Large-Scale Antenna Systems with Hybrid Analog and Digital Beamforming for Millimeter Wave 5G

Shuangfeng Han, Chih-Lin I, Zhikun Xu, and Corbett Rowell

ABSTRACT

With the severe spectrum shortage in conventional cellular bands, large-scale antenna systems in the mmWave bands can potentially help to meet the anticipated demands of mobile traffic in the 5G era. There are many challenging issues, however, regarding the implementation of digital beamforming in large-scale antenna systems: complexity, energy consumption, and cost. In a practical large-scale antenna deployment, hybrid analog and digital beamforming structures can be important alternative choices. In this article, optimal designs of hybrid beamforming structures are investigated, with the focus on an N (the number of transceivers) by M (the number of active antennas per transceiver) hybrid beamforming structure. Optimal analog and digital beamforming designs in a multi-user beamforming scenario are discussed. Also, the energy efficiency and spectrum efficiency of the $N \times$ M beamforming structure are analyzed, including their relationship at the green point (i.e., the point with the highest energy efficiency) on the energy efficiency-spectrum efficiency curve, the impact of N on the energy efficiency performance at a given spectrum efficiency value, and the impact of N on the green point energy efficiency. These results can be conveniently utilized to guide practical LSAS design for optimal energy/spectrum efficiency trade-off. Finally, a reference signal design for the hybrid beamform structure is presented, which achieves better channel estimation performance than the method solely based on analog beamforming. It is expected that large-scale antenna systems with hybrid beamforming structures in the mmWave band can play an important role in 5G.

INTRODUCTION

Studies on the fifth generation (5G) wireless communication system are gaining more momentum worldwide in an attempt to provide solutions for the exponential increase of mobile data traffic by 2020. Although 5G is in its embryonic stage, some trends have appeared in how to design 5G networks, such as to make it green and soft, as proposed by China Mobile [1]. Multiple research topics have been identified as 5G candidates, including large-scale antenna systems (LSAS), non-orthogonal multiplex access, full duplex, spectrum sharing, high-frequency bands (e.g., millimeter-wave, mmWave), high-density networks, new network architecture, new waveform design, and so on. These technologies may have great potential in system performance improvement in 5G.

The fundamental premise of LSAS [2] is that the number of base station antennas is much larger than the number of single antenna terminals. Theoretically, LSAS with full digital beamforming (BF) can yield the optimal performance, significantly increasing the system energy efficiency (EE) and spectrum efficiency (SE) via multiuser BF [3]. However, implementing LSAS in lower frequency bands is difficult since the much larger physical footprints of LSAS base stations will not only bring significant tower construction challenges, but also lead to increasing concerns about possible health effects. Higher frequency bands like mmWave [4] are appealing for LSAS, since the physical array size of LSAS can be greatly reduced due to the decrease in wavelength. Another differentiating feature that makes mmWave especially attractive in 5G is that large chunks of underutilized spectrum are available, which can be potentially utilized to provide significant system capacity improvements. Recently, mmWave outdoor channel measurement results [5], advances in hardware design [6], a successful outdoor trial [7], and the availability of abundant spectrum in mmWave have encouraged the wireless industry to consider mmWave for cellular systems [8].

When a large number of antennas are implemented to achieve better BF gains, however, implementing the same number of transceivers may not be feasible due to excessive demand on real-time signal processing for high BF gains [11], high power consumption, and high cost (especially the high cost and power consumption of mixed-signal devices in mmWave systems). A BF structure with a much lower number of digital transceivers than the total antenna number will therefore be more practical and cost effective to deploy. One interesting approach to reducing the transceiver number is via analog BF [6, 7], where each transceiver is connected with multiple active antennas, and the signal

The authors are with China Mobile Research Institute.

0163-6804/15/$25.00 © 2015 IEEE

phase on each antenna is controlled via a network of analog phase shifters. Generally, each transceiver generates one beam toward one user in analog BF. When the user number simultaneously served is much smaller than the antenna number (this is generally true in an LSAS system), the transceiver number can be designed to be much smaller than the antenna number. However, there may be severe inter-user interference when users are not adequately spatially separated. Digital BF over transceivers can then be utilized to achieve multiple data stream precoding on top of analog BF [9–12] to further enhance the performance.

Two hybrid BF structures that have drawn much attention of researchers are shown in Fig. 1, with N being the transceiver number and NM being the antenna number. In structure 1 each transceiver is connected with all antennas, such that the transmitted signal on each of the N digital transceivers goes through NM RF paths (mixer, power amplifier, phase shifter, etc.) and summed up before being connected with each antenna element [9, 10], as shown in Fig. 1a. Analog BF is performed over NM RF paths per transceiver, and digital BF can then be performed over N transceivers. This structure is a natural combination of analog BF and digital BF, and achieves full BF gain for each transceiver. However, the complexity of this structure is rather high; for example, the total number of RF paths is N^2M.

An $N \times M$ hybrid BF structure is shown in Fig. 1b, where each of the N transceivers is connected to M antennas. Analog BF is performed over only M RF paths in each transceiver, and digital BF is performed over N transceivers [8, 11]. This structure is more practical for base station antenna deployment in the current cellular systems, where each transceiver is generally connected to a column of antennas. Compared to structure 1, the BF gain per transceiver is $1/N$ the gain in structure 1, but with much reduced complexity, the total number of RF paths being NM.

Recently, there has been growing interest in hybrid BF structure design for mmWave communication. For structure 1, a simple precoding solution was proposed assuming only partial channel knowledge at the base station and mobile station in the form of angle of arrival (AoA) and angle of departure (AoD) knowledge [9]. The spatial structure of mmWave channels was further exploited in [10] to formulate the single-user precoding/combining problem as a sparse reconstruction problem. AoA estimation and beamforming algorithms were proposed in [11] for structure 2. A successful mmWave outdoor trial was carried out in Korea, in which the hybrid BF of structure 2 was implemented [7]. There are still many challenging issues regarding hybrid BF structures:

- What is the performance gap compared to digital BF?
- Does a larger transceiver number N always lead to better EE, and what is the EE-SE optimal design?
- How can we design efficient reference signals (RSs) for better availability of the channel state information at the transmitter side?

