Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Non-ideal space division multiple access and its application

Ji-ying XIANG
ZTE Corporation, Shenzhen 518054, China
E-mail: xiang.jiying@zte.com.cn

Received Dec. 11, 2017; Revision accepted Mar. 1, 2018; Crosschecked Mar. 12, 2018

Abstract: We study the performance of space division multiple access (SDMA) under a non-ideal engineering situation. When the SDMA channel is of high inter-layer correlation and the condition number is large, the multiple user multiple input and multiple output user equipment (MUMIMOUE) grouping should be optimized, and in some cases further dimension-reduction should be applied. As the channel measuring is always non-ideal, we use two methods, feedback mode and non-feedback mode, in terms of performance and overhead. It is proposed that the non-feedback mode is preferable even for some non-reciprocal channels. Principle analysis and test results are given.

Key words: Fifth generation (5G); Condition number; Channel reciprocity; Feedback mode; Non-feedback mode
https://doi.org/10.1631/FITEE.1700827 **CLC number:** TN802

1 Introduction

A classical communication system takes combination of time, space, and frequency as the resource to carry information. Unfortunately, the available resources are always in short supply. Therefore, high spectrum efficiency is extremely important for any wireless communication system. The single channel spectrum efficiency is specified by the Shannon limit (Karasawa, 2016). Different kinds of modulation/demodulation (Jarry and Beneat, 2015) and coder/encoder technologies (Melo et al., 2016; Sodha, 2016) are used to approach the Shannon limit.

Time division multiplex (TDM), code division multiplex (CDM) (Zhang et al., 2016), and frequency division multiplex (FDM) are used to aggregate data from multiple access users into a single Shannon channel (Dahlman et al., 2016).

As the Shannon limit and noise limit have been closely approached, space division multiple access (SDMA) becomes the most effective way to further multiply the spectral efficiency. There are different kinds of SDMA, such as single user multiple input multiple output (SUMIMO), multiple user MIMO (MUMIMO), massive MIMO, and distributed MIMO (Adian, 2014; Adian and Adyan, 2014). SDMA segments logically the channel into multiples, and the eventual throughput satisfies the multi-channel expansion of the Shannon limit, which can be many times higher than that of the single channel (Adian and Adyan, 2014). SDMA has become one of the most important fifth generation wireless systems (5G) technologies (Xiang, 2015). The sub-6G 5G depends on SDMA to improve the capacity (Lu et al., 2014; Xiang, 2015), and the mmWave 5G also depends on SDMA to improve coverage (Jiang et al., 2016).

The real situation in wireless communication is always non-ideal, while previous studies usually assumed that the channel can be measured in real time and that the measurement result is precise. How to optimize the SDMA in non-ideal situations, however, was not considered. In real engineering situations, measurement is affected at least by the quantization error during analog-to-digital (AD) or digital-to-analog (DA) conversion. It is also affected by the lack of precision in the local clock. With more and more antennas in place, non-ideal situations of the local

clock cannot be ignored. In addition, the measuring process always takes time to a certain extent, and then the real channel will be mismatched with the measurement. In addition, in some cases, the channel even cannot be directly measured. In frequency division duplex (FDD), for instance, if the measurement is taken in the downlink, then an unacceptable overhead will take place for massive MIMO. Or, if the measurement is taken on the uplink, then the phase of the signal cannot be obtained, while downlink and uplink are presenting at different frequencies.

Besides this, channel conditions are more or less non-ideal, and this might introduce inter-layer interference during SDMA, particularly when forcing a high-order SDMA. The inter-layer interference will significantly impact the performance for low signal-to-interference-plus-noise ratio (SINR) cases.

We analyze the non-ideal factors, propose solutions to reconstruct the channel and maximize the SDMA performance, and develop an adaptive mechanism which complies with the 3rd Generation Partnership Project (3GPP) standard. Furthermore, we do a field test to verify the optimization, particularly under the non-ideal channel conditions.

2 Space division multiple access under non-ideal channel conditions

In a multiple antenna communication system with n transmitters and m receivers, the communication process can be described as Fig. 1a. After measuring the received signals, receivers need to reverse and deduce the transmitted data.

If the signals can be jointly processed by transmitters and receivers, singular value decomposition (SVD) becomes the optimal solution to SDMA. The SVD result can be taken as a pre-coding and equalization matrix. In this way, we try to map the mesh-like channel into linear ones (Fig. 1b).

By using linear algebra, the channel matrix $\boldsymbol{H}_{m\times n}$ can be decomposed as

$$\boldsymbol{H}_{m\times n}=\boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}_{n\times n}^{\mathrm{H}}, \tag{1}$$

where $\boldsymbol{\Sigma}_{m\times n}$ is a diagonal matrix and its elements are the singular values of $\boldsymbol{H}_{m\times n}$, and $\boldsymbol{U}_{m\times m}$ and $\boldsymbol{V}_{n\times n}$ are unitary matrices. In this way, the inter-layer interference is reduced greatly as follows:

$$\begin{aligned}\boldsymbol{H}_{m\times n}&=\begin{pmatrix}h_{11}&h_{12}&\cdots&h_{1n}\\h_{21}&h_{22}&\cdots&h_{2n}\\\vdots&\vdots&&\vdots\\h_{m1}&h_{m2}&\cdots&h_{mn}\end{pmatrix}\\&\Rightarrow\boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}_{n\times n}^{\mathrm{H}}\\&=\boldsymbol{U}_{m\times m}\begin{pmatrix}\sigma_1&0&\cdots&0\\0&\sigma_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&\sigma_{mn}\end{pmatrix}\boldsymbol{V}_{n\times n}^{\mathrm{H}}.\end{aligned} \tag{2}$$

