Chapter 12
Sparse Code Multiple Access (SCMA)

Zheng Ma and Jinchen Bao

12.1 General Description

Overloaded systems, in which the number of users is greater than the dimension of signal-space, are of practical interest in bandwidth-efficient multi-user communications. One kind of such systems is sparse code multiple access (SCMA), which is a promising code-domain non-orthogonal multiple access technique to address the challenges for the fifth-generation (5G) mobile networks [1–4]. Non-orthogonal multiple access has the potential to accommodate more users with limited resources, which provides many advantages over orthogonal multiple access including multi-user capacity, supporting overloaded transmission, enabling reliable and low latency grant-free transmission, enabling flexible service multiplexing, etc. Applications of non-orthogonal signaling for multi-user communications have been investigated several years ago, significant efforts were paid to the optimal signaling design and intensive multi-user detection techniques, to suppress the multiple access interference (MAI) for lowering probability of error or increasing capacity. Hoshyar and Guo suggest the low-density signature (LDS)-based multiple access [5], or sparsely spread code-division multiple access (CDMA) [6], which intentionally arranges each user to spread its data over a fraction of the chips, instead of all chips, to reduce both the MAI and the complexity of multi-user detection. Inspired by the overloading capability and the low-complexity feature of LDS, SCMA is developed by inheriting from LDS the sparse sequence structure, such that the message-passing algorithm (MPA) is available in multi-user detection to achieve near-optimal performance. In contrast to the LDS scheme, multi-dimensional signal constellations, instead of the spreading, are utilized in SCMA to combat the channel fading and MAI. As a result,

Z. Ma (✉) · J. Bao
Southwest Jiaotong University, West Section, High-tech Zone, Chengdu, Sichuan, China
e-mail: zma@swjtu.edu.cn

J. Bao
e-mail: jinchen_bao@my.swjtu.edu.cn

M. Vaezi et al. (eds.), *Multiple Access Techniques for 5G Wireless Networks and Beyond*, https://doi.org/10.1007/978-3-319-92090-0_12

the larger coding gain and better spectrum efficiency are achievable for SCMA due to the improved codebooks, compared to LDS.

As one of NOMA family, SCMA is capable of supporting overloaded access over the coding domain, hence increasing the overall rate and connectivity. By carefully designing the codebook and multi-dimensional modulation constellations, the coding and shaping gain can be obtained simultaneously. In an SCMA system, users occupy the same resource blocks in a low-density way, which allows affordable low multi-user joint detection complexity at receiver. The sparsity of signal guarantees a small collision even for a large number of concurrent users, and the spread-coding like codes design brings good coverage and anti-interference capability due to spreading gain as well.

12.1.1 System Model

12.1.1.1 Multiple Access Procedure

An SCMA transmission system can be simply illuminated in Fig. 12.1. Suppose that there are J synchronous users multiplexing over K shared orthogonal resources, e.g., K time slots or orthogonal frequency division multiplexing (OFDM) tones, and each user employs one SCMA layer.[1] The forward error control (FEC) coding scheme can be low-density parity-check (LDPC) codes or polar codes which have been adopted for 5G recently. Each SCMA modulator/encoder maps the coded bits to a K-dimensional complex codeword, and the resulted J codewords constitute an SCMA block, as is shown in Fig. 12.1 ($J = 6$, $K = 4$ in the figure). The multi-user codewords in each SCMA block are multiplexed over the air transmissions in uplink multiple access channel (MAC), or they are superimposed at the transmitter of the downlink broadcast channel (BC). Since each SCMA block occupies K resources for codeword transmitting, the resulted *overloading factor* is J/K. This multiple access process is similar to that of CDMA, where the spread signals in CDMA are replaced with the SCMA codewords. Multi-user detection is carried out at the receiver to recover the colliding codewords.

For the uplink MAC, the received signal vector after the synchronous user multiplexing is expressed as

$$\mathbf{y} = \sum_{j=1}^{J} \mathrm{diag}(\mathbf{h}_j)\mathbf{x}_j + \mathbf{n} \tag{12.1}$$

where $\mathbf{x}_j = \left[x_j[1], \ldots, x_j[K]\right]^t$ and $\mathbf{h}_j = [h_j[1], \ldots, h_j[K]]^t$, are the K-dimensional codeword and the corresponding channel gain for the jth user, respectively, and $\mathrm{diag}(\mathbf{h}_j)$ denotes the diagonal matrix with $h_j[k]$ being the kth diagonal

[1] In practical scenarios, each user employs one or multiple layers.

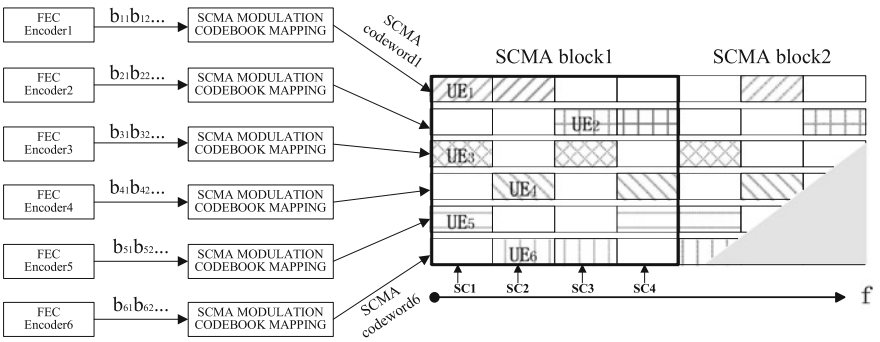


Fig. 12.1 The system model for SCMA

element. The K-vector $\mathbf{n}$ is the additive white Gaussian noise (AWGN) with zero mean and variance N_0 per dimension. It is convenient to view the MAC model as an equivalent "MIMO" communication system, and the received vector in (12.1) becomes

$$\mathbf{y} = \mathbf{HX} + \mathbf{n} \tag{12.2}$$

where $\mathbf{H} = \left[\text{diag}(\mathbf{h}_1), \text{diag}(\mathbf{h}_2), \ldots, \text{diag}(\mathbf{h}_J)\right]$, is the equivalent "MIMO" channel matrix, and $\mathbf{X} = \left[\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_J^t\right]^t$, is the combined multi-user codeword representing an SCMA block.

For the downlink BC, the codewords from multiple users are superimposed before the transmission, so that they experience the same fading. In the case of absence of interference between K resources, the received signal vector is given by

$$\mathbf{y} = \text{diag}(\mathbf{h}) \sum_{j=1}^{J} \mathbf{x}_j + \mathbf{n} = \text{diag}(\mathbf{h})\mathbf{X} + \mathbf{n} \tag{12.3}$$

where a single receiver is considered here for simplicity, and $\mathbf{X} = \sum_{j=1}^{J} \mathbf{x}_j$, is the superimposed codeword of J users at the input of a BC, which also represents an SCMA codeword block.

In the following, the upper case $\mathbf{X}$ always denotes the combined multi-user codeword of J users in the MAC model, or the superimposed codeword in the BC model.

12.1.1.2 SCMA Codebook Mapping

Unlike the modulation used for 3G and 4G, the modulation and codebook mapping in SCMA are designed jointly in a multi-dimensional and sparsely spread way. An SCMA modulator/encoder maps the input bits to a K-dimensional sparse codeword,

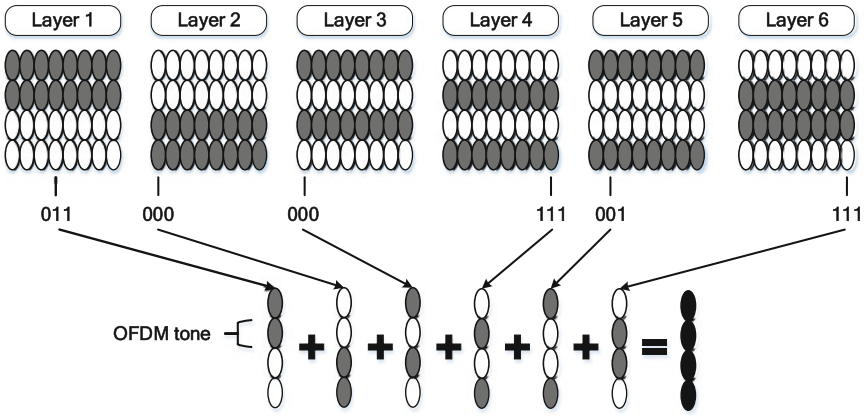


Fig. 12.2 Illustration of SCMA codebooks and bits to codeword mapping

which is selected from a layer-specific codebook of size M. The K-dimensional complex codewords of the codebook are sparse vectors with $N < K$ nonzero entries, and all the codewords contain 0 in the same dimensions. Then, the codebook is sparse, and this is where the "sparse code multiple access" is named from.

The codebooks are constructed by a mapping from an N-dimensional complex constellation with a mapping matrix. Denote the constellation for the jth layer/user with $\mathbb{C}_j$, which contains M_j constellation points of length N_j. The mapping matrix $\mathbf{V}_j$ maps the N_j-dimensional constellation points to SCMA codewords to form the codebook $\mathbb{X}_j$. To simplify our illumination and analysis, we assume that all layers have the same constellation size and length, i.e., $M_j = M$, $N_j = N$, $\forall j$. In summary, the resulting codebook for the jth user contains M codewords, each codeword consists of K complex values from which only N are nonzero specified by the mapping matrix $\mathbf{V}_j$.

An example of the codebook mapping is shown in Fig. 12.2, where a codebook set containing 6 codebooks for transmitting 6 SCMA layers is illustrated ($J = 6$). Each codebook contains 8 four-dimensional codewords ($M = 8$, $K = 4$), and two of the four entries in the codewords are nonzero ($N = 2$). Upon transmission, the codeword of each layer is selected based on the labeling of the bit sequence.

12.1.1.3 Factor Graph Representation

The low-density structure of SCMA codewords can be efficiently characterized by a factor graph, which is analogous to that for LDPC codes. A binary column vector $\mathbf{f}_j$ of length K is used to indicate the positions of zero (with digit 0) and nonzero (with digit 1) entries of the jth codebook. Then, a $K \times J$ sparse matrix $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_J]$, called *factor graph matrix*, can be used to indicate the relationships between the layers and resources. The rows of $\mathbf{F}$ indicate the resources and the columns indicate

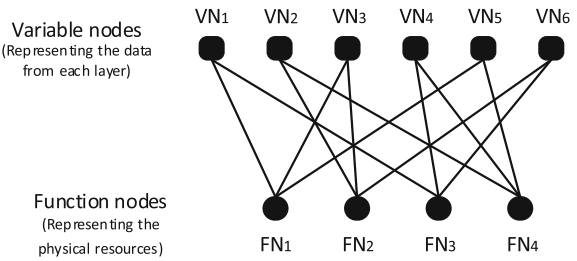


Fig. 12.3 Factor graph representation for SCMA

the layers. The (k, j)th element of $\mathbf{F}$, denoted as $f_{k,j}$, is 1 if the jth layer contributes its data to the kth resource.

Correspondingly, let the J variable nodes (VNs) and K function nodes (FNs) in the factor graph represent the layers and resources, respectively, and the jth VN is connected to the kth FN if and only if $f_{k,j} = 1$. In the following, we denote

$$\begin{aligned} \phi_k &= \{j : 1 \le j \le J, f_{k,j} = 1\}, \\ \varphi_j &= \{k : 1 \le k \le K, f_{k,j} = 1\} \end{aligned} \tag{12.4}$$

the index set of layers contributing to the kth resource, and the index set of resources occupied by the jth layer, respectively. For a regular factor graph matrix, $|\phi_1| = \cdots = |\phi_K|$ and $|\varphi_1| = \cdots = |\varphi_J|$, and let $d_f = |\phi_k|$ and $d_v = |\varphi_j|$.

Example 1 Consider a 6-user SCMA transmission system with $J = 6$, $K = 4$, such a system permits a transmission overloading 150%, and the system model is depicted in Fig. 12.1. If we carefully design the factor graph matrix $\mathbf{F}$ to allow the users to collide over only one nonzero element, then a choice of $\mathbf{F}$ is given by

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{12.5}$$

In the sparse matrix settings, matrix (12.5) has $d_f = 3$ and $d_v = 2$, which means that each FN is connected to three VNs and each VN is connected to two FNs. The corresponding factor graph is shown in Fig. 12.3, and an example of a codebook (with size $M = 4$) is listed in Table 12.1.

In summary, the main features of SCMA lie in:

- Code-domain non-orthogonal multiplexing: SCMA allows superposition of multiple codewords from different users over several resources, which supports overloading. The superposition pattern on each resource is defined in codebooks.
- Sparse spreading: SCMA uses sparse spreading to reduce inter-layer interference, so that more codewords collisions can be tolerated with low receiver complexity.

Table 12.1 An Example of SCMA Codebook ($K = M = 4$, $N = 2$, $J = 6$)

SCMA codebook index	SCMA Codebook for each layer
Codebook 1	$\begin{bmatrix} 0.7851 & -0.2243 & 0.2243 & -0.7851 \\ 0 & 0 & 0 & 0 \\ -0.1815-0.1318i & -0.6351-0.4615i & 0.6351+0.4615i & 0.1815+0.1318i \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Codebook 2	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.1815-0.1318i & -0.6351-0.4615i & 0.6351+0.4615i & 0.1815+0.1318i \\ 0 & 0 & 0 & 0 \\ 0.7851 & -0.2243 & 0.2243 & -0.7851 \end{bmatrix}$
Codebook 3	$\begin{bmatrix} -0.6351+0.4615i & 0.1815-0.1318i & -0.1815+0.1318i & 0.6351-0.4615i \\ 0.1392-0.1759i & 0.4873-0.6156i & -0.4873+0.6156i & -0.1392+0.1759i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Codebook 4	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.7851 & -0.2243 & 0.2243 & -0.7851 \\ -0.0055-0.2242i & -0.0193-0.7848i & 0.0193+0.7848 & 0.0055+0.2242i \end{bmatrix}$
Codebook 5	$\begin{bmatrix} -0.0055-0.2242i & -0.0193-0.7848i & 0.0193+0.7848 & 0.0055+0.2242i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -0.6351+0.4615i & 0.1815-0.1318i & -0.1815+0.1318i & 0.6351-0.4615i \end{bmatrix}$
Codebook 6	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.7851 & -0.2243 & 0.2243 & -0.7851 \\ 0.1392-0.1759i & 0.4873-0.6156i & -0.4873+0.6156i & -0.1392+0.1759i \\ 0 & 0 & 0 & 0 \end{bmatrix}$

- Multi-dimensional modulation: SCMA employs multi-dimensional constellations for better spectral efficiency.

12.1.2 Multi-user Detection

This subsection discusses multi-user detection schemes for SCMA, including the optimal detection, the MPA receiver and other advanced receivers.

12.1.2.1 Optimal/Quasi-optimal Multi-user Detection

A. Optimal Multi-user Detection

Assume that channel state is perfectly estimated at the receiver, given the received signal vector $\mathbf{y}$, the joint optimum maximum a posteriori probability (MAP) detection, for multi-user codeword $\mathbf{X}$ and for the jth user's codeword $\mathbf{x}_j$, can be written as

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{x}_j \in \mathbb{X}_j, \forall j} p(\mathbf{X}|\mathbf{y}), \quad \text{and} \quad \hat{\mathbf{x}}_j = \arg \max_{\mathbf{x}_j \in \mathbb{X}_j} \sum_{\mathbf{x}_i \in \mathbb{X}_i, \forall i \neq j} p(\mathbf{X}|\mathbf{y}) \tag{12.6}$$

respectively. Using Bayes' rule

$$p(\mathbf{X}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{y})} \propto p(\mathbf{y}|\mathbf{X})P(\mathbf{X})$$

where

$$p(\mathbf{X}) = \prod_{j=1}^{J} p(\mathbf{x}_j), \quad \text{and} \quad p(\mathbf{y}) = \sum_{\mathbf{x}_j \in \mathbb{X}_j, \forall j} p(\mathbf{y}|\mathbf{X})p(\mathbf{X})$$

are the joint a prior probability[2] for each user's codeword, and the probability of the received signal vector, respectively.

By assuming that the noise components over the K resources are identically independently distributed (i.i.d.), it holds

$$p(\mathbf{y}|\mathbf{X}) = \prod_{k=1}^{K} p(y[k]|\mathbf{X})$$

and considering only d_f users actually collided over the kth resource, we have

$$p(y[k]|\mathbf{X}) = \frac{1}{\pi N_0} \exp\left(-\frac{1}{N_0} \left| y[k] - \sum_{j \in \phi_k} h_j[k] x_j[k] \right|^2 \right) \tag{12.7}$$

Thus, the MAP decision for the jth user's codeword is given by

$$\hat{\mathbf{x}}_j = \arg \max_{\mathbf{x}_j \in \mathbb{X}_j} \sum_{\mathbf{x}_i \in \mathbb{X}_i, \forall i \neq j} P(\mathbf{X}) \prod_{k=1}^{K} p(y[k]|\mathbf{X}), \quad \forall j \tag{12.8}$$

With the codeword probability for each user, it is straightforward to calculate the log-likelihood rate (LLR) for each coded bit, so that they can serve as the input for

[2] Without feedback from the FEC decoder, $p(\mathbf{x}_j) = \frac{1}{M}$ for all the users.

the FEC decoder. For the jth user, the LLR considering the mth bit $b_{j,m}$ is calculated by

$$\begin{aligned}\Lambda(b_{j,m}) &= \log \frac{\Pr\{b_{j,m}=1|\mathbf{y}\}}{\Pr\{b_{j,m}=0|\mathbf{y}\}} \\ &= \log \frac{\sum_{\mathbf{x}_j \in \mathbb{X}_{j,m}^1} \sum_{\mathbf{x}_i \in \mathbb{X}_i, \forall i \neq j} P(\mathbf{X}) \prod_{k=1}^{K} p(y[k]|\mathbf{X})}{\sum_{\mathbf{x}_j \in \mathbb{X}_{j,m}^0} \sum_{\mathbf{x}_i \in \mathbb{X}_i, \forall i \neq j} P(\mathbf{X}) \prod_{k=1}^{K} p(y[k]|\mathbf{X})}\end{aligned} \tag{12.9}$$

where $\mathbb{X}_{j,m}^1$ and $\mathbb{X}_{j,m}^0$ are subsets of $\mathbb{X}_j$ for which the mth bit of the jth user $b_{j,m}=1$ and $b_{j,m}=0$, respectively. Note that solving (12.8) is equivalent to solve the marginal product of functions (MPF) problem, which is of exponential complexity with *brute-force* searching, and is prohibitive to employ when the number of users increases.

