ACCEPTED FROM OPEN CALL

Massive MIMO for Next Generation Wireless Systems

Erik G. Larsson, ISY, Linköping University, Sweden
Ove Edfors and Fredrik Tufvesson, Lund University, Sweden
Thomas L. Marzetta, Bell Labs, Alcatel-Lucent, United States

ABSTRACT

Multi-user MIMO offers big advantages over conventional point-to-point MIMO: it works with cheap single-antenna terminals, a rich scattering environment is not required, and resource allocation is simplified because every active terminal utilizes all of the time-frequency bins. However, multi-user MIMO, as originally envisioned, with roughly equal numbers of service antennas and terminals and frequency-division duplex operation, is not a scalable technology. Massive MIMO (also known as large-scale antenna systems, very large MIMO, hyper MIMO, full-dimension MIMO, and ARGOS) makes a clean break with current practice through the use of a large excess of service antennas over active terminals and time-division duplex operation. Extra antennas help by focusing energy into ever smaller regions of space to bring huge improvements in throughput and radiated energy efficiency. Other benefits of massive MIMO include extensive use of inexpensive low-power components, reduced latency, simplification of the MAC layer, and robustness against intentional jamming. The anticipated throughput depends on the propagation environment providing asymptotically orthogonal channels to the terminals, but so far experiments have not disclosed any limitations in this regard. While massive MIMO renders many traditional research problems irrelevant, it uncovers entirely new problems that urgently need attention: the challenge of making many low-cost low-precision components that work effectively together, acquisition and synchronization for newly joined terminals, the exploitation of extra degrees of freedom provided by the excess of service antennas, reducing internal power consumption to achieve total energy efficiency reductions, and finding new deployment scenarios. This article presents an overview of the massive MIMO concept and contemporary research on the topic.

GOING LARGE: MASSIVE MIMO

Massive multiple-input multiple-output (MIMO) is an emerging technology that scales up MIMO by possibly orders of magnitude compared to the current state of the art. In this article, we follow up on our earlier exposition [1], with a focus on the developments in the last three years; most particularly, energy efficiency, exploitation of excess degrees of freedom, time-division duplex (TDD) calibration, techniques to combat pilot contamination, and entirely new channel measurements.

With massive MIMO, we think of systems that use antenna arrays with a few hundred antennas simultaneously serving many tens of terminals in the same time-frequency resource. The basic premise behind massive MIMO is to reap all the benefits of conventional MIMO, but on a much greater scale. Overall, massive MIMO is an enabler for the development of future broadband (fixed and mobile) networks, which will be energy-efficient, secure, and robust, and will use the spectrum efficiently. As such, it is an enabler for the future digital society infrastructure that will connect the Internet of people and Internet of Things with clouds and other network infrastructure. Many different configurations and deployment scenarios for the actual antenna arrays used by a massive MIMO system can be envisioned (Fig. 1). Each antenna unit would be small and active, preferably fed via an optical or electric digital bus.

Massive MIMO relies on spatial multiplexing, which in turn relies on the base station having good enough channel knowledge, on both the uplink and the downlink. On the uplink, this is easy to accomplish by having the terminals send pilots, based on which the base station estimates the channel responses to each of the terminals. The downlink is more difficult. In conventional MIMO systems such as the Long Term Evolution (LTE) standard, the base station sends out pilot waveforms, based on which the terminals estimate the channel responses, quantize the thus obtained estimates, and feed them back to the base station. This will not be feasible in massive MIMO systems, at least not when operating in a high-mobility environment, for two reasons. First, optimal downlink pilots should be mutually orthogonal between the antennas. This means that the amount of time-frequency resources needed for downlink pilots scales with the number of antennas, so a massive MIMO system would require up to 100 times more such resources than a conventional system. Second,

0163-6804/14/$25.00 © 2014 IEEE

the number of channel responses each terminal must estimate is also proportional to the number of base station antennas. Hence, the uplink resources needed to inform the base station of the channel responses would be up to 100 times larger than in conventional systems. Generally, the solution is to operate in TDD mode, and rely on reciprocity between the uplink and downlink channels, although frequency-division duplext (FDD) operation may be possible in certain cases [2].

While the concepts of massive MIMO have been mostly theoretical so far, stimulating much research particularly in random matrix theory and related mathematics, basic testbeds are becoming available [3], and initial channel measurements have been performed [4, 5].

The Potential of Massive MIMO

Massive MIMO technology relies on phase-coherent but computationally very simple processing of signals from all the antennas at the base station. Some specific benefits of a massive MU-MIMO system are:

• Massive MIMO can *increase the capacity* 10 times or more and *simultaneously improve the radiated energy efficiency* on the order of 100 times. The *capacity increase* results from the aggressive spatial multiplexing used in massive MIMO. The fundamental principle that makes the dramatic increase in *energy efficiency* possible is that with a large number of antennas, energy can be focused with extreme sharpness into small regions in space (Fig. 2). The underlying physics is *coherent superposition* of wavefronts. By appropriately shaping the signals sent out by the antennas, the base station can make sure that all wavefronts collectively emitted by all antennas add up constructively at the locations of the intended terminals, but *destructively* (*randomly*) almost everywhere else. Interference between terminals can be suppressed even further by using, for example, zero-forcing (ZF). This, however, may come at the cost of more transmitted power, as illustrated in Fig. 2.

More quantitatively, Fig. 3 (from [6]) depicts the fundamental trade-off between *energy efficiency* in terms of the total number of bits (sum rate) transmitted per Joule per terminal receiving service of energy spent, and *spectral efficiency* in terms of total number of bits (sum rate) transmitted per unit of radio spectrum consumed. The figure illustrates the relation for the uplink, from the terminals to the base station (the downlink performance is similar). The figure shows the trade-off for three cases:

- A reference system with one single antenna serving a single terminal (purple)
- A system with 100 antennas serving a single terminal using conventional beamforming (green)
- A massive MIMO system with 100 antennas simultaneously serving multiple (about 40 here) terminals (red, using maximum ratio combining, and blue, using ZF).

