggest challenge in this setting is coordination/synchronization of different things like
    a. Time – to make sure every AP agrees on a schedule on when they would transmit their control packets in channel measurement phase.
    b. Oscillator frequency - to track beamforming within a single packet (difference in frequency results in difference of phase after some time)

<u>MIMO Behavior</u>

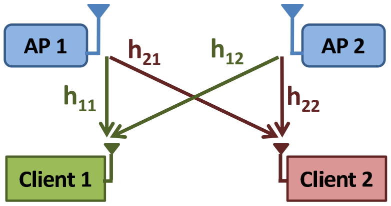


Figure 2: **Channel matrix with 2 APs transmitting to 2 clients.**

Consider a simple scenario of two transmitter and two receivers with synchronized oscillators for the receivers and synchronized oscillators for the transmitters. Receivers are all on the same device. In the given setup, there are four channels. Let h_{ij}, where, $i, j \in \{1,2\}$ be the channel to client i from AP j. Also let $x_j(t)$ be the symbol[1] that needs to be delivered to client j at time t, and $y_j(t)$ the symbol that is received by client j at time t.

In the simplest case, if there was only 1 transmitter and 1 receiver, ideally the received function at any specific time should be a linear function of input, something like $y_i = hx_i + n$ but in practice, the received signal is also dependent on all the previous signals: $y_i = |h|e^{j\theta}x_i + |h'|e^{j'\theta'}x_{i-1} + \cdots + n$. This is because there is never a single path and the same signal travels through multiple paths. This phenomenon is also called multipath. However, the weight of the previous terms keeps going down and noise also cancels out and thus effectively it turns out the case that $y_i = hx_i$

Now in the case of two transmitter and two receivers, since each client has only one antenna, client 1 receives $y_1 = h_{11}x_1 + h_{12}x_2$ and client 2 receives $y_2 = h_{21}x_1 + h_{22}x_2$. This can be written in the form of matrix as below:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$, which equivalently can be written as $\boldsymbol{Y} = \boldsymbol{HX}$

<u>In the regular MIMO case</u>, since all the receivers are on the same node, the node has all the information - y_1 and y_2 and the $\boldsymbol{H}$ matrix. To retrieve x_1 and x_2, assuming $\boldsymbol{H}$ is invertible, we have $\boldsymbol{X} = \boldsymbol{H}^{-1}\boldsymbol{Y}$.

---

[1] [Symbol is an analog quantity i.e. voltage level that represents a bit pattern. In other words, the voltage level can be abstractly thought of as a real number standing for a certain bit pattern and the number of bits in the bit pattern is log_2(# of voltage levels) The arrangement of voltage levels is called a constellation and noise (or signal-to-noise ratio) determines the tightness of packing of symbols. In practice, two voltages are sent at once on orthogonal waves (sine and cos). Also, it is notation to represent symbol as a complex number with a real voltage level and an imaginary voltage level. Also, in polar notation it is represented as exponential. That is a symbol $z = x + iy = Re^{j\theta}$ with $R = \sqrt{x^2 + y^2}$]

In the multi user MIMO case, all the transmitters are on the same node, but receivers are on different nodes. Here, the clients have partial information (client 1 has $y_1, h_{11} and\ h_{12}$and client 2 has $y_2, h_{21} and\ h_{22}$) and hence beamforming is used. In beamforming, the APs measure all the channel coefficients from the transmitters to the receivers at time 0. Then, instead of transmitting x_1 and x_2 directly, the AP transmit

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = H^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$ (the sender is one centralized entity and has the H matrix available to it)

The two receivers receive $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = H \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$, which is same as $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = HH^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Since $HH^{-1} = I$, each receiver effectively gets a scaler multiple of its symbol (without interference from the signal intended for the other client) and they can handle.