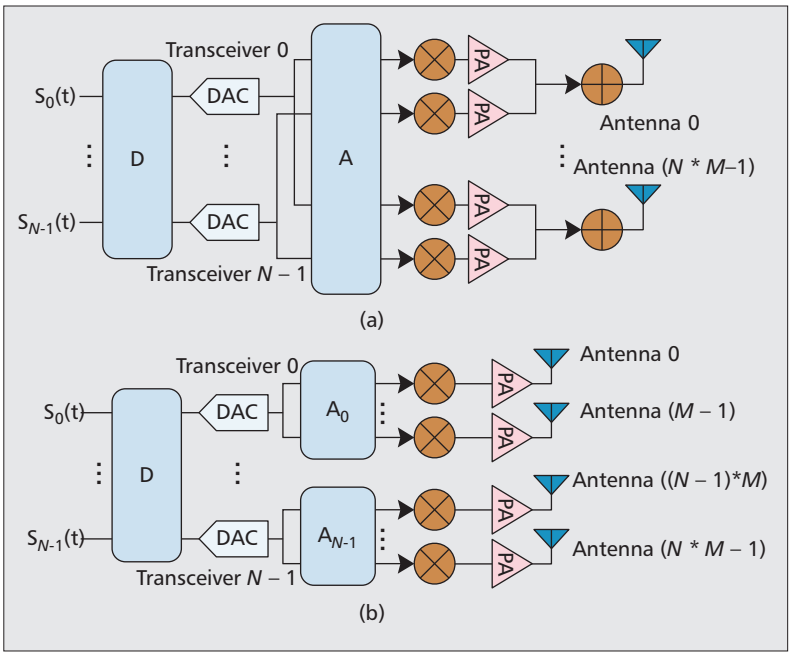

Figure 1. Hybrid BF structures.

This article addresses the above important issues on how LSAS with hybrid BF structure can potentially be utilized in mmWave systems with a focus on the N by M hybrid BF structure. The optimal analog BF and digital BF design are investigated for a multi-user BF scenario. The EE-SE relationship of the $N \times M$ hybrid BF structure is analyzed, paving a path to an EE-SE optimized design. A downlink RS design is discussed. Beam domain RSs based on hybrid BF are presented, which outperform RSs based on analog BF. This article is finally summarized.

Hybrid Analog BF and Digital BF Design

This section studies the hybrid BF strategies for the two structures shown in Fig. 1, especially for the case of a sufficiently large antenna number. Suppose there are N downlink users, each having a single antenna. Denote $\mathbf{H}$ as the downlink channel with the size of $N \times NM$, $\mathbf{A}$ as the analog BF matrix with the size of $NM \times N$, $\mathbf{D}$ as the digital BF matrix with the size of $N \times N$, $\mathbf{s}$ as the N user data vector with the size of $N \times 1$, and $\mathbf{n}$ as the noise vector, respectively. The received signal in the downlink is expressed as $\mathbf{y} = \mathbf{HADs} + \mathbf{n}$.

Design for Structure 1

As shown in Fig. 1a, the digital signal from each transceiver can be delivered to any antenna after analog weighting. The ith column of $\mathbf{A}$ is thus the analog BF vector for all NM antennas on the ith transceiver, $i = 0, \ldots, N-1$. From matrix theory, it can be derived that the channel capacity can be achieved when the hybrid BF transmits signals over the largest N eigenvectors of $\mathbf{H}$. The power allocation can be carried out following the water-filling method. A good reference on joint $\mathbf{D}$ and $\mathbf{A}$ design can be found in [10].

The analytical goal of the first case is to find the optimal number of transceivers and, correspondingly, the optimal number of antennas per transceiver given the total number of antennas, while the investigation in the second case is to explore how the independent N *and* M *should be jointly optimized.*

Note that the upper bound on the channel capacity with effective channel **HA** can be achieved when **HA** has N equal singular values [2]. When M is large enough, **HA** turns out to be a diagonal matrix with equal diagonal elements if we take **A** to be exactly the normalized conjugate transpose of **H**. By doing so, there is negligible inter-user interference, and hence the upper bound of N-user capacity can be achieved. Correspondingly, the optimal **D** is an identity matrix, indicating that there is actually no need to implement digital BF across the transceivers. Although the optimal N user channel capacity can be achieved under this structure, the analog weighting should be done N^2M times in total, which brings high implementation complexity.

Design for Structure 2

For the $N \times M$ hybrid BF structure in Fig. 1b, the digital signal from each transceiver can only be delivered to M antennas. Therefore, analog BF matrix **A** has a different structure from that in structure 1, $\mathbf{A} = diag[\mathbf{A}_0, \ldots, \mathbf{A}_{N-1}]$, where $\mathbf{A}_i$, $i = 0, \ldots, N-1$ is the $M \times 1$ analog precoder on the ith transceiver. Suppose **H** and **A** are known; the digital BF matrix **D** can then be designed with traditional multiple-input multiple-output (MIMO) theories (e.g., to maximize the sum rate). To the best of the authors' knowledge, the joint optimal design of **D** and **A** is still an open issue.

When M is large enough, by partitioning the ith row of **H** (denoted by $\mathbf{h}_i$) into N consecutive parts with equal number of elements, and designing $\mathbf{A}_i$ to be exactly the normalized conjugate transpose of the ith part of $\mathbf{h}_i$, again **HA** becomes exactly a diagonal matrix with equal diagonal elements. The capacity upper bound can then be achieved. Correspondingly, the optimal **D** is also the identity matrix. Note that the diagonal elements of **HA** are actually the BF gains, which is N times higher in structure 1 than that in structure 2. This leads to roughly $Nlog_2N$ b/s/Hz capacity loss in structure 2.

EE-SE Analysis of the $N \times M$ BF Structure

From earlier, when the user number equals the transceiver number, structure 1 can achieve the same channel capacity as that achieved by the traditional digital BF structure; whereas the capacity achieved by structure 2 is lower than that achieved in structure 1. For both structures, the required number of transceivers is much smaller than that in the digital BF structure. This section continues to investigate the optimal transceiver number N in terms of EE-SE co-design [1]. The analysis is focused on structure 2 (i.e., the $N \times M$ BF structure) since its complexity is much lower than that of structure 1, and the capacity loss is not significant with a practical N.