The attractiveness of maximum ratio combining (MRC) compared with ZF is not only its computational simplicity — multiplication of the received signals by the conjugate channel

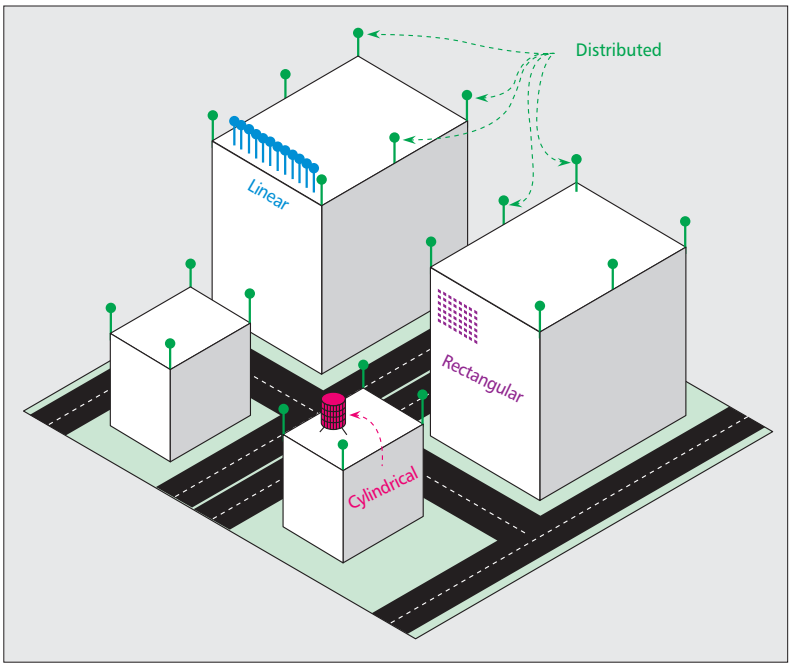

Figure 1. *Some possible antenna configurations and deployment scenarios for a massive MIMO base station.*

responses — but also that it can be performed in a distributed fashion, independently at each antenna unit. While ZF also works fairly well for a conventional or moderately sized MIMO system, MRC generally does not. The reason that MRC works so well for massive MIMO is that the channel responses associated with different terminals tend to be nearly orthogonal when the number of base station antennas is large.

The prediction in Fig. 3 is based on an information-theoretic analysis that takes into account intracell interference, as well as the bandwidth and energy cost of using pilots to acquire channel state information in a high-mobility environment [6]. With the MRC receiver, we operate in the nearly noise-limited regime of information theory. This means providing each terminal with a rate of about 1 b/complex dimension (1 b/s/Hz). In a massive MIMO system, when using MRC and operating in the "green" regime (i.e., scaling down the power as much as possible without seriously affecting the overall spectral efficiency), multiuser interference and effects from hardware imperfections tend to be overwhelmed by the thermal noise. The reason that the overall spectral efficiency still can be 10 times higher than in conventional MIMO is that many tens of terminals are served simultaneously, *in the same time-frequency resource*. When operating in the 1 b/dimension/terminal regime, there is also some evidence that intersymbol interference can be treated as additional thermal noise [7], hence offering a way of disposing with orthogonal frequency-division multiplexing (OFDM) as a means of combatting intersymbol interference.

To understand the scale of the capacity gains massive MIMO offers, consider an array consisting of 6400 omnidirectional antennas (total form factor $6400 \times (\lambda/2)^2 \approx 40$ m^2) transmitting with a total power of 120 W (i.e., each antenna radiat-

ing about 20 mW) over a 20 MHz bandwidth in the personal communications services (PCS) band (1900 MHz). The array serves 1000 fixed terminals randomly distributed in a disk of radius 6 km centered on the array, each terminal having an 8 dB gain antenna. The height of the antenna array is 30 m, and the height of the terminals is 5 m. Using the Hata-COST231 model, we find that the path loss is 127 dB at 1 km range, and the range-decay exponent is 3.52. There is also log-normal shadow fading with 8 dB standard deviation. The receivers have a 9 dB noise figure. One-quarter of the time is spent on transmission of uplink pilots for TDD channel estimation, and it is assumed that the channel is substantially constant over intervals of 164 ms in order to estimate the channel gains with sufficient accuracy. Downlink data is transmitted via maximum ratio transmission (MRT) beamforming combined with power control, where the 5 percent of terminals with the worst channels are excluded from service. We use a capacity lower bound from [8] extended to accommodate slow fading, near/far effects and power control, which accounts for receiver noise, channel estimation errors, the overhead of pilot transmission, and the imperfections of MRT beamforming. We use optimal max-min power control, which confers an equal signal-to-interference-plus-noise ratio on each of the 950 terminals and therefore equal throughput. Numerical averaging over random terminal locations and the shadow fading shows that 95 percent of the terminals will receive a throughput of 21.2 Mb/s/terminal. Overall, the array in this example will offer the 1000 terminals a total downlink throughput of 20 Gb/s, resulting in a sum-spectral efficiency of 1000 b/s/Hz. This would be enough, for example, to provide 20 Mb/s broadband service to each of 1000 homes. The max-min power control provides equal service *simultaneously* to 950 terminals. Other types of power control combined with time-division multiplexing could accommodate heterogeneous traffic demands of a larger set of terminals.

The MRC receiver (for the uplink) and its counterpart MRT precoding (for the downlink) are also known as matched filtering (MF) in the literature.

- Massive MIMO can be built with *inexpensive*, *low-power components*.

Massive MIMO is a game changing technology with regard to theory, systems, and implementation. With massive MIMO, expensive ultra-linear 50 W amplifiers used in conventional systems are replaced by hundreds of low-cost