Note- In all of these matrix operations, H is assumed invertible. (In case it is not invertible, it's the case when Determinant is 0. This condition physically corresponds to the case when the receiver has same signal from both the transmitter. That is, there isn't enough independence in the system to use MIMO to increase throughput or gain diversity. This happens when the antennas are very close by, they must be separated by an order of few wavelengths. Hence an essential condition for MIMO to work is that the antennas needs to be reasonable separated. With a high SNR, the separation could be lower but with moderate SNR, the separation must be reasonable)

In the distributed MIMO case, the H matrix keeps changing and hence beamforming based on the value of Hat $t = 0$ doesn't work. At any time t,

$$\begin{bmatrix} y_1(t) \\ y_2\,(t) \end{bmatrix} = \begin{bmatrix} h_{11}e^{j(\omega T1 - \omega R1)t} & h_{12}e^{j(\omega T2 - \omega R1)t} \\ h_{21}e^{j(\omega T1 - \omega R2)t} & h_{22}e^{j(\omega T2 - \omega R2)t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ where } j\ =\ sqrt(-1)$$

The terms $\omega T_i - \omega R_j$ represent the phase difference between transmitter *i* and receiver *j* (the drift between their oscillators will make the signals rotate at different speeds relative to each other, causing the phases to diverge). Here we can clearly see, performing beamforming as before and transmitting $\boldsymbol{H^{-1}X}$ (based on the channel value computed at $t = 0$) would not work because $\boldsymbol{H(t)H^{-1}}$is no longer diagonal and hence the receivers cannot decode their intended signal

The channel matrix $\boldsymbol{H}$ at time t is decomposed as $\boldsymbol{H(t)} = \boldsymbol{R(t)HT(t)}$, where $\boldsymbol{H}$ is time invariant, and $\boldsymbol{R(t)}$ and $\boldsymbol{T(t)}$ are diagonal matrices defined as

$$\boldsymbol{R(t)} = \begin{pmatrix} e^{-j(\omega R1)t} & 0 \\ 0 & e^{-j(\omega R2)t} \end{pmatrix}$$

and

$$\boldsymbol{T(t)} = \begin{pmatrix} e^{j(\omega T1)t} & 0 \\ 0 & e^{j(\omega T2)t} \end{pmatrix}$$

Now, if the transmitters transmit the modified signal $T(t)^{-1}H^{-1}x$ at time t, then the received signal can be written as:

$$\boldsymbol{Y = R(t)HT(t)T(t)^{-1}H^{-1}X}$$

which reduces to

$$\boldsymbol{Y = R(t)X}$$

Since $\boldsymbol{R(t)}$ is a diagonal matrix, the receiver can now easily decode their intended signal and thus effective beamforming behavior is achieved.

Also, note that $\boldsymbol{T(t)}$ is also diagonal, and as a result the transmitter phase correction matrix $\boldsymbol{T(t)^{-1}}$is also diagonal.

$$\boldsymbol{T}(\boldsymbol{t})^{-\mathbf{1}} = \begin{pmatrix} e^{-j(\omega T1)t} & 0 \\ 0 & e^{-j(\omega T2)t} \end{pmatrix}$$

There is benefit to it because as we can see, the phase correction entry for each transmitter depends only on the oscillator phase of that transmitter and hence these transmitters' phase correction can be applied locally. But measuring exact value of phase directly is very hard. On further processing,

$$\boldsymbol{H}(\boldsymbol{t}) = e^{j(\omega T1)t}\boldsymbol{R}(\boldsymbol{t})\boldsymbol{H}\boldsymbol{T}(\boldsymbol{t})e^{-j(\omega T1)t}$$

which converts to

$$\boldsymbol{H}(\boldsymbol{t}) = \begin{pmatrix} e^{j(\omega T1-\omega R1)t} & 0 \\ 0 & e^{j(\omega T-\omega R2)t} \end{pmatrix}\boldsymbol{H}\begin{pmatrix} 1 & 0 \\ 0 & e^{j(\omega T2-\omega T1)t} \end{pmatrix}$$

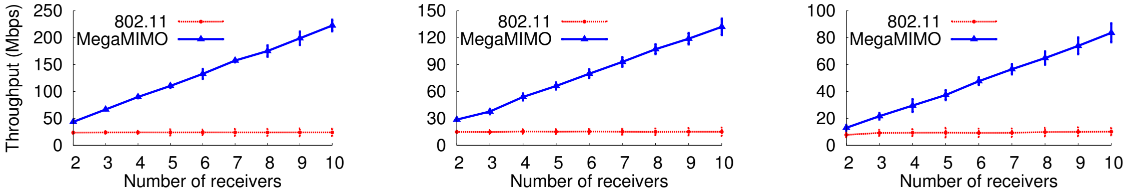
Here, instead of each transmitter measuring its phase value, the transmitters now need to estimate phase difference between themselves and some fixed leader transmitter, which is a much easier thing to do. Similarly, at the receiver side, each receiver needs to estimate the phase difference between themselves and some fixed leader transmitter which can be done using standard wifi techniques.

