Coherent and non-coherent combining in distributed millimetre wave beamforming

(Invited Paper)

Shammi Farhana Islam, Alister Burr and David Grace
Dept of Electronic Engineering, University of York, York Y010 5DD, U.K.
{shammi.islam, alister.burr, david.grace}@york.ac.uk

Abstract—This paper considers the application of distributed beamforming to millimeter-wave (mmWave) communications, in the context of 5G wireless communications. We discuss both the potential applications of 5G and the key technologies involved, including mmWave and Massive MIMO, including distributed massive MIMO and 3D beamforming. We point out that in the context of mmWave propagation, which is usually dominated by line-of-sight propagation over relatively short distances, distributed rather than centralized beamforming is more applicable. We then present a numerical evaluation of the performance of distributed beamforming with different numbers of access points carrying $10 \times 10$ arrays of mmWave antenna elements, assuming that the signals combine either non-coherently or coherently. We show that even in the non-coherent case a gain is available that is greater than the increase in transmission power. This gain is very significantly increased in the coherent case. We then discuss practical implementation issues, especially in the context of OFDM systems.

Keywords— Distributed massive MIMO, millimeter-wave, 3D beamforming, coherent, non-coherent

I. INTRODUCTION

Massive MIMO (Multi-Input-Multi-Output) is a key technology in implementing 5G for future wireless communication. Each year innumerable advanced, smart devices are launched in the market to satisfy the public demand. Studies showed that the use of smart devices and the number of connections has increased exponentially over the period [1]. Moreover the use of video has greatly increased user data rates, and this trend will continue with holographic transmission and other advanced applications. The advent of the Internet of Things (IoT) will further massively increase the number of connections. Different approaches have been investigated for future broadband cellular communication systems. Recently massive MIMO techniques have gained considerable attention because of the high spectral efficiencies, energy efficiencies and extended coverage to serve this increasing data traffic. The concept of Massive MIMO is to use an array containing a large number of antenna elements at the base station (BS) which can serve the multi-antenna terminals of many users at the same time and over the same frequency resource [2] [3]. In 2010, Thomas L. Marzetta first introduced this concept of large antenna arrays in his paper which basically proposed multi-user MIMO with large BS antenna arrays. The paper concluded that if the number of antenna elements at the base station is very large compared with the numbers of users being served, simple linear precoding becomes possible on the forward link, and linear combining on the reverse link.

This led to a large body of subsequent research on massive MIMO. The concept is also named large-scale antenna systems (LSAS), very large multiuser MIMO and large-scale MIMO [4] [5]. In massive MIMO, to increase user density rather than reducing the cell size, more antennas are added to the current cell sites thus increasing the network capacity. In addition the transmit powers in uplink and downlink are reduced in large scale antenna systems as the powers are combined coherently and the antenna aperture is reduced [6]. With the transmission of data streams from multiple antennas and utilization of diversity gains offered by channel propagation effects, the data rate can be increased as well as the reliability of data reception. Massive MIMO can be implemented using large antenna arrays, with perhaps hundreds or thousands of elements and these large antenna systems concentrate the energy over a small area/short time which enhances the antenna throughput and the efficiency of radiated energy with dense BS. Massive MIMO thus offers many benefits compared to conventional MIMO such as increased data rate, enhanced reliability, improved energy efficiency and reduced interference [7].

The greatly increased channel number of channel observations provided by the massive array leads to the phenomenon of channel hardening: the small-scale randomness of the channel gain is reduced, and the channel becomes almost deterministic. Further the quasi-orthogonal nature of the channels between each BS and the active users using the same time/frequency resources gives rise to favourable propagation, enabling user signals to be readily separated without complex signal processing. For some assigned numbers of users, with large antenna arrays, the orthogonality is sharpened and simple linear transceivers even with single-user beamforming show performance close to optimal level [8].

The implementation of massive MIMO is however a huge challenge. With the increasing number of antennas, the complexity of the system also rises. To resolve this issue proper coordination between hundreds or thousands of antennas and terminals is needed. Sophisticated channel estimation and synchronisation are also required [9]. Moreover massive MIMO faces other difficulties such as the cost of the equipment and maintenance, reduction of the internal consumption of power, synchronization of terminals, extra degree of freedom offered by large antenna elements etc. [10].