B. MPA Detection

As the SCMA encoding can be represented by a factor graph with sparse property, the low-complexity MPA can be used to solve the MPF problem with near-optimum performance.

Let $I_{k\to j}^{(t)}$ be the extrinsic information to be passed from FN k to VN j at the tth iteration, and $I_{j\to k}^{(t)}$ be the extrinsic information to be passed from VN j to FN k. Given the *a prior* probability $p(\mathbf{x}_j)$, the probability that $\mathbf{x}_j$ is transmitted by the jth user given the channel sample is updated as

$$I_{j\to k}^{(t)}(\mathbf{x}_j) = p(\mathbf{x}_j) \prod_{l \in \varphi_j \backslash k} I_{l\to j}^{(t)}(\mathbf{x}_j)$$

Then, for any $\mathbf{x}_j \in \mathbb{X}_j$, the probability of the received signal $y[k]$ given that $\mathbf{x}_j$ is transmitted by the jth user, marginalized over all possible codewords of the other users, is given by

$$I_{k\to j}^{(t)}(\mathbf{x}_j) = \sum_{\mathbf{x}_i \in \mathbb{X}_i, \forall i \in \phi_k \backslash j} p\left(y[k]|\mathbf{X}\right) \prod_{i \in \phi_k \backslash j} I_{i\to k}^{(t-1)}(\mathbf{x}_i) \tag{12.10}$$

After a number of iterations, the posterior probability of $\mathbf{x}_j$ produced by the MPA is proportional to

$$I_j(\mathbf{x}_j) = p(\mathbf{x}_j) \prod_{k \in \varphi_j} I_{k\to j}^{(T)}(\mathbf{x}_j), \quad \mathbf{x}_j \in \mathbb{X}_j, j = 1, \ldots, J \tag{12.11}$$

where T is the number of iterations at the termination.

Similar to that for MAP detection, the LLR of the mth bit of the jth user $b_{j,m}$ is calculated by

$$\Lambda(b_{j,m}) = \log \frac{\sum_{\mathbf{x}_j \in \mathbb{X}_{j,m}^1} I_j(\mathbf{x}_j)}{\sum_{\mathbf{x}_j \in \mathbb{X}_{j,m}^0} I_j(\mathbf{x}_j)} \tag{12.12}$$

where $\mathbb{X}_{j,m}^{1}$ and $\mathbb{X}_{j,m}^{0}$ are the same as that in (12.9).

The main complexity of MPA comes from the calculation of (12.10), the summation over $\mathbf{x}_i$ adds up $M^{|\phi_k|-1}$ terms while M probabilities should be calculated in each iteration, which leads to a complexity order $O(TKM^{d_f})$, and is far below that of the optimal MAP detection. In practical implementations, the exponential function in MPA algorithm may cause large dynamic ranges and very high storage burden, then the logarithmic domain MPA is preferred to avoid the exponential operations. For the log-MPA operation, the information exchanged between the FNs and VNs can be expressed as

$$I_{j\to k}^{(t)}(\mathbf{x}_j) = \log p(\mathbf{x}_j) + \sum_{l\in\varphi_j\backslash k} I_{l\to j}^{(t)}(\mathbf{x}_j)$$

$$I_{k\to j}^{(t)}(\mathbf{x}_j) = \max_{\mathbf{x}_i\in\mathbb{X}_i,\forall i\in\phi_k\backslash j} \left\{ \log p\left(y[k]|\mathbf{X}\right) + \sum_{i\in\phi_k\backslash j} I_{i\to k}^{(t-1)}(\mathbf{x}_i) \right\}$$

where Jacobi's logarithm formula $\log\left(\sum_i e^{p_i}\right) \approx \max_i p_i$ is applied for a complexity reduction to a certain degree, which results in the max-log-MPA detection. The output LLR of the MPA detector is given by

$$\Lambda(b_{j,m}) = \max_{\mathbf{x}_j\in\mathbb{X}_{j,m}^{1}} I_j(\mathbf{x}_j) - \max_{\mathbf{x}_j\in\mathbb{X}_{j,m}^{0}} I_j(\mathbf{x}_j)$$

where

$$I_j(\mathbf{x}_j) = \log p(\mathbf{x}_j) + \sum_{k\in\varphi_j} I_{k\to j}^{(T)}(\mathbf{x}_j)$$

12.1.2.2 Other Advanced Detectors

The MPA detector still has exponential complexity with respect to the codebook size (M) and the number of accessed users at each resource (d_f), which may become impractical for the implementation of very large codebook size (e.g., $M \geq 64$) and very high overload (e.g., $d_f \geq 8$). Some other advanced detectors can harness the potential gain of SCMA while provide sufficient flexibility for a good trade-off between the performance and detection complexity [7, 8].

A. EPA Detector

Expectation propagation algorithm (EPA) is an approximate Bayesian inference method in machine learning for estimating sophisticated posterior distributions with simple distributions through distribution projection, and it turns out to be an efficient iterative multi-user detector for SCMA as well as some other multiple access schemes [8]. It approximates the discrete message in MPA as continuous Gaussian message using the minimum Kullback–Leibler (KL) divergence criterion, and use the a posteriori probabilities fed back from the FEC decoder to compute the approx-

imate symbol belief and the approximate message, such that the message passing reduces to mean and variance parameters update. The detailed algorithm is given in [8]. With EPA, the complexity order of SCMA detection is reduced to linear complexity, i.e., it only scales linearly with the codebook size M and the average degree of the factor nodes d_f, while simulation results show that the EPA detector shows nearly the same error performance as MPA for SCMA with receiver diversity. As a result, the computation burden of the SCMA receiver is significantly alleviated and is no longer a problem for implementation in real systems.

B. SIC-MPA Detector

Successive interference cancelation (SIC) receiver is a kind of multi-user receiver that treats all the other undecoded users as interference when decoding a target user, and can be implemented as either symbol level or codeword level. It works well when the received SNR among users are quite different from each other. However, the detecting performance deteriorates when the SNR difference is not obvious between users, in which case error propagation happens.

To strike a good balance between link performance and implementation complexity, it is reasonable to combine SIC with an MPA (SIC-MPA) receiver. More specifically, MPA is applied to a limited number of users firstly, so that the number of colliding users over each resource does not exceed a given threshold value (e.g., d_s users). Then, the successfully decoded users are removed by SIC and the procedure continues until all users are successfully decoded or no new user gets successfully decoded in MPA. In the case of $d_s = d_f$, full MPA is realized, and when $d_s = 1$, it becomes a pure SIC receiver. Due to the fact that MPA is used for a very limited number of users instead of all the users, the decoding complexity is greatly reduced, which is of the order $O(TKM^{d_s})$.

12.2 Performance Evaluation

The error performance and capacity are excellent measures that indicating the goodness of a system, and more importantly, they serve as powerful tools for the practical system design. For SCMA, the multi-user codebook plays a key role in the system performance improvement, and it is necessary to establish performance criterion to guide the codebooks design. In this section, error performance and capacity analysis for uplink and downlink SCMA systems are provided, and independent Rayleigh fadings are assumed.

12.2.1 Average Error Probability

The error probability, e.g., the average codeword error probability (ACEP), is one of the most important performance criteria, since it is most revealing about the nature

of a system behavior. However, it is quite difficult to evaluate the exact ACEP for SCMA systems, since one needs to average over several fading statistics due to the multi-channel transmissions, and the integration involves a decision cell of a multi-dimensional signal point. As an alternative approach, it is convenient to resort to an upper bound or approximation on the ACEP [9]. In this subsection, we use union bound to evaluate the error performance of uplink and downlink SCMA under joint maximum likelihood (ML) multi-user detection.

12.2.1.1 PEP over Uplink MACs

Consider the equivalent "MIMO" channel (12.2). Under the assumption of perfect channel estimation at the receiver, the joint ML detection of multi-user codewords is equivalent to the joint minimum distance decoding

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{y} - \mathbf{H}\mathbf{X}\|$$

The pairwise error probability (PEP), defined as the probability that received signal vector $\mathbf{y}$ is detected into $\mathbf{X}_b$ given that $\mathbf{X}_a$ is transmitted, is given by [10]

$$P\{\mathbf{X}_a \to \mathbf{X}_b\} = \mathbb{E}_{\mathbf{H}}\left[Q\left(\sqrt{\frac{\|\mathbf{H}(\mathbf{X}_a - \mathbf{X}_b)\|^2}{2N_0}} \right) \right] \tag{12.13}$$

where, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} \mathrm{d}t$, is the well-known Gaussian function [11], and $\mathbb{E}_{\mathbf{H}}[\cdot]$ denotes the mean.

Let $x_{j,a}[k]$ and $x_{j,b}[k]$ be the kth entries of the jth user's codewords $\mathbf{x}_{j,a}$ and $\mathbf{x}_{j,b}$, corresponding to $\mathbf{X}_a$ and $\mathbf{X}_b$, respectively. Due to the sparseness of the codewords, $x_{j,a}[k] = x_{j,b}[k] = 0$ whenever $j \notin \phi_k$. Now we define a distance for the MAC.

Definition 1 The kth dimension-wise distance, between the multi-user combined codewords $\mathbf{X}_a$ and $\mathbf{X}_b$, for the uplink MAC is defined as

$$\lambda_k^2 = \sum_{j=1}^{J} |x_{j,a}[k] - x_{j,b}[k]|^2 = \sum_{j \in \phi_k} |\delta_j[k]|^2, \quad \forall k \tag{12.14}$$

where $\delta_j[k] = x_{j,a}[k] - x_{j,b}[k]$.

Assume that there are repeated values among the set $\{\lambda_1^2, \ldots, \lambda_K^2\}$, such that they can be divided into V $(1 \le V \le K)$ groups, and each group contains the collection of a certain value $\hat{\lambda}_v^2$. Let $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1^2, \ldots, \hat{\lambda}_V^2]^t$, be the vector of V distinct elements among $\{\lambda_1^2, \ldots, \lambda_K^2\}$, and $\mathbf{r} = [r_1, \ldots, r_V]^t$, where r_v is the number of elements in $\{\lambda_1^2, \ldots, \lambda_K^2\}$ that equals to $\hat{\lambda}_v^2$, such that $\sum_{v=1}^{V} r_v = K$.

Definition 2 Define the parameter

$$A_{\mathbf{r},\boldsymbol{\lambda}} = \prod_{k=1}^{K} \lambda_k^{-2} = \prod_{v=1}^{V} \hat{\lambda}_v^{-2r_v} \tag{12.15}$$

as the reciprocal of the product of the dimension-wise distances.

Definition 3 For positive integers l, v and vectors $\mathbf{r}$ and $\hat{\boldsymbol{\lambda}}$, define the parameter

$$B_{v,l,\mathbf{r},\hat{\boldsymbol{\lambda}}} = (-1)^{l+1} \sum_{\boldsymbol{\eta}\in\Omega_{v,l}} \prod_{j=1, j\neq v}^{V} \binom{\eta_j + r_j - 1}{\eta_j} \left(\frac{1}{\hat{\lambda}_j^2} - \frac{1}{\hat{\lambda}_v^2}\right)^{-(r_j+\eta_j)} \tag{12.16}$$

where the vector $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_V]^t$ is created from the set $\Omega_{v,l}$ of all nonnegative integer partitions of $l-1$ (with $\eta_v = 0$). The set $\Omega_{v,l}$ is defined as

$$\Omega_{v,l} = \left\{ \boldsymbol{\eta} = [\eta_1, \ldots, \eta_V]^t \in \mathbb{Z}^V ; \sum_{j=1}^{V} \eta_j = l-1, \eta_v = 0, \eta_j \geq 0 \; \forall j \right\}$$

Next, we provide the main result regarding the PEP.

Theorem 1 *For J users using K-dimensional codebooks in the uplink SCMA systems, the PEP between $\mathbf{X}_a$ and $\mathbf{X}_b$ is given by*

$$\begin{aligned} P\{\mathbf{X}_a \to \mathbf{X}_b\} = A_{\mathbf{r},\hat{\boldsymbol{\lambda}}} \sum_{v=1}^{V} \sum_{L=1}^{r_v} \hat{\lambda}_v^{2L} B_{v,r_v-L+1,\mathbf{r},\hat{\boldsymbol{\lambda}}} \\ \times \left(\frac{1-\mu_v}{2}\right)^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} \left(\frac{1+\mu_v}{2}\right)^k \end{aligned} \tag{12.17}$$

where

$$\mu_v = \sqrt{\frac{\hat{\lambda}_v^2}{4N_0 + \hat{\lambda}_v^2}}$$

Proof Consider the metric in (12.13),

$$\|\mathbf{H}(\mathbf{X}_a - \mathbf{X}_b)\|^2 = \sum_{k=1}^{K} \|\mathbf{h}[k]^\dagger (\mathbf{x}_a[k] - \mathbf{x}_b[k])\|^2$$

where $\mathbf{x}_a[k] = [x_{1,a}[k], \ldots, x_{J,a}[k]]'$, is the vector of the kth component for J users, and $\mathbf{h}[k]^\dagger = [h_1[k], \ldots, h_J[k]]$, are the corresponding channel gains, and $[\cdot]^\dagger$ denotes conjugate transpose. Using the matrix decomposition, it holds that

$$(\mathbf{x}_a[k] - \mathbf{x}_b[k])(\mathbf{x}_a[k] - \mathbf{x}_b[k])^\dagger = \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^\dagger$$

where $\mathbf{U}_k$ is unitary and $\boldsymbol{\Lambda}_k$ is a diagonal matrix, i.e., $\boldsymbol{\Lambda}_k = \text{diag}(\tilde{\lambda}_{k,1}^2, \ldots, \tilde{\lambda}_{k,J}^2)$, with $\tilde{\lambda}_{k,j}^2$ being the ordered singular values of the matrix $(\mathbf{x}_a[k] - \mathbf{x}_b[k])(\mathbf{x}_a[k] - \mathbf{x}_b[k])^\dagger$. Note that the matrix $(\mathbf{x}_a[k] - \mathbf{x}_b[k])(\mathbf{x}_a[k] - \mathbf{x}_b[k])^\dagger$ is of rank 1 and the unique nonzero singular value in $\boldsymbol{\Lambda}_k$ is

$$\tilde{\lambda}_{k,1}^2 = \|\mathbf{x}_a[k] - \mathbf{x}_b[k]\|^2 \overset{(a)}{=} \sum_{j \in \phi_k} |x_{j,a}[k] - x_{j,b}[k]|^2$$

where (a) is due to the sparseness of the codebooks. Obviously, the nonzero eigenvalue is equal to the dimension-wise distances defined in Definition 1, namely $\tilde{\lambda}_{k,1}^2 = \lambda_k^2$. Hence,

$$\begin{aligned} \|\mathbf{h}[k]^\dagger(\mathbf{x}_a[k] - \mathbf{x}_b[k])\|^2 &= \mathbf{h}[k]^\dagger \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^\dagger \mathbf{h}[k] \\ &= \tilde{\mathbf{h}}[k]^\dagger \boldsymbol{\Lambda}_k \tilde{\mathbf{h}}[k] = \lambda_k^2 |\tilde{h}_1[k]|^2 \end{aligned}$$

where we define, $\tilde{\mathbf{h}}[k]^\dagger = \mathbf{h}[k]^\dagger \mathbf{U}_k = [\tilde{h}_1[k], \ldots, \tilde{h}_J[k]]$. Thus, $\tilde{\mathbf{h}}[k]$ has the same distribution as $\mathbf{h}[k]$, since multiplying with unitary matrix $\mathbf{U}_k$ doesn't change the amplitudes. Thus, the average PEP in (12.13) is equal to

$$P\{\mathbf{X}_a \to \mathbf{X}_b\} = \mathbb{E}_{\tilde{h}[1],\ldots,\tilde{h}[K]} \left[Q\left(\sqrt{\frac{\sum_{k=1}^K \lambda_k^2 |\tilde{h}[k]|^2}{2N_0}} \right) \right] \tag{12.18}$$

where the index 1 is dropped here for $\tilde{h}_1[k]^2$.