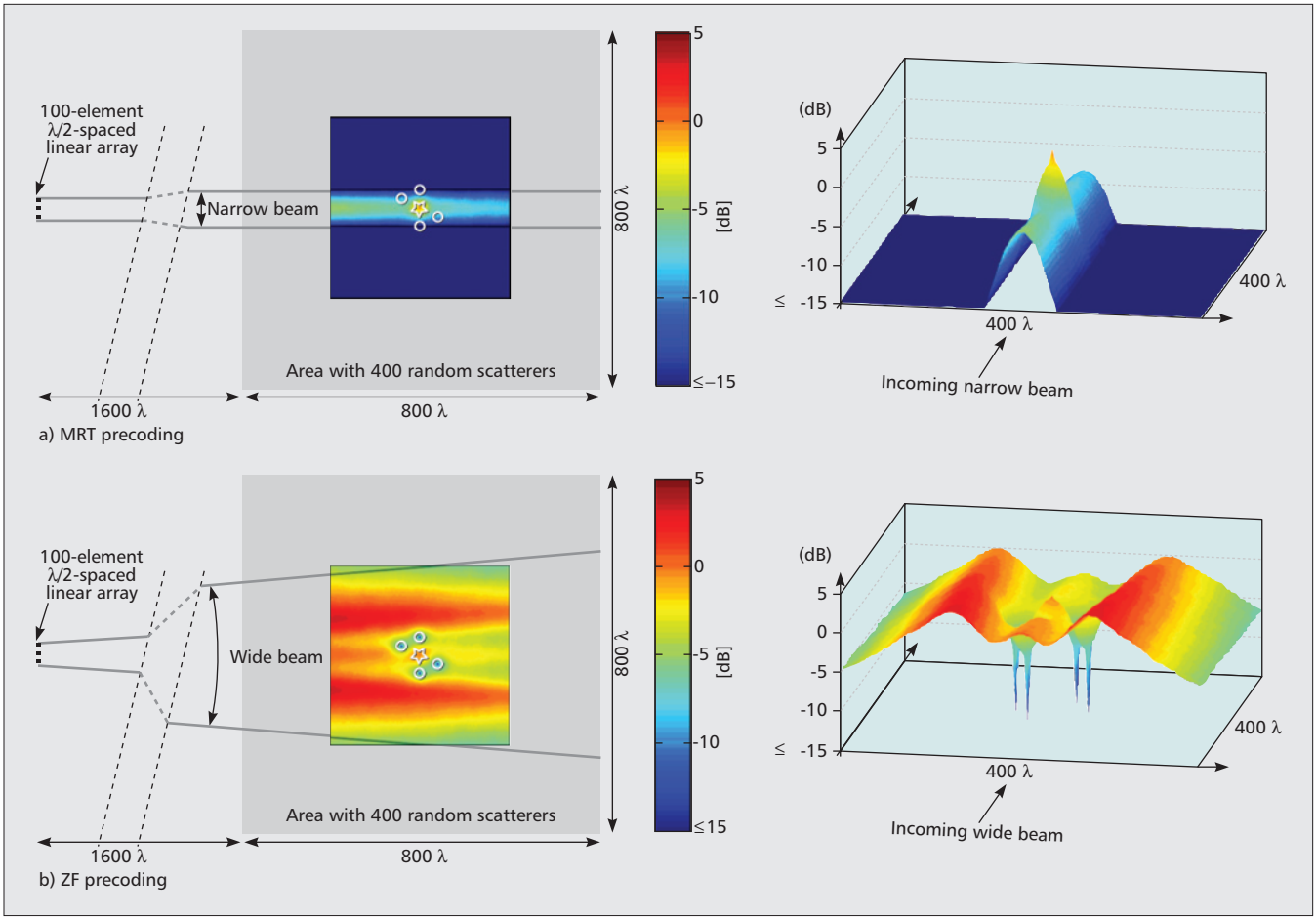

Figure 2. *Relative field strength around a target terminal in a scattering environment of size 800 λ × 800 λ when the base station is placed 1600 λ to the left. Average field strengths are calculated over 10,000 random placements of 400 scatterers when two different linear precoders are used: a) MRT precoders; b) ZF precoders. Left: pseudo-color plots of average field strengths, with target user positions at the center (⋆) and four other users nearby (○). Right: average field strengths as surface plots, allowing an alternate view of the spatial focusing.*

amplifiers with output power in the milli-Watt range. The contrast to classical array designs, which use few antennas fed from high-power amplifiers, is significant. Several expensive and bulky items, such as large coaxial cables, can be eliminated altogether. (The typical coaxial cables used for tower-mounted base stations today are more than 4 cm in diameter!)

Massive MIMO reduces the constraints on accuracy and linearity of each individual amplifier and RF chain. All that matters is their combined action. In a way, massive MIMO relies on the law of large numbers to make sure that noise, fading, and hardware imperfections average out when signals from a large number of antennas are combined in the air. The same property that makes massive MIMO resilient against fading also makes the technology extremely robust to failure of one or a few of the antenna unit(s).

A massive MIMO system has a large surplus of degrees of freedom. For example, with 200 antennas serving 20 terminals, 180 degrees of freedom are unused. These degrees of freedom can be used for hardware-friendly signal shaping. In particular, each antenna can transmit signals with very small peak-to-average ratio [9] or even constant envelope [10] at a very modest penalty in terms of increased total radiated power. Such (near-constant) envelope signaling facilitates the use of extremely cheap and power-efficient RF amplifiers. The techniques in [9, 10] must not be confused with conventional beamforming techniques or equal-magnitude-weight beamforming techniques. This distinction is explained in Fig. 4. With (near) constant-envelope multiuser precoding, no beams are formed, and the signals emitted by each antenna are not formed by weighing a symbol. Rather, a wavefield is created such that when this wavefield is sampled at the spots where the terminals are located, the terminals see precisely the signals we want them to see. The fundamental property of the massive MIMO channel that makes this possible is that the channel has a large nullspace: almost anything can be put into this nullspace without affecting what the terminals see. In particular, components can be put into this nullspace that make the transmitted waveforms satisfy the desired envelope constraints. Notwithstanding, the *effective channels* between the base station and each of the terminals can take any signal constellation as input and do not require the use of phase shift keying (PSK) modulation.

The drastically improved energy efficiency enables massive MIMO systems to operate with a total output RF power two orders of magnitude less than with current technology. This matters, because the energy consumption of cellular base stations is a growing concern worldwide. In addition, base stations that consume many orders of magnitude less power could be powered by wind or solar, and hence easily deployed where no electricity grid is available. As a bonus, the total emitted power can be dramatically cut, and therefore the base station will generate substantially less electromagnetic interference. This is important due to the increased concerns regarding electromagnetic exposure.