Accordingly, the original interconnecting mesh channels between transmitters and receivers (Fig. 1a) can be equivalent to the linear combination of the following sections (Fig. 1b), namely, raw data (multi-streaming data) $\boldsymbol{S}_{n\times 1}$, an r-dimensional vector (extended to an n-dimensional vector with the high-dimensional elements filled with 0), and the raw data is transformed to an n-dimensional (high-dimensional) vector after a pre-coding process (pre-multiplied by the n-dimensional vector $\boldsymbol{V}_{n\times n}$). Then the dimension of the vector is reduced to m after traveling $m\times n$ air channels. Eventually, this becomes an m-dimensional receiving vector by an m-dimensional equalizer (pre-multiplied by $\boldsymbol{U}_{m\times m}$).

$$\begin{aligned}\boldsymbol{Z}_{m\times 1}&=\boldsymbol{U}_{m\times m}^{\mathrm{H}}(\boldsymbol{H}_{m\times n}\boldsymbol{V}_{n\times n}\boldsymbol{S}_{n\times 1}+\boldsymbol{N}_{m\times 1})\\&=\boldsymbol{U}_{m\times m}^{\mathrm{H}}(\boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}_{n\times n}^{\mathrm{H}}\boldsymbol{V}_{n\times n}\boldsymbol{S}_{n\times 1}+\boldsymbol{N}_{m\times 1}),\end{aligned} \tag{3}$$

where $\boldsymbol{Z}_{m\times 1}$ and $\boldsymbol{S}_{n\times 1}$ are the transmitted and received matrices, respectively, and $\boldsymbol{N}_{m\times 1}$ is the noise matrix. Since $\boldsymbol{U}_{m\times m}$ and $\boldsymbol{V}_{n\times n}$ are unitary matrices, we can obtain

$$\boldsymbol{Z}_{m\times 1}=\boldsymbol{\Sigma}_{m\times n}\boldsymbol{S}_{n\times 1}+\boldsymbol{U}_{m\times m}^{\mathrm{H}}\boldsymbol{N}_{m\times 1}\ . \tag{4}$$

It can be found that multiple raw data are recovered at the receivers after pre-coding (increased from r-dimensional to n-dimensional), dimension-reduction (reduced from n-dimensional to m-dimensional), and jointing equalization processes. In addition, the air channels are mapped into logic channels, so it is important to analyze $\boldsymbol{\Sigma}_{m\times n}$.

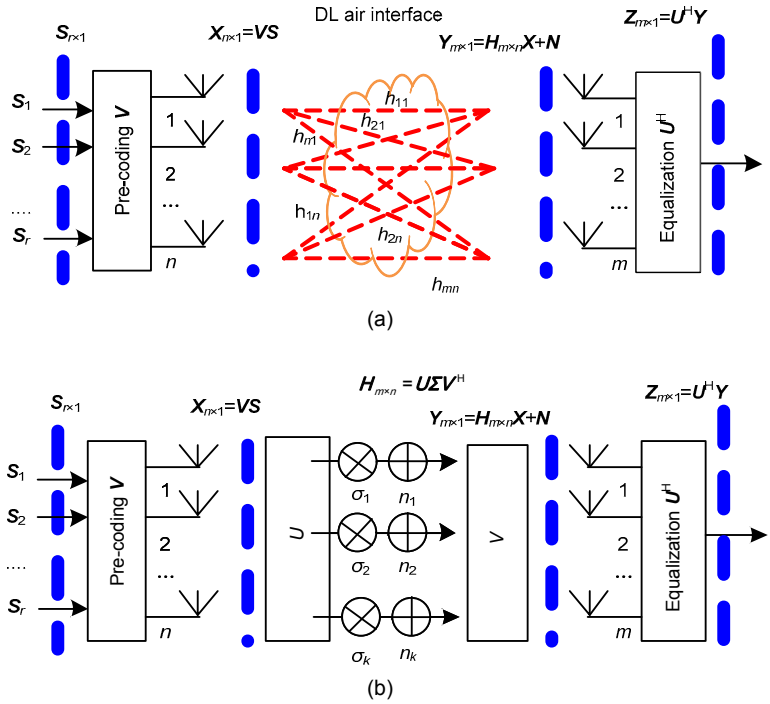

Fig. 1 Channel mapping of space division multiple access: (a) original channels; (b) mapped channels

$$\boldsymbol{Z}_{m\times1}=\boldsymbol{\Sigma}_{m\times n}\boldsymbol{S}_{n\times1}+\boldsymbol{N}_{m\times1}=\begin{pmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_{mn}\end{pmatrix}_{m\times n}\cdot\boldsymbol{S}_{n\times1}+\boldsymbol{N}_{m\times1}. \tag{5}$$

Note that $\boldsymbol{\Sigma}_{m\times n}$ is a quasi-diagonal matrix, and the mapped logic channels are independent of each other and can be regarded as independent transmission channels without interference. The gain value of each logic channel is σ_k, namely the singular value of $\boldsymbol{H}$.