MegaMIMO PROTOCOL:

1. Channel measurement phase – the APs measure two types of channels: 1) the channels from themselves to the receivers (i.e., the matrix H), which is the beamforming channel whose inverse the APs use to transmit data concurrently to their clients; and 2) the channels from the lead AP to lead the slave Aps which enables each slave AP to determine its relative oscillator offset from the lead AP.
2. Data transmission phase – the APs transmit jointly to deliver concurrent packets to multiple receivers. Data transmission uses beamforming after having each slave AP corrects for its frequency offset with respect to the lead AP.

Results:

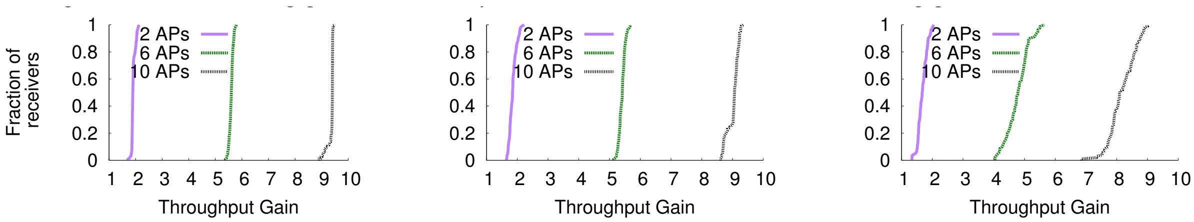
Below are the plots that show a comparison of how throughput scales in total 802.11 network and MegaMIMO's network, with an increase in number of transmitters in three different scenarios of high, medium and low SNR. High SNR is achieved when receiver is pretty close to the transmitter while low SNR is when they are far. The below plots show a linear increase of throughput with increase in number of Aps in all three scenarios.



(a) High SNR (> 18 dB) (b) Medium SNR (12-18 dB) (c) Low SNR (6-12 dB)

Figure 9: **Scaling of throughput with the number of APs.** In this experiment, the number of APs equals the number of receivers. At all SNRs, MegaMIMO's network throughput increases linearly with the number of APs while total 802.11 throughput remains constant.

Below are the plots that show the comparison of CDF of per-receiver throughput gain in three different scenarios of high, medium and low SNR. The steep cliff in CDF shows that the throughput gain is roughly the same across all receivers.



(a) High SNR (> 18 dB) (b) Medium SNR (12-18 dB) (c) Low SNR (6-12 dB)

Figure 10: **Fairness.** CDFs of per-client throughput gain. Across all SNRs, MegaMIMO provides all clients with very similar gains.

Below graph shows the diversity gains' comparison. At a particular value of SNR, as the number of transmitters increase, the throughput goes up because there is more redundancy for each client providing increased diversity gains. 802.11 doesn't provide diversity gains, hence even though the throughput scale up with SNR but it is still less as compared to a MegaMIMO network.

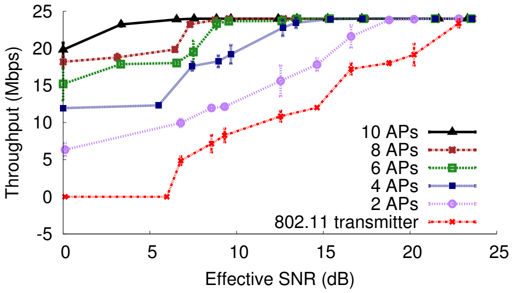


Figure 11: **Diversity Throughput.** Throughput of a MegaMIMO client when using diversity with 2, 4, 6, 8 and 10 APs. MegaMIMO can achieve close to maximum rate even to a client unable to receive any packets with 802.11.