For i.i.d. Rayleigh fading, $\tilde{h}[1], \ldots, \tilde{h}[K]$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. Thus, $\sum_{k=1}^K \lambda_k^2 |\tilde{h}[k]|^2$ is the sum of K exponential random variables with different means, or a linear combination of V independent χ^2-distributed random variables with $2r_1, \ldots, 2r_V$ degrees of freedom, which follows Gamma or Generalized chi-squared distribution [12], with PDF given by

$$f(x; \mathbf{r}, \hat{\boldsymbol{\lambda}}) = A_{\mathbf{r},\hat{\boldsymbol{\lambda}}} \sum_{v=1}^{V} \sum_{l=1}^{r_v} \frac{B_{v,l,\mathbf{r},\hat{\boldsymbol{\lambda}}}}{(r_v - l)!} x^{r_v - l} e^{-\frac{x}{\hat{\lambda}_v^2}}$$

Then, the PEP can be obtained as

$$\begin{aligned} P\{\mathbf{X}_a \to \mathbf{X}_b\} &= \int_0^\infty Q\left(\sqrt{\frac{x}{2N_0}}\right) f(x;\mathbf{r},\hat{\boldsymbol{\lambda}})\mathrm{d}x \\ &= A_{\mathbf{r},\hat{\lambda}} \sum_{v=1}^{V}\sum_{l=1}^{r_v} \frac{B_{v,l,\mathbf{r},\hat{\lambda}}}{(r_v-l)!} \int_0^\infty Q\left(\sqrt{\frac{x}{2N_0}}\right) x^{r_v-l} e^{-\frac{x}{\hat{\lambda}_v^2}}\mathrm{d}x \\ &= A_{\mathbf{r},\hat{\lambda}} \sum_{v=1}^{V}\sum_{l=1}^{r_v} B_{v,l,\mathbf{r},\hat{\lambda}} \left(\frac{(1-\mu_v)\hat{\lambda}_v^2}{2}\right)^{r_v-l+1} \\ &\quad \times \sum_{k=0}^{r_v-l} \binom{r_v-l+k}{k} \left(\frac{1+\mu_v}{2}\right)^k \end{aligned} \tag{12.19}$$

where the last step follows from (13.4-15) in [11]. Substituting l with $r_v - L + 1$, (12.17) is proved. This concludes the proof.

The PEP is uniquely determined by the set of all λ_k^2s and the SNR, which is valid for any multi-dimensional codebooks and an arbitrary number of users. By reducing the number of users to 1, $P\{\mathbf{X}_a \to \mathbf{X}_b\}$ becomes the PEP between two multi-dimensional constellation points for a single-user transmission system. Therefore, the PEP of a joint multi-user detector is actually identical to that of the PEP of a single-user transmitting over a fading channel, where an equivalent K-dimensional constellation is employed such that the dimension-wise distances between the two constellation points are $\lambda_1^2, \ldots, \lambda_K^2$.

12.2.1.2 PEP over Downlink BCs

Consider the received signal vector of the downlink BC in (12.3), where $\mathbf{X} = \sum_{j=1}^{J} \mathbf{x}_j$ is the superimposed codeword of multiple users at the transmitter. Obviously, the model is exactly the same with that in the single-user communications, where $\mathbf{X}$ is used as the K-dimensional transmitted codeword. The ML multi-user detection for the superimposed codeword $\mathbf{X}$ becomes

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{y} - \mathrm{diag}(\mathbf{h})\mathbf{X}\|$$

Similar to that in uplink SCMA, we define the distances for downlink BC model.

Definition 4 Let $\mathbf{X}_a$ and $\mathbf{X}_b$ be two superimposed codewords, and $x_{j,a}[k]$ and $x_{j,b}[k]$ are the kth entries of the jth user's codeword corresponding to $\mathbf{X}_a$ and $\mathbf{X}_b$, respectively. The kth dimension-wise distance between $\mathbf{X}_a$ and $\mathbf{X}_b$, for the downlink broadcast channel, is defined as

$$\tau_k^2 = \left| \sum_{j=1}^{J} \left(x_{j,a}[k] - x_{j,b}[k] \right) \right|^2 = \left| \sum_{j \in \phi_k} \delta_j[k] \right|^2, \quad \forall k \tag{12.20}$$

where $\delta_j[k] = x_{j,a}[k] - x_{j,b}[k]$.

Theorem 2 *The PEP of a Rayleigh broadcast channel is the same as that in (12.17), after the substitution of λ_k^2 with τ_k^2.*

Proof Similar to that of the uplink case, the average PEP between $\mathbf{X}_a$ and $\mathbf{X}_b$ is equal to

$$\begin{aligned} P\{\mathbf{X}_a \to \mathbf{X}_b\} &= \mathbb{E}_{\mathbf{h}} \left[Q \left(\sqrt{\frac{\| \text{diag}(\mathbf{h}) \, (\mathbf{X}_a - \mathbf{X}_b) \, \|^2}{2N_0}} \right) \right] \\ &= \mathbb{E}_{\mathbf{h}} \left[Q \left(\sqrt{\frac{\sum_{k=1}^{K} \tau_k^2 |h[k]|^2}{2N_0}} \right) \right]. \end{aligned} \tag{12.21}$$

As $h[1], \ldots, h[K]$ are independent Rayleigh distributed random variables, the integral has been solved in (12.18), and the PEP has the similar expression as that in the MAC case, after the substitution of λ_k^2 with τ_k^2. This completes the proof.

It should be noted that, while the PEP of a BC can be evaluated through the same expression as that in the MAC case, τ_k^2 is different from the dimension-wise distance λ_k^2 in MAC, due to the absence of cross components $\delta_j[k] \times \delta_i[k]$, $j \neq i$, between different users. This is because in the MAC, the receiver distinguishes the multi-user signals by exploiting the differences among the channel coefficients, and only the amplitude of $\delta_j[k]$ contributes to the PEP. However, in the broadcast channel case, since the receiver exploits the differences among the multiple users' signals to perform the joint detection, both the amplitude and signs of $\delta_j[k]$ will influence the result of PEP.

12.2.1.3 PEP over the AWGN Channel

For the AWGN channel, where $h_j[k]$ is a constant for all j and k (assume that $|h_j[k]| = c$), the expressions of the received signal vector in (12.1) for uplink channels and (12.3) for downlink channels are the same. Then, according to (12.21), it is straightforward to derive the PEP as

$$\begin{aligned} P\{\mathbf{X}_a \to \mathbf{X}_b\} &= Q \left(\sqrt{\frac{c^2 \sum_{k=1}^{K} \tau_k^2}{2N_0}} \right) \\ &= Q \left(\sqrt{\frac{c^2 \sum_{k=1}^{K} \left| \sum_{j \in \phi_k} \delta_j[k] \right|^2}{2N_0}} \right) \end{aligned} \tag{12.22}$$

where τ_k^2 is the dimension-wise distance defined in (12.20), and $\delta_j[k] = x_{j,a}[k] - x_{j,b}[k]$.

12.2.1.4 Upper Bounds on PEP

In the codebook design, sometimes it is sufficient and easier to optimize the performance through a bound or an approximation of PEP. The exact PEP in (12.17) is a little complicated for large K, due to the large number of enumerations in $\Omega_{v,l}$, when calculating $B_{v,r_v-L+1,\mathbf{r},\boldsymbol{\lambda}}$. An alternative way to evaluate (12.17) is to use an upper bound for the Q-function as [13]

$$Q(x) \leq \sum_{i=1}^{N} a_i e^{-b_i x^2}, \quad \text{for } x > 0,$$

where N, a_i, b_i are constants. Note that the upper bound in Sect. 12.2.1.4 tends to the exact value as N increases.

For the multiple access and broadcast channels, since $X = |\hat{h}[k]|^2$ is an exponential random variable with unit mean, holds that $\mathbb{E}_X[e^{tX}] = \int_0^\infty e^{tx} e^{-x} dx = \frac{1}{1-t}$, for $t \leq 1$, and

$$\begin{aligned} P\{\mathbf{X}_a \to \mathbf{X}_b\} &\leq \mathbb{E}_{\tilde{h}[1],\ldots,\tilde{h}[K]} \left[\sum_{i=1}^{N} a_i \exp\left(-\frac{b_i \sum_{k=1}^{K} \lambda_k^2 |\tilde{h}[k]|^2}{2N_0} \right) \right] \\ &= \sum_{i=1}^{N} a_i \prod_{k=1}^{K} \mathbb{E}_{\tilde{h}[k]} \left[\exp\left(-\frac{b_i \lambda_k^2 |\tilde{h}[k]|^2}{2N_0} \right) \right] \\ &= \sum_{i=1}^{N} a_i \prod_{k=1}^{K} \frac{2N_0}{2N_0 + b_i \lambda_k^2} \end{aligned}$$

By choosing $N = 1, a_1 = b_1 = \frac{1}{2}$, we get the Chernoff bound with a scaling factor of 0.5 as

$$P_{\text{ch}}\{\mathbf{X}_a \to \mathbf{X}_b\} \leq \frac{1}{2} \prod_{k=1}^{K} \frac{4N_0}{4N_0 + \lambda_k^2} \tag{12.23}$$

In general, the Chernoff bound may be a little loose, but this does not affect the optimization criteria in the constellation design. It is obvious from (12.23) that a good direction is to design multi-dimensional multi-user codebooks, such that λ_k^2 to span in as many dimensions as possible (maximizing the diversity) and to make the maximum PEP or maximum of $P_{\text{ch}}\{\mathbf{X}_a \to \mathbf{X}_b\}$ as small as possible. If for any

codeword pair $\mathbf{X}_a$ and $\mathbf{X}_b$, all the λ_k^2 are positive, then the maximal diversity order of K can be achieved. Due to the sparseness of the codebooks, the diversity is always less than K. A tight and simple bound (or approximation) is to choose $N = 2, a_1 = \frac{1}{12}, a_2 = \frac{1}{4}, b_1 = \frac{1}{2}, b_2 = \frac{2}{3}$, which is denoted as $P_{\text{ub}}\{\mathbf{X}_a \to \mathbf{X}_b\}$.

12.2.1.5 A Universal Bound of ACEP for Joint ML Detection of Multiple Signals

A commonly used approach for the error performance analysis is the evaluation of the ACEP by using a union bound, assuming that the codewords are equiprobable transmitted. In general, the ACEP is dominated by the nearest neighbors of codewords, which result in a tight upper bound. However, it is quite difficult (if not impossible) to find the nearest neighbors in multi-user scenarios. To deal with this, we take into account all possible codewords that contribute to the ACEP.

Let $M_1, \ldots, M_J$ be the codebook size for J users, respectively. We define $\{\mathbb{X}_j\}_{j=1}^J$ as the set of all $\prod_{j=1}^J M_j$ possible combined codewords of J users, and let $\mathbf{X}_a, \mathbf{X}_b \in \{\mathbb{X}_j\}_{j=1}^J$ be two different elements of $\{\mathbb{X}_j\}_{j=1}^J$. Here, the combined codeword $\mathbf{X}_a$ and $\mathbf{X}_b$ are a JK-dimensional vector for the MAC, or the sum of J K-dimensional codewords for the BC. Denote $\mathbf{x}_{j,a}$ and $\mathbf{x}_{j,b}$ the transmitted codewords of the jth user corresponding to $\mathbf{X}_a$ and $\mathbf{X}_b$. Then, there are M_j possible values for $\mathbf{x}_{j,a}$ and $\mathbf{x}_{j,b}$. Note that $\mathbf{x}_{j,a}$ and $\mathbf{x}_{j,b}$ are K-dimensional vector with complex entries, i.e., SCMA codeword. Following the approach in [14] for multiple signals and [15] for MIMO channels, the ACEP for the jth user with joint ML detection of J users' signals is upper bounded by

$$P_j(e) \le \frac{1}{\prod_{j=1}^J M_j} \sum_{\mathbf{X}_a} \left(\sum_{\mathbf{X}_b, \mathbf{x}_{j,b} \ne \mathbf{x}_{j,a}} P\{\mathbf{X}_a \to \mathbf{X}_b\} \right) \tag{12.24}$$

The ACEP of the system can be obtained by taking the mean of all the single-user ACEPs, namely $P(e) = \frac{1}{J} \sum_{j=1}^J P_j(e)$.

On the right-hand side of (12.24), the summation over $\mathbf{X}_a$ will add up $\prod_{j=1}^J M_j$ terms and the summation over $\mathbf{X}_b$ will add up $(M_1 - 1) \prod_{j=2}^J M_j$ terms. Thus, there will be up to $(M_1^2 - M_1) \prod_{j=2}^J M_j^2$ PEPs in (12.24), which is intractable for a large constellation size and number of users. However, we can simplify it by using the symmetry of the dimension-wise distances. For example, consider the ACEP for the first user $P_1(e)$ here. The upper bound of $P_1(e)$ can be decomposed into the summation of two parts as

$$\begin{aligned} &\frac{1}{\prod_{j=1}^{J} M_j} \sum_{\mathbf{X}_a} \sum_{\substack{\mathbf{X}_b, \mathbf{x}_{1,b} \neq \mathbf{x}_{1,a}, \\ [\mathbf{x}_{2,b}, \ldots, \mathbf{x}_{j,b}] = [\mathbf{x}_{2,a}, \ldots, \mathbf{x}_{j,a}]}} P\{\mathbf{X}_a \to \mathbf{X}_b\} \\ &+ \frac{1}{\prod_{j=1}^{J} M_j} \sum_{\mathbf{X}_a} \sum_{\substack{\mathbf{X}_b, \mathbf{x}_{1,b} \neq \mathbf{x}_{1,a}, \\ [\mathbf{x}_{2,b}, \ldots, \mathbf{x}_{j,b}] \neq [\mathbf{x}_{2,a}, \ldots, \mathbf{x}_{j,a}]}} P\{\mathbf{X}_a \to \mathbf{X}_b\}. \end{aligned} \tag{12.25}$$

The first part in (12.25) is the union bound of the probability of the event that all users' signals are correctly detected except for the first user, namely the ACEP for the first user with single-user detection in the absence of interference. This part is a summation of $(M_1 - 1)\prod_{j=1}^{J} M_j$ PEPs, while only $\frac{1}{2}M_1(M_1 - 1)$ different PEP values should be calculated, due to the symmetry of the dimension-wise distance for the first user. The second part is the probability of the event that the errors happen for the first user and for at least one user among $\{2, \ldots, J\}$, which is the summation of $(M_1 - 1)(\prod_{j=2}^{J} M_j - 1)\prod_{j=1}^{J} M_j$ PEPs, but only one-fourth of them should be considered. A further simplification can be achieved for the MAC by considering more decompositions of the second part.

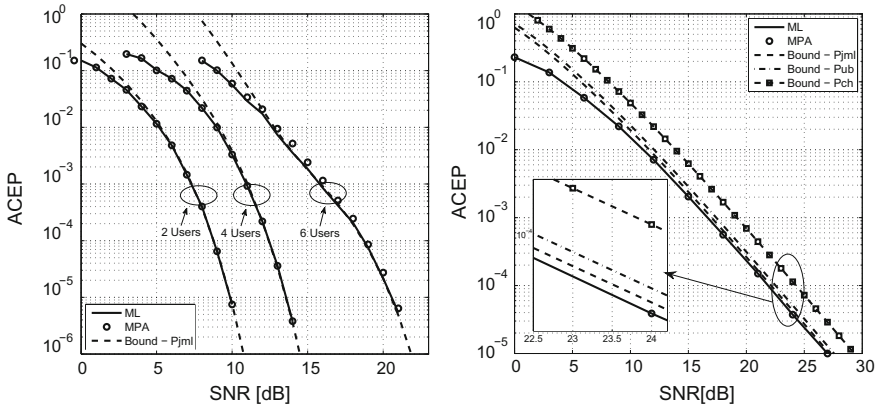
In general, SCMA codebooks of all users are constructed from a common mother constellation [16], with some layer-specific operations over this constellation to get their own layer's codebook. These layer-specific operations do not change the fundamental properties of the mother constellation, such as the Euclidean distance. The layer operation losses their efficiency in the uplink multiple access fading channels, due to the distinctness of each user's channel gain. If the factor graph matrix is regular as that in (12.5), every user will suffer from the same interference from other users. Then, the system results in the same performance for all users, while for other cases, the ACEP is asymmetric for each user.

The MPA detection is believed to be an efficient approach for SCMA systems. Theoretically, the MPA detector is asymptotically equivalent to the optimal MAP detector [17, 18] (or ML conditioned on equal probably transmissions) for a sparsely spread system with long signatures. The analytical bounds, proposed in this subsection, work for ML detector as well as for the MPA detector.

12.2.1.6 Numerical Results and Simulations

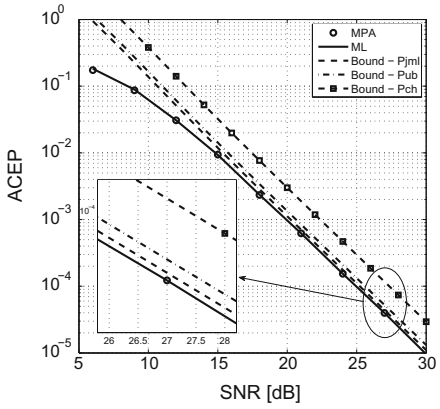
We consider an SCMA system illustrated in Example 1, and the four-dimensional four-ary codebooks are listed in Table 12.1. The ACEP of SCMA over AWGN and uplink Rayleigh fading channels for 2, 4, and 6 user cases are evaluated. For the Rayleigh fading channel, we give analytical results of the union bound on the ACEP, corresponding to exact PEP (denoted as P_{jml}), the upper bound on PEP P_{ub}, and the scaled Chernov bound P_{ch}, respectively.