The second key technology for 5G is undoubtedly millimetre-wave (mmWave) transmission, because of the greatly increased bandwidth available in this band. This also
means that the size of an antenna element at these high frequencies is much reduced, and the array element spacing (usually half of the desired wavelength of carrier frequency) can also be reduced, making large-scale antenna arrays feasible within a limited physical size. The increased free-space path loss (FSPL) at these frequencies also requires increased antenna gain in compensation, along with steerable beams, which also calls for large-scale arrays.

In this paper however we consider the application of distributed massive MIMO to mmWave 5G communications, assuming that in place of a single centralized large antenna array, a number of smaller arrays are distributed at access points (APs) across the service area. These then perform cooperative beamforming: they jointly steer coordinated beams in the direction of the user terminal (UT). We consider especially the distinction between coherent and non-coherent combining: whether or not it is possible to control the phase of the signals from different APs such that they combine in phase at the UT. The contribution of the paper is to outline and to provide initial performance analysis for the distributed mmWave beamforming concept for both the coherent and non-coherent cases. It is structured as follows: in the next section the distributed mmWave beamforming concept, while Section III describes our system model by which we will evaluate its performance, and gives numerical results for coherent and non-coherent schemes. Section IV includes discussion of the results and of the practical implementation of the system, including of both coherent and non-coherent combining. It also proposes further work to develop the system and further evaluate it. Finally Section V concludes the paper.

II. DISTRIBUTED MILLIMETRE-WAVE BEAMFORMING

A. Distributed Massive MIMO

In a distributed MIMO system, multiple APs are connected to a central server and the combined system operates as a large distributed multi-antenna access point. Massive MIMO with a collocated antenna array (CAA), improves the user separation (as a result of favourable propagation) but there are practical limits on the number of antenna elements, and the performance suffers because of the characteristics of the propagation environment. While antenna gain is increased, there may still be a severe path loss, especially for cell-edge users. On the other hand it has been shown that distributed antenna arrays (DAAs) exhibit improved user separation capability compared with CAA in the indoor line-of-sight (LoS) environment [11] [12] [13]. In many other cases also distributed massive MIMO shows better performance than co-located massive MIMO, especially because APs are typically nearer to UTs, reducing the path loss.

B. 3-D Beamforming

MIMO wireless communication provides several ways to transmit the information signal, two of which are Spatial Multiplexing and Beamforming (BF). In spatial multiplexing, data signals are transmitted independently over separately encoded streams, which are also known as “layers”, from each of the multiple antennas [14], [15]. On the other hand beamforming is an important radio wave technology/technique incorporating adaptive antenna array systems in which directional antenna beam patterns are produced using multiple antennas to steer the transmitted signal toward a desired user’s location. In this technique, signal strength is increased to improve SINR through coordinating the phases of transmitted signals and reduction of interference [16]. Thus BF uses only one mode unlike spatial multiplexing. BF proved to be a better approach in terms of low complexity but in general spatial multiplexing yields the maximum channel capacity [17]: however note that when the channel is low-rank, which is typically the case on line-of-sight channels, beamforming is in fact optimal [37]. Beamforming becomes effective with larger numbers of transmit antennas (as in massive MIMO) as it reduces interference when there are many users. Progressive works have been done to study different aspects of beamforming. Other terms like precoding, spatial filtering are also used for techniques which are essentially the same [18], [19]. The basic principle is well known [38]: for a uniform linear array of \( M \) elements the signal on the \( j \)-th element is multiplied by a weight:

\[
w_i = \exp \left( j \frac{2\pi i l \sin \theta}{\lambda} \right) \tag{1}\]

where \( l \) is the antenna element spacing, \( \lambda \) is the wavelength, and \( \theta \) is the angle of the beam centre from the broadside direction, as illustrated in Fig 1. The principle is that the signal at each element is phase-shifted so that the signals combine in phase in the required direction, and thus create a beam-pattern with a maximum in that direction.