EE-SE Relationship

Consider the $N \times M$ BF structure: perfect analog BF is assumed within M antennas per transceiver for one user (in total there are N users). There is zero inter-user interference with the assumption of a large enough M or via proper user scheduling even when M is not so large. For example, the beamwidth of the analog beam generated by a 50-element linear antenna array with half wavelength spacing can be as small as 2°. It would not be difficult to schedule, say, $N = 10$ users with negligible inter-user interference in one cell. The N user sum capacity of this structure is derived as $C = WNlog_2(1 + MP\eta_{SE}/WN_0)$, where W is the system bandwidth, P is the total power of M power amplifiers (PAs) per transceiver, η_{PA} is the PA efficiency, and N_0 is the thermal noise density. Channel gain is assumed to be 1. The SE of this structure is written as $\eta_{SE} = C/W$.

An accurate power model is needed to calculate the EE. However, this is not straightforward, since base stations have different types (macro, pico, femto) and are generally produced by different vendors with various implementation technologies. In this section, the following simple power model is used — $P_{total} = NP + P_{static} = NP + NP_0 + P_{common} + NMP_{rf\_circuit}$ — where P_{total} is the total power, NP is the RF power of total N transceivers, P_{static} is the static power of the base station, including the part of power NP_0 that scales with the number of transceivers N, P_{common}, which is common for any transceiver number, and $NMP_{rf\_circuit}$, which scales with total antenna number NM. The EE-SE relationship can be written as

$$\eta_{EE} = C / P_{total} = \frac{\eta_{SE}}{\left(2^{\frac{\eta_{SE}}{N}} - 1\right)\frac{N_0}{\eta_{PA}}\frac{N}{M} + \frac{NMP_{rf\_circuit}}{W}} \quad (1)$$

EE-SE Relationship at the Green Points

As shown in [13], the EE-SE relationship based on Shannon's theory is monotonic, where a higher SE will always lead to a lower EE. When the circuit power is considered, however, there is a green point on the EE-SE curve where the maximum EE, η^*_{EE}, is achieved. Two cases are discussed here for the $N \times M$ hybrid BF structure, the case when $NM = L$ (i.e. the total antenna number is fixed to be L, but N and M are variable), and the case when N and M are independent. The analytical goal of the first case is to find the optimal number of transceivers and, correspondingly, the optimal number of antennas per transceiver given the total number of antennas, while the investigation on the second case is to explore how the independent N and M should be jointly optimized.

It can be derived [12] based on Eq. 1 that there is only one green point on the EE-SE curve for each case (i.e., there is only one η^*_{SE} that maximizes the EE performance). The relationship between η^*_{EE} and η^*_{SE} is further given as $lg(\eta^*_{EE}) = -\eta^*_{SE}\, lg2/N + lg(M\eta_{PA}/N_0 ln2)$; that is, $lg(\eta^*_{EE})$ scales with η^*_{SE} linearly with a slope of $-lg2/N$. Similar to the EE-SE relationship with classic Shannon theory, a higher η^*_{SE} will always lead to a lower η^*_{EE}. Interestingly, the relationship between η^*_{EE} and η^*_{SE} is independent of P_0, P_{common}, $P_{rf\_circuit}$, and W, although, as can be seen from Eq. 1, η^*_{SE} and η^*_{EE} are determined

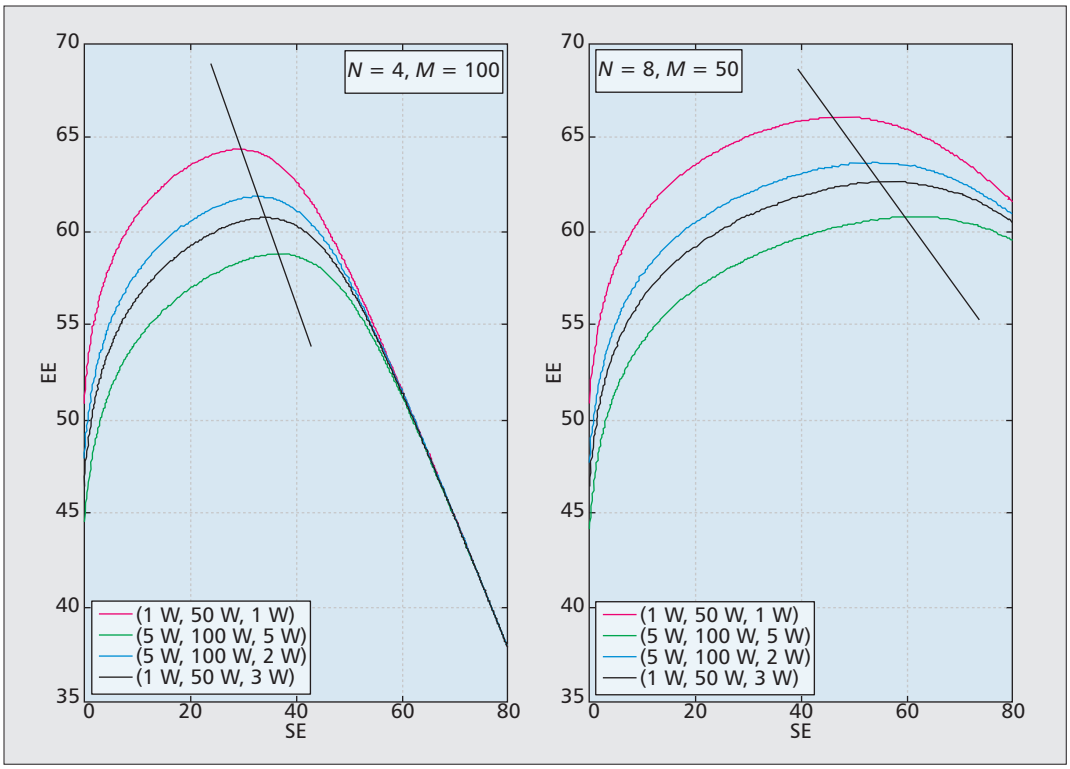

Figure 2. EE-SE relationship at the green points.

based on all the other parameters. The EE-SE relationship was also investigated in [14] for MIMO systems, which also showed a similar trend.