This is the case where both transmitters and receivers can do joint processing. If neither of them can jointly process signals, a sub-optimal solution of SDMA can be obtained through zero forcing or other methods.

There are some non-ideal factors in communication, particularly in the field, such as measurement latency, quantitative errors caused by AD/DA conversion, jitter of the clock, thermal noise or other inherent noise, and interference. In addition, the air interface channels themselves are non-ideal and might not be suitable for solving SDMA. These non-ideal factors lead to a deviation from the optimized solutions and even a larger error (error amplification).

Therefore, it is necessary to measure the stability of the solutions. For this purpose, the concept of conditional number (CN) is introduced. One example of a 2-CN is described as

$$\mathrm{CN}_2(\boldsymbol{H})=\sqrt{\frac{\lambda_{\max}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})}{\lambda_{\min}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})}}=\frac{\sigma_{\max}}{\sigma_{\min}}. \tag{6}$$

Since channels are mapped with different gains, if CN is large, then the gains of different channels differ a lot. Thus, any small disturbance can be amplified on the high gain channel, which could cause a large impact on the estimated values. The matrix with a large CN is qualitatively called an ‘ill-conditioned’ matrix, and the corresponding system of equations is called an ‘ill-conditioned system’; otherwise, $\boldsymbol{H}$ is called a ‘well-conditioned matrix’.

If CN is very large, it is not recommended to force higher rank ill-conditioned equations; instead, the rank and the dimension should be artificially reduced using the initiative of wireless scheduling to reduce CN to a normal level.

Assuming that the gain of logic channel i is σ_i and the transmitted power distributed on the transmitter of channel i is P_i, then the power measured at the receiver of the logic channel is $\sigma_i^2 P_i$. A similar conclusion can be drawn no matter whether the power is distributed using the water injection method or the equal-power method. Taking the equal-power method as an example, according to the Shannon formula, we can obtain

$$\begin{aligned}\frac{C}{W}&=\sum_{i=1}^{m}\log_2(1+\sigma_i^2 P_i/N)\\&=\sum_{i=1}^{m}\log_2[1+\sigma_i^2 P/(mN)].\end{aligned}\tag{7}$$

Further analysis indicates that the system capacity varies with the signal-to-noise ratio (SNR) depending on the CN value. In some cases (a smaller SNR and a larger CN), the system capacity with a larger number of SDMA streams (a higher rank) is even smaller than a system with a smaller number of SDMA streams (a lower rank). Fig. 2 shows a simple example of a second-order case to explain the concept.

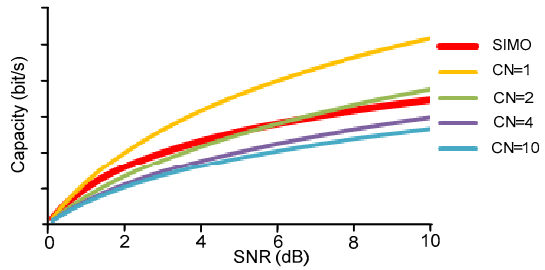

Fig. 2 Condition number impact on capacity (References to color refer to the online version of this figure)

In an actual network, CN is related to the configuration of eNodeB antennas, the configuration of terminal antennas, and the reflectors around the eNodeBs and terminals (angle extension). Generally speaking, the larger the antenna spacing, the more abundant scattering nearby, and a smaller CN is more favorable for a higher SDMA rank. Conversely, the smaller the antenna spacing, the less scattering nearby, and a smaller spread-angle is not good for SDMA. In this case, channels tend to be ill-conditioned (Xiang, 2015).

If channels are ill-conditioned, the CN can be optimized selectively by grouping multiple users. In addition to achieving the optimal capacity, it is sometimes necessary to decrease the number of streams (rank reduction). Thus, it is less sensitive to non-ideal factors (for example, noise) and SDMA is more stable. Otherwise, the noise is greatly amplified and as the number of streams increases, the performance decreases.

3 Space division multiple access under non-ideal channel measurement

In SDMA, the transmitters need to be pre-coded correctly according to the channel situation. However, channels continue to change because of motion of the terminal and surrounding objects. Therefore, it is necessary to measure the instantaneous values of channels from time to time, and then calculate the pre-coding and equalizing matrices based on the measured values.

The channel status is usually measured by two methods: non-feedback mode and feedback mode. The following example describes only the downlink channels (from the eNodeB to a terminal), while the uplink channels are similar.