Fig. 1. Beamforming from uniform linear array

A (horizontal) ULA forms a beam only in the azimuth direction: a two-dimensional (2D) uniform array in the vertical plane can also create and steer the beam in elevation, thus further concentrating the radiation in that direction and increasing the gain. The weights for the 2D array can then be defined as:

\[
w_{ik} = \exp \left( j \frac{2\pi}{\lambda} (il_x \sin \theta + kl_v \sin \phi) \right) \tag{2}\]

where \( \phi \) is the elevation angle, \( k \) is the index of the antenna element in the vertical direction, and \( l_x, l_v \) are the element spacings in the horizontal and vertical directions. This spacing is frequently set to \( \lambda/2 \) to avoid grating lobes in directions other than those intended.

It is easy to see that the gain of an array antenna compared to the ideal isotropic antenna transmitting the same power is simply the total number of elements \( M \) in the array case, the power per element is inversely proportional to \( M \), and hence the amplitude to \( \sqrt{M} \), and hence the total amplitude is proportional to \( M/\sqrt{M} \), which increases as \( \sqrt{M} \). Hence the power is proportional to \( M \).
C. Millimetre-wave spectrum

One of the main goals of 5G wireless communication is the desired higher data rate of the order of 10 Gbps to support ultra-high data rate applications such as ultra-high definition video (UHDTV). The specifications for the 5G frequency range are published in the 3GPP Release 15 and beyond. The ranges for frequency are defined for sub 6 GHz (5G macro optimized), 3-30 GHz (5G E small cells) and 30-100 GHz (5G Ultra Dense) [23]. The frequency bands that fall between 30 GHz to 300 GHz are known as the mmWave band. This is because the wavelength of electro-magnetic waves in this range is in the millimetre (1-10 mm) range. According to [24] this spectrum can be used for high-speed wireless communications with the latest 802.11ad Wi-Fi standard (operating at 60 GHz). Moreover the shortage of spectrum below 6 GHz has obliged mobile operators to explore mmWave as mobile frequency spectrum due to its much wider bandwidth. However it is important to consider the advantages and disadvantages of the mmWave frequency band before implementation of these systems.

The benefits, as described above, primarily arise from the substantial bandwidth available: of course the total bandwidth available between 30 and 300 GHz is nine times the total bandwidth available in all bands below 30 GHz. Secondly the wavelength is much smaller than at sub-6GHz frequencies, and hence the physical size of the antenna for given gain is much reduced.

The disadvantages are primarily related to propagation losses, as touched upon above. Unlike lower frequencies mmWave suffers from poor scattering and attenuation due to rain, fog/cloud and gasses such as oxygen and water vapour, and penetration loss is high as signals suffer severe attenuation through walls of buildings and other blockages. [25] [26] [27] [31]. For example, in an indoor environment, a 1.9 cm thick whiteboard attenuates a 60 GHz signal by 9.6 dB [29] [30]. mmWaves are also attenuated by rain droplets since the wavelengths of mmWave signals range between 1 mm and 10 mm which is of the same order as the size of a raindrop, hence causing scattering. For example light and heavy rain-rates can induce an attenuation of up to 2.55 dB/km and 20 dB/km, respectively [25] [29].

Because of the reduced wavelength, Doppler spread is more significant at mmWave frequencies, causing increased time variation of the channel and posing additional problems for channel estimation. In addition for the same reason diffraction losses are much larger, implying that diffracted signals are negligible compared to line of sight or reflected signals.

Most fundamentally however the free space path loss increases with frequency according to the Friis equation:

\[ P_r = P_t G_r G_t \left( \frac{\lambda}{4\pi d} \right)^2 \]  

where \( P_r, P_t \) are the received and transmitted powers, \( G_r, G_t \) are the receiver and transmitter antenna gains, \( d \) is the path length, and \( \lambda = \frac{c}{f_c} \) is the wavelength, with \( f_c \) the carrier frequency and \( c \) the speed of light.