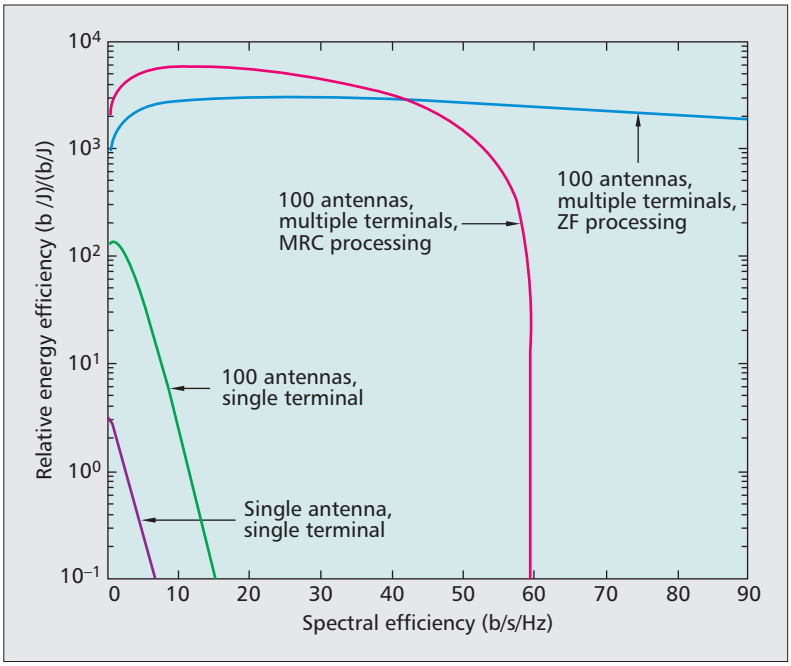

Figure 3. *Half the power — twice the force (from [6]): Improving uplink spectral efficiency 10 times and simultaneously increasing the radiated power efficiency 100 times with massive MIMO technology, using extremely simple signal processing, taking into account the energy and bandwidth costs of obtaining channel state information.*

- Massive MIMO enables a significant r*eduction of latency* on the air interface.

The performance of wireless communications systems is normally limited by fading. Fading can render the received signal strength very small at certain times. This happens when the signal sent from a base station travels through multiple paths before it reaches the terminal, and the waves resulting from these multiple paths interfere destructively. It is this fading that makes it hard to build low-latency wireless links. If the terminal is trapped in a fading dip, it has to wait until the propagation channel has sufficiently changed until any data can be received. Massive MIMO relies on the law of large numbers and beamforming in order to avoid fading dips, so fading no longer limits latency.

- Massive MIMO *simplifies the multiple access layer*.

Due to the law of large numbers, the channel *hardens* so that frequency domain scheduling no longer pays off. With OFDM, each subcarrier in a massive MIMO system will have substantially the same channel gain. Each terminal can be given the whole bandwidth, which renders most of the physical layer control signaling redundant.

- Massive MIMO increases the *robustness* against both unintended man-made interference and intentional jamming.

Intentional jamming of civilian wireless systems is a growing concern and a serious cybersecurity threat that seems to be little known to the public. Simple jammers can be bought off the Internet for a few hundred dollars, and equipment that used to be military-grade can be put together using off-the-shelf software radio-based platforms for a few thousand dollars.

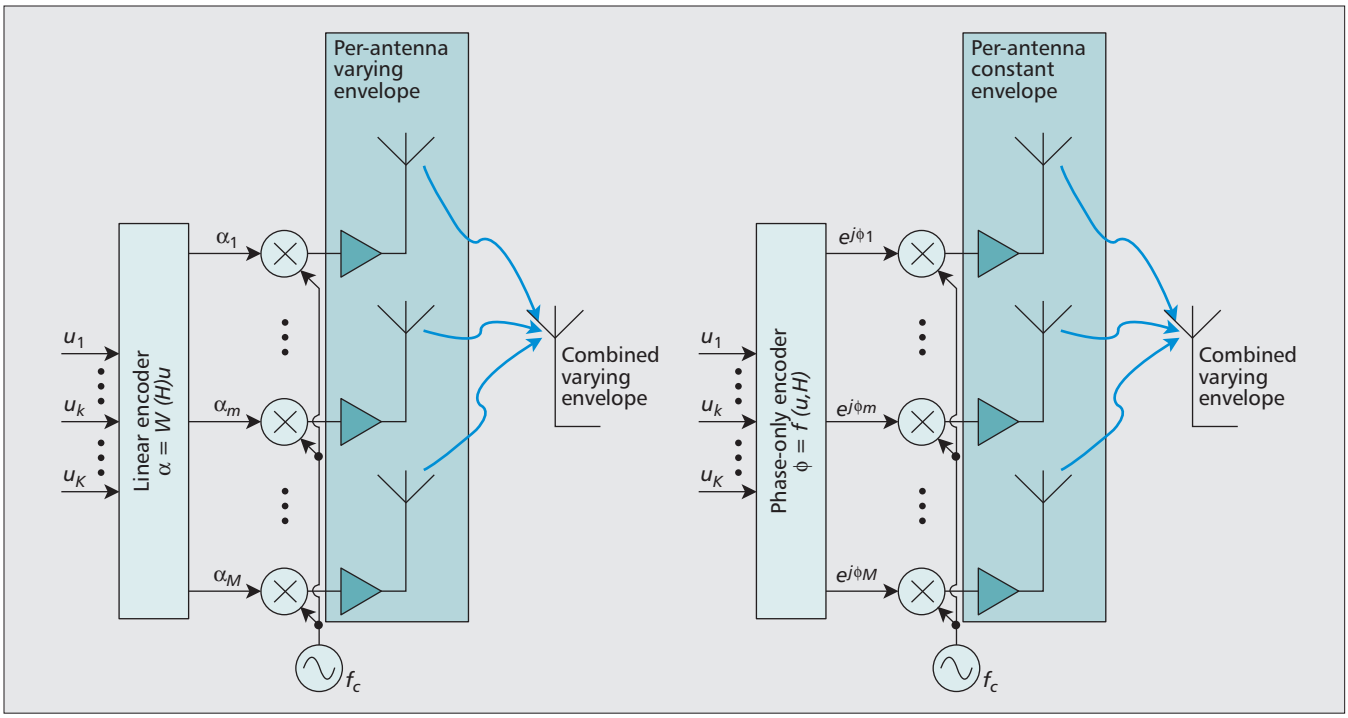

Figure 4. *Conventional MIMO beamforming contrasted with per-antenna constant envelope transmission in massive MIMO. Left: conventional beamforming, where the signal emitted by each antenna has a large dynamic range. Right: per-antenna constant envelope transmission, where each antenna sends out a signal with a constant envelope.*

Numerous recent incidents, especially in public safety applications, illustrate the magnitude of the problem. During the EU summit in Gothenburg, Sweden, in 2001, demonstrators used a jammer located in a nearby apartment, and during critical phases of riots, the chief commander could not reach any of the 700 police officers engaged [11].