Assume $W = 2 \times 10^7$ Hz, $N_0 = 10^{-17}$ dBm/Hz, and a channel gain of –100 dB. The EE-SE relationship is depicted for two scenarios, the scenario with $N = 4$ and $M = 100$, and the scenario with $N = 8$ and $M = 50$. For each scenario, four ($P_{rf\_circuit}$, P_{common}, P_0) combinations are simulated, including (1 W, 50 W, 1 W), (5 W, 100 W, 5 W), (5 W, 100 W, 2 W), and (1 W, 50 W, 3 W). EE (bits per Joule) is depicted in log-scale, and SE (bits per second per Hertz) is depicted in linear scale. As shown in Fig. 2, the EE-SE curve is different with different ($P_{rf\_circuit}$, P_{common}, P_0) combinations. On each curve, there is only one green point. Also, the green points are actually in a straight line for both scenarios, with a large slope in the first scenario. This indicates that in the EE optimal design, a smaller transceiver number N brings more EE performance improvement with a given SE reduction.

With the above analysis, therefore, it is expected that the system operates at the green point. Also, it is important that η_{SE}^* satisfies the system SE requirement. Besides, η_{EE}^* should be high enough. These require careful design of the following parameters: P_0, P_{common}, $P_{rf\_circuit}$, W, η_{PA}, N, and M. For example, when other parameters are given, N can be designed to maximize η_{EE}^*. The optimal N in terms of EE-SE co-design is discussed in the following.

How Does *N* Affect EE-SE?

When the required SE is predetermined, it is desirable that the transceiver number N is optimized, yielding the highest EE performance with the minimum transceiver number. Based on Eq. 1, it is found that in the cases $NM = L$ and independent N and M, for any given SE, there is only one optimal N to yield the best EE. A detailed proof can be found in [12]. The practical meaning of the existence of the optimal N is that with a given SE, a system designer does not need to implement too many transceivers to achieve the best EE performance.

Assume $P_{rf\_circuit} = 1$ W, $P_{common} = 50$ W, $P_0 = 1$ W, $\eta_{PA} = 0.375$, $W = 2 \times 10^7$ Hz, $N_0 = 10^{-17}$ dBm/Hz, and a channel gain of –100 dB. Considering the $NM = 500$ case, the impact of N (from 1 to 10) on EE performance is shown in subplot 1 of Fig. 3, where five SE values are simulated. Note that since N and M are integers, there are only five valid (N, M) combinations for $NM = 500$, that is, (1,500), (2,250), (4,125), (5,100), and (10,50). It can be observed that on each curve there is one optimal N that yields the highest EE. For example, when SE is 20 b/s/Hz, the optimal N is 4. When SE is 8 b/s/Hz, the optimal N is just 1. The case when N is larger than 10 is not shown in the figure, since it may be difficult to schedule users with negligible inter-user interference when M is very small.

When N and M are independent, the impact of N on EE performance is shown in subplot 2 in Fig. 3, where $M = 50$, and other parameters are the same as those in subplot 1. Similar to the fixed NM case, on each curve there is one optimal N. For example, when SE is 40 b/s/Hz, the optimal N is 6. Different from the fixed NM case, the EE performance is very sensitive to N, because the total antenna number scales with N.

In a practical system operation, the SE requirement may vary according to the traffic load and service types. For example, as shown in

This indicates that in the EE optimal design, a smaller transceiver number N brings more EE performance improvement with a given SE reduction.

The intuition is that given one required SE value, we need to optimize the system parameters such that the EE achieved via the optimal transceiver number N is exactly at the green point of its EE-SE curve.

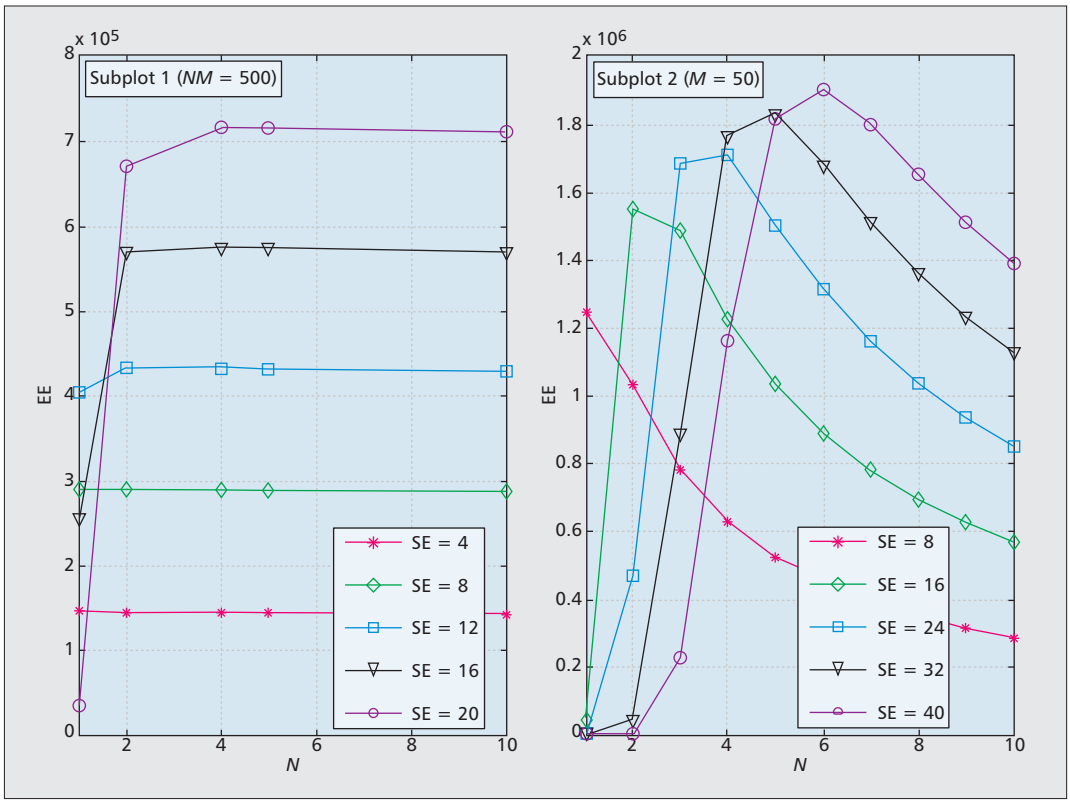

Figure 3. N vs. EE with different SE values.

subplot 2 in Fig. 3, with the maximum 40 b/s/Hz SE requirement, the optimum N should be designed to be 6. But when the SE requirement is reduced to 8 b/s/Hz, the optimal N should be 1. Therefore, it is important that for the possible SE range, the system can be designed with the largest optimal N, and selects the best N according to the SE requirement via transceiver on/off. This can help to further enhance the EE performance according to the system traffic load.

