

# Preamble Design for Collision Detection and Channel Estimation in Machine-Type Communication

Shilei Zheng<sup>(✉)</sup>, Fanggang Wang, and Xia Chen

Beijing Jiaotong University, Beijing, China  
14120188@bjtu.edu.cn

**Abstract.** Preamble is widely used for initial synchronization, channel estimation, user identification, and collision detection in communication systems. For machine-type communication (MTC), there are massive machines within one cell. Contention-based random access could be a candidate protocol in this scenario. However, simultaneous transmission of multiple users can lead to signal aliasing. This paper designs a novel structure of preamble. Collisions of multiple users can be detected based on well-designed structure of the preamble. Furthermore, channel state information (CSI) can be estimated as a byproduct in the process of collision detection. We claim that the proposed preambles applies in flat-fading channel, multipath channel and asynchronous scenario. The simulation results validate the accuracy and robustness of the proposed scheme.

**Keywords:** Asynchronization · Contention-based random access · Collision detection · Channel estimation · Preamble

## 1 Introduction

In wireless systems, synchronization and channel estimation are generally accomplished by a particular signal, which is usually composed of two parts, i.e., a short training sequence for synchronization followed by a long training sequence mainly for channel estimation [1–3]. Optimal preamble sequences lead to high access probability and enhance the system performance [4]. Previous research of preamble design focused on scheduling-based communications. In this paper, we investigate preamble design for contention-based communications, which can detect collision as well in addition to synchronization acquisition and channel estimation.

Contention-based communications attract much interest due to its potential to serve a large number of wireless access terminals in machine-type communication (MTC) in 5G. So far, the cellular network is mainly based on scheduling, in which each user is assigned to a time/frequency/space unit orthogonal to other

---

F. Wang—This work was supported in part by the National Natural Science Foundation under Grant 61571034 and under Grant U1334202, the Fundamental Research Funds for the Central Universities under Grant 2015JBM112, and the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2016ZT013.

users. In future, machines and/or things will be also connected into the network rather than human only. With massive terminals which generally requires random sparse traffic, contention-based communication becomes a good candidate. However, traditional contention-based protocol allows no collision. In [5], it employed a fully-distributed random access protocol for collision avoidance mechanism and in [6] it proposed using a novel medium access protocol Carrier Sense Multiple Access/Collision Detection with Reservation (CSMA/CDR) to resolve collision. In addition, adopting orthogonal spreading sequences having a high autocorrelation to detect the collision is another choice, such as using Zadoff-Chu (ZC) sequence [7, 8]. For example, in the random access process of LTE, the network access point detects the collision by a bank of correlators, each of which is matched to a particular ZC sequence in the set [9].

However, the method of ZC sequence requires the correlation of different sequences to be zero and thus it is very sensitive to asynchronization. In this paper, we proposed a novel structure of preamble, with which we can detect collision and estimate CSI regardless of correlation issue at the receiver. Specifically, a Discrete Fourier Transformation (DFT) matrix is used to design the preamble, preamble matrices of different users can be obtained by alternating rows of a basis DFT matrix. Accordingly, a method of collision detection is provided which applies in practical channel conditions and channel state information can be estimated as a byproduct. The receivers can utilize the structure of the preamble to detect collision. And a collision detection algorithm which aims at asynchronous and multipath scenarios in practical communication is proposed. With the results of the algorithm, CSI can be estimated.

The rest of the paper is organized as follow. Section 2 introduces the system model. Sections 3 and 4 we propose method for collision detection, channel estimation in practical channel conditions. In Sect. 5 simulation results are provided. Finally, Sect. 6 concludes this paper.

## 2 System Model and Preamble Structure

Consider a multiple access channel with one access point and  $M$  users. The transmission of the users are contention-based, in which collision occurs with a large probability when there are massive users, which is a typical scenario in 5G communication.

Then, we introduce the structure of our proposed preamble by using a Discrete Fourier Transformation (DFT) matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{L-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(L-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^M & w^{2M} & \cdots & w^{M(L-1)} \end{bmatrix} \quad (1)$$

where  $w = e^{j\frac{2\pi}{L}}$ , and  $M, L \in \mathbb{Z}^+$ . For illustration brevity, we define a vector as

$$\mathbf{g}_i = [1, w^i, w^{2i}, \cdots, w^{i(L-1)}]^T, \quad i \in \mathcal{I}_M \quad (2)$$

where  $\mathcal{I}_M = \{0, 1, \dots, M\}$ . In this paper, we propose to use  $\mathbf{G}$  as a basis matrix. The number of rows is equal to maximum users which can be detected in a collision, and the number of columns is related to the diversity order which is further elaborated later. Accordingly, the basis matrix is rewritten by

$$\mathbf{G} = [\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_M]^T. \tag{3}$$