Due to the scarcity of bandwidth, spreading information over frequency just is not feasible, so the only way of improving robustness of wireless communications is to use multiple antennas. Massive MIMO offers many excess degrees of freedom that can be used to cancel signals from intentional jammers. If massive MIMO is implemented using uplink pilots for channel estimation, smart jammers could cause harmful interference with modest transmission power. However, more clever implementations using joint channel estimation and decoding should be able to substantially diminish that problem.

Limiting Factors of Massive MIMO

Channel Reciprocity

Time-division duplexing operation relies on channel reciprocity. There appears to be a reasonable consensus that the propagation channel itself is essentially reciprocal unless the propagation is affected by materials with strange magnetic properties. However, the hardware chains in the base station and terminal transceivers may not be reciprocal between the uplink and the downlink. Calibration of the hardware chains does not seem to constitute a serious problem, and there are calibration-based solutions that have already been tested to some extent in practice [3, 12]. Specifically, [3] treats reciprocity calibration for a 64-antenna system in some detail and claims a successful experimental implementation.

Note that calibration of the terminal uplink and downlink chains is not required in order to obtain the full beamforming gains of massive MIMO: if the base station equipment is properly calibrated, the array will indeed transmit a coherent beam to the terminal. (There will still be some mismatch within the receiver chain of the terminal, but this can be handled by transmitting pilots through the beam to the terminal; the overhead for these supplementary pilots is very small.) Absolute calibration within the array is not required. Instead, as proposed in [3], one of the antennas can be treated as a reference, and signals can be traded between the reference antenna and each of the other antennas to derive a compensation factor for that antenna. It may be possible to entirely forgo reciprocity calibration within the array; for example if the maximum phase difference between the uplink and downlink chains were less than 60°, coherent beamforming would still occur (at least with MRT beamforming), albeit with a possible 3 dB reduction in gain.

Pilot Contamination

Ideally, every terminal in a massive MIMO system is assigned an orthogonal uplink pilot sequence. However, the maximum number of orthogonal pilot sequences that can exist is upper-bounded by the duration of the coherence interval divided by the channel delay spread. In

[13], for a typical operating scenario, the maximum number of orthogonal pilot sequences in a 1 ms coherence interval is estimated to be about 200. It is easy to exhaust the available supply of orthogonal pilot sequences in a multicellular system.

The effect of reusing pilots from one cell to another and the associated negative consequences is termed *pilot contamination*. More specifically, when the service array correlates its received pilot signal with the pilot sequence associated with a particular terminal, it actually obtains a channel estimate that is contaminated by a linear combination of channels with other terminals that share the same pilot sequence. Downlink beamforming based on the contaminated channel estimate results in interference directed at those terminals that share the same pilot sequence. Similar interference is associated with uplink transmissions of data. This directed interference grows with the number of service antennas at the same rate as the desired signal [13]. Even partially correlated pilot sequences result in directed interference.

Pilot contamination as a basic phenomenon is not really specific to massive MIMO, but its effect on massive MIMO appears to be much more profound than in classical MIMO [13, 14]. In [13] it was argued that pilot contamination constitutes an ultimate limit on performance when the number of antennas is increased without bound, at least with receivers that rely on pilot-based channel estimation. While this argument has been contested recently [15], at least under some specific assumptions on the power control used, it appears likely that pilot contamination must be dealt with in some way. This can be done in several ways:

- The allocation of pilot waveforms can be optimized. One possibility is to use a less aggressive frequency reuse factor for the pilots (but not necessarily for the payload data); say, 3 or 7. This pushes mutually contaminating cells farther apart. It is also possible to coordinate the use of pilots or adaptively allocate pilot sequences to the different terminals in the network [16]. Currently, the optimal strategy is unknown.
- Clever channel estimation algorithms [15], or even blind techniques that circumvent the use of pilots altogether [17], may mitigate or eliminate the effects of pilot contamination. The most promising direction seems to be blind techniques that jointly estimate the channels and the payload data.
- New precoding techniques that take into account the network structure, such as pilot contamination precoding [18], can utilize cooperative transmission over a multiplicity of cells — outside of the beamforming operation — to nullify, at least partially, the directed interference that results from pilot contamination. Unlike coordinated beamforming over multiple cells, which requires estimates of the actual channels between the terminals and the service arrays of the contaminating cells, pilot contamination precoding requires only the corresponding slow-fading coefficients. Practical pilot contamination precoding remains to be developed.

Radio Propagation and Orthogonality of Channel Responses

Massive MIMO (and especially MRC/MRT processing) relies to a large extent on a property of the radio environment called *favorable propagation*. Simply stated, favorable propagation means that the propagation channel responses from the base station to different terminals are sufficiently different. To study the behavior of massive MIMO systems, channel measurements have to be performed using realistic antenna arrays. This is so because the channel behavior using large arrays differs from that usually experienced using conventional smaller arrays. The most important differences are that:

- There might be large-scale fading over the array.
- The small-scale signal statistics may also change over the array. Of course, this is also true for physically smaller arrays with directional antenna elements pointing in various directions.

Figure 5 shows pictures of the two massive MIMO arrays used for the measurements reported in this article. On the left is a compact circular massive MIMO array with 128 antenna ports. This array consists of 16 dual-polarized patch antenna elements arranged in a circle, with 4 such circles stacked on top of each other. Besides having the advantage of being compact, this array also provides the possibility to resolve scatterers at different elevations, but it suffers from worse resolution in azimuth due to its limited aperture. To the right is a physically large linear (virtual) array, where a single omnidirectional antenna element is moved to 128 different positions in an otherwise static environment to emulate a real array with the same dimensions.

One way of quantifying how different the channel responses to different terminals are is to look at the spread between the smallest and largest *singular values* of the matrix that contains the channel responses. Figure 6 illustrates this for a case with 4 user terminals and a base station having 4, 32, and 128 antenna ports, respectively, configured as either a physically large single-polarized linear array or a compact dual-polarized circular array. More specifically, the figure shows the cumulative density function (CDF) of the difference between the smallest and largest singular values for the different measured (narrowband) frequency points in the different cases. As a reference, we also show simulated results for ideal independent identically distributed (i.i.d.) channel matrices, often used in theoretical studies. The measurements were performed outdoors in the Lund University campus area. The center frequency was 2.6 GHz and the measurement bandwidth 50 MHz. When using the cylindrical array, the RUSK Lund channel sounder was employed, while a network analyzer was used for the synthetic linear array measurements. The first results from the campaign were presented in [4].