Results for AWGN channels are shown in Fig. 12.4a. The analytical bound of a joint ML detector closely coincides with the simulation curves for large SNR. The bound is quite tight for values of ACEP below 10^{-3}, even for six users. Thus, this

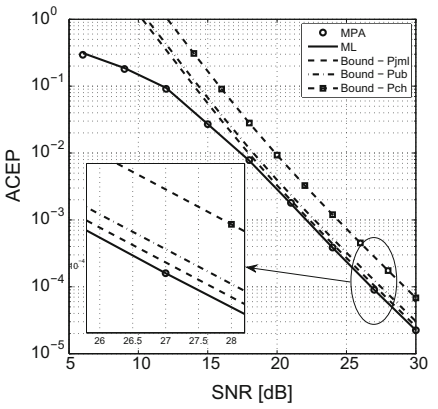


(a) SCMA over AWGN channel.

(b) SCMA with 2 users in Rayleigh fading.



(c) SCMA with 4 users in Rayleigh fading.



(d) SCMA with 6 users in Rayleigh fading.

Fig. 12.4 ACEP of uplink SCMA over AWGN and Rayleigh fading channels

bound is sufficient for the analysis and design of a signal constellation in AWGN. Surprisingly, there are bends for the ACEP curves of an ML detector and analytical bound for the six user case. The performance turns better than expected within the SNR region from 12 to 18 dB, which is due to the sparse codebooks. This phenomenon happens if the distance profile of the multi-user codebooks is uneven. For example, a quite small distance exists while the others are very large. In general, the ACEP of a constellation in AWGN channels is proportional to the summation of Q-function of the distances d and SNR, i.e., $P(\text{SNR}) \propto \sum_d Q(\sqrt{d\text{SNR}})$, where d is the set of distances among the constellation points. If there is a large difference between two distance components, $P(\text{SNR})$ is not a convex function and a bend appears in the $P(\text{SNR})$ vs SNR curves in log–log scale at low SNR. The theoretical bound is still quite close to the actual ACEP within the bend region. It can be seen from Fig. 12.4a that there is nearly 0.4 dB gap between the performance of the MPA detector and the

ML detector at the SNR of 14 dB. The performance of the MPA detector is improved asymptotically and approaches that of ML detector at high SNRs.

Figure 12.4b–d present the performance for 2, 4, and 6 users over Rayleigh fading channels, respectively. All the bounds are asymptotically tight as SNR increases. The analytical bound P_{jml} is quite tight for values of ACEP below 10^{-3} for all numbers of users, and the gap between P_{jml} and the exact ACEP is almost constant at high SNRs, when the number of users increases. Moreover, the bounds become looser at low SNRs as the number of users increases. The upper bound P_{ub} shows superiority over all the other bounds, since it is much easier to calculate than P_{jml} while it has only a little difference. It should be noted that the scaled P_{ch} is much looser compared to P_{ub}. As expected, the MPA detector shows exactly the same performance as the ML detector for any number of users and any values of SNR over Rayleigh fading channels.

12.2.2 Capacity and Cutoff Rate

This subsection discusses the sum rate analysis of SCMA systems. The channel capacity characterizes the limit information rate that can be reliably transmitted over a channel. It is well known that the sum rate of multi-channel transmissions is simply the sum of per channel rate, and in the uplink SCMA, the communications over each SCMA resource constitutes a multiple access process, then, the sum rate of uplink SCMA is

$$\begin{aligned} C &= \sum_{k=1}^{K} \mathbb{E}_{h_1[1],\ldots,h_J[K]} \left[\log_2 \left(1 + \rho \sum_{j \in \phi_k} |h_j[k]|^2 \right) \right] \\ &= \frac{K e^{1/\rho}}{\rho^{d_f} \ln 2} \sum_{i=1}^{d_f} \sum_{j=0}^{d_f - i} \frac{(-1)^{d_f - j - i} \rho^{i+j}}{j!(d_f - i - j)!} \Gamma(j, 1/\rho) \end{aligned} \tag{12.26}$$

where ρ is the SNR, and $\Gamma(a, x) = \int_x^{\infty} t^{a-1} e^{-t} \mathrm{d}t$, is the incomplete Gamma function. In the above sum rate evaluation, we assume that the users have the same transmitting power, and each SCMA resource carries the same number of users, i.e., $d_f = |\phi_k|$. To achieve any point on the sum rate curve, codebooks with Gaussian distributions and successive interference cancelation (SIC) receivers are generally required.

In practical cases, it is more valuable to investigate the capacity restricted by specific codebooks, i.e., the discrete codebook-constrained capacity (DCCC). Consider the equivalent linear system of uplink SCMA in (12.2). Assuming that perfect channel knowledge is available at the receiver. The conditional probability density function (PDF) of the received signal vector is

$$f(\mathbf{y}|\mathbf{X},\mathbf{H}) = \frac{1}{(\pi N_0)^K}\exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{X}\|^2}{N_0}\right) \tag{12.27}$$

The mutual information $I(\mathbf{X};\mathbf{y})$ between the discrete input $\mathbf{X}$ and the continuous output $\mathbf{y}$, or the DCCC, is given by [11]

$$\begin{aligned} I(\mathbf{X};\mathbf{y}) &= \log_2 M^J - \mathbb{E}_{\mathbf{y},\mathbf{X}_a,\mathbf{H}}\left[\log_2 \frac{\sum_{\mathbf{X}_b} f(\mathbf{y}|\mathbf{X}_b,\mathbf{H})}{f(\mathbf{y}|\mathbf{X}_a,\mathbf{H})}\right] \\ &= \log_2 M^J - \mathbb{E}_{\mathbf{H}}\left[\frac{1}{M^J}\int_{\mathbf{y}}\sum_{\mathbf{X}_a} f(\mathbf{y}|\mathbf{X}_a,\mathbf{H})\log\frac{\sum_{\mathbf{X}_b} f(\mathbf{y}|\mathbf{X}_b,\mathbf{H})}{f(\mathbf{y}|\mathbf{X}_a,\mathbf{H})}\mathrm{d}\mathbf{y}\right] \end{aligned} \tag{12.28}$$

where $\mathbf{X}_a, \mathbf{X}_b \in \{\mathbb{X}_j\}_{j=1}^J$ are two combined codewords for J users. Obviously, it is quite difficult—if not impossible—to deal with the expression for the mutual information, and a closed-form solution is unattainable. In the following, we resort to the cutoff rate analysis.

12.2.2.1 Cutoff Rate of Uplink MAC

The channel cutoff rate R_0, which is a lower bound on the channel capacity, is another commonly used metric characterizing the channel rate. The cutoff rate is more informative than the DCCC, since it provides a good estimate of the capacity as well as a tight upper bound on the error probability of an optimal detector.

The cutoff rate can be defined by [11]

$$R_0 = -\log_2\left[\sum_{\mathbf{X}_a}\sum_{\mathbf{X}_b} p(\mathbf{X}_a)p(\mathbf{X}_b)\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b}\right] \tag{12.29}$$

where $p(\mathbf{X}_a) = p(\mathbf{X}_b) = \frac{1}{M^J}$, and $\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b}$ is the Bhatacharyya bound on the PEP between $\mathbf{X}_a$ and $\mathbf{X}_b$, which is given by [11]

$$\begin{aligned} \boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b} &= \mathbb{E}_{\mathbf{H}}\left[\int \sqrt{p\left(\mathbf{y}\mid\mathbf{X}_a,\mathbf{H}\right)p\left(\mathbf{y}\mid\mathbf{X}_b,\mathbf{H}\right)}\,\mathrm{d}\mathbf{y}\right] \\ &= \mathbb{E}_{\mathbf{H}}\left[e^{-\frac{1}{4N_0}\|\mathbf{H}(\mathbf{X}_a-\mathbf{X}_b)\|^2}\int\frac{1}{(\pi N_0)^K}e^{-\frac{1}{N_0}\left\|\mathbf{y}-\frac{\mathbf{H}(\mathbf{X}_a+\mathbf{X}_b)}{2}\right\|^2}\mathrm{d}\mathbf{y}\right] \\ &= \mathbb{E}_{\mathbf{H}}\left[e^{-\frac{1}{4N_0}\|\mathbf{H}(\mathbf{X}_a-\mathbf{X}_b)\|^2}\right] \end{aligned} \tag{12.30}$$

Note that $\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_a} = 1$, then the cutoff rate can be written as

$$R_0 = \log_2 M^J - \log_2 \left(1 + \frac{1}{M^J} \sum_{\mathbf{X}_a} \sum_{\mathbf{X}_b \neq \mathbf{X}_a} \boldsymbol{\Delta}_{\mathbf{X}_a, \mathbf{X}_b}\right) \tag{12.31}$$

It is observed that, the term $\frac{1}{M^J} \sum_{\mathbf{X}_a} \sum_{\mathbf{X}_b \neq \mathbf{X}_a} \boldsymbol{\Delta}_{\mathbf{X}_a, \mathbf{X}_b}$, inside the bracket of (12.31), is the union-Bhatacharyya bound on the joint codeword error probability for multiple users. Therefore, optimizing the mean cutoff rate is equivalent to the optimization of the error probability, and cutoff rate can be used as a good performance criterion for the system design.

For the uplink MAC, according to the analysis in Theorem 1,

$$\|\mathbf{H}(\mathbf{X}_a - \mathbf{X}_b)\|^2 = \sum_{k=1}^{K} \lambda_k^2 |\tilde{h}[k]|^2$$

where λ_k^2 is the kth dimension-wise distance in the MAC defined in Definition 1, and $|\tilde{h}[1]|, \ldots, |\tilde{h}[K]|$ are independent Rayleigh distributed random variables. Therefore,

$$\boldsymbol{\Delta}_{\mathbf{X}_a, \mathbf{X}_b} = \prod_{k=1}^{K} \mathbb{E}_{\tilde{h}[k]} \left[e^{-\frac{1}{4N_0} \lambda_k^2 |\tilde{h}[k]|^2} \right] = \prod_{k=1}^{K} \left(1 + \frac{\lambda_k^2}{4N_0}\right)^{-1}$$

and thus the cutoff rate for uplink MAC is given by

$$R_0 = \log M^J - \log \left[1 + \frac{1}{M^J} \sum_{\mathbf{X}_a} \sum_{\mathbf{X}_b, b \neq a} \prod_{k=1}^{K} \left(1 + \frac{\lambda_k^2}{4N_0}\right)^{-1} \right] \tag{12.32}$$

The average sum rate and cutoff rate for uplink SCMA in Rayleigh fading are depicted in Fig. 12.5, where the 4-ary codebook in Table 12.1 is adopted. The sum rates of SCMA with 2, 4 and 6 users are represented by the uppermost curves, which increase almost linearly with the SNR when SNR becomes very large. Due to the discrete codebooks, the DCCC and the cutoff rate are upper bounded by $\frac{J}{K} \log_2(M)$. However, significant rate improvement can be achieved by overloaded access for moderate to large SNRs. As it is observed, the cutoff rate establish a lower bound to the DCCC, and it asymptotically approaches the DCCC with increasing SNRs.

12.2.2.2 Cutoff Rate of Downlink BC

Consider the downlink BC model in (12.3), for the jth user, the cutoff rate corresponding to the mutual information $I(\mathbf{x}_j; \mathbf{y})$ is given by [19]

$$R_0 = \log_2 M - \log_2 \left(1 + \frac{1}{M} \sum_{\mathbf{x}_{j,a}} \sum_{\mathbf{x}_{j,b} \neq \mathbf{x}_{j,a}} \boldsymbol{\Delta}_{\mathbf{x}_{j,a}, \mathbf{x}_{j,b}}\right)$$

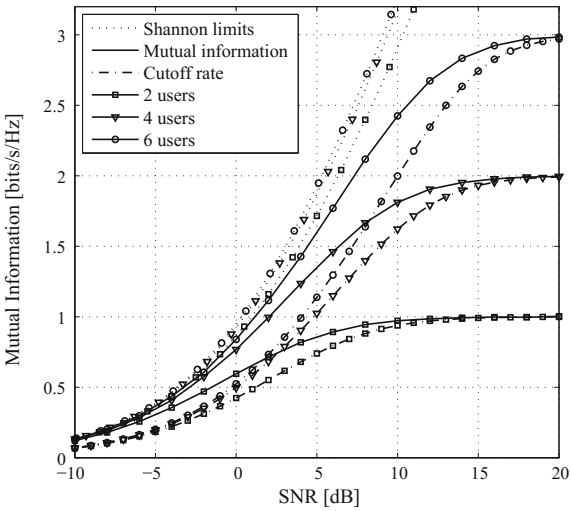


Fig. 12.5 Capacity and cutoff rate of uplink SCMA in Rayleigh fading

and the Bhatacharyya parameter is

$$\begin{aligned}\boldsymbol{\Delta}_{\mathbf{x}_{j,a},\mathbf{x}_{j,b}} &= \mathbb{E}_{\mathbf{h}}\left[\int \sqrt{p\left(\mathbf{y} \mid \mathbf{x}_{j,a}, \mathbf{h}\right) p\left(\mathbf{y} \mid \mathbf{x}_{j,b}, \mathbf{h}\right)}\mathrm{d}\mathbf{y}\right] \\ &= \mathbb{E}_{\mathbf{h}}\left[\frac{1}{M^{J-1}} \int \sqrt{\sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_a, \mathbf{h}\right) \sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_b, \mathbf{h}\right)}\mathrm{d}\mathbf{y}\right]\end{aligned} \tag{12.33}$$

In the integral of (12.33) a square root of the double sum of the products "$p\left(\mathbf{y}|\mathbf{X}_a, \mathbf{h}\right)\ p\left(\mathbf{y}|\mathbf{X}_b, \mathbf{h}\right)$" is involved, which makes it excessively complex for a large number of users. Hence, we will attempt to obtain reasonable bounds for the cutoff rate. To deal with the expression, we first calculate $\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b}$. According to the PEP analysis in (12.21) for downlink SCMA, the channel-dependent metric is equal to

$$\|\mathrm{diag}(\mathbf{h})(\mathbf{X}_a - \mathbf{X}_b)\|^2 = \sum_{k=1}^{K} \tau_k^2 |h[k]|^2$$

where τ_k^2 is the dimension-wise distance defined in Definition 2, and $|h[k]|$ is the Rayleigh distributed random variables. Similar to that in the uplink case, the Bhatacharyya parameter considering the superimposed codewords $\mathbf{X}_a$ and $\mathbf{X}_b$ is given by

$$\begin{aligned}\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b} &= \mathbb{E}_{\mathbf{h}}\left[e^{-\frac{1}{4N_0}\|\mathrm{diag}(\mathbf{h})(\mathbf{X}_a-\mathbf{X}_b)\|^2}\right] \\ &= \mathbb{E}_{\mathbf{h}}\left[e^{-\frac{1}{4N_0}\sum_{k=1}^{K}\tau_k^2|h[k]|^2}\right] = \prod_{k=1}^{K}\left(1+\frac{\tau_k^2}{4N_0}\right)^{-1}\end{aligned}$$

By applying Holder's inequality,

$$\sqrt{\sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_a,\mathbf{h}\right)\sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_b,\mathbf{h}\right)} \geq \sum_{(\mathbf{x}_{i,a},\mathbf{x}_{i,b})\in\mathscr{P}_i,\forall i\neq j}\sqrt{p\left(\mathbf{y}|\mathbf{X}_a,\mathbf{h}\right)p\left(\mathbf{y}|\mathbf{X}_b,\mathbf{h}\right)}$$

where $\mathscr{P}_i$ is the set of a point-to-point pairing of codewords $(\mathbf{x}_{i,a},\mathbf{x}_{i,b})$ for all $i\neq j$, which contains M elements.[3] Then it holds that

$$\begin{aligned}\boldsymbol{\Delta}_{\mathbf{x}_{j,a},\mathbf{x}_{j,b}} &\geq \mathbb{E}_{\mathbf{h}}\left[\frac{1}{M^{J-1}}\sum_{(\mathbf{x}_{i,a},\mathbf{x}_{i,b})\in\mathscr{P}_i,\forall i\neq j}\int\sqrt{p\left(\mathbf{y}|\mathbf{X}_a,\mathbf{h}\right)p\left(\mathbf{y}|\mathbf{X}_b,\mathbf{h}\right)}\mathrm{d}\mathbf{y}\right]\\ &= \frac{1}{M^{J-1}}\sum_{(\mathbf{x}_{i,a},\mathbf{x}_{i,b})\in\mathscr{P}_i,\forall i\neq j}\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b}\end{aligned}$$

Thus, an upper bound on the cutoff rate of downlink SCMA is

$$R_0|^{\text{upper}} = \log_2 M - \log_2\left[1+\frac{1}{M^J}\sum_{\mathbf{x}_{j,a}}\sum_{\mathbf{x}_{j,b}\neq\mathbf{x}_{j,a}}\sum_{(\mathbf{x}_{i,a},\mathbf{x}_{i,b})\in\mathscr{P}_i,\forall i\neq j}\prod_{k=1}^{K}\left(1+\frac{\tau_k^2}{4N_0}\right)^{-1}\right] \tag{12.34}$$

For the sake of deriving the lower bound of the cutoff rate, we may invoke the following simple inequality

$$\begin{aligned}&\sqrt{\sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_a,\mathbf{h}\right)\sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j} p\left(\mathbf{y}|\mathbf{X}_b,\mathbf{h}\right)}\\ &\leq \sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j}\sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j}\sqrt{p\left(\mathbf{y}|\mathbf{X}_a,\mathbf{h}\right)p\left(\mathbf{y}|\mathbf{X}_b,\mathbf{h}\right)}\end{aligned}$$