The preamble matrix  $\mathbf{G}_k$  for the  $k$ th user is obtained by alternating the first row and the  $(k + 1)$ th row of the basis matrix  $\mathbf{G}$ , which is written by

$$\mathbf{G}_k = [\mathbf{g}_k, \mathbf{g}_2, \dots, \mathbf{g}_0, \dots, \mathbf{g}_M]^T \tag{4}$$

Note that the preamble matrix is sent column by column in practice. It is stacked up into an equal dimension matrix at the access point for detecting collision and estimating channels<sup>1</sup>.

### 3 Collision Detection and Channel Estimation

#### 3.1 Toy Model of Collision Detection

In this toy model, we assume that, in a noiseless channel the  $i$ th user and the  $k$ th user send their preambles in exactly the same slot. The overlapped preamble signal  $\mathbf{Y}_t = \mathbf{G}_i + \mathbf{G}_k$  at the access point can be expressed as

$$\mathbf{Y}_t = [\mathbf{g}_i + \mathbf{g}_k, 2\mathbf{g}_1, \dots, \mathbf{g}_0 + \mathbf{g}_i, \dots, \mathbf{g}_k + \mathbf{g}_0, \dots, 2\mathbf{g}_M]^T. \tag{5}$$

From the definition of the preamble matrix in (1) and (4), each element of the first column of  $\mathbf{Y}_t$  is 2 currently, which is equal to the number of collision users. In practice, the preamble is contaminated by noises. Then, we average over all elements in the first column to determine the number of collisions. We further notice that the number of collisions can also be estimated by the sum of all elements of  $\mathbf{Y}_t$  divided by that of  $\mathbf{G}$ , which is straightforward since  $\mathbf{Y}_t$  is obtained by alternating different rows of  $\mathbf{G}$  only. After that, we can further determine which users collide from a decision matrix  $\mathbf{Y}'_t$  calculated by

$$\begin{aligned} \mathbf{Y}'_t &= \mathbf{Y}_t - K\mathbf{G} = \mathbf{G}_i + \mathbf{G}_k - 2\mathbf{G} \\ &= \begin{bmatrix} 0 & w^i + w^k - 2 & \dots & w^{i(L-1)} + w^{k(L-1)} - 2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 - w^i & \dots & 1 - w^{i(L-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 - w^k & \dots & 1 - w^{k(L-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}. \end{aligned} \tag{6}$$

---

<sup>1</sup> There could be misalignment among multiple users, which induces extra elements for reconstructing the preamble matrix. The extra elements can be simply ignored. Details are provided in Sect. 3.3.

The collision users are determined by the indices of nonzero rows in  $\mathbf{Y}'_t$  except for the first row.

### 3.2 Collision Detection in Flat-Fading Channel

Consider a noiseless flat fading channel, and the  $i$ th user and the  $k$ th user still transmit in the same slot. The channel responses of the  $i$ th user and the  $k$ th user are  $h_i$  and  $h_k$ , respectively. Then the received preamble matrix  $\mathbf{Y}_f$  can be expressed as

$$\begin{aligned} \mathbf{Y}_f &= h_i \mathbf{G}_i + h_k \mathbf{G}_k \\ &= [h_i \mathbf{g}_i + h_k \mathbf{g}_k, \dots, h_i \mathbf{g}_0 + h_k \mathbf{g}_i, \dots, h_i \mathbf{g}_k + h_k \mathbf{g}_0, \dots, (h_i + h_k) \mathbf{g}_M]^T. \end{aligned} \tag{7}$$

Similarly to the previous toy model, we can average over the first column to get the sum of the channel fading of the two users, i.e.,  $h_i + h_k$ . An alternative method is to sum up over all element divided by the sum of all element of  $\mathbf{G}$ . Then the detection matrix  $\mathbf{Y}'_f$  is calculated by

$$\begin{aligned} \mathbf{Y}'_f &= \mathbf{Y}_f - (h_i + h_k) \mathbf{G} = h_i \mathbf{G}_i + h_k \mathbf{G}_k - (h_i + h_k) \mathbf{G} \\ &= \begin{bmatrix} 0 & h_i w^i + h_k w^k - (h_i + h_k) & \dots & h_i w^{i(L-1)} + h_k w^{k(L-1)} - (h_i + h_k) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & h_i(1 - w^i) & \dots & h_i(1 - w^{i(L-1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & h_k(1 - w^k) & \dots & h_k(1 - w^{k(L-1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}. \end{aligned} \tag{8}$$