Relationship between the Green Point EE and N

Given the transceiver number N, there is only one green point on EE-SE curve. We have shown that for any SE, there is one optimal N. But it is still unclear whether $\eta_{EE}^{*}(N)$ at the green point is the highest EE performance of all N. Based on Eq. 1, it can be derived that when N and M are independent, a larger N always results in a higher η_{EE}^{*}. When $NM = L$, however, the impact of N on $\eta_{EE}^{*}(N)$ is determined based on the parameters in Eq. 1 [12].

One example is shown to illustrate how N impacts $\eta_{EE}^{*}(N)$. Assume $P_{rf\_circuit} = 1$ W, $P_{common} = 50$ W, $P_0 = 1$ W, $\eta_{PA} = 0.375$, $W = 2 \times 108$ Hz, $N_0 = 10^{-17}$ dBm/Hz, and a channel gain of –100 dB. Note that a much larger bandwidth is considered here for a smaller optimal N in the simulation. The EE-SE curves with different N are shown in subplot 1 of Fig. 4 for $NM = 800$, with N being 1, 2, 4, 5, 8, 10, and 16. It can be found that as N increases from 1 to 8, $\eta_{EE}^{*}(N)$ increases, but when N increases from 8 to 16, $\eta_{EE}^{*}(N)$ decreases. The optimal transceiver number is therefore 8. Transceiver numbers larger than 16 are not shown in this figure. For the independent N and M case, the impact of N (from 1 to 10) on $\eta_{EE}^{*}(N)$ is shown in subplot 2, with $M = 50$. As N increases, the green point EE also increases monotonically. The intuition is that given one required SE value, we need to optimize the system parameters such that the EE achieved via the optimal transceiver number N is exactly at the green point of its EE-SE curve.

Reference Signal Design

Based on the earlier analysis, hybrid BF structures can achieve optimal channel capacity with a much smaller number of transceivers N optimized in terms of joint EE-SE co-design. This section discusses possible reference signal design, which determines to what extent the LSAS with hybrid BF structure helps to enhance system performance.

Generally, there are two methods to obtain downlink CSI: via uplink sounding in a time-division duplex (TDD) system or downlink RSs. The first method is not trivial in LSAS, because the downlink/uplink channel reciprocity calibration using traditional methods at a base station may not function well due to high complexity of the calibration circuits. In addition, the $N \times M$ structure makes it difficult to accurately measure the CSI per antenna. The second method has issues as well, since the antennas connected to each transceiver are actually one logical antenna, and how the reference signals should be transmitted is still not well understood. Therefore, one prac-

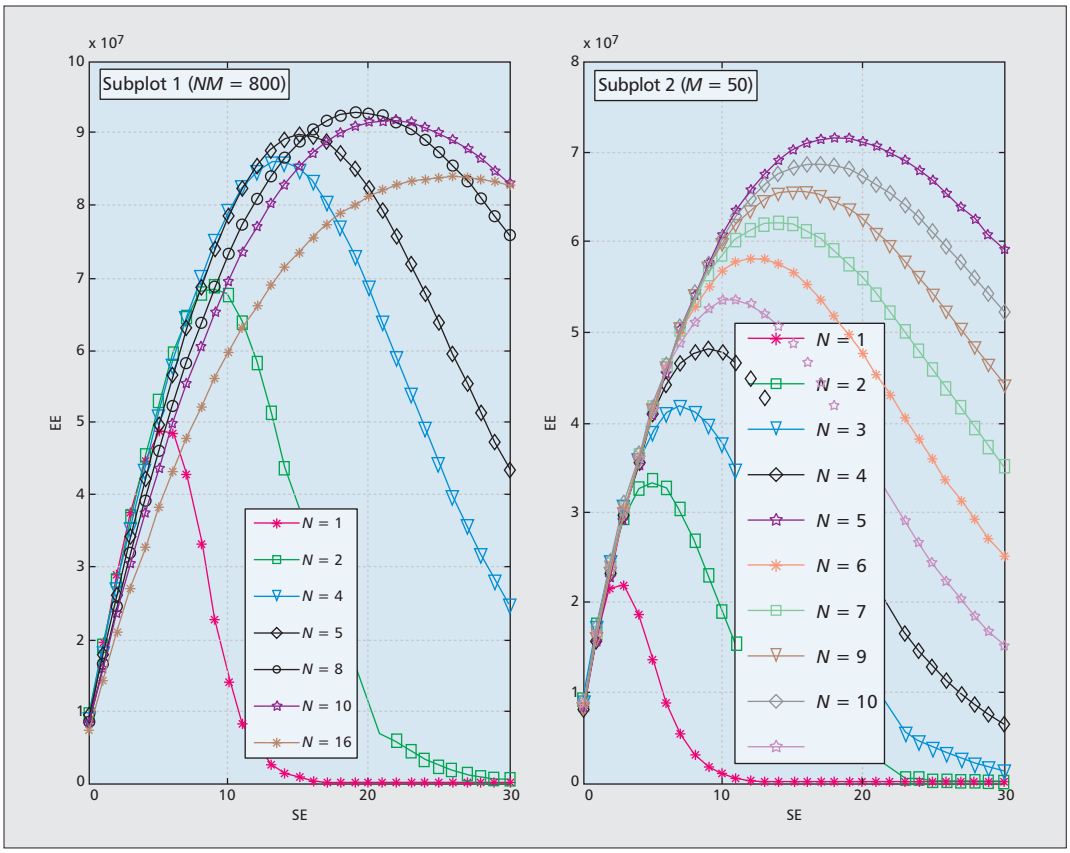

Figure 4. EE-SE curves with different N.