Full Duplex Radios

This paper takes a different approach (relying on analog electronics) towards improving Wifi capacity. The main idea here is to make wireless duplex medium i.e. to be able to receive and transmit signals at the same channel and at the same time using a single antenna. Full Duplex is hard to achieve due to self-interference (which is basically the overwhelming self-echo from the transmitted signal). It is difficult to get rid of self-interference. Below figure shows the self-interference clearly

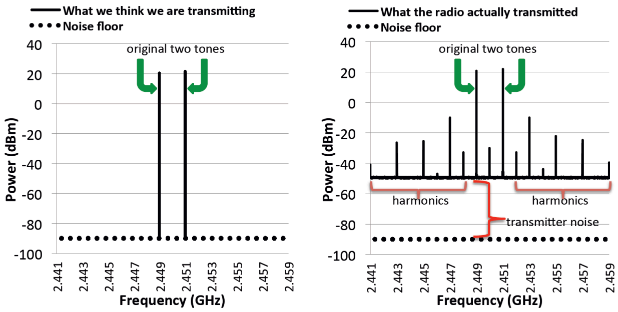


Figure 1: What we think we are transmitting in digital on the left side, and what the radio actually transmitted on the right side. The actual transmitted signal differs significantly from the two tones generated in digital baseband. Note transmitter noise and harmonics are generated in addition to the two main transmitter tones.

The main components in self-interference can be classified into three major categories:

1. Linear Components- basically the linear distortion of original symbol x
2. Non-Linear Components- basically the non-linear cubic and higher order distortion of original symbol x
3. Analog Interference- mostly caused by transmitter noise, also known as Broadband noise and is generated from high power components in the radio transmitter such as power amplifiers. This noise is significantly higher than the normal receiver noise we talk about in general sense.

At a conceptual level, there is a difference between analog and digital cancellation. Analog cancellation doesn't really assume anything about how the signal is transformed from input (x) to output (y) because the noise term is anyway random and can't be modeled. Digital cancellation, however does assume a model about how the signal is transformed from the input (x) to output (y)

The nonstructural and strong analog distortion is also the reason why a naive model of subtracting the linear and non-linear component doesn't work for cancelling out self-interference.

So, the authors have tried different strategies to cancel out all the above components. Following two step approach is used-

1. Analog Cancellation – a model/structure free cancellation and doesn't depend on input signal
2. Digital Cancellation – based on the model on the input signal and has two parts
    a. Linear Components
    b. Non-Linear Components

Analog cancellation must be done before the digital cancellation. This is because the receivers will get saturated if the power of input signal is beyond a particular level that is determined by their ADC resolution. So earlier analog cancellation prevents receiver saturation from strong self-interference, allowing usage of commodity radios.

Analog Cancellation

In one of the previous work, the approach of analog cancellation using antenna separation was explored. In the setup, there are two transmitters and one receiver and the distances between them are arranged in such a way that the signals from both the transmit antennae destructively interfere at receiver.

In this paper, the authors carry out analog cancelation without separate antennas. Basically, they recognized that the self-interference signal is some signal x(t) that have been delayed by a certain amount d which is unknown. As shown in the diagram below, they create an analog cancellation circuit consisting of a set of delay lines with fixed delays but programmable attenuations. They set the attenuators by solving an optimization problem online to minimize error between reconstructed and original signal. They tap the TX chain to obtain a small copy of the transmitted signal, pass it through the Analog Cancellation Circuit and the output signal from this circuit is then subtracted from the signal on the receive path to facilitate analog cancellation. The authors achieve at least 60dB of self-interference cancellation using this approach.