Similarly, we determine the collision users by finding nonzero rows in  $\mathbf{Y}'_f$  except for the first row. Furthermore, channel fading of the users can be estimated (by averaging method) as

$$\hat{h}_j = \frac{1}{L-1} \cdot \sum_{\ell=1}^{L-1} \frac{h_j(1 - w^{j\ell})}{1 - w^{j\ell}}, \quad j = i, k \tag{9}$$

where  $\hat{h}_j$  is the estimated channel fading of user  $j$ , and  $j = i, k$ .

### 3.3 Collision Detection in Asynchronous Scenario

In this section, we consider two users collides in an asynchronous manner. First, we denote the received matrix by a delay sum operation, which is expressed as  $\mathbf{C} = \text{sum}(\mathbf{A}, \mathbf{B}; d)$ , where the matrix  $\mathbf{A}$  is summed up with a  $d$ -element-delay version of  $\mathbf{B}$ , and the elementary delay occurs column by column from the first

elementary of the matrix signal is sent column by column. With the definition of the delay sum, the received signal matrix is expressed as

$$\mathbf{Y}_a = \text{sum}(h_i \mathbf{G}_i, h_k \mathbf{G}_k; d) \tag{10}$$

where  $h_i$  and  $h_k$  are the channel state information of the  $i$ th user and the  $k$  user respectively, and here in this example  $d$  is the number of symbols between the two users. From the structure obtained by the delay sum, it can be seen that the previous preamble cannot be applied to detect the two collided users directly. Next, we proposed an enhanced structure of preamble. Define  $f = e^{\frac{j2\pi}{N}}$ , where  $N$  is the number of signal sampling points ( $N > M + 1$ ), matrix  $\mathbf{E}$  can be expressed as

$$\mathbf{E} = \text{diag} [f, f^2, \dots, f^{M+1}]. \tag{11}$$

The preamble matrix  $\mathbf{F}_i$  of the  $i$ th user can be designed by

$$\mathbf{F}_i = \mathbf{E} \mathbf{G}_i = [f \mathbf{g}_i, f^2 \mathbf{g}_1, \dots, f^{i+1} \mathbf{g}_0, \dots, f^{M+1} \mathbf{g}_M]^T \tag{12}$$

The fundamental interpretation of this procedure is to modulate the symbols in each row to a different carrier, which can be realized by fast Fourier transform (FFT). Then, the received signal  $\mathbf{Y}_a$  is expressed as

$$\begin{aligned} \mathbf{Y}_a &= \text{sum}(h_i \mathbf{F}_i, h_k \mathbf{F}_k; d) \\ &= \begin{bmatrix} fh_i & fh_i w^i + f^{M-d+2} w^k & \dots \\ \vdots & \vdots & \dots \\ f^{d+1} h_i + fh_k & f^{d+1} h_i w^d + fh_k w^k & \dots \\ \vdots & \vdots & \dots \\ f^{M+1} h_i + f^{M-d+1} h_k & f^{M+1} h_i w^M + f^{M-d+1} h_k w^{M-d} & \dots \\ fh_i w^{i(L-1)} + f^{M-d+2} h_k w^{(M-d+1)(L-2)} & f^{M-d+2} h_k w^{(M-d+1)(L-1)} & \vdots \\ \vdots & \vdots & \vdots \\ f^{d+1} h_i w^{d(L-1)} + fh_k w^{k(L-1)} & 0 & \vdots \\ \vdots & \vdots & \vdots \\ f^{M+1} h_i w^{M(L-1)} + f^{M-d+1} h_k w^{(M-d)(L-1)} & 0 & \vdots \end{bmatrix} \end{aligned} \tag{13}$$

The access point needs to demodulate each row of the received matrix  $\mathbf{Y}_a$  with corresponding frequency. Then, the matrix after demodulation goes through low-pass filtering (LPF), the receiver gets a new matrix that eliminates the interference of other users. Collision detection and channel estimation can be proceeded as the previous scheme in flat-fading channel. Due to the fact that

$$\frac{1}{N} \sum_{q=0}^{N-1} e^{\frac{j2\pi qm}{N}} e^{-\frac{j2\pi qn}{N}} = \delta(m - n) \tag{14}$$

where  $m, n = 1, 2, \dots, M + 1$ , and  $q = 0, 1, \dots, N - 1$ . For example,

$$\frac{1}{N} \sum_{q=0}^{N-1} (w^{m-1} e^{\frac{j2\pi qm}{N}} + w^{n-1} e^{\frac{j2\pi qn}{N}}) e^{-\frac{j2\pi qn}{N}} = \begin{cases} w^{n-1}, & m \neq n \\ w^{m-1} + w^{n-1}, & m = n \end{cases}. \quad (15)$$

For two elements on the same frequency, both of them are not eliminated. For two elements on different frequency, only one will remain. So the demodulation and low-pass filtering can be substituted for the above process in this paper. Here we express the process as  $\frac{1}{T} \int_t \mathbf{E}^H \mathbf{Y}_a$ .