Hence the path loss increases as the square of carrier frequency, or by 20 dB/decade of frequency increase. It is worth noting, however, that for given effective aperture \( A_{eff} \) (which is proportional to the physical area), the antenna gain also increases with frequency, since:

\[ G_{t,r} = \frac{4 \pi r}{\lambda^2} \]

It is again easy to see that in a 2D array antenna with \( \frac{\lambda}{2} \) element spacing the number of elements, and hence the gain, is inversely proportional to \( \lambda^2 \), for given array area. Thus in our proposed system it is possible to compensate for the increased FSPL by increasing the AP antenna gain: the disadvantage is that this increases the complexity of the AP antenna.

D. Distributed millimetre-wave beamforming

These propagation characteristics of mmWave signals mean that longer paths tend to be heavily attenuated, both due to the atmospheric and rain losses on the line of sight, and because in urban areas at least, the line of sight is very likely to be blocked over longer paths. Hence we propose distributing the antenna array over the service area in multiple smaller arrays at multiple APs, which can be placed much nearer the UTs. However we allow these arrays to cooperate in serving a UT, so that the ensemble of AP arrays may (in principle) form a single large array, thus increasing the gain in proportion to the number of cooperating arrays.

This however assumes that the signals from different arrays can be phase-controlled to the same accuracy as if they were collocated, so that their signals combine coherently at the UT. In this case the received signal amplitude is the sum of the amplitudes of the signals from each AP array, and the power is the square of this. Since this may not be feasible in practice, we also consider the case of non-coherent combining, where the received signal power is simply the sum of the powers from each of the AP arrays, which does not assume coherent combining.

III. SYSTEM MODEL

We next evaluate the potential performance of the distributed millimetre-wave beamforming system by defining a simulation model for the system (see Fig. 2). In this model we consider a 50x50m area within which antenna arrays at APs are randomly distributed, pointing in random directions, with a UT at the centre in free space. Each AP transmits 1 mW in a bandwidth of 500 MHz, and the receiver has a noise figure of 10 dB.

In 3D beamforming the beam is steerable in both azimuth and elevation angle. The transmitting antenna arrays are 1.5 metres in height and the UE is at 1 metre so the elevation is almost constant in every trial. However we restricted the azimuth angle to within +90° to -90°, since if the antenna arrays are not facing the UE then their contributions will be negligible. In this proposed model we worked with three, five and ten distributed antenna arrays (10x10 elements each) respectively and observed the difference in performance.

Given the size of the simulation area we assume that a line-of-sight is available between all APs and the UT, and assume that propagation follows the inverse square law, as given by the Friis equation. Again given the short link lengths we neglect atmospheric and other losses.
In the case of coherent combining we calculate the signal amplitude at the UT from each AP, add and square it to obtain the power. For non-coherent combining we simply add the powers from each AP.

This raises the question of how the proposed system might be implemented in practice. We begin with the non-coherent case, bearing in mind that a broadband communication system is likely to use OFDM, which would also fit with the 5G New Radio standard. Consider one subcarrier of the OFDM multiplex, transmitted at a frequency $f$ simultaneously from $M_{AP}$ access points, each having path gain and relative time difference $h_i, \tau_i, i = 1 \ldots M_{AP}$. Then (neglecting noise) the received signal has complex amplitude:

$$y = \sum_{i=1}^{M_{AP}} h_i \exp(-j2\pi f \tau_i)$$

(5)

The power of the received signal, averaged over the bandwidth of the multiplex, is then:

$$|y|^2 = \sum_{i=1}^{M_{AP}} \sum_{i' \neq i} h_i^* h_{i'} \exp(-j2\pi f (\tau_i - \tau_{i'})) = \sum_{i=1}^{M_{AP}} |h_i|^2$$

(6)