1. Non-feedback mode. Channels are directly measured in uplink, using uplink-downlink reciprocity (Fig. 3), thus obtaining an estimation of downlink channels. Each antenna of the terminal sends sounding signals at a given frequency and time, and then the eNodeB measures the sounding signals on each channel to obtain the uplink channel and estimate the downlink channel with the assumption of reciprocity. Using this method, the terminal does not need to transmit channel measurement result messages in the uplink, the method is also called ‘standard transparency’. Two characteristics are used, uplink/downlink reciprocity and channel continuous across time. The two characteristics indicate that the method generally applies only to time division duplex (TDD) systems instead of FDD systems, because different frequencies are used in the uplink and downlink in FDD systems. Therefore, the phases of uplink and downlink channels are not the same. The phase difference between uplink and downlink depending on the scattering objects and distance is an unknown quantity.

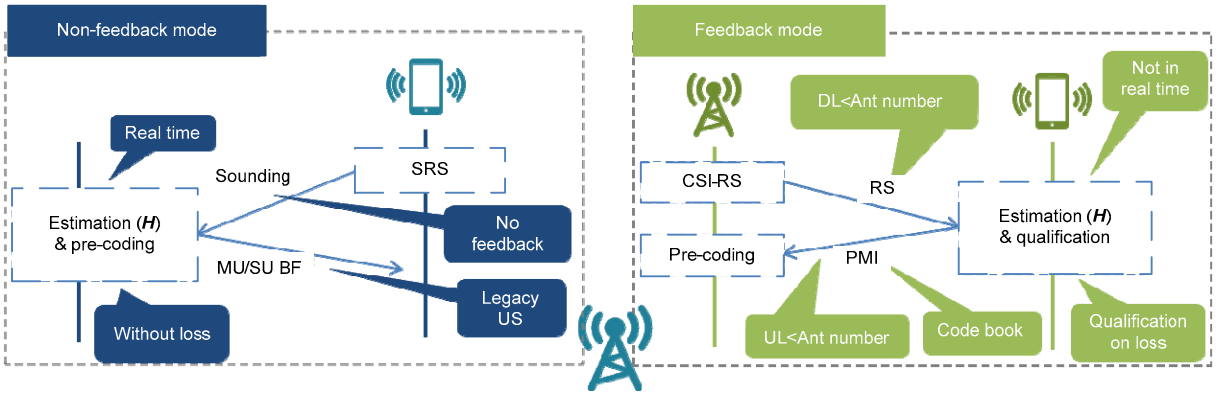

Fig. 3 Feed-back mode and non-feedback mode

2. Feedback mode (Xu and Guan, 2011; Jiang et al., 2016). First, the eNodeB sends a set of channel-state-indicator reference-symbol (CSI-RS) to each channel and the reference symbols are orthogonal to each other (Fig. 3). The terminal measures the reference symbol of each eNodeB channel on all the antennas to directly obtain the downlink channel situation, generates messages, and sends the pre-coding matrix indicator (PMI) to the eNodeB over the uplink channel. In feedback mode, standard transparency is impossible because terminals are involved deeply and the measurement messages should be pre-defined.

When the downlink channels are directly measured, the measurement result seems more accurate in principle. However, the measurement result needs to be reported in the uplink message; thus, the measurement data need to be quantified to reduce the amount of feedback. For example, according to 3GPP specifications, amplitude information is directly removed and the phase is matched among a pre-defined codebook. Eventually, only the codebook index needs to be reported in the uplink message (3GPP, 2009; 3GPP R1-050889, MIMO for Long Term Evolution).

Even though the codebooks are well matched, the overhead of feedback is still very high if there are a large number of eNodeB channels. In addition, the latency for message feedback is large.

In feedback mode, the eNodeB needs to send an independent reference symbol to each channel. According to 3GPP 4G specifications, the overhead of the reference symbol of one channel is about 4% (Fig. 4) (3GPP, 2009). When there are dozens of channels, the overhead can become unacceptable.

In non-feedback mode, the downlink channel is not directly measured. However, with the uplink-downlink reciprocity, direct measurement of the uplink can be treated as direct measurement of the downlink, and this can be absolutely precise. In addition, the process does not depend on the terminal, and it can be standard transparent. It does not suffer from the quantification error. Therefore, normally non-feedback mode performs better.

One limitation of non-feedback mode is its dependence on downlink/uplink (DL/UL) reciprocity. Therefore, the non-feedback method is typically applicable to only TDD but not FDD systems. In the rest of the study, the possibility of the FDD non-feedback mode is investigated.

4 Overhead reduction in feedback mode

With the increase of the number of massive MIMO antennas, more reference symbols are mapped to the physical antennas. However, user equipment (UE) resources are limited, and therefore excessive reference signal overhead can totally block the system. In terms of the feedback mode, it is important to reduce the overhead of the reference symbol for massive antennas.

The sub-dimension-based method is a way to reduce overhead. The reference symbols are measured on the row and the column antennas depending on the antenna topology. Fig. 5 shows an example of a 32-antenna configuration.