The accuracy of this design depends on the RSs coverage and the transceiver number N. With a given transceiver number N, it is difficult to improve AoD estimation accuracy. In the next subsection, one beam domain RSs design with hybrid BF is presented.

tical assumption on channel information availability in LSAS with hybrid BF structure is that partial channel information like AoA and AoD are known, as used in [9]. Two beam domain downlink RS designs for the AoD estimation, RSs with analog BF and RSs with hybrid BF, are discussed below.

Beam Domain RSs with Analog BF

Since per antenna channel state information (CSI) is difficult to measure accurately, beam domain RS is more practical for this $N \times M$ BF structure, especially in line of sight (LOS) environments. One straightforward design is that each transceiver transmits one predefined beam via analog BF over M antennas, such that there are N simultaneous beams with different main beam directions (i.e., AoD in the downlink). Mobile users then feed back the index of the received beam with the highest signal power, and the base station will transmit to each user with the corresponding beam. The accuracy of this design depends on the RSs' coverage and transceiver number N. With a given transceiver number N, it is difficult to improve AoD estimation accuracy. In the next subsection, one beam domain RSs design with hybrid BF is presented.

Beam Domain RSs with Hybrid BF

Consider the $N \times M$ structure with a linear antenna array in an orthogonal frequency-division multiplex (OFDM) system. As shown in Fig. 5a, on the kth subcarrier the reference signal s_k first passes through the digital BF, and then, after the digital-to-analog conversion, passes through the analog BF. The same analog BF with a main beam direction ϕ_0 is applied to all transceivers, while the digital BF matrix $\mathbf{D}_k$ is diagonal with the ith ($i = 1, \ldots, N-1$) diagonal element being $\exp(j\alpha_i)$, where α_i is the phase shift on the ith transceiver. By choosing proper phase shift on each transceiver, the sum array factor (AF) of all NM antennas in a direction ϕ around ϕ_0 can be maximized, with the corresponding maximum value being N times that of a single transceiver analog BF. More details on hybrid BF design can be found in [15] for both the linear array and planar array cases. Therefore, for a given analog BF, the main beam direction of the hybrid BF is determined by the digital BF weights. This leads to the following beam domain RS design based on hybrid BF.

As shown in Fig. 5b, the RSs occupy $2K + 1$ consecutive subcarriers, which are assumed to be well within the channel's coherent bandwidth. In the sth OFDM symbol, the main beam direction of the analog BF is set to be $\phi_{0,s}$ for all transceivers. While the main beam direction of the hybrid analog and digital BF on the kth ($k = 0, 1, \ldots 2K$) subcarrier is $\phi_{0,s} + (K - k)\,\Delta$, with Δ being the beam spacing. The digital BF weight on the ith transceiver on the kth subcarrier in the sth OFDM symbol can readily be calculated accordingly. With the feedback of the subcarrier index on which the signal power is the largest, the base station knows the AoD information, such as the kth subcarrier corresponds to an AoD of $\phi_{0,s} + (K - k)\,\Delta$. This is because within the coherent bandwidth the channel response is the same for all subcarriers, and therefore, the

The design based on hybrid BF is featured by same analog BF on each transceiver, on top of which digital BF is designed to maximize the gain in a certain direction around the main beam direction of the analog BF. Much better AoD estimation performance can be achieved compared to beam domain RSs with analog BF.

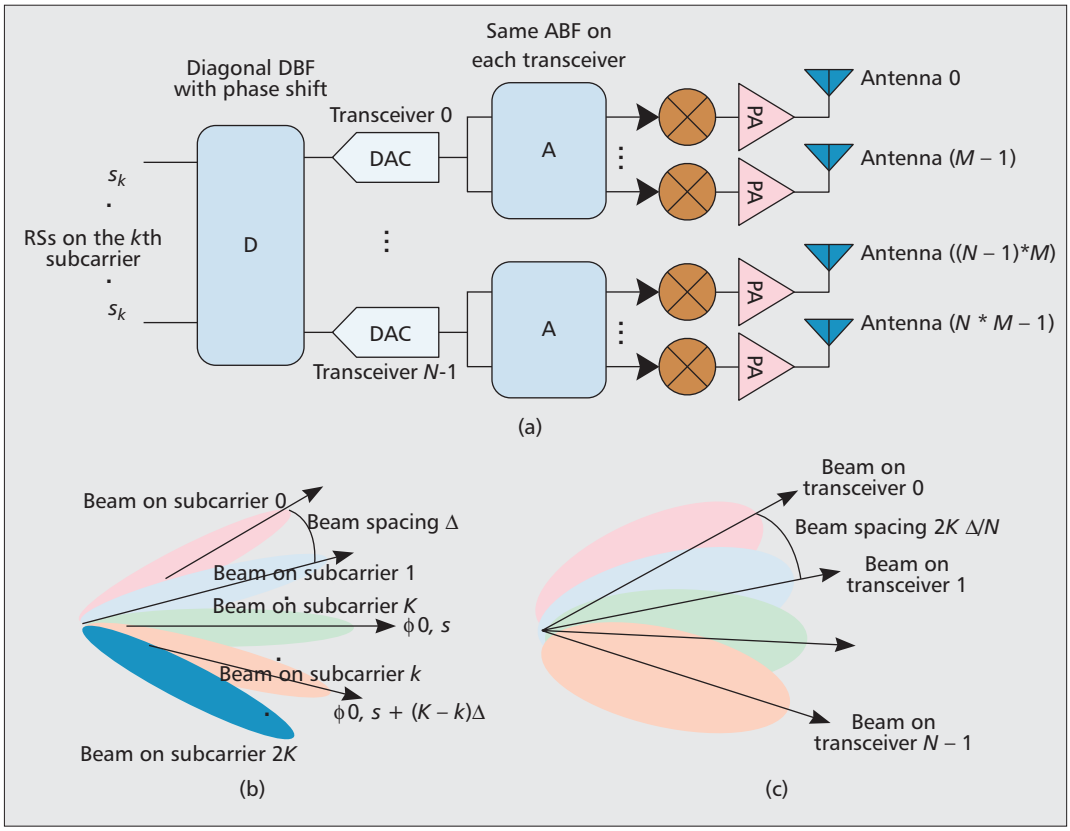

Figure 5. RS design for the $N \times M$ hybrid BF structure: a) hybrid BF for beam domain RSs; b) beam domain RSs with hybrid BF on different subcarriers; c) beam domain RSs with M antenna analog BF.

difference in received signal power on different subcarriers only comes from different digital BF weights.