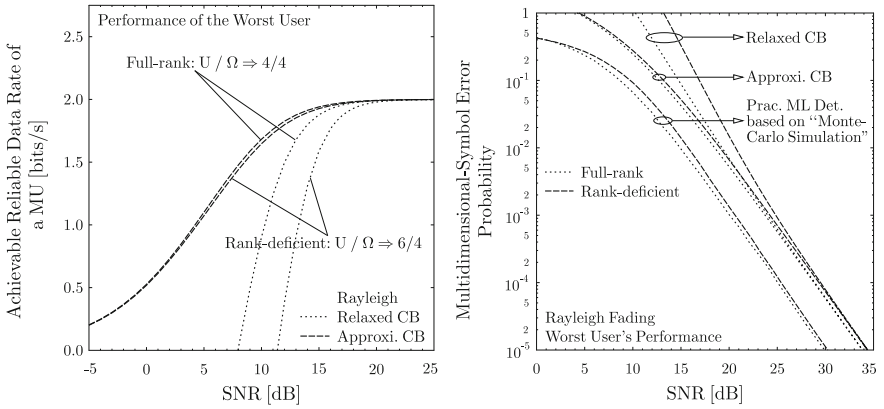
we get that

$$\boldsymbol{\Delta}_{\mathbf{x}_{j,a},\mathbf{x}_{j,b}} \leq \frac{1}{M^{J-1}}\sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j}\sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j}\boldsymbol{\Delta}_{\mathbf{X}_a,\mathbf{X}_b}$$

and a lower bound on the cutoff rate is obtained

$$\begin{aligned}R_0|^{\text{lower}} &= \log_2 M - \log_2\left[1+\frac{1}{M^J}\sum_{\mathbf{x}_{j,a}}\sum_{\mathbf{x}_{j,b}\neq\mathbf{x}_{j,a}}\sum_{\mathbf{x}_{i,a}\in\mathbb{X}_i,\forall i\neq j}\sum_{\mathbf{x}_{i,b}\in\mathbb{X}_i,\forall i\neq j}\prod_{k=1}^{K}\left(1+\frac{\tau_k^2}{4N_0}\right)^{-1}\right]\\ &= \log_2 M - \log_2\left[1+\frac{1}{M^J}\sum_{\mathbf{X}_a}\sum_{\mathbf{X}_b,\mathbf{x}_{j,b}\neq\mathbf{x}_{j,a}}\prod_{k=1}^{K}\left(1+\frac{\tau_k^2}{4N_0}\right)^{-1}\right]\end{aligned} \tag{12.35}$$

[3] There are $M!$ possible pairing patterns for $(\mathbf{x}_{i,a},\mathbf{x}_{i,b})$, hence $M!$ choices for $\mathscr{P}_i$. The tightness of the bound is determined by the specific selection of the pairing patterns. A detailed seek for the appropriate pairing pattern can be found in [19].



(a) Cutoff rate. (b) ACEP.

Fig. 12.6 Cutoff rate and ACEP of the worst user for downlink SCMA in Rayleigh fading

As discussed in (12.24), the expression inside the square bracket of (12.35) is the union-Bhatacharyya bound on the ACEP for the jth user.

With the derived upper and lower bound on R_0 or $\boldsymbol{\Delta}_{\mathbf{x}_{j,a},\mathbf{x}_{j,b}}$, the corresponding bounds to the union-Bhatacharyya bound on ACEPs, i.e., $\frac{1}{M}\sum_{\mathbf{x}_{j,a}}\sum_{\mathbf{x}_{j,b}\neq\mathbf{x}_{j,a}}\boldsymbol{\Delta}_{\mathbf{x}_{j,a},\mathbf{x}_{j,b}}$, can be obtained straightforwardly.

We will now verify the cutoff rate bounds and the corresponding bounds for ACEPs by simulations. If $R_0|^{\text{upper}}$ is sufficiently tight, it may be regarded as satisfactory approximation of R_0. In order to emphasize the primary common characteristic between $R_0|^{\text{upper}}$ and $R_0|^{\text{lower}}$, we can readily refer to $R_0|^{\text{upper}}$ as the approximated Chernov bound (Approxi. CB), and $R_0|^{\text{lower}}$ as relaxed Chernov bounds (Relaxed CB). The results for the cutoff rate of the worst user in downlink SCMA and the corresponding ACEPs are plotted in Fig. 12.6a and Fig. 12.6b, respectively. The curve of "Approxi. CB" and that of "Relaxed CB" merge with each other in the high-SNR region, while the "Approxi. CB" bound gets significantly close to the associated practical performance. Hence we may claim that $R_0|^{\text{upper}}$ indeed represents a satisfactory approximation of R_0. This conclusion may be verified by the associated simulation results shown in Fig. 12.6b, where the "Approxi. CB" over ACEP gives a better estimation of the practical ACEP than that of "Relaxed CB", for both full-rank (with 4 users) and rank-deficient (with 6 users) SCMA systems.

12.3 Codebook Design

As the performance of SCMA strongly depends on the multi-dimensional codebooks, codebook design constitutes one of the most important issues for SCMA, and it is what distinguishes SCMA from other non-orthogonal multiple access schemes.

Incorporating a sophisticated codebook design into SCMA has the potential of significantly improving the spectrum efficiency, and reducing the detection complexity.

12.3.1 General Design Rules

The design of SCMA codebook is a joint optimization of the sparse mapping matrix and the multi-dimensional constellations. Assume that all layers have the same constellation size and length. An SCMA codebook can be represented by structure $\mathfrak{S}(\mathscr{V}, \mathscr{C}; J, M, N, K)$, where $\mathscr{V} = \{\mathbf{V}_j\}_{j=1}^{J}$, is the set of mapping matrices, and $\mathscr{C} = \{\mathbb{C}_j\}_{j=1}^{J}$, is the set of signal constellations for J layers. Thus, the SCMA codebook designing is equivalent to solve the optimization problem [1]

$$\mathscr{V}^{+}, \mathscr{C}^{+} = \arg\max_{\mathscr{V},\mathscr{C}} \Upsilon(\mathfrak{S}(\mathscr{V}, \mathscr{C}; J, M, N, K)) \tag{12.36}$$

where the function $\Upsilon(\cdot)$ is somehow the design criterion.

Unfortunately, for a given criterion $\Upsilon(\cdot)$ and such a multi-dimensional problem, the optimum solution cannot be found. In practice, a suboptimal multi-stage optimization approach is adopted, by optimizing the mapping matrices and constellations separately. The set of mapping matrices $\mathscr{V}$ is generally selected in order to meet the maximum overloading, while the design of J multi-dimensional constellations is simplified to the design of a *mother constellation* and multiple layer-specific operators.

12.3.1.1 Mapping Matrices

The set of mapping matrices $\mathscr{V}$ should be pre-determined before the constellation design, since it determines the number of users/layers interfering at each resource node and complexity of the multi-user detection. As $\mathscr{V}$ can be characterized and uniquely determined by the factor graph matrix, the design of $\mathscr{V}$ can borrow the idea from the design of LDPC codes. However, here we introduce general rules for the designing:

- $\mathbf{V}_j \in \mathbb{B}^{K\times N}$, and $\mathbf{V}_i \neq \mathbf{V}_j, \forall i \neq j$
- $\mathbf{V}_j^{[\phi]} = \mathbf{I}_N$, where $\mathbf{V}_j^{[\phi]}$ is obtained by removing all-zero rows in $\mathbf{V}_j$

Thus we may insert $K - N$ all-zero row vectors into rows of $\mathbf{I}_N$ to obtain the unique solution $\mathscr{V}^{+}$ for problem (12.36).

If we take Example 1 as the illumination, we have following properties and relations for SCMA encoding parameters

- Choose the constellation length $N = 2$, and the codebook length $K = 4$
- The maximum number of layers $J = \binom{K}{N} = 6$

- The number of multiplexed layers over each resource $d_f = \frac{JN}{K} = 3$
- Overloading factor $\lambda = \frac{J}{K} = 1.5$
- $\max(0, 2N - K) \leq l \leq N - 1$, where l is the number of the overlapping elements of any two distinct $\mathbf{f}_j$ vectors. Thus $0 \leq l \leq 1$ if $K = 4$ means that the codeword are either orthogonal or collide at only one overlap nonzero element over any two rows.

The resulting factor graph matrix $\mathbf{F}$ is the same as (12.5) and the factor graph is shown in Fig. 12.3.

12.3.1.2 Multi-dimensional Constellations

Having the mapping set $\mathscr{V}^+$, the optimization problem of an SCMA is reduced to

$$\mathscr{C}^+ = \arg\max_{\mathscr{C}} \Upsilon\left(\mathfrak{S}(\mathscr{V}^+, \mathscr{C}; J, M, N, K)\right) \tag{12.37}$$

which is to find J different N-dimensional complex constellations, each contains M signal points. In general, the joint design of multiple multi-dimensional constellations is challenging, a further simplification of (12.37) can be conducted by dividing the problem into the design of a *mother constellation* and J layer-specific operators, and optimizing them separately. Without loss of generality, define $\mathbb{C}_j \equiv \boldsymbol{\Theta}_j(\mathbb{C}), \forall j$, where $\boldsymbol{\Theta}_j(\cdot)$ denotes a *constellation operator*. Thus the optimization problem in (12.37) becomes

$$\mathbb{C}^+, \left\{\boldsymbol{\Theta}_j^+\right\}_{j=1}^J = \arg\max_{\mathbb{C}, \{\boldsymbol{\Theta}_j\}_{j=1}^J} \Upsilon\left(\mathfrak{S}(\mathscr{V}^+, \mathscr{C} \equiv \left\{\boldsymbol{\Theta}_j(\mathbb{C})\right\}_{j=1}^J; J, M, N, K)\right) \tag{12.38}$$

A. Mother Constellation

In general, a constellation with large minimum Euclidean distance achieves good performance when no collisions occur among users/layers over a tone. With increasing number of users/layers, the collisions are unavoidable and the multi-user interference will be introduced. To mitigate such interference, it is required to induce dependency among the nonzero elements of the codewords, such that the receiver can recover colliding codewords from other tones. In general, the mother constellation can be any form of a multi-dimensional constellation with a maximized minimum Euclidean distance. To control dimensional dependency and power variation without destroying the Euclidean distance profile, a unitary rotation can be applied to the mother constellation. For transmission over fading channels, the performance is dominated by the product distance of a constellation at high-SNR region. Thus the goal of designing a good mother constellation for SCMA is trying to optimize both the minimum Euclidean distance and product distance. Fortunately, the optimization of the product distance could be realized by unitary rotation as well. Thus the two types of distances can be optimized separately. In [20], using the Chernoff bounding

technique, it is shown that for Rayleigh fading channels, the error probability of a multi-dimensional signal set is essentially dominated by four factors. To improve performance is necessary to

- minimize the average energy per constellation point;
- maximize the modulation or signal-space diversity;
- maximize the minimum product distance

$$d_{p,\min} = \min_{\mathbf{x}_a, \mathbf{x}_b} \prod_{x_a[k] \neq x_b[k]} |x_a[k] - x_b[k]| \tag{12.39}$$

  between any two points $\mathbf{x}_a$ and $\mathbf{x}_b$ in the constellation;
- minimize the product *kissing number* for the minimum product distance, i.e., the total number of points at the minimum product distance.

For low rates, constellation design can be done by brute-force searching, however, this is not necessarily the case for higher rates and a larger number of users/layers due to the prohibitive searching complexity. Under this circumstance, the structured construction is required. Lattice constellation construction can be considered as a possible way to design good mother constellations. If we construct a constellation from the lattice $\mathbb{Z}^{2N}$ with gray labeling, the construction could be done effectively by forming orthogonal QAM constellations on different complex planes. To maximize the minimum product distance of rotated lattice, the unitary rotations of QAM lattice constellations might be optimized as in [20, 21].

B. Constellation Operators

After obtaining the mother constellation $\mathbb{C}^+$, layer-specific operators should be designed to guarantee the unique decodability of the multi-layer signals at the receiver, and also lower the multi-user interference. The optimization problem for the operators can be formulated as

$$\{\boldsymbol{\Theta}_j^+\}_{j=1}^J = \arg \max_{\{\boldsymbol{\Theta}_j\}_{j=1}^J} \Upsilon\left(\mathfrak{S}(\mathscr{V}^+, \mathscr{C} \equiv \{\boldsymbol{\Theta}_j(\mathbb{C}^+)\}_{j=1}^J; J, M, N, K)\right) \tag{12.40}$$

Note that here, the design criterion $\Upsilon(\cdot)$ are not necessarily the same as that in (12.38) for the joint design of mother constellation and constellation operators.

The constellations for different SCMA layers might be constructed with different operators $\boldsymbol{\Theta}_j(\cdot)$, and the constellation operators generally include complex conjugate, phase rotation and dimensional permutation. Generally speaking, if the different users have different power levels, the interfering codewords would be easily separated at receiver due to the power diversity. To do this, it is obliged to have a diverse average power level over the constellation dimensions when designing the mother constellations, which could be done by an appropriate rotation of the lattice constellation as discussed in [16]. Thus the task of optimization problem can be the permutation operators which enable the SCMA codebooks to capture as much power diversity as possible over the interfering users. The optimization for power variation

over users can be designed to permute each codebook set to avoid interfering with the same dimensions of a mother constellation over a resource node.

As discussed in Sect. 12.2.1.2, the constellation operators is unnecessary for the uplink SCMA in fading channels. On the one hand, in MAC, the fading itself takes the role of constellation operations, and the receiver exploiting the differences among the channel fadings to separate the multi-user signals. On the other hand, the constellation operators like phase rotation and complex conjugate don't change λ_k^2 in Definition 1, hence don't change the error probability. However, it is important to design layer-specific operator for downlink SCMA, because all users experience the same channel condition and the destructive codeword collision can be avoided by careful design of $\boldsymbol{\Theta}_j(\cdot)$ in the downlink.

12.3.1.3 Constellations for Lower Receiver Complexity

This part introduces two kinds of multi-dimensional constellations for SCMA, that allow MPA receiving with reduced complexity.

A. Shuffled Multi-dimensional Constellation

The dependency among the complex dimensions of the mother constellation guarantees an efficient detection and diversity for fading channels. It is possible to construct a mother constellation such that the real part and imaginary part are independent with each other, while the complex dimensions are still dependent. One kind of approach is the *shuffling* [16], which enables the MPA to reduce the complexity from M^{d_f} to $M^{d_f/2}$. The shuffling method rotates two independent N-dimensional real constellations to maximize the minimum product distances, with the same or different unitary rotations, then generates an N-dimensional complex mother constellation by concatenation of the two N-dimensional rotated real constellations. One of the two N-dimensional real constellations corresponds to the real part of the points of

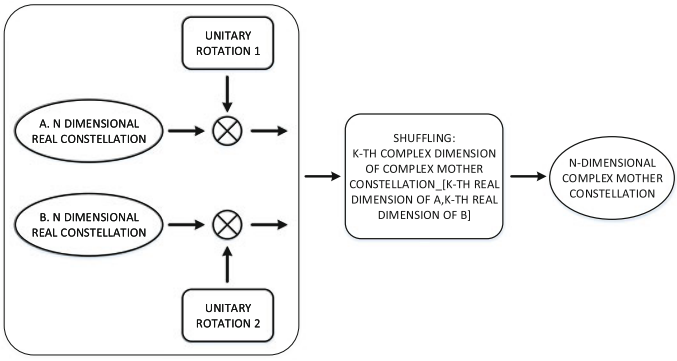


Fig. 12.7 Shuffling construction of the mother constellation [16]

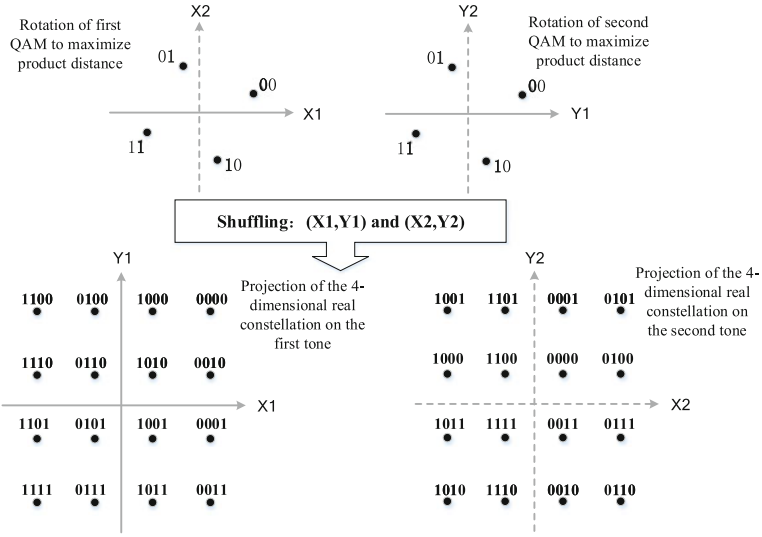


Fig. 12.8 An example of shuffling construction of two-dimensional 16-ary SCMA constellation [16]

the mother constellation, and the other one corresponds to the imaginary part. The construction is illustrated in Fig. 12.7.

Example 2 The construction of a 16-point SCMA mother constellation applicable to codebooks with two nonzero position ($N = 2$) by shuffling is illuminated in Fig. 12.8. Its optimum rotation angle is $\tan^{-1}\left(\frac{1+\sqrt{5}}{2}\right)$, which maximizes the minimum product distance.