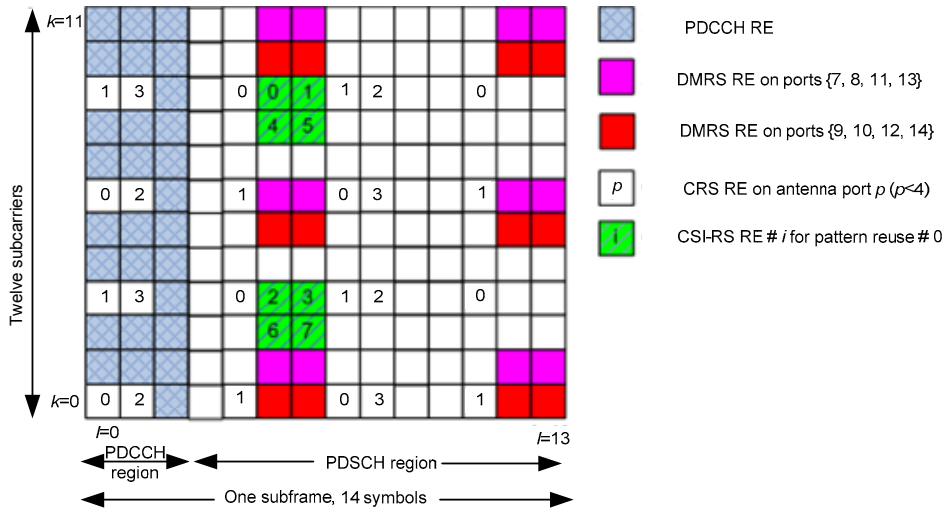

Fig. 4 3GPP long term evolution (LTE) reference symbol mapping (References to color refer to the online version of this figure)

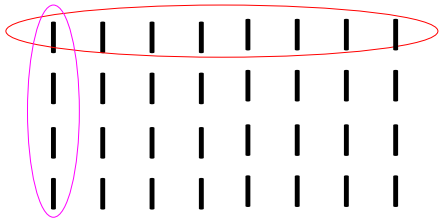

Fig. 5 One option for a 32-antenna configuration

An antenna array can be abstracted into vertical and horizontal antennas, and the channel elements are the Kronecker product of the corresponding horizontal vectors and vertical vectors. In this way, the CSI-RS reference signals are sent in the horizontal and vertical directions, respectively, during RS transmission. A CSI-RS reference signal can be divided into a vertical 4Tx CSI-RS component and a horizontal 8Tx CSI-RS component, greatly reducing the RS overhead. For example, a 32-antenna RS is reduced to (8+4) RS. In such a topology, the existing CSI-RS pilot patterns do not need to be modified for the (8+4)-antenna. Instead, they can be used directly, measured at a different time or on different frequencies, and then combined on the eNodeB.

The disadvantage of the sub-dimension-based method is that, since the Kronecker product is used during reference signal transmission, channels are considered to be highly correlated by default. Therefore, the sub-dimension-based feedback method is applicable only to lower rank scenarios.

The feedback-based design is similar to the pilot-based design in principle. The codebook for different directions can be designed given the reference signal in the horizontal and vertical directions. Based on the antenna topology previously described, when a UE feeds back the measurement, the code index can be reported in the horizontal and vertical directions, respectively (C1 for the vertical dimension and C2 for the horizontal dimension).

C1 and C2 indicate the codes for periodic feedback, and the period of C1 feedback is longer than that of C2.

C1 is an R84 Tx code word (W_3) and C2 is an 8Tx code word (W_1W_2). W_3 of C1 and W_1 of C2 are reported at the same time.

The following description focuses on how to feed back the channel quality indicator (CQI), PMI, and rank-indicator (RI), if all-dimensional pilot information is obtained through interpolation using reference signal compression instead of sub-dimension.

If the UE directly reports ready-made channel information and the measured interference+noise (I+N) statistics to the eNodeB instead of the RI, PMI, and CQI, the eNodeB can obtain ready-made channel-situation and I+N information and accurate precoding information (MUMIMO), but the CQI information is obtained before this reception, without considering the performance of the receiver.

A receiver influence factor G can be given. However, it indicates only part of the receiver's performance.

In addition to the sub-dimension method, the compressed sensing method can be used. Multi-antenna channel response is a function of frequency, time, and space, and is theoretically a random variable.

$$\boldsymbol{H}(\omega,t)=\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nm} \end{pmatrix}. \tag{8}$$

According to many real communication statistics, channels change slowly and this is called 'sparsity of the channel'.

Fig. 6 shows the Matlab simulation. The simulation indicates how the channel varies with frequency and time. The frequency response is indicated for a sub-urban macro-site, urban-macro site, and urban-micro site, respectively (Fig. 6a). The channel varies from 1820 to 1840 MHz. The channel is sparse with frequency, particularly for the sub-urban-macro and urban-micro. The time response is indicated, and it can be seen that the channel is sparse with time, particularly for the sub-urban macro and urban-macro (Fig. 6b). The channel is sparse with space. In addition, in many cases, frequency-domain sparsity, time-domain sparsity, and space-domain sparsity coexist. Due to the existence of the sparsity, channels can be sampled sparsely instead of continuously.

In view of this, the compressed sensing theory is developed for the feedback mode to reduce feedback overhead. However, a major disadvantage of this technology is that it depends strongly on feedback and the system performance cannot be guaranteed because of the lack of prior information.

5 Using non-feedback method in non-reciprocal channel

For non-reciprocal channels (for example, in an FDD system), the feedback method seems to be the only choice. However, the feedback method shows both disadvantages in performance (codebook matching-related degradation, feedback delay-related degradation, and high overhead) and ease of use (standard non-transparency). Therefore, it is necessary to investigate the feasibility of the non-feedback method (standard transparency) for non-reciprocal channels.