Note that $\exp(-j2\pi f (\tau_i - \tau_{i'}))$ is a sinusoid in the frequency domain with period $1/(2\pi (\tau_i - \tau_{i'}))$. Hence the average over the multiplex is zero provided the bandwidth is sufficient to contain a (preferably large) integer number of cycles for the minimum value of $(\tau_i - \tau_{i'})$, $i' \neq i$. Then the average power over the multiplex is given by the sum of the squared magnitudes of the channel gains, as given in Fig. 3. Note however that in order to achieve a BER performance corresponding to the predicted average SNR we must combine the signals from multiple subcarriers across the multiplex: otherwise the BER will tend to be dominated by the performance of the poorest subcarrier, at which the signals from several APs cancel one another out causing severe fading. This however can readily be achieved using a forward error correction code with Hamming distance sufficient to enable a frequency diversity gain, along with interleaving sufficient to ensure that adjacent code symbol are subject to uncorrelated fading.

Coherent transmission, on the other hand, requires that the transmitting APs have knowledge at least of their own channel to the receiver. They can then adjust the phase and/or timing of their transmissions in each subcarrier, in a form of precoding, to ensure that the signals in each subcarrier arrive in phase. In principle this requires that:

- a) The channel from each AP is estimated at the UT, and the resulting channel state information (CSI) is signalled back to each AP; and

- b) Both the channel and the phase/frequency of the oscillators at the APs are stable enough that this CSI remains sufficiently accurate for the period between the estimation of the channel and the transmission time.
The required CSI might involve the amplitude/phase response in each subcarrier of the multiplex, to some degree of precision; however there are factors that mitigate this requirement. We note that it is primarily the phase information that determines whether signals combine coherently, and also that the accuracy does not need to be very high to enable most of the gains of coherent transmission: probably an accuracy of 45° is sufficient, requiring only 3 bits. Also, assuming that only the line-of-sight component is significant, it is sufficient to measure the time delay of the signal from each AP, which will require less information to be fed back than the phase of every subcarrier.

The channel stability is influenced by the wavelength of the signal, which in the mmWave band is less than 1 cm. This implies that movements of a few mm, such as human users are likely to make even when stationary, are sufficient to affect the coherent combining. Moreover oscillator technology in this frequency range is subject to larger phase noise, which also affects the stability of the phase of the combining signals. Further work is likely to be required here: but with the more widespread introduction of mmWave devices in 5G it is likely that technology improvements will continue to occur.

There are some potential approaches that may simplify the process, at the cost of some performance reduction. For example, rather than adjusting phases at the APs, the UT may signal which subcarriers receive reduced powers due to signal cancellation, so that the APs do not transmit in these subcarriers. This means that only subcarriers that combine coherently are used. This requires only one bit of CSI to be transmitted on the reverse channel. The network may also track location of UTs and use this information to predict channel delays, and also to steer beams.

V. CONCLUSIONS AND FURTHER WORK

We have considered the application of distributed beamforming to mmWave communications, and in particular have compared coherent and non-coherent combining. We have shown that significant gains are available in both cases, though they are significantly larger in the case of coherent transmission. Even with non-coherent transmission gains occur in comparison with centralised beamforming not simply due to the increased power that may be available from multiple access points, but also because the expected distance between user terminals and access points is reduced. With coherent combining the gains available are significantly larger: of the order of an additional 10 dB with up to 10 cooperating APs. We also discuss how these schemes might be implemented in practice, with particular reference to OFDM transmission: while non-coherent implementation is likely to be straightforward, coherent systems will face significant challenges, due in particular to the stability of the channel and the local oscillators in APs and UTs, but also to the requirement for feedback of channel state information.

Further work is thus required to develop schemes that will be feasible in practice, including simplified schemes that reduce the amount of CSI feedback required, and also the evaluation of the stability of the channel and the oscillators in practical applications. Channel prediction methods may assist in overcoming channel variations and reducing the CSI requirements.

REFERENCES

[5] https://ma-mimo.ellintech.se
[9] https://www.cambiumnetworks.com/blog/what-is-massive-mimo/
[10] https://enterpriseiotinsights.com/20160805/5g/massive-mimo-3g-tag31-tag99
[23] https://www.cambiumnetworks.com/blog/what-is-massive-mimo/
[24] https://enterpriseiotinsights.com/20160805/5g/massive-mimo-3g-tag31-tag99