B. Low-Projected Multi-dimensional Constellation

A key feature of SCMA codebooks is that the multi-dimensional constellation allows a few constellation points to collide over some of the dimensions, as they can still be separated through other components. An example is shown in Fig. 12.9, in which the constellation points corresponding to 01 and 10 collide over the first dimension, but are separated over the second tone, making the number of projection points equal to 3 instead of 4. By employing this low-projected constellation, the MPA receiver is able to reduce the number of probability calculations at the FNs during each iteration. As a result, the complexity is reduced to $O(M_p^{d_f})$, where $M_p \leq M$ is the number of projection points.

To do this, it is obliged to let the minimum "product distance"[4] be zero during the mother constellations design by rotation. However, the zero minimum product

[4]This is the relaxed product distance that takes the product of all the dimension-wise distance between two points into consideration.

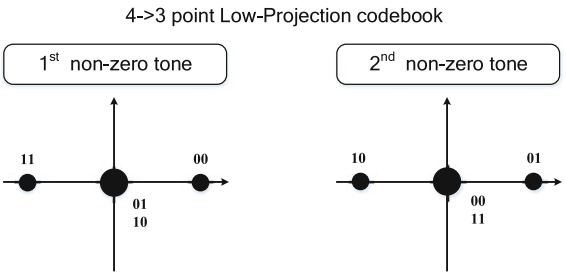


Fig. 12.9 An example of a 4-ary constellation with 3 projections per complex plane

distance would cause the performance degradation at high SNR, thus the trade-off between the performance and complexity should be considered for different scenarios.

12.3.2 Multi-user Codebooks Design for Uplink SCMA Systems

In this subsection, we introduce a practical codebook design approach for uplink SCMA systems over Rayleigh fading channels. Instead of optimizing the mother constellation and constellation operators separately, we address the joint design of multi-user constellations for small constellation size and number of users [22].

12.3.2.1 Design Criterion

To address the design of good codebooks, we need to establish appropriate performance criteria for a given system, i.e., determine the $\Upsilon(\cdot)$ in (12.37). It is straightforward to use the DCCC $I(\mathbf{X}, \mathbf{y})$ in (12.28), or the ACEP in (12.24), as the criterion, for increasing capacity or lowering probability of error. However, it is inefficient to use the DCCC or the ACEP as the cost function directly, since the evaluation of $I(\mathbf{X}, \mathbf{y})$ involves either Monte Carlo simulations or a large amount of numerical integration, and the calculation of the union bound on the ACEP is a little bit complicated.

As an alternative metric, the cutoff rate also gives an approximated evaluation for the capacity as well as the error probability and allows us to optimize the codebooks at a target value of SNR. Therefore, we can formulate the criterion for the multi-user codebooks design, by making the cutoff rate as large as possible, or equivalently the union-Bhatacharyya bound on the ACEP as small as possible. According to the cutoff rate analysis in (12.32) for MAC, maximizing R_0 is equivalent to choose the combined codewords such that

$$\{\mathbb{X}_j^+\}_{j=1}^J = \arg \min_{\{\mathbb{X}_j\}_{j=1}^J} \sum_{\mathbf{X}_a} \sum_{\mathbf{X}_b, b\neq a} \prod_{k=1}^{K} \left(1 + \frac{\lambda_k^2}{4N_0}\right)^{-1} \quad \text{s.t.} \quad \mathbb{E}\left[\|\mathbf{X}\|^2\right] = E_s \tag{12.41}$$

where E_s is the total power constraint on the multi-user codebooks. It is expected that the criterion is optimal in the sense of designing codebooks with large DCCC or small ACEP, since it involves all pairs of possible codewords for multiple users.

12.3.2.2 Signal-Space Diversity Scheme

Recall that the problem of designing multi-dimensional constellations for fading channels has been solved by using signal-space diversity (SSD), which rotates the QAM constellations with a unitary matrix, constructed either from the algebraic number theory or by computer search [21, 23]. Therefore, here, we use the rotated constellation to build the multi-user codebooks. In particular, we obtain the rotation matrices through computer search over compact parameterizations of unitary matrices. Note, that the null dimensions of codebooks are discarded before the rotations. The $N \times N$ unitary matrix can be written as [23]

$$\mathbf{R} = \prod_{m=1}^{N-1} \prod_{n=m+1}^{N} \mathbf{T}_{m,n} \tag{12.42}$$

where $\mathbf{T}_{m,n}$ is a complex Givens matrix given by [24]

$$\mathbf{T}_{m,n} = \begin{bmatrix} \ddots & & & & \\ & \cos\theta_{m,n} & \cdots & e^{-i\eta_{m,n}}\sin\theta_{m,n} & \\ & \vdots & \ddots & \vdots & \\ & -e^{i\eta_{m,n}}\sin\theta_{m,n} & \cdots & \cos\theta_{m,n} & \\ & & & & \ddots \end{bmatrix}$$

which changes the identity matrix by replacing its (m, m)th, (n, n)th, (m, n)th and (n, m)th elements with $\cos\theta_{m,n}$, $\cos\theta_{m,n}$, $e^{-i\eta_{m,n}}\sin\theta_{m,n}$ and $-e^{i\eta_{m,n}}\sin\theta_{m,n}$, respectively. The angles satisfy $\theta_{m,n} \in [-\pi, \pi]$ and $\eta_{m,n} \in [-\pi/2, \pi/2]$. So, the search for $\mathbf{R}$ is reduced to the search for a sequence of phase of

$$\begin{aligned} \boldsymbol{\theta} &= \{\theta_{1,2}, \ldots, \theta_{1,N}, \theta_{2,3}, \ldots, \theta_{2,N}, \ldots, \theta_{N-1,N}\}, \\ \boldsymbol{\eta} &= \{\eta_{1,2}, \ldots, \eta_{1,N}, \eta_{2,3}, \ldots, \eta_{2,N}, \ldots, \eta_{N-1,N}\}. \end{aligned} \tag{12.43}$$

It seems that optimizing the constellations using the above criterion is intractable even for a moderate number of users and codebook size, since we need to search

$N(N-1)J$ angles, and the summation in the right-hand side of (12.41) will add up $M^J(M^J-1)$ terms. However, searching results for two-dimensional constellations with a small number of users show that the rotation matrices are the same for all codebooks, and are independent of the number of users. Therefore, we simplify the optimization process by searching over a single rotation matrix, and reducing the number of accessed users, even though this is suboptimal. Furthermore, we can use the approach developed in [25], where all the entries of the rotation matrix are equal in magnitude. Therefore, by expanding the product in (12.42), we get $\boldsymbol{\theta} = \left\{\frac{\pi}{4}\right\}$ and $\boldsymbol{\theta} = \left\{\frac{\pi}{4}, 0.6155, \frac{\pi}{4}\right\}$ for $N = 2$ and $N = 3$, respectively. Exhaustive search is computationally feasible, provided, that each user occupies a moderate number of resources such that $N \leq 3$.

In the signal-space diversity scheme, the constellations are restricted to lattice constellations such that the rotated QAMs are suggested. In practice, the rotation can be done over any multi-dimensional constellations to improve the cutoff rate, e.g., the rotated spherical codebook [26] and rotations over the product of other low-dimensional constellations.

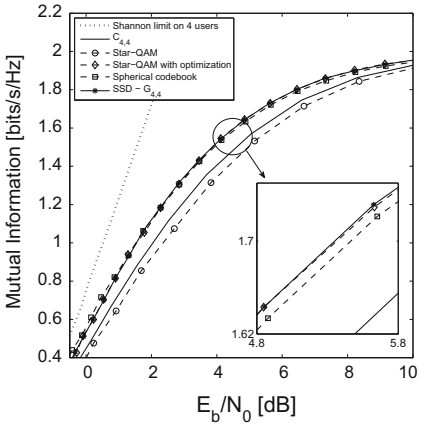
12.3.2.3 Simulations and Discussions

Consider the SCMA system in Example 1, which supports an overloading factor 150%. Simulation results of packet error rate (PER) for uncoded SCMA and DCCC are provided, which are performed over i.i.d Rayleigh fading channels for 4-ary and 16-ary codebooks. Four kinds of codebooks including the codebooks through SSD scheme discussed in this subsection (named as $G_{4,4}/G_{16,16}$), the codebook from [16] (named as $C_{4,4}/C_{16,16}$), spherical codebook [26], star-QAM-based codebooks [27], are employed, and we also provide the results of the star-QAM-based codebooks after optimization using the criterion (12.41), for which we extend α to complex numbers and get $\beta = 1, \alpha = -i$ and $\alpha = 0.8 - 0.8i$ for 4-ary and 16-ary codebooks, respectively.[5]

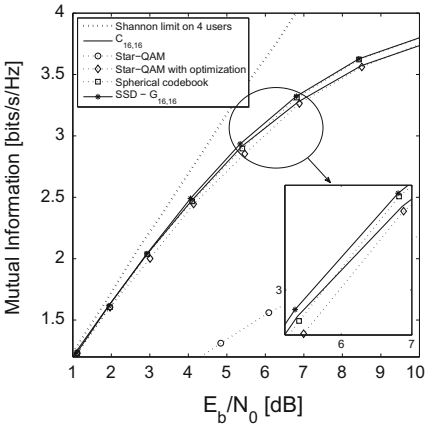
Figure 12.10 plots the DCCC of 4-ary and 16-ary codebooks for four users, together with the theoretical limit rates of i.i.d Gaussian inputs. As it is evident, the SSD scheme outperforms all the other codebooks in the high rate region for both 4-ary and 16-ary codebooks, while the mutual information gain is more clear for 4-ary case. While the rate of the star-QAM scheme is small, a significant gain is achieved after optimization with the criterion in (12.41), and it becomes as good as the SSD scheme for 4-ary codebook.

Figure 12.11 compares the PER performance of different codebooks for uplink SCMA with six users, where two antennas are employed for receive diversity, and the MPA detector is used with six iterations all the time. As it is observed, the SSD scheme has a gain about 0.8 dB over $C_{4,4}$ and 0.6 dB over $C_{16,16}$, and a gain about 0.5 dB and 0.3 dB over the spherical codebook for 4-ary and 16-ary cases, respectively.

[5] The star-QAM-based codebook targets on downlink channels, while its performance deteriorates in the uplink and for large constellation size.

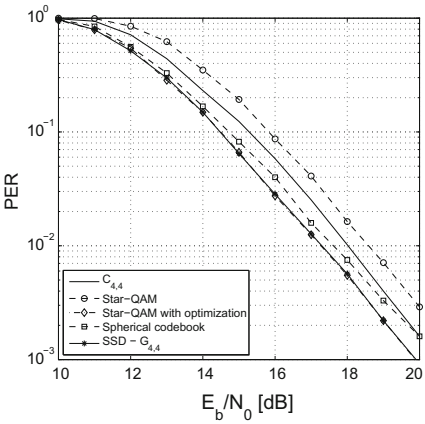


(a) 4-ary SCMA codebooks.

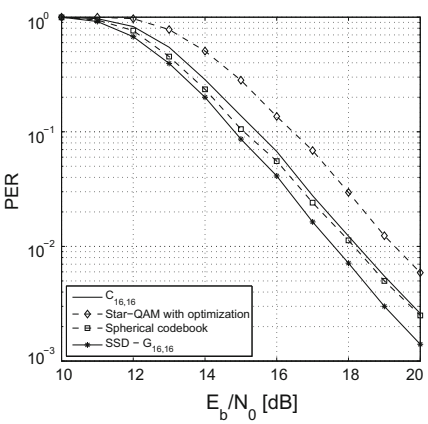


(b) 16-ary SCMA codebooks.

Fig. 12.10 Mutual information of uplink SCMA with 4 users



(a) 4-ary SCMA codebooks.



(b) 16-ary SCMA codebooks.

Fig. 12.11 PER of uplink SCMA systems over Rayleigh fading channels

Without optimization, the star-QAM scheme yields the worst error performance. However, it performs much better after optimization, which coincides with the result of mutual information in Fig. 12.10.

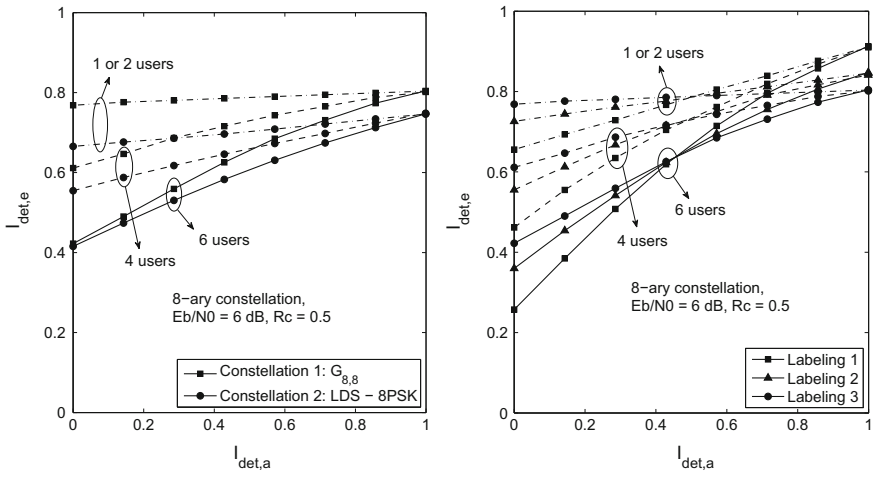
12.3.3 Low-Projected Multi-dimensional Constellations Design

As is discussed above, by employing the multi-dimensional constellations with low projections, the MPA receiver is able to utilize the constellation structure to reduce the receiver complexity. This subsection introduces an approach of constructing low-projected multi-dimensional constellations for uplink coded SCMA. In particular, constellation optimization for bit-interleaved convolutional coded SCMA with iterative multi-user detection is considered.

12.3.3.1 Transfer Characteristics of Turbo-MPA Detector

Extrinsic information transfer (EXIT) characteristics are investigated to find the effect of multi-user constellations on the performance of the MPA detector, and give us insights on the constellation optimization criteria. For each user, the EXIT chart analysis computes the average mutual information (AMI) between the extrinsic LLR (L_e), or the a priori LLR (L_a), and each coded bit. Thus, the extrinsic AMI is calculated as [28]

$$I_{\text{det,e}} = 1 - \frac{1}{\sqrt{2\pi\sigma_e^2}} \int_{-\infty}^{+\infty} \exp\left[\frac{-\left(l - \sigma_e^2/2\right)^2}{2\sigma_e^2}\right] \log_2(1 + e^{-l}) \mathrm{d}l$$



(a) EXIT charts for different constellations. (b) EXIT charts for different labelings.

Fig. 12.12 Impacts of constellations and labelings on the detector's transfer characteristics ($E_b/N_0 = 6$ dB, MPA detector with 3 iterations)

where σ_e^2 is the variance of the extrinsic LLR. It is worth noting that due to the multi-user interference, the *a priori* AMI ($I_{\text{det,a}}$) and extrinsic AMI ($I_{\text{det,e}}$) for each user will be influenced by the other users, and hence, a J-dimensional EXIT chart is necessary to characterize the transfer function. Here, the AMI is averaged over all the users such that the EXIT curves can be depicted on a one-dimensional complex plane.

Now, we investigate the transfer characteristics of the turbo-MPA detector for uplink SCMA over i.i.d. Rayleigh fading, and a factor graph matrix as in (12.5) is considered. Figure 12.12 presents the detector's transfer characteristics of two-dimensional 8-ary constellations for different number of users (J) at $E_b/N_0 = 6$ dB, where the MPA detector performs 3 iterations. Note that when $J = 2$, the signals from the two users are orthogonal with each other, so that they have the same AMI as that in the single-user case.

The impact of different constellations on the detector's transfer characteristics is shown in Fig. 12.12a, where the detector's EXIT curves of two different 8-ary constellations[6] with the same labelings are provided. Obviously, constellation 1 outperforms constellation 2, and the superiority of constellation 1 over constellation 2 is independent with the number of users. This implies that the effect of the constellation on the single-user system agrees with its multi-user counterpart, even though the EXIT curves become steeper as the number of users increases. In Fig. 12.12b, the results for the same 8-ary constellation with different mappings/labelings are demonstrated. It is observed that different labelings result in transfer characteristics curves of different slopes, for all the number of users cases, and the labeling with a steeper EXIT curve in the single-user case shows the larger slope in its multi-user counterpart.

The conclusion implies that the influence of constellations and labelings on the single-user system is consistent with that on the multi-user case. More precisely, a constellation or a labeling that is good for single-user systems will be beneficial to the multi-user systems. Therefore, we suggest to simplify the complicated multi-user constellations optimization in SCMA to the suboptimal single-user system design. It is expected that the constellation designing criteria for the single-user system is efficient for multi-user cases.

12.3.3.2 Design Criteria of Multi-dimensional Constellations

A. Links Between EXIT Charts and Constellation Design

The EXIT chart is a good tool to guide the system design. For iteratively decoded systems, given an outer convolutional code, the constellation should be designed to form a tunnel between the transfer curves of the detector and the decoder, and the

[6]The constellation 1 is constructed by rotation over the product of a binary phase-shift keying (BPSK) and a quadrature phase-shift keying (QPSK) constellation with Gray labelings, using the approach in Sect. 12.3.2.2 ($G_{8,8}$), and constellation 2 is the repetition over an 8PSK constellation with Gray labeling, i.e., the LDS scheme [29].

starting point of the detector curve and the intersection point between the detector curve and the decoder curve should be as high as possible, to guarantee a low threshold as well as a low error floor.