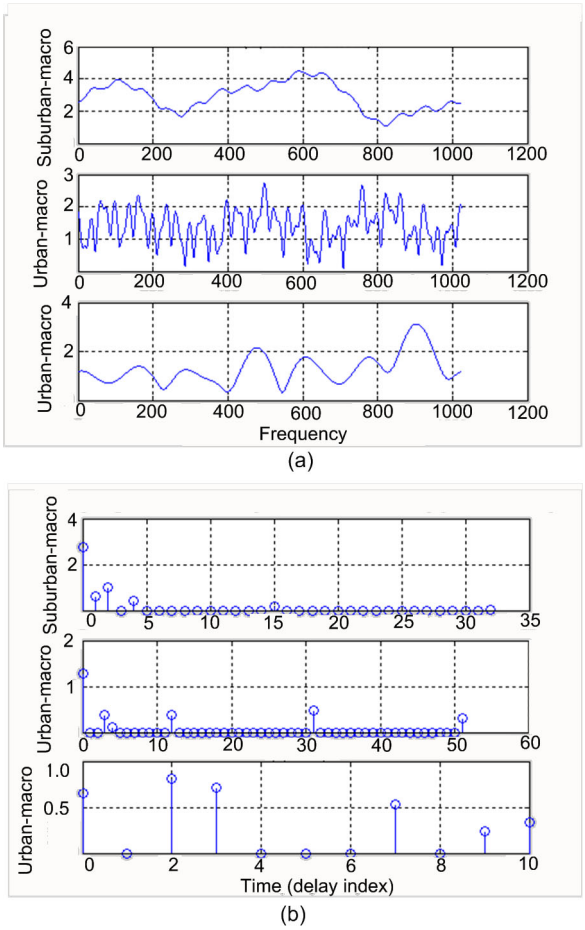

Fig. 6 Sparsity of channels: (a) frequency magnitude in a symbol at different frequencies and in different scenarios (6 path/SCM models); (b) delay magnitude in the same symbol in different scenarios (6 path/SCM models)

We study the statistical characteristics of channels at first. The n-dimensional channel vector (h_1, h_2, …, h_n) is sampled multiple times. For example, for 3GPP a 20 MHz channel is segmented into 100 resource blocks. Then it is sampled at least 100 times. Also, in the time domain, it is sampled multiple times, for example, for one millisecond, at least 14 times for 3GPP. So, (h_{11}, h_{21}, …, h_{n1}), (h_{12}, h_{22}, …, h_{n2}), ..., (h_{1k}, h_{2k}, …, h_{nk}) are obtained.

The second-order mixing center moment of this vector sampling (i, j=1, 2, …, n) is

$$c_{ij}=\mathrm{Cov}(h_i,h)=E\{[h_i-E(h_i)][h_j-E(h_j)]\}. \quad (9)$$

Then the covariance matrix of this vector sampling is

$$\boldsymbol{\Sigma}=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}. \quad (10)$$

Analyzing each element of the covariance matrix, we can obtain

$$\begin{aligned} c_{ij} &=E\{[h_i-E(h_i)][h_j-E(h_j)]\} \\ &=E[h_ih_j-h_iE(h_j)-h_jE(h_i)+E(h_i)E(h_j)] \\ &=E(h_ih_j)-E[h_iE(h_j)]-E[h_jE(h_i)]+E[E(h_i)E(h_j)]. \end{aligned} \quad (11)$$

In terms of a set of samples, considering that the mathematical expectation of each h is a constant instead of a random number, the following equation holds:

$$\begin{aligned} c_{ij} &=E(h_ih_j)-E[h_iE(h_j)]-E[h_jE(h_i)]+E[E(h_i)E(h_j)] \\ &=E(h_ih_j)-E(h_i)E(h_j)-E(h_j)E(h_i)+E(h_i)E(h_j) \\ &=E(h_ih_j)-E(h_i)E(h_j). \end{aligned} \quad (12)$$

Therefore, we have Eq. (13), which is shown on the bottom of this page.

The channel covariance matrix is the mathematical expectation of the channel correlation matrix minus the correlation matrix of the mathematical expectation of the channel.

We analyze the correlation matrix of the mathematical expectation of the channel. In an actual communication system, the pre-coding correlation matrix has certain characteristics. Taking a 4G long term evolution (LTE) system with multi-point sampling on the frequency domain as an example, one LTE carrier bandwidth is 20 MHz. When h-sampling traverses 20 MHz, the arrival phases of the subcarriers are different even if the transmission path is the same, and the phases rotate with the frequencies. The rotation speed is related to the subcarrier spacing and transmission distance.

Suppose the distance from the UE to the eNodeB antenna is d and the center frequency is f. In a simplified scenario with a single path, the absolute transmission phase is $2\pi df/c$, and accordingly the relative phase difference generated on (f+Δf) is

$$\Delta\varphi=\frac{2\pi d(f+\Delta f)}{c}-\frac{2\pi df}{c}=\frac{2\pi d\Delta f}{c}. \quad (14)$$

If d=100 m and Δf=15 kHz per subcarrier, then $\Delta\phi$=1.8°; if d=100 m and Δf=180 kHz per resource block (RB), then $\Delta\phi$=21.6°; if d=100 m and Δf=20 MHz per LTE carrier, then $\Delta\phi$=6.67×360°.