AoD Estimation Performance Comparison

Consider the cell area covered by $2K + 1$ hybrid analog and digital BF beams shown in Fig. 5b, which is roughly $2K\Delta$. The AoD estimation accuracy of RSs with hybrid BF is therefore Δ. As shown in Fig. 5c, however, for RS beams with analog BF per transceiver (totally N beams) to cover the same area, the AoD estimation accuracy is $2K\Delta/N$. In LOS environments where the channel response is rather flat in frequency domain, $2K$ can be very large, indicating that RSs design with hybrid BF has much better AoD estimation performance than RSs with analog BF. According to the measurement results on 38 GHz [5], more than 80 percent of the non-LOS (NLOS) links had root mean square (RMS) delay spreads under 20 ns, and 90 percent of the NLOS links had RMS delay spreads under 40 ns. This indicates that the coherent bandwidth in NLOS environments can also be very large. Therefore, a large $2K/N$ is possible even in NLOS environments.

Illustration of Hybrid BF

The effect of hybrid BF is shown in Fig. 6. A 32-element linear antenna array (with half wavelength antenna spacing) is simulated, where the transceiver number is 4, and antenna number per transceiver is 8. The max array gain is achieved at azimuth 90°. The first curve is AF of 8 antenna elements analog BF with main beam direction at azimuth 90°. The second curve is the AF envelope of the 32-element hybrid analog and digital BF. It can be seen that the amplitude of the second curve is exactly four times that of the first curve. With same analog BF per transceiver, seven digital BF designs are shown, with main beam direction of azimuth 84°, 86°, 88°, 90°, 92°, 94, and 96°, respectively. As shown in Fig. 6, the main beam direction can be controlled with hybrid BF design.

Some observations are summarized:

- The further the main beam direction of the hybrid BF is away from the analog BF main beam direction, the smaller the gain. Therefore, the coverage of the beams is limited.
- The hybrid BF designed for 90° is actually the analog BF with all antennas.
- Users located at certain AoD ϕ within the range of 84° to 96° will find the hybrid BF designed for direction ϕ has the largest signal power.

Conclusions

The marriage of LSAS and mmWave technologies is expected to bring significant EE and SE enhancement to 5G. The high cost and power consumption of mixed-signal devices in mmWave systems, however, makes hybrid BF structure with a much reduced transceiver number a possible solution. In an attempt to shed some insight on how hybrid BF can be potentially utilized in mmWave 5G, this article has addressed some

important issues. We have discussed the optimal digital and analog BF design, showing that the investigated hybrid BF structures can achieve the multi-user channel capacity. The EE-SE relationship of the $N \times M$ hybrid BF structure have been analyzed, paving a path for an EE-SE optimized design. Finally, we have discussed beam domain RSs. The design based on hybrid BF is featured by the same analog BF on each transceiver, on top of which digital BF is designed to maximize the gain in a certain direction around the main beam direction of the analog BF. Much better AoD estimation performance can be achieved compared to beam domain RSs with analog BF.

Future work may include more efficient single-user and multi-user BF schemes, RS design and feedback mechanism, frame structure, protocol, and signaling design. Also, more effort is needed to investigate the impact of system parameters like W, P_0, $P_{rf\_circuit}$, P_{common}, and antenna spacing on optimal N and M design in terms of joint EE-SE optimization. It is expected that LSAS with mmWave can play an important role in future cellular communication systems.

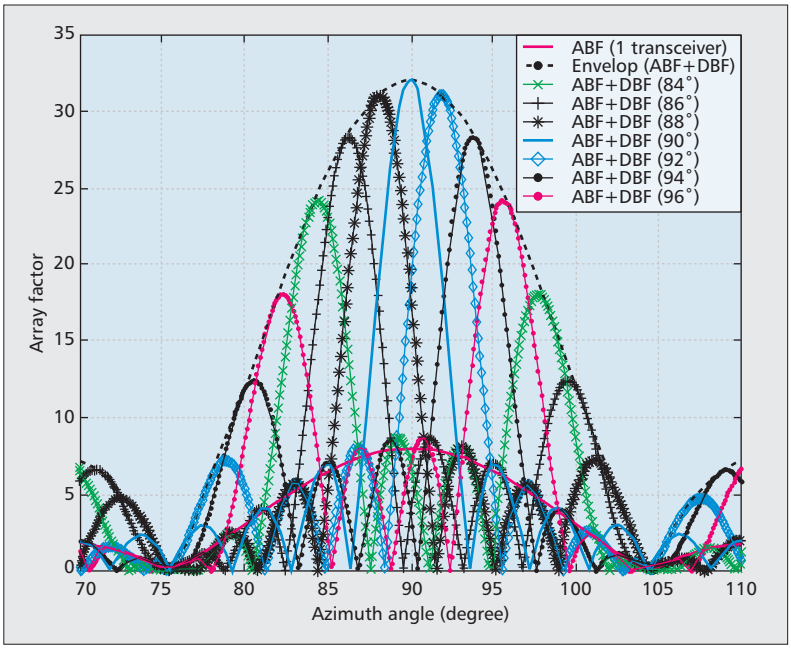

Figure 6. Hybrid BF with a 32-element antenna array.

ACKNOWLEDGMENT

The authors would like to thank the editors and reviewers for their very helpful comments and reviews. Their suggestions greatly improved the presentation of this article. The authors also would like to express their gratitude to Profs. Jing Wang and Shidong Zhou from Tsinghua University for their enlightening comments. The authors are also grateful to the team members in the Green Communication Research Center of the China Mobile Research Institute, particularly Drs. Sen Wang and Zhengang Pan.