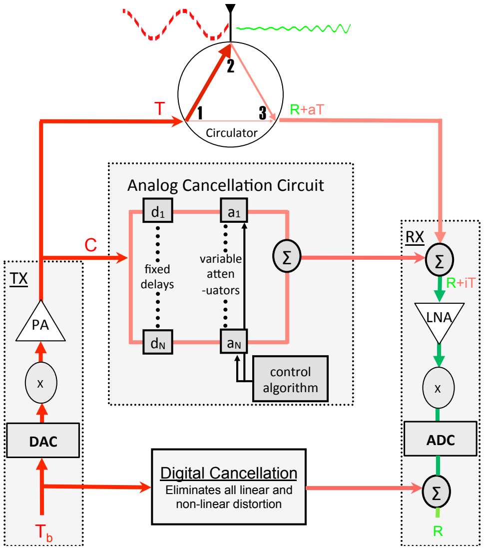


Figure 3: Full duplex radio block diagram. T_b is intended baseband signal we think we are transmitting, but in fact the transmit signal is T (red). The intended receive signal is R (green), however we see strong components of the red signal the RX side. Some of these red signals are undesirably leaked through the circulator. The analog cancellation circuit is trying to recreate a signal that matches the leaked interference signal for cancellation. The digital cancellation stage eliminates any residual self interference.

Digital Cancellation

In the digital cancellation stage, the authors perform separate cancellation for the linear and non-linear components. They model the linear components as a linear and non-causal function of the transmitted signal. And they use a general model to approximate the non-linear function using Taylor series expansion. They succeed in their aim to clean out any residual self-interference (i.e. at least 50dB of linear main signal component and at least 20dB of the non-linear component)

Results:

Below result shows for two different radio platforms, that in order to achieve the required 110dB cancellation, both analog and digital cancellation are required. Pushing exceptionally hard on one or the other is not going to work.

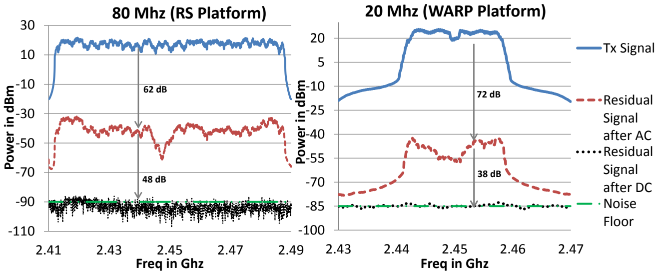


Figure 8: Spectrum Response for our cancellation with the Rohde-Schwarz (RS) radios and the WARP radios. The figure shows the amount of cancellation achieved by different stages of our design. It also shows that our design provides the same 110dB of cancellation even with WARP radios.

Also, it can be seen from the below graph that full duplex provides median gain of 87.

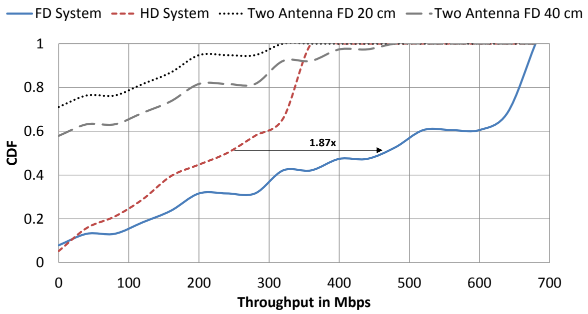


Figure 14: CDF of throughput for full duplex link using TX power = 20 dBm, bandwidth = 80MHz. We see a median gain of 87% using full duplex as compared half duplex. Further, prior full duplex with two antenna's separated by 40cm show gains, only in 8% of cases.

Comparing MegaMIMO and Full Duplex

Both the approaches have their pros and cons. With Full duplex, one doesn't need more antennas, but it can't scale beyond a factor of 2. Smaller changes at physical layer are required for its implementation. MIMO, on the other hand can scale indefinitely, but more antennas are needed which is difficult at receiver side due to size constraints. Multiuser MIMO has taken off because more antenna can be easily fit at transmitter side. Distributed MIMO is still not production ready.

MIMO is applicable in most crowded WiFi scenarios (e.g., malls). While Full duplex is less applicable (a bit rarer to require simultaneous transmit and receive). One application for full duplex is wireless relays.

References

1. Full Duplex Radio, Dinesh Bharadia, Emily Mcmilin, Sachin Katti, Stanford
   https://cs.nyu.edu/~anirudh/CSCI-GA.2620-001/papers/fullduplex.pdf

2. MegaMIMO: Scaling Wireless Capacity with User Demands, Hariharan Rahul, Swarun Kumar, Dina Katabi
   https://cs.nyu.edu/~anirudh/CSCI-GA.2620-001/papers/megamimo.pdf