At a given value of SNR, the transfer characteristics of the detector are affected by both the constellation itself and the labeling, as shown in Fig. 12.12. In terms of the constellation, it is known that the area under the detector's EXIT curve is approximately equal to the DCCC per number of bits of a constellation point [30]. Based on this property, once a constellation with a larger DCCC is constructed, a larger area is obtained and then it has the potential of providing a wider EXIT tunnel, or equivalently, it would be easier to let the detector's curve to be above the decoder's curve. In terms of the labeling, for a given constellation, the detector curves corresponding to distinct labelings are rotations with each other, since the labeling does not change the DCCC and hence the area below the detector curve. A good labeling rotates the detector curve such that a large AMI is produced when $I_{\text{det,a}} = 1$, which provides an error floor that reaches the BER range of practical interest, and at the same time, to make the tunnel between the transfer curves of the detector and the decoder still open.

Based on these facts, we divide the constellation optimization framework into two steps. First, try to design the multi-dimensional constellation by maximizing the DCCC; Second, optimize the labeling by EXIT curve-fitting. In the following, we introduce two figures of merits for the constellation and the labeling.

B. Constellation Figure of Merit

As discussed above, the cutoff rate, corresponding to the DCCC, is a good criterion that allows us to optimize the constellation at a target value of SNR. Considering the received signal $\mathbf{y} = \text{diag}(\mathbf{h})\mathbf{x} + \mathbf{n}$, the cutoff rate constrained by an M-ary K-dimensional signal set $\mathbb{C}$ in i.i.d. Rayleigh fading, is given by [25]

$$\Psi_{\text{CFM}}(\mathbb{C}) = \log_2 M - \log\left[1 + \frac{1}{M}\sum_{\mathbf{x}_a \in \mathbb{C}} \sum_{\mathbf{x}_b \in \mathbb{C}, \mathbf{x}_b \neq \mathbf{x}_a} \prod_{k=1}^{K} \left(1 + \frac{\delta_k^2}{4N_0}\right)^{-1}\right] \quad (12.44)$$

where $\delta_k = |x_a[k] - x_b[k]|$, is the dimension-wise distance between any two distinct K-dimensional symbols $\mathbf{x}_a$ and $\mathbf{x}_b$. We take the quantity $\Psi_{\text{CFM}}(\mathbb{C})$ as the SNR-dependent constellation figure of merit, which is a function of SNR and the constellation $\mathbb{C}$, or the set of all pairwise distances between the constellation points. It involves all pairs of multi-dimensional symbols, and is independent of the labeling or any channel codes.

C. Labeling Figure of Merit

The constellation labeling is a crucial design parameter to achieve a high coding gain over the iterations for iteratively decoded bit-interleaved coded modulation (BICM) systems. To obtain an optimization criterion for the labelings, we resort to the error performance of multi-dimensional constellations under ideal interleaving. Let $\tilde{\mathbf{x}}_{(i)} = [\tilde{x}_{(i)}[1], \ldots, \tilde{x}_{(i)}[K]]^t$, be the symbol having the same label with that of $\mathbf{x}$

except at the ith bit position. The effect of labeling μ on the performance of BICM with iterative decoding (BICM-ID) systems employing multi-dimensional signal constellation can be characterized by [31]

$$\Psi_{\mathrm{LFM}}(\mu) = \frac{1}{mM} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{\mathbf{x} \in \mathbb{C}_i^b} \prod_{k=1}^{K} \left(1 + \frac{1}{4N_0} \delta_k^2\right)^{-1} \tag{12.45}$$

where $\delta_k = |x[k] - \tilde{x}_{(i)}[k]|$, and $\mathbb{C}_i^b$ is the subset of $\mathbb{C}$ that consisting of symbols whose label has the value b in the ith bit position. The SNR-dependent object function $\Psi_{\mathrm{LFM}}(\mu)$ is able to characterize the influence of both the constellation $\mathbb{C}$ and the labeling μ to the bit error rate (BER) performance of BICM-ID systems. With this criterion, one can optimize the bit labeling when fixing the signal constellation, or optimize the constellation for a given labeling, or optimize them jointly. Since optimizing the labeling μ by decreasing $\Psi_{\mathrm{LFM}}(\mu)$ improves the BER performance, we take $\frac{1}{\Psi_{\mathrm{LFM}}(\mu)}$ as the labeling figure of merit to guide the labeling design for a given multi-dimensional constellation.

12.3.3.3 Design Multi-dimensional Constellations

The multi-dimensional constellation with the same projections over each dimension can be viewed as a multi-modulation scheme [32], where the data bits are modulated into multiple one-dimensional symbols that are chosen from a one-dimensional complex constellation $\mathbb{A}$, called subconstellations in the following. The difference among the modulations for each dimension is that they have different labelings. This implies that the multi-dimensional constellation can be constructed by permutations of the one-dimensional subconstellation $\mathbb{A}$, dimensionally. Therefore, the problem is to design an M-ary subconstellation $\mathbb{A}$ with M_p distinct signal points, and the specific mapping or permutation for each dimension. In the following, we propose a multi-stage optimization, and the K-dimensional constellation is constructed by three steps:

(a) Determine the desired number of projection points M_p such that $M_p \leq M$, choose a one-dimensional M-ary subconstellation $\mathbb{A}$ with M_p projections;
(b) Based on the one-dimensional subconstellation $\mathbb{A}$, construct a K-dimensional constellation $\mathbb{C}$ using permutations;
(c) Design a labeling for the K-dimensional constellation $\mathbb{C}$.

A. Design One-Dimensional Subconstellation $\mathbb{A}$

Different from the traditional constellation design, the M-ary constellation with M_p projections imply that there are $M - M_p$ signal points that overlap with others. We first choose an M_p-ary constellation $\mathbb{A}_p$ without overlappings, then allocate the M_p signal points with M labels to obtain $\mathbb{A}$. The choice of $\mathbb{A}_p$ is various, any one-dimensional complex constellation, e.g., quadrature amplitude modulation (QAM)

or phase-shift keying (PSK), is available. Here, we construct $\mathbb{A}_p$ using the amplitude phase-shift keying (APSK) constellation, since it is able to provide good DCCC compared to other conventional modulations [33, 34].

An M_p-APSK constellation is composed of L concentric rings, each with uniformly spaced PSK points. The M_p-APSK constellation can be expressed as [33]

$$\mathbb{A}_p = \left\{ r_1 e^{j\theta_1} \mathbb{P}(m_1), \ldots, r_L e^{j\theta_L} \mathbb{P}(m_L) \right\}$$

where $\mathbb{P}(m_l)$ is an m_l-ary PSK constellation with unit average energy, and r_l, θ_l are the radius and phase offset of the lth ring, respectively. Let $\mathbf{m} = [m_1, \ldots, m_L]^t$, be the vector of the number of points over each ring so that $M_p = \sum_{l=1}^{L} m_l$. To guarantee a good distance profile, it is preferred to locate fewer constellation points on the inner rings than that on the outer rings. Then, for a set of ordered radius $r_1 < \cdots < r_L$, it is suggested that $m_1 \leq \cdots \leq m_L$.

Following the general APSK design procedure proposed in [34], the M_p-APSK constellation can be constructed as

- Select the number of rings L and the number of constellation points on each ring m_l, such that $M_p = \sum_{l=1}^{L} m_l$;
- Determine the radius of each ring r_l,

$$r_l = \sqrt{-\ln\left[1 - \frac{1}{M_p}\left(\sum_{i=1}^{l-1} m_i + \frac{m_l}{2}\right)\right]};$$

- Set θ_l as 0 or π/m_l.

Given the designed M_p-APSK constellation $\mathbb{A}_p$, we allocate the M-ary constellation with the M_p signal points. The problem can be formulated as how to put M numbers, $0, 1, \ldots, M-1$, into M_p sets, where each set represents a signal point in $\mathbb{A}_p$. The allocation strategy is preferred to follow the rules:

- The numbers that are allocated to a set should be less than or equal to M_p, and greater than or equal to 1, such that the overlapped points can be separated through other dimensions;
- The numbers in each set should be as less as possible, such that the resulted multi-dimensional constellation has a good distance profile;
- Symmetry of the constellation $\mathbb{A}$ is preferred so that it has a zero mean;
- The sets with low power levels may be allocated with more numbers, such that the constellation has a small average energy.

Note that in some cases, the allocation yields a constellation $\mathbb{A}$ with nonzero mean, then we shift $\mathbb{A}$ toward the origin, such that the mean of all signals is zero and therefore more energy-efficient.

Now, we give an example to illustrate the allocations. For a given 9-APSK $\mathbb{A}_p$ that is constructed with 3 rings and $\mathbf{m} = [1, 3, 5]^t$, a 16-ary subconstellation $\mathbb{A}$ can be obtained by allocating 16 numbers into 9 sets. Some possible allocations are given

in Fig. 12.13. Among the four strategies A, B, C, and D, while the strategy A is the most energy-efficient, it shows the worst performance when used to construct multi-dimensional constellations. This is because too many points overlap with each other, leading to a very poor distance profile for the multi-dimensional constellation. Numerical results show that the strategies B and C are equally efficient, and the strategy D is the best one, since the largest number of overlappings is only two.

B. Construct K-dimensional Constellation $\mathbb{C}$

Given the designed M-ary subconstellation $\mathbb{A}$, denote $\pi_k(\mathbb{A})$ a column vector of the kth permutation of the signals in $\mathbb{A}$, and let $\pi_1(\mathbb{A}) = \mathbb{A}$. The K-dimensional constellation through the permutation construction can be expressed with a $K \times M$ matrix as

$$\mathbb{C} = [\pi_1(\mathbb{A}), \ldots, \pi_K(\mathbb{A})]^t$$

where each column of the matrix corresponds to a K-dimensional symbol. Then, constructing a K-dimensional constellation requires to find $K-1$ permutations $\pi_2, \ldots, \pi_K$, such that the constellation figure of merit $\Psi_{\text{CFM}}(\mathbb{C})$ in (12.44) is maximized.

We focus on two-dimensional constellations ($K = 2$), by maximizing the constellation figure of merit, the unique permutation function π is selected as

$$\pi = \arg\min_{\hat{\pi}} \sum_{\mathbf{x}_a \in \mathbb{C}} \sum_{\mathbf{x}_b \in \mathbb{C}, \mathbf{x}_b \neq \mathbf{x}_a} \prod_{k=1}^{K} \left(1 + \frac{|x_a[k] - x_b[k]|^2}{4N_0}\right)^{-1}$$

There are $M!$ different choices for the permutations, for small constellation size where $M \leq 8$, the optimum solution can be solved by exhaustive search with a reasonable complexity. However, it becomes intractable for high order constellations. Note that this problem is similar to the labeling map of a constellation in bit-interleaved coded modulation with iterative decoding (BICM-ID) systems, which can be efficiently solved by using the binary switching algorithm (BSA) [35], or iteratively searching inside a randomly selected list, and a local optimum permutation can be found for a given cost function.

As for a larger dimension where $K > 2$, the search for $K-1$ permutations is challenging. A suboptimal solution can be used by successively optimizing the multi-dimensional constellation from lower dimensions to higher dimensions, such that only one permutation is needed to be checked in every round.

C. Labeling the K-dimensional Constellation

When a multi-dimensional constellation is found, we should choose an appropriate labeling for the constellation. In terms of EXIT chart, optimizing the labeling is to adjust the slope of the detector's curve. Our approach is to obtain a set of labelings with various slopes in their EXIT curves, firstly. Then, to choose a labeling from the set such that the detector EXIT curve matches with the decoder curve of a given convolutional code.

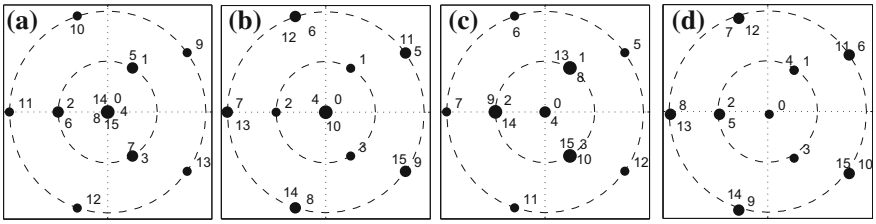


Fig. 12.13 Examples of 16-ary subconstellation $\mathbb{A}$ based on 9-APSK with $\mathbf{m} = [1, 3, 5]^t$

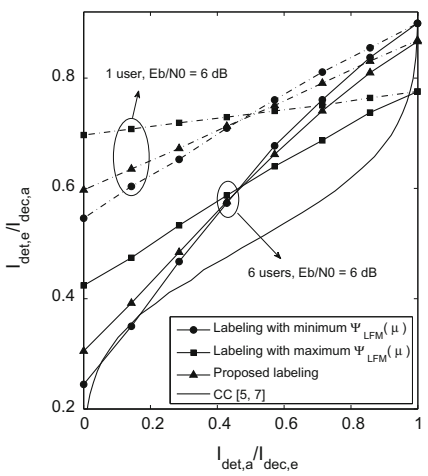
The labeling figure of merit $\frac{1}{\Psi_{\mathrm{LFM}}(\mu)}$ in (12.45) represents the ultimate performance with perfectly known *a priori* AMI, and in EXIT charts, it corresponds to the maximum achievable value of $I_{\mathrm{det,e}}$ with $I_{\mathrm{det,a}} = 1$, denoted as I^*, and I^* becomes larger with decreasing $\Psi_{\mathrm{LFM}}(\mu)$. It is observed in Fig. 12.12a, b that for single-user systems, the detector's EXIT curves corresponding to distinct labelings can be approximated to straight lines, with a common intersection around the point with $I_{\mathrm{det,a}} = 0.5$. Then, the slope of the EXIT curve corresponding to a labeling can thus be determined by I^*, approximately. A labeling with a larger I^* can have a steeper transfer curve. Therefore, we can use $\Psi_{\mathrm{LFM}}(\mu)$ to approximately control the slope of the EXIT curves, and the detector's curve becomes steeper with decreasing $\Psi_{\mathrm{LFM}}(\mu)$.

Denote the set of labelings as Ω. For constellations with small sizes ($M \leq 8$), Ω is chosen to be the set of all possible labelings with distinct $\Psi_{\mathrm{LFM}}(\mu)$. For higher order constellations, the BSA can be used once again to obtain Ω. Begin with a given original labeling, by minimizing $\Psi_{\mathrm{LFM}}(\mu)$ using the BSA, new labelings with increasing slopes may be obtained during the search, we output these labelings and store them into Ω. Similarly, new labelings with decreasing slopes may be obtained by maximizing $\Psi_{\mathrm{LFM}}(\mu)$ with the same original labeling. Then, the labelings in Ω are sorted with increasing values of $\Psi_{\mathrm{LFM}}(\mu)$. The set Ω can also be obtained by iteratively searching inside a randomly selected list.

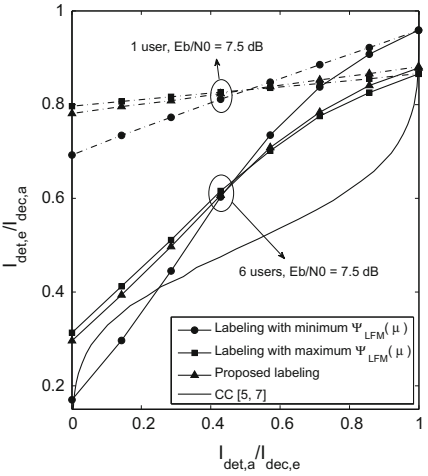
Now, we choose a labeling from Ω, with the aid of EXIT chart analysis. At an appropriate SNR, the following two conditions have to be fulfilled for the labeling:

(a) the slope of either the single-user or multi-user detector EXIT curve should be as steep as possible, to achieve a low BER error floor;
(b) the tunnel between the decoder and the multi-user detector curves should be open, or the intersection point between them should be as high as possible, to guarantee a low threshold.

As an illustrative example, Fig. 12.14 shows the choices of labelings for a two-dimensional 8-ary constellation (with 3 projections) and a 16-ary constellation (with 9 projections) for SCMA. The detector curves of several labelings, with distinct $\Psi_{\mathrm{LFM}}(\mu)$, as well as the decoder curve are provided, where a half-rate four-state non-recursive convolutional code with generator $[5, 7]_8$ is employed as the outer channel code, and $I_{\mathrm{dec,a}}(I_{\mathrm{dec,e}})$ denotes the AMI between the *a priori* (extrinsic) LLR and the transmitted coded bit at the input (output) of the convolutional decoder.



(a) Labeling for an 8-ary constellation with 3 projections.



(b) Labeling for a 16-ary constellation with 9 projections.

Fig. 12.14 Examples of labelings for two-dimensional constellations

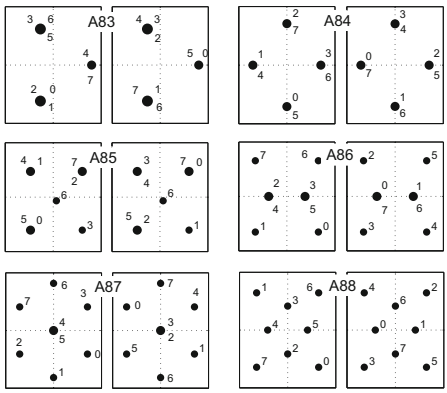
As it is observed in Fig. 12.14, the labeling with the maximum $\Psi_{\mathrm{LFM}}(\mu)$ yields a relatively flat slope in the 6 users case, and the one with the minimum $\Psi_{\mathrm{LFM}}(\mu)$ closes the tunnel between the decoder and the multi-user detector curves. In contrast, the proposed labeling, which shows a very steep slope in the EXIT curve while still keeps the tunnel open, achieves a good trade-off between the threshold and the BER performance.