We come to a conclusion that the phase rotates 6.67 times when h-sampling traverses one LTE carrier (20 MHz) even at a shorter transmission distance (100 m). Therefore, the correlation matrix of $\boldsymbol{h}$ can be treated as a constant matrix. In other words, the covariance matrix is the difference between the mathematical expectation of the $\boldsymbol{h}$-vector correlation matrix and a constant matrix.

According to the above analysis, particularly for Eqs. (13) and (14), the preliminary conclusion can be: if the covariance matrix of the h-sampling in a certain bandwidth is independent of the center frequency, the

$$\begin{aligned} \boldsymbol{\Sigma} &= E\begin{pmatrix} h_1h_1 & h_2h_1 & \cdots & h_nh_1 \\ h_1h_2 & h_2h_2 & \cdots & h_nh_2 \\ \vdots & \vdots & & \vdots \\ h_1h_n & h_2h_n & \cdots & h_nh_n \end{pmatrix}-\begin{pmatrix} E(h_1)E(h_1) & E(h_2)E(h_1) & \cdots & E(h_n)E(h_1) \\ E(h_1)E(h_2) & E(h_2)E(h_2) & \cdots & E(h_n)E(h_2) \\ \vdots & \vdots & & \vdots \\ E(h_1)E(h_n) & E(h_2)E(h_n) & \cdots & E(h_n)E(h_n) \end{pmatrix} \\ &= E(h\cdot h^{\mathrm{H}})-\begin{pmatrix} E(h_1)E(h_1) & E(h_2)E(h_1) & \cdots & E(h_n)E(h_1) \\ E(h_1)E(h_2) & E(h_2)E(h_2) & \cdots & E(h_n)E(h_2) \\ \vdots & \vdots & & \vdots \\ E(h_1)E(h_n) & E(h_2)E(h_n) & \cdots & E(h_n)E(h_n) \end{pmatrix}. \end{aligned} \quad (13)$$

mathematical expectation of the $\boldsymbol{h}$-vector correlation matrix can be treated as independent of the center frequency. This constitutes the theoretical basis for the non-feedback method in a non-reciprocal channel scenario.

6 Field test of space division multiple access

The performance of massive MIMO SDMA with 64 channels was field-tested and compared with that of conventional eight-channel eNodeBs. Both massive MIMO and conventional eNodeB were connected with the existing core network. Fig. 7 shows the test site setup.

The UE and eNodeBs fully meet 3GPP R12 and are backward compatible with R8/9. The test was done in Guangzhou, China, in a dense urban situation, where the wireless channel situation is complex, with rich scatter, reflection, and deflection around.

The previous study indicated that, if possible, the non-feedback mode is always preferred. This test verifies the benefits:

1. For this 64-antenna, 8-stream setup, the feedback mode took around 60% resource to do channel indication and feedback. There was more overhead for 16 or 24 streams. However, with the non-feedback mode, the only overhead was uplink sounding which takes 2% resource, and was thus greatly reduced. This agrees with Sections 4 and 5.

2. The feedback mode took at least 5–10 ms, waiting until the next UL time-slot to make a report, and then another 10 ms to make it work. In contrast, in the non-feedback mode, the tested channel in the uplink could be directly used in the downlink, and the latency could be reduced to under 5 ms. This agrees with Section 3.

3. The feedback mode used the 3GPP codebook, which suffers from a large quadrature error, and is mismatched with the complex non-ideal channel. In contrast, the non-feedback mode measured directly the amplitude and phase, and the precision could be around 14 bits for amplitude and 1 degree for phase. Thus, the performance is greatly improved. This again agrees with Section 3.

In addition, as addressed in Section 2, in the situation where the CN and SNR are not that good, the more layers there are; the less the capacity will be. So, the test took an adaptive mechanism to reduce the number of layers according to the TM mode, CN, SNR, and so on. The purpose of this adaptation is to make sure that the performance is always optimized, even when the situation is not that ideal.

In the first test case, eight UEs were roughly uniformly distributed (test points P1 to P8). With a low user correlation, the distribution ratio of users with good, medium, and poor signal quality is 2:4:2, and massive MIMO shows an obvious gain in terms of capacity: (1) the uplink capacity gain of macro-coverage massive MIMO is up to 164%; (2) the downlink capacity gain of macro-coverage massive MIMO is up to 350%.