For the 4-element array, the median of the singular value spread is about 23–18 dB. This number is a measure of the fading margin, the additional power that has to be used in order to serve all users with a reasonable received signal

Massive MIMO (and especially MRC/MRT processing) relies to a large extent on a property of the radio environment called favorable propagation. Simply stated, favorable propagation means that the propagation channel responses from the base station to different terminals are sufficiently different.

Figure 5. *Massive MIMO antenna arrays used for the measurements.*

power. With the massive linear array, the spread is less than 3 dB. In addition, note that none of the curves has any substantial tail. This means that the probability of seeing a singular value spread larger than 3 dB anywhere over the measured bandwidth is essentially negligible.

To further illustrate the influence of different numbers of antenna elements at the base station and antenna configuration, we plot in Fig. 7 the sum rate for 4 closely spaced users (less than 2 m between users at a distance of about 40 m from the base station) in a non line-of-sight (NLOS) scenario when using MRT as precoding. The transmit power is normalized so that on average, the interference-free signal-to-noise-ratio at the terminals is 10 dB.

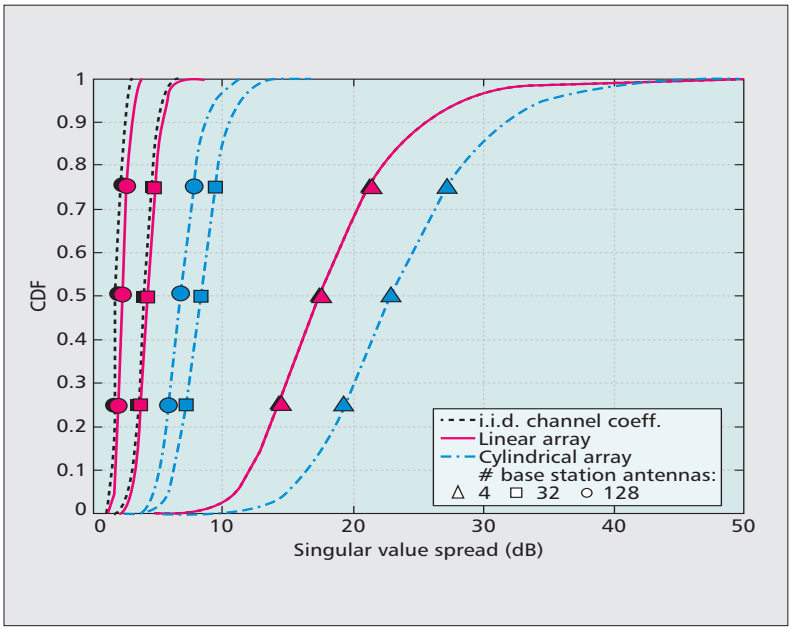

Figure 6. *CDF of the singular value spread for MIMO systems with 4 terminals and three different numbers of base station antennas: 4, 32, and 128. The theoretical i.i.d. channel is shown as a reference, while the other two cases are measured channels with linear and cylindrical array structures at the base station. Note that the curve for the linear array coincides with that of the i.i.d. channel for four base stations.*

As can be seen in Fig. 7, the sum rate approaches that of the theoretical interference-free case as the number of antennas at the base station increases. The shaded areas in red (for the linear array) and blue (for the circular array) shows the 90 percent confidence intervals of the sum rates for the different narrowband frequency realizations. As before, the variance of the sum rate decreases as the number of antennas increases, but slowly for the measured channels. The slow decrease can, at least partially, be attributed to the shadow fading occurring across the arrays: for the linear array in the form of shadowing by external objects along the array, and for the cylindical array in the form of shadowing caused by directive antenna elements pointing in the wrong direction. The performance of the physically large array approaches that of the theoretical i.i.d. case when the number of antennas grows large. The compact circular array has inferior performance compared to the linear array due to its smaller aperture — it cannot resolve the scatterers as well as the physically large array — and its directive antenna elements sometimes pointing in the wrong direction. Also, due to the fact that most of the scatterers are seen at the same horizontal angle, the possibility to resolve scatters at different elevations gives only marginal contributions to the sum rate in this scenario.

It should be mentioned here that when using somewhat more complex, but still linear, precoding methods such as ZF or minimum mean square error, the convergence to the i.i.d. channel performance is faster and the variance of the sum rate lower as the number of base station antennas is increased; see [4] for further details. Also, another aspect worth mentioning is that for a very tricky propagation scenario, such as closely spaced users in line-of-sight conditions, it seems that the large array is able to separate the users to a reasonable extent using the different spatial signatures the users have at the base station due to the enhanced spatial resolution. This would not be possible with conventional MIMO. These conclusions are also in line with the observations in [5], where another outdoor measurement campaign is described and analyzed.

Overall, there is compelling evidence that the assumptions on favorable propagation underpinning massive MIMO are substantially valid in practice. Depending on the exact configuration of the large array and the precoding algorithms used, the convergence toward the ideal performance may be faster or slower as the number of antennas is increased. However, with about 10 times more base station antennas than the number of users, it seems that it is possible to get stable performance not far from the theoretically ideal performance also under what are normally considered very difficult propagation conditions.

Massive MIMO: A Gold Mine of Research Problems

While massive MIMO renders many traditional problems in communication theory less relevant, it uncovers entirely new problems that need research.

Fast and distributed coherent signal processing: Massive MIMO arrays generate vast amounts of baseband data that must be processed in real time. This processing will have to be simple, and simple means linear or nearly linear. Fundamentally, this is good in many cases (Fig. 3). Much research needs be invested in the design of optimized algorithms and their implementation. On the downlink, there is enormous potential for ingenious precoding schemes. Some examples of recent work in this direction include [19].

The challenge of low-cost hardware: Building hundreds of RF chains, up/down converters, analog-to-digital (A/D)-digital-to-analog (D/A) converters, and so forth, will require economy of scale in manufacturing comparable to what we have seen for mobile handsets.

Hardware impairments: Massive MIMO relies on the law of large numbers to average out noise, fading and to some extent, interference. In reality, massive MIMO must be built with low-cost components. This is likely to mean that hardware imperfections are larger: in particular, phase noise and I/Q imbalance. Low-cost and power-efficient A/D converters yield higher levels of quantization noise. Power amplifiers with very relaxed linearity requirements will necessitate the use of per-antenna low peak-to-average signaling, which, as already noted, is feasible with a large excess of transmitter antennas. With low-cost phase locked loops or even free-running oscillators at each antenna, phase noise may become a limiting factor. However, what ultimately matters is how much the phase will drift between the point in time when a pilot symbol is received and the point in time when a data symbol is received at each antenna. There is great potential to get around the phase noise problem by design of smart transmission physical layer schemes and receiver algorithms.