With the proposed approach, it is possible to construct constellations with any projections. Figure 12.15a and 12.15b show the examples of the designed two-dimensional 8-ary constellations with various projections, and a four-dimensional 64-ary constellation with 16 projections, respectively.

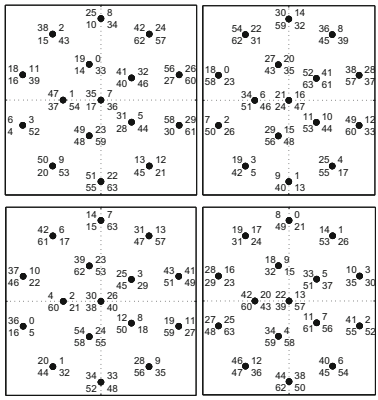
12.3.3.4 Simulations and Discussions

In the following, simulations are conducted to evaluate the performance of low-projected constellation for an uplink convolutional coded SCMA system. The detailed simulation configuration is given in Table 12.2. The SCMA system follows a factor graph matrix as (12.5), which supports an effective system loading of 75% (JR_c/K). For the sake of simplicity, let A_{M,M_p} denote the APSK-based M-ary constellation with M_p projections. The constellations in [16, 36, 37] are named as C_{M,M_p} (the SSD scheme in Sect. 12.3.2 is denoted as G_{M,M_p}), which are used as the benchmark.

The simulated BER performances of 4-ary, 8-ary and 16-ary codebooks are depicted in Figs. 12.16 and 12.17, respectively. It is obvious that the A_{M,M_p} code-



(a) Two-dimensiona l8-ary constellations with various projections.



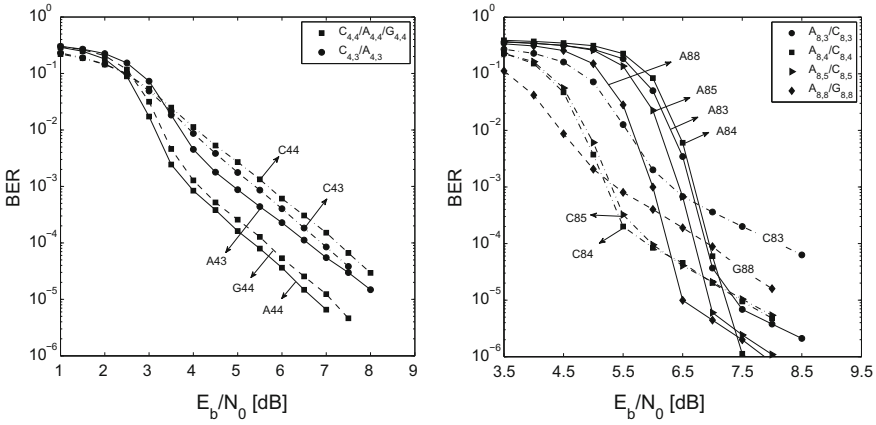
(b) A four-dimensional 64-ary constellation with 16 projections.

Fig. 12.15 Examples of APSK-based low-projected SCMA constellations

Table 12.2 Simulation parameters

Parameters	Values
Channel model	Uplink Rayleigh fading channel
Target spectral efficiency	1.5, 2.25, 3 bits/resource
Number of users	6
FEC coding	1/2-rate convolutional code with generator $[5, 7]_8$
Interleaving	Random interleaver, interleave length: 1024 bits
Codebooks	C_{43}/A_{43}, $C_{44}/A_{44}/G_{44}$, C_{83}/A_{83}, C_{84}/A_{84}, C_{85}/A_{85}, A_{88}/G_{88} C_{169}/A_{169}, $C_{1616}/G_{1616}/A_{1616}$
Receiver	Turbo-MPA (3 MPA iterations + 6 BICM iterations)

books outperform others in the large SNR region, for almost all the simulation cases. In Fig. 12.16a, the BER floor of $A_{4,3}$ is lower than $C_{4,3}$ when SNR is less than 8 dB, and $A_{4,4}$ shows much better performance than $C_{4,4}$, and has a gain about 0.25 dB over $G_{4,4}$. Note that $A_{4,4}$ outperforms $G_{4,4}$ in the whole SNR region. This is because $A_{4,4}$ has the same labeling with $G_{4,4}$ but the larger DCCC. For the 8-ary codebooks shown in Fig. 12.16b, the error floors of the A_{M,M_p} codebooks happen at the BER level below 10^{-5}, which are much lower than the other codebooks. Thus, much smaller values of SNR are required to achieve a BER value of 10^{-6}. Similar results are also obtained for 16-ary codebooks, which are shown in Fig. 12.17. The gain of the A_{M,M_p} codebooks over C_{M,M_p} is smaller than that in the 8-ary codebooks case. The BER curve of $C_{16,16}$ degrades earlier than the others, but it arrives the error



(a) 4-ary codebooks. (b) 8-ary codebooks.

Fig. 12.16 BER performance of uplink coded SCMA for 4-ary and 8-ary codebooks

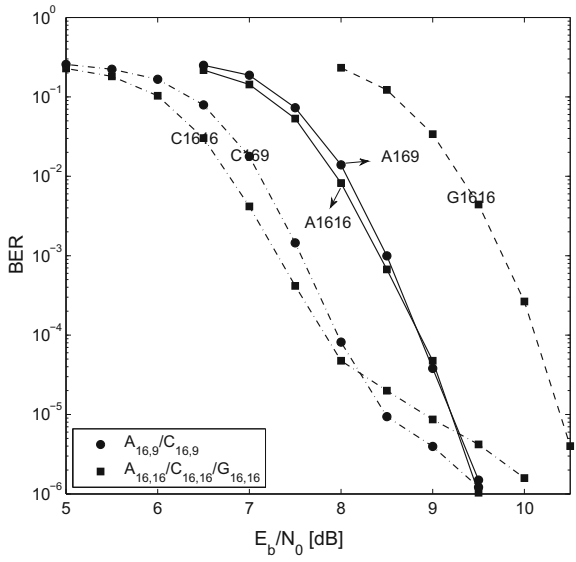


Fig. 12.17 BER performance of uplink coded SCMA for 16-ary codebooks

floor quickly. The codebook $G_{16,16}$ shows a very large threshold, and $A_{16,16}$ achieves a good trade-off. Even though the BER threshold of $A_{16,9}$ is larger than $C_{16,9}$, $A_{16,9}$ shows almost the same BER performance with $C_{16,9}$ at the SNR around 9.5 dB. To achieve the BER performance of 10^{-6}, equal or less SNR are required for the A_{M,M_p} codebooks.

12.4 SCMA for 5G Radio Transmission

12.4.1 Application Scenarios for 5G Networks

In addition to achieving higher transmission rates, faster access, supporting of larger user density and better user experience in enhanced mobile broadband (eMBB), the 5G air interface connects to new vertical industries and new devices, creating new application scenarios such as massive machine type communications (mMTC) and ultra reliable low latency communications (URLLC) services, by supporting massive number of devices and enabling mission-critical transmissions with ultra high reliability and ultra-low latency requirement, respectively. This presents new challenges and considerations for the radio multiple access to be fully scalable to support these diverse service requirements. The current orthogonal multiple access might not be able to fulfill some of the requirements, such as services for dense MTC devices deployments, and SCMA can be considered as a promising candidate to meet the 5G performance requirements. In particular, SCMA is proposed for 5G to achieve the following benefits:

- for eMBB: larger capacity region by non-orthogonal multiplexing; robustness to fading and interference with code-domain design; robust link adaptation with relaxed CSI accuracy.
- for URLLC: higher reliability through diversity gain achieved by multi-dimensional constellations, and robustness to collision by carefully design the codebooks; latency reduction and more transmission opportunities by enabling grant-free access; Non-Orthogonal Multiplexing of mixed traffic types.
- for mMTC: higher connection density with high overloading; reduction of signaling overhead and power consumption by enabling grant-free access.

Moreover, it is also possible to extend SCMA application to unlicensed spectrum and V2X systems, since the non-orthogonal transmissions can help to increase the system efficiency and deal with the interference.

The link-level performance evaluation for some uplink SCMA scenarios is provided in [38], which compares SCMA with orthogonal frequency division multiple access (OFDMA) in typical scenarios and investigate the robustness of SCMA to overloading and codebook collision. Results show that SCMA achieves significant gain over orthogonal multiple access with good codebook design, and the gain increases as the supported number of users and target spectrum efficiency increases. Moreover, high overloading with stable performance is feasible with SCMA design, which enables robust overloaded transmission, and the performance loss with codebook collision is negligible with SCMA design, which enables robust grant-free transmission.

12.4.2 Challenges and Future Works

While SCMA is able to greatly enhance the system capability for 5G networks, some further issues on design and implementation of SCMA remain to be resolved, which can be listed as follows:

- Reduced complexity receiver design: Even though MPA or EPA receiver is able to significantly reduce the complexity of SCMA, the complexity of MPA is still very high and iterative multi-user receiver is usually required, which brings several challenging issues for practical implementation:
  - It limits the capability of SCMA to support massive connectivity;
  - The iterative multi-user detection brings a large processing delay;
  - The complexity makes it difficult for SCMA to employ constellations with large sizes, hence limits the transmission rate.

  Sophisticated multi-user detection schemes should be developed to address the high complexity.
- Theoretical analysis: Further theoretical analysis of SCMA is needed to get more insights on the practical system design. For example, the capacity or error performance with randomly codebook allocations. Also, interference cancelation may be incorporated into the MPA detection for lower complexity, then it is desirable to determine the performance and capacity under practical detectors.
- Codebook design: The codebook design is complicated, especially for high-dimensional codebooks and that with large size. Advanced multi-dimensional constellation construction is necessary, and the joint design of factor graph matrix and constellations is to be developed, for further performance improvement. Moreover, the design for the scenario that all the overloaded users have different transmission rates (codebook sizes) is to be investigated, to improve the link adaptation.
- Other issues: System scalability of supporting various loading, SCMA in both uplink and downlink transmissions, supporting of other techniques such as MIMO, resource/codebook allocation, channel estimation for uplink SCMA, etc.

References

1. H. Nikopour, H. Baligh, Sparse code multiple access, in *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'13)* (2013), pp. 332–336
2. S. Zhang, X. Xu, L. Lu, Y. Wu, G. He, Y. Chen, Sparse code multiple access: an energy efficient uplink approach for 5G wireless systems, in *IEEE Global Communications Conference (GLOBECOM'14)* (2014), pp. 4782–4787
3. Y. Wu, S. Zhang, Y. Chen, Iterative multiuser receiver in sparse code multiple access systems, in *Proceedings of IEEE International Conference on Communications (ICC'15)* (2015), pp. 2918–2923
4. J. Zhang, L. Lu, Y. Sun et. al., PoC of SCMA-Based Uplink Grant-Free Transmission in UCNC for 5G. IEEE J. Sel. Areas Commun. **35**, 1353–1362 (2017)

5. R. Hoshyar, F.P. Wathan, R. Tafazolli, Novel low-density signature for synchronous CDMA systems over AWGN channel. IEEE Trans. Signal Process. **56**, 1616–1626 (2008)
6. D. Guo, C. Wang, Multiuser detection of sparsely spread CDMA. IEEE J. Sel. Areas Commun. **26**, 421–431 (2008)
7. R1-166098: Discussion on feasibility of advanced MU-detector. Huawei, HiSilicon, 3GPP TSG RAN WG1 Meeting #86 (2016)
8. X. Meng, Y. Wu, Y. Chen, M. Cheng M, Low complexity receiver for uplink SCMA system via expectation propagation, in *Proceedings of Wireless Communications and Networking Conference (WCNC' 17)*, (2017), pp. 1–5
9. J. Bao, Z. Ma, G.K. Karagiannidis, M. Xiao, Z. Zhu, Joint multiuser detection of multidimensional constellations over fading channels. IEEE Trans. Commun. **65**, 161–172 (2017)
10. D. Tse, P. Viswanath, *Fundamentals of Wireless Communications* (Cambridge University Press, 2005)
11. J.G. Proakis, M. Salehi, *Digital Communications* (McGraw-Hill, New York, 2008)
12. E. Björnson, D. Hammarwall, B. Ottersten, Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems. IEEE Trans. Signal Process. **57**, 4027–4041 (2009)
13. M. Chiani, D. Dardari, M.K. Simon, New exponential bounds and approximations for the computation of error probability in fading channels. IEEE Trans. Wirel. Commun. **2**, 840–845 (2003)
14. S.J. Grant, J.K. Cavers, Performance enhancement through joint detection of cochannel signals using diversity arrays. IEEE Trans. Commun. **46**, 1038–1049 (1998)
15. X. Zhu, R.D. Murch, Performance analysis of maximum likelihood detection in a MIMO antenna system. IEEE Trans. Commun. **50**, 187–191 (2002)
16. M. Taherzadeh, H. Nikopour, A. Bayesteh, H. Baligh, SCMA codebook design, in *Proceedings of IEEE 80th Conference on Vehicular Technology (VTC Fall'14)* (2014), pp. 1–5
17. A. Montanari, D. Tse, Analysis of belief propagation for nonlinear problems: the example of CDMA (or: How to prove Tanaka's formula), in *Proceeding IEEE Information Theory Workshop (ITW)* (2006), pp. 122–126
18. C.C. Wang, D. Guo, Belief propagation is asymptotically equivalent to MAP estimation for sparse linear systems, in *Proceedings of 44th Annual Allerton Conference on Communication, Control, and Computing* (2006), pp. 926–935
19. L. Li, Z. Ma, L. Wang, P. Fan, L. Hanzo, Cutoff rate of sparse code multiple access in downlink broadcast channels. IEEE Trans. Commun. **65**, 3328–3342 (2017)
20. J. Boutros, E. Viterbo, C. Rastello, J.C. Belfiore, Good lattice constellations for both Rayleigh fading and Gaussian channels. IEEE Trans. Inf. Theory. **42**, 502–518 (1996)
21. J. Boutros, E. Viterbo, Signal space diversity: a power- and bandwidth-efficient diversity technique for the rayleigh fading channel. IEEE Trans. Inf. Theroy **44**, 1453–1467 (1998)
22. J. Bao, Z. Ma, Z. Ding, G.K. Karagiannidis, Z. Zhu, On the design of multiuser codebooks for uplink SCMA systems. IEEE Commun. Lett. **20**, 1920–1923 (2016)
23. Y. Xin, Z. Wang, G.B. Giannakis, Space-time diversity systems based on linear constellation precoding. IEEE Trans. Wireless Commun. **2**, 294–309 (2003)
24. G.H. Golub, C.F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, 1996)
25. S.P. Herath, N.H. Tran, T. Le-Ngoc, Rotated multi-D constellations in rayleigh fading: mutual information improvement and pragmatic approach for near-capacity performance in high-rate regions. IEEE Trans. Commun. **60**, 3694–3704 (2012)
26. J. Bao, Z. Ma, M.A. Mahamadu, Z. Zhu, D. Chen, Spherical codes for SCMA codebook, in *Proceedings of IEEE 83th Conference on Vehicular Technology (VTC Spring'16)* (2016), pp. 1–5
27. L. Yu, X. Lei, P. Fan, D. Chen, An optimized design of SCMA codebook based on star-QAM signaling constellations, in *Proceedings of International Conference on Wireless Communications & Signal Processing (WCSP'15)* (2015), pp. 1–5
28. S.T. Brink, Convergence behavior of iteratively decoded parallel concatenated codes. IEEE Trans. Commun. **49**, 1727–1737 (2001)

29. J.V.D. Beek, B.M. Popović, Multiple access with low-density signatures, in *Proceedings of IEEE Conference on Global Communications (GLOBECOM)* (2009)
30. A. Ashikhmin, G. Kramer, S.T. Brink, Extrinsic information transfer functions: model and erasure channel properties. IEEE Trans. Inf. Theory. **50**, 2657–2673 (2004)
31. N.H. Tran, H.H. Nguyen, Design and performance of BICM-ID systems with hypercube constellations. IEEE Trans. Wirel. Commun. **5**, 1169–1179 (2006)
32. A. Seyedi, Multi-QAM modulation: a low-complexity full rate diversity scheme, in *Proceedings of IEEE International Conference on Communications (ICC)* (2006), pp. 1470–1475
33. C.M. Thomas, M.Y. Weidner, S.H. Durrani, Digital amplitude-phase keying with M-ary alphabets. IEEE Trans. Commun. **22**, 168–180 (1974)
34. Q. Xie, Z. Yang, J. Song, L. Hanzo, EXIT-chart-matching-aided near-capacity coded modulation design and a BICM-ID design example for both gaussian and rayleigh channels. IEEE Trans. Veh. Tech. **62**, 1216–1227 (2013)
35. F. Schreckenbach, N. Görtz, J. Hagenauer, G. Bauch, Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding. IEEE Commun. Lett. **7**, 593–595 (2003)
36. M.T. Boroujeni, A. Bayesteh, H. Nikopour, M. Baligh, System and method for generating codebooks with small projections per complex dimension and utilization thereof, U.S. Patent 0,049,999, 18 Feb 2016
37. A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, J. Ma, Low complexity techniques for SCMA detection, in *Proceedings of IEEE Globecom Workshops* (2015), pp. 1–6
38. R1-164037: LLS results for uplink multiple access. Huawei, HiSilicon, 3GPP TSG RAN WG1 Meeting #85 (2016)