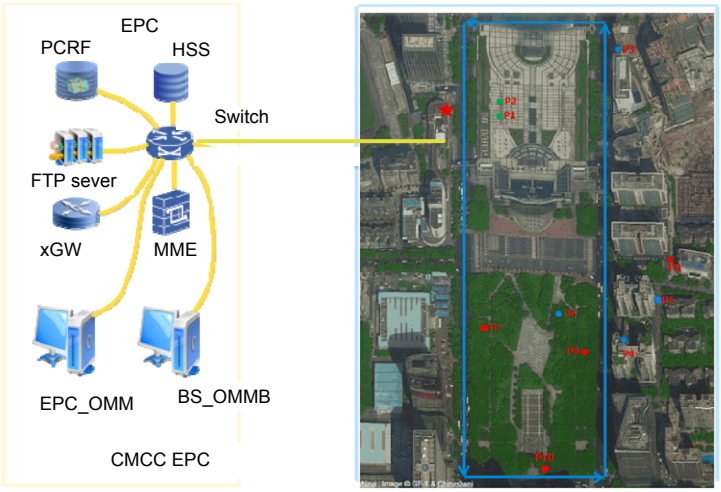

Fig. 7 Tested massive multiple input and multiple output (MIMO) field map

This indicates that even in a complex situation, the adaptive mechanism can achieve very good performance.

The second test case was set to test the extreme worse condition, i.e., to put some UEs very close to each other, which is not good for SDMA because of the high user correlation. The CN was large, and it was ill-conditioned for channel matrices. UE1 to UE4 were located at P3, and UE5 to UE8 were located at P6. The uplink capacity gain of macro-coverage massive MIMO was up to 147%. The downlink capacity gain of macro-coverage massive MIMO was up to 126%.

Although the inter-user interference was large and the CN was also large, if taking the traditional method, then based on physics, there was no way to reach eight streams; two streams would be the most. Massive MIMO took dynamic methods, and still led to up to 4–8 streams, depending on how bad the situation is. The overall throughput is still very good.

7 Conclusions

In field deployment, there are many non-ideal factors that reduce SDMA performance. For example, when the channel situation is not ideal, a high inter-layer correlation will be introduced, and the CN value for SDMA will be large. The measurement of the channel situation is not precise and not in real time. The overhead of the channel measurement is too high. We have analyzed these non-ideal factors. The impact of these non-ideal factors was studied, and methods to optimize the performance were proposed. When the SDMA channel is of high inter-layer correlation and the CN is large, the MUMIMOUE grouping should be optimized accordingly, and in some cases the dimension should be further reduced.

Even for some non-reciprocal channels, we still prefer non-feedback mode, because it takes less overhead. This is very critical for very large-scale antennas.

A field test for the massive MIMO SDMA has been given. The results indicated that with the adaptive-rank mechanism, the performance was optimized in all cases.

References

3GPP, 2009. Physical layer procedures. 3GPP TS 36.211 V8.6.0 (2009-05).

Adian MG, 2014. Beamforming with reduced complexity in MIMO cooperative cognitive radio networks. *J Optim*, 2014:325217. https://doi.org/10.1155/2014/325217

Adian MG, Adyan MG, 2014. Optimal and suboptimal resource allocation in MIMO cooperative cognitive radio networks. *J Optim*, 2014:190-196. https://doi.org/10.1155/2014/190196

Dahlman E, Parkvall S, Sköld J, 2016. New 5G radio-access technology. In: Dahlman E, Parkvall S, Sköld J (Eds.), 4G LTE-Advanced Pro and the Road to 5G (3rd Ed.). Elsevier Ltd., Amsterdam. https://doi.org/10.1016/B978-0-12-804575-6.00024-8

Jarry P, Beneat JN, 2015. 12-Modulations and demodulations without noise. In: Jarry P, Beneat JN (Eds.), Digital Communications. Elsevier Ltd., Amsterdam. https://doi.org/10.1016/B978-1-78548-037-9.50012-0

Jiang J, Kong DT, Chattha HT, 2016. Joint user scheduling and MU-MIMO hybrid beamforming algorithm for mmWave FDMA massive MIMO system. *Int J Antennas Propag*, 2016:4341068. https://doi.org/10.1155/2016/4341068

Karasawa Y, 2016. On physical limit of wireless digital transmission from radio wave propagation perspective. *Radio Sci*, 51(9):1600-1612. https://doi.org/10.1002/2016RS006040

Lu ZH, Zhang XD, Xiao HH, et al., 2014. Key technology research on large-scale antenna array system. *Video Eng*, 38(5):132-135 (in Chinese). https://doi.org/10.3969/j.issn.1002-8692.2014.05.035

Melo BLR, Cunha DC, Pimentel C, 2016. Optimal power distribution in non-binary LDPC code-based cooperative wireless networks. *Comput Netw*, 100:157-165. https://doi.org/10.1016/j.comnet.2016.02.022

Sodha J, 2016. Check node LDPC decoder synchronization. *Int J Electron Lett*, 4(3):287-295. https://doi.org/10.1080/21681724.2015.1034190

Xiang JY, 2015. Evolution of wireless new technologies. *ZTE Technol J*, 21(3):50-54 (in Chinese). https://doi.org/10.3969/j.issn.1009-6868.2015.03.012

Xu RF, Guan JX, 2011. Discrimination on the principles of beamforming in multiple antenna systems. *Ship Electron Eng*, 31(9):73-76 (in Chinese). https://doi.org/10.3969/j.issn.1627-9730.2011.09.022

Zhang TQ, Qian WR, Zhang G, et al., 2016. Parameter estimation of MC-CDMA signals based on modified cyclic autocorrelation. *Dig Signal Process*, 54:46-53. https://doi.org/10.1016/j.dsp.2016.03.007