REFERENCES

[1] C.-L. I *et al.*, "Toward Green & Soft: A 5G Perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 66–73.
[2] F. Rusek *et al.*, "Scaling up Mimo: Opportunities and Challenges with very Large Arrays," *IEEE Sig. Proc. Mag.*, vol. 30, no. 1, Jan 2013, pp. 40–60.
[3] H. Q. Ngo, E. G. Larsson, and T. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Trans. Commun.*, vol. 61, no. 4, Apr. 2013, pp. 1436–49.
[4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," *Proc. IEEE*, vol. 102, no. 3, 2014, pp. 366–85.
[5] T. Rappaport *et al.*, "Broadband Millimeter-wave Propagation Measurements and Models Using Adaptive-beam Antennas for Outdoor Urban Cellular Communications," *IEEE Trans. Antennas and Propagation*, vol. 61, no. 4, 2013, pp. 1850–59.
[6] C. Doan *et al.*, "Design Considerations for 60 GHz CMOS Radios," *IEEE Commun. Mag.*, vol. 42, no. 12, 2004, pp. 132–40.
[7] W. Roh *et al.*, "Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 106–13.
[8] Z. Pi and F. Khan, "An Introduction to Millimeter-Wave Mobile Broadband Systems," *IEEE Commun. Mag.*, vol. 49, no. 6, 2011, pp. 101–07.
[9] A. Alkhateeb *et al.*, "Hybrid Precoding for Millimeter Wave Cellular Systems with Partial Channel Knowledge," *Info. Theory and Applications Wksp.*, 2013.
[10] O. El Ayach *et al.*, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, Mar. 2014, pp.1499–1513.
[11] X. Huang, Y. Jay Guo, and J. Bunton, "A Hybrid Adaptive Antenna Array," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, May 2010, pp.1770–79.
[12] S. Han *et al.*, "Large Scale Antenna System with Hybrid Digital and Analog Beamforming Structure," *ICC Wksp. 2014.*
[13] G. Y. Lie*t al.*, "Energy-Efficient Wireless Communications: Tutorial, Survey, and Open Issues," *IEEE Wireless Commun.*, vol. 18, no. 6, Dec. 2011, pp. 28–35.
[14] Z. Xu, Z. Pan, and C.-L. I, "Fundamental Properties of the EE-SE Relationship," *IEEE WCNC 2014.*
[15] S. Han *et al.*, "Reference Signals Design for Hybrid Analog and Digital Beamforming," *IEEE Commun. Letters*, vol. 18, no. 7, July 2014, pp. 1191–93.

BIOGRAPHIES

SHUANGFENG HAN (hanshuangfeng@chinamobile.com) is currently a senior project manager in the Green Communication Research Center of the China Mobile Research Institute (CMRI). His research interests are mainly focused on green technologies R&D in 5G wireless communication systems, including large-scale antenna systems, active antenna systems, co-frequency co-time full duplex, non-orthogonal multiple access schemes, energy efficiency and spectrum efficiency co-design, and beyond cellular green generation solutions. Prior to joining CMRI, he was a senior engineer at Samsung Electronics' Research Center from 2006 to 2012. His research interests include MultiBS MIMO, MIMO codebook design, small cell/HetNet, millimeter-wave communication, D2D, and distributed radio over fiber. He graduated from Tsinghua University, Beijing, in 2006, and majored in information and communication systems. He is an inventor on 37 patent applications, and the author on over 20 peer-reviewed conference and journal publications. He has been reviewer for various IEEE journals and conferences.

CHIH-LIN I (icl@chinamobile.com) is the chief scientist of China Mobile Wireless Technologies, in charge of advanced wireless communication R&D efforts of CMRI. She established the Green Communications Research Center of China Mobile, spearheading major initiatives including 5G Key Technologies R&D; high energy efficiency system architecture, technologies, and devices; green energy; C-RAN and soft base station. She received her Ph.D. degree in electrical engineering from Stanford University, and has almost 30 years of experience in wireless communications. She has worked in various world-class companies and research institutes, including the wireless communication fundamental research department of AT&T Bell Labs; the headquarters of AT&T as director of Wireless Communications Infrastructure and Access Technology; ITRI of Taiwan as director of Wireless Communication Technology; and Hong Kong ASTRI as VP and the founding general director of the Communications Technology Domain. She received the *IEEE*

Transactions on Communications Stephen Rice Best Paper Award, and is a winner of the CCCP National 1000 Talent program. She was an elected Board Member of IEEE ComSoc, Chair of ComSoc Meetings and Conference Board, and Founding Chair of the IEEE WCNC Steering Committee. She is currently Chair of FuTURE Forum 5G SIG, an Executive Board Member of GreenTouch, a Network Operator Council Member of ETSI NFV, a Steering Board Member of Wireless World Research Forum, and an adjunct professor of Beijing University of Posts and Telecommunications.

ZHIKUN XU (xuzhikun@chinamobile.com) received his B.S.E. and Ph.D. degrees in signal and information processing from Beihang University (BUAA), Beijing, China, in 2007 and 2013, respectively. He was a visiting researcher in the School of Electrical and Computer Engineering, Georgia Institute of Technology, from 2009 to 2010. After graduation, he joined the Green Communication Research Center (GCRC) of CMRI as a project manager. His current interests include green technologies, the fundamental relationships between energy efficiency and spectral efficiency, energy-efficient network deployment and operation, cross-layer resource allocation in cellular networks, and advanced signal processing and transmission techniques.

CORBETT ROWELL (corbett.rowell@nu.edu.kz, corbettrowell@chinamobile.com) is a professor of electronic and electrical engineering at Nazarbayev University. He received his B.A. degree (honors) in physics from the University of California Santa Cruz, and his M.Phil. and Ph.D. degrees in electrical and electronic engineering from Hong Kong University of Science and Technology and Hong Kong University, respectively. He has worked for over 18 years in industry inside startups, research institutes, antenna manufacturers, and operators, designing a wide variety of products including cellular antennas, digital repeaters, radio units, MRI, NFC, MIMO, and base station RF systems. He designed large-scale antenna systems for 4G/5G as a research director at CMRI. He has more than 1400 citations for over 30 patents and 20 journal papers, and is the Technical Program Co-Chair for IEEE MTT IWS 2015. His research interests are miniature antennas, active antenna arrays, LSAS, metamaterials, massive MIMO, and sensor arrays.