Internal power consumption: Massive MIMO offers the potential to reduce the radiated power 1000 times and at the same time drastically scale up data rates. However, in practice, the total power consumed must be considered, including the cost of baseband signal processing. Much research must be invested in highly parallel, perhaps dedicated, hardware for the baseband signal processing.

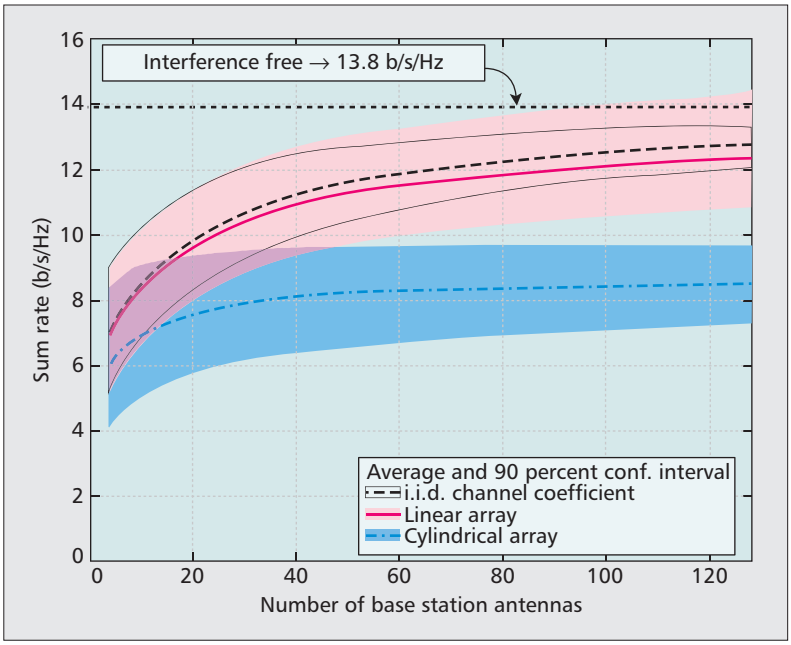

Figure 7. *Achieved downlink sum rates using MRT precoding, with 4 single-antenna terminals and between 4 and 128 base station antennas.*

Channel characterization: There are additional properties of the channel to consider when using massive MIMO instead of conventional MIMO. To facilitate a realistic performance assessment of massive MIMO systems, it is necessary to have channel models that reflect the true behavior of the radio channel (i.e., the propagation channel including effects of realistic antenna arrangements). It is also important to develop more sophisticated analytical channel models. Such models need not necessarily be correct in every fine detail, but they must capture the essential behavior of the channel. For example, in conventional MIMO the Kronecker model is widely used to model channel correlation. This model is not an exact representation of reality, but provides a useful model for certain types of analysis despite its limitations. A similar way of thinking could probably be adopted for massive MIMO channel modeling.

Cost of reciprocity calibration: TDD will require reciprocity calibration. How often must this be done, and what is the best way of doing it? What is the cost, in terms of time- and frequency resources needed to do the calibration, and in terms of additional hardware components needed?

Pilot contamination: It is likely that pilot contamination imposes much more severe limitations on massive MIMO than on traditional MIMO systems. We have discussed some of the issues in detail and outlined some of the most relevant research directions earlier.

Non-CSI@TX operation: Before a link has been established with a terminal, the base station has no way of knowing the channel response to the terminal. This means that no array beamforming gain can be harnessed. In this case, probably some form of space-time block coding

Continued testbed development is highly desired to both prove the massive MIMO concept with even larger numbers of antennas and discover potentially new issues that urgently need research.

is optimal. Once the terminal has been contacted and sent a pilot, the base station can learn the channel response and operate in coherent MU-MIMO beamforming mode, reaping the power gains offered by having a very large array.

New deployment scenarios: It is considered extraordinarily difficult to introduce a radical new wireless standard. One possibility is to introduce dedicated applications of massive MIMO technology that do not require backward compatibility. For example, as discussed earlier, in rural areas, a billboard-sized array could provide 20 Mb/s service to each of 1000 homes using special equipment that would be used solely for this application. Alternatively, a massive array could provide the backhaul for base stations that serve small cells in a densely populated area. Thus, rather than thinking of massive MIMO as a competitor to LTE, it can be an enabler for something that was just never before considered possible with wireless technology.

System studies and relation to small-cell and heterogeneous network solutions: The driving motivation of massive MIMO is to simultaneously and drastically increase data rates and overall energy efficiency. Other potential ways of reaching this goal are network densification by the deployment of small cells, resulting in a heterogeneous architecture, or coordination of the transmission of multiple individual base stations. From a purely fundamental perspective, the ultimately limiting factor of the performance of any wireless network appears to be the availability of good enough channel state information (CSI) to facilitate phase-coherent processing at multiple antennas or multiple access points [20]. Considering factors like mobility, Doppler shifts, phase noise, and clock synchronization, acquiring high-quality CSI seems to be easier with a collocated massive array than in a system where the antennas are distributed over a large geographical area. But at the same time, a distributed array or small cell solution may offer substantial path loss gains and would also provide some diversity against shadow fading. The deployment costs of a massive MIMO array and a distributed or small cell system are also likely to be very different. Hence, both communication-theoretic and techno-economic studies are needed to conclusively determine which approach is superior. However, it is likely that the winning solution will comprise a combination of all available technologies.

Prototype development: While massive MIMO is in its infancy, basic prototyping work on various aspects of the technology is going on in different parts of the world. The Argos testbed [3] was developed at Rice University in cooperation with Alcatel-Lucent, and shows the basic feasibility of the massive MIMO concept using 64 coherently operating antennas. In particular, the testbed shows that TDD operation relying on channel reciprocity is possible. One of the virtues of the Argos testbed in particular is that it is entirely modular and scalable, and built around commercially available hardware (the WARP platform). Other test systems around the world have also demonstrated the basic feasibility of scaling up the number of antennas. The Ngara testbed in Australia [21] uses a 32-element base station array to serve up to 18 users simultaneously with true spatial multiplexing. Continued testbed development is highly desired to both prove the massive MIMO concept with even larger numbers of antennas and discover potentially new issues that urgently need research.

Conclusions and Outlook

In this article we have highlighted the large potential of massive MIMO systems as a key enabling technology for future beyond fourth generation (4G) cellular systems. The technology offers huge advantages in terms of energy efficiency, spectral efficiency, robustness, and reliability. It allows for the use of low-cost hardware at both the base station and the mobile unit side. At the base station the use of expensive and powerful, but power-inefficient, hardware is replaced by massive use of parallel low-cost low-power units that operate coherently together. There are still challenges ahead to realize the full potential of the technology, for example, computational complexity, realization of distributed processing algorithms, and synchronization of the antenna units. This gives researchers in both academia and industry a gold mine of entirely new research problems to tackle.

Acknowledgments

The authors would like to thank Xiang Gao, doctoral student at Lund University, for her analysis of the channel measurements presented in Fig. 6 and Fig. 7, and the Swedish organizations ELLIIT, VR, and SSF for their funding of parts of this work.

References

[1] F. Rusek *et al.*, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Sig. Proc. Mag.*, vol. 30, Jan. 2013, pp. 40–60.
[2] J. Nam *et al.*, "Joint Spatial Division and Multiplexing: Realizing Massive MIMO Gains with Limited Channel State Information," *46th Annual Conf. Information Sciences and Systems*, 2012.
[3] C. Shepard *et al.*, "Argos: Practical Many-Antenna Base Stations," *ACM Int'l. Conf. Mobile Computing and Networking*, Istanbul, Turkey, Aug. 2012.
[4] X. Gao *et al.*, "Measured Propagation Characteristics for Very-Large MIMO at 2.6 GHz," *Proc. 46th Annual Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2012.
[5] J. Hoydis *et al.*, "Channel Measurements for Large Antenna Arrays," *IEEE Int'l. Symp. Wireless Commun. Systems*, Paris, France, Aug. 2012.
[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Trans. Commun.*, vol. 61, Apr. 2013, pp. 1436–49.
[7] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the Optimality of Single-Carrier Transmission in Large-Scale Antenna Systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, Aug. 2012, pp. 276–79.
[8] H. Yang and T. L. Marzetta, "Performance of Conjugate and Zero-Forcing Beamforming in Large-Scale Antenna Systems," *IEEE JSAC*, vol. 31, no. 2, Feb. 2013, pp. 172–79.
[9] C. Studer and E. G. Larsson, "PAR-Aware Large-Scale Multi-User MIMO-OFDM Downlink," *IEEE JSAC*, vol. 31, Feb. 2013, pp. 303–13.
[10] S. K. Mohammed and E. G. Larsson, "Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems," *IEEE Trans. Commun.*, vol. 61, Mar. 2013, pp. 1059–71.
[11] P. Stenumgaard *et al.*, "An Early-Warning Service for Emerging Communication Problems in Security and Safety Applications," *IEEE Commun. Mag.*, vol. 51, no. 5, Mar. 2013, pp. 186–92.

[12] F. Kaltenberger *et al.*, "Relative Channel Reciprocity Calibration in MIMO/TDD Systems," *Proc. Future Network and Mobile Summit*, 2010, 2010.
[13] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, Nov. 2010, pp. 3590–600.
[14] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE JSAC*, vol. 31, no. 2, Feb. 2013, pp. 160–71.
[15] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind Pilot Decontamination," *Proc. ITG Wksp. Smart Antennas*, Stuttgart, Mar. 2013.
[16] H. Yin *et al.*, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE JSAC*, vol. 31, no. 2, Feb. 2013, pp. 264–73.
[17] H. Q. Ngo and E. G. Larsson, "EVD-Based Channel Estimations for Multicell Multiuser MIMO with Very Large Antenna Arrays," *Proc. IEEE Int'l. Conf. Acoustics, Speed and Sig. Proc.*, Mar. 2012.
[18] A. Ashikhmin and T. L. Marzetta, "Pilot Contamination Precoding in Multi-Cell Large Scale Antenna Systems," *IEEE Int'l. Symp. Information Theory*, Cambridge, MA, July 2012.
[19] J. Zhang, X. Yuan, and L. Ping, "Hermitian Precoding for Distributed MIMO Systems with Individual Channel State Information," *IEEE JSAC*, vol. 31, no. 2, Feb. 2013, pp. 241–50.
[20] A. Lozano, R.W. Heath Jr, and J. G. Andrews, "Fundamental Limits of Cooperation," *IEEE Trans. Info. Theory*, vol. 59, Sept. 2013, pp. 5213–26.
[21] H. Suzuki *et al.*, "Highly Spectrally Efficient Ngara Rural Wireless Broadband ACCESS Demonstrator," *Proc. IEEE Int'l. Symp. Commun. and Information Technologies*, Oct. 2012.

BIOGRAPHIES

ERIK G. LARSSON is a professor and head of the Division for Communication Systems in the Department of Electrical Engineering at Linköping University, Sweden. He has published some 100 journal papers on signal processing and communications, and is a co-author of the textbook *Space-Time Block Coding for Wireless Communications*. He is an Associate Editor for *IEEE Transactions on Communications*, and he received the *IEEE Signal Processing Magazine* Best Column Award 2012.

OVE EDFORS is a professor of radio systems in the Department of Electrical and Information Technology, Lund University, Sweden. His research interests include statistical signal processing and low-complexity algorithms with applications in wireless communications. In the context of massive MIMO, his research focus is on how realistic propagation characteristics influence system performance and baseband processing complexity.

FREDRIK TUFVESSON received his Ph.D. in 2000 from Lund University. After almost two years at a startup company, Fiberless Society, he is now an associate professor in the Department of Electrical and Information Technology, Lund University. His main research interests are channel measurements and modeling for wireless communication, including channels for both MIMO and UWB systems. Besides these, he also works on distributed antenna systems and radio-based positioning.

THOMAS L. MARZETTA [F'03] received his Ph.D. in electrical engineering from the Massachusetts Institute of Technology. He joined Bell Laboratories in 1995. Within the former Mathematical Sciences Research Center he was director of the Communications and Statistical Sciences Department. He was an early proponent of massive MIMO, which can provide huge improvements in wireless spectral efficiency and energy efficiency over 4G technologies. He received the 1981 ASSP Paper Award and received the 2013 IEEE Guglielmo Marconi Best Paper Award.

There are still challenges ahead to realize the full potential of the technology, for example, when it comes to computational complexity, realization of distributed processing algorithms, and synchronization of the antenna units.