### 3.4 Influence of Multipath Effect and Noise

In a communication system, multipath and noise should be considered in practice. For the receiver, multipath can be interpreted as a signal with multiple copies with different attenuations and delays. The received signal is a delay sum of multiple components with different delays. The access point only regards these copies as some other collision users. This phenomenon shows clearly that the multipath does not affect the collision detection of the proposed preamble.

Because of the noise, the access point can not use nonzero rows to detect collision directly any more. Therefore, we should set a threshold based on the noise. In general, it is set to twice or three times the noise power. In the collision detection, if the average power of a row is greater than the threshold, this row will be regarded as nonzero row. So the problem of noise is solved.

## 4 Collision Detection Algorithm

Figure 1 is a flow-chart showing the process of collision detection. Where  $\mathbf{Y}$  is a matrix that stacked up by the received signal, and symbols  $m, \mathbf{Y}', \mathbf{Y}'', \mathbf{Y}''', H, h$  are defined to save the intermediate results. Ignore the influence of noise temporarily, and  $d_i$  represents the delay between the  $i$ th user and the  $(i + 1)$ th user. The process of detection is as follow

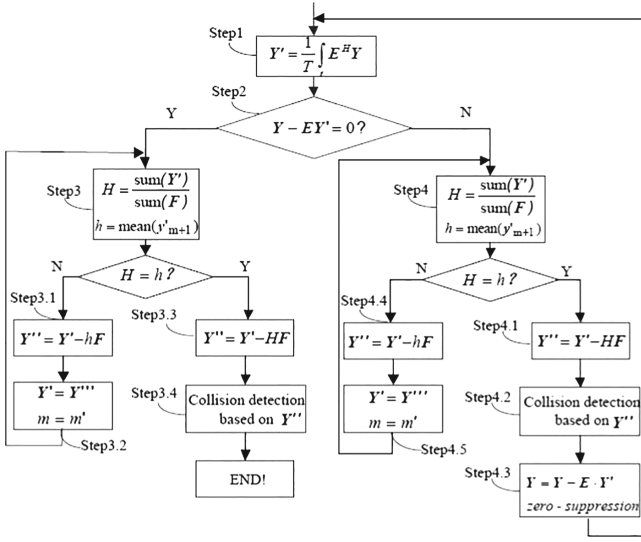
Step 1: Every single row of matrix  $\mathbf{Y}$  needs down-conversion (DC) and LPF. Then the receivers can get the matrix  $\mathbf{Y}' = \frac{1}{T} \int_t \mathbf{E}^H \mathbf{Y}$ .

Step 2: Testing  $\mathbf{Y} - \mathbf{E}\mathbf{Y}' = 0$ . If it is correct, that means every  $d_i$  is integral multiple of  $M + 1$ . Then execute step 3. If it is false, jump to step 4.

Step 3: Here we define a new variable named  $m$ , and  $m = 0$ . Calculating  $H = \frac{\text{sum}(\mathbf{Y}')}{\text{sum}(\mathbf{F})}$ , where  $\text{sum}(\cdot)$  denotes the sum of all elements of matrix  $(\cdot)$ .  $h = \text{mean}(\mathbf{y}'_{m+1})$ , where  $\text{mean}(\mathbf{y}'_{m+1})$  denotes the mean value of  $(m + 1)$ th column of matrix  $\mathbf{Y}'$ . If  $H = h$ , it means the rest of users after DC and LPF are synchronical completely, then execute step 3.1. If  $H \neq h$ , execute step 3.3.

Step 3.1: Calculating detection matrix  $\mathbf{Y}''$  as follow

$$\mathbf{Y}'' = [\mathbf{y}'_{m+1}, \mathbf{y}'_{m+2}, \dots, \mathbf{y}'_{m+L}] - H\mathbf{F} \quad (16)$$



**Fig. 1.** Collision detection algorithm in asynchronous scenario.

where  $[\mathbf{y}'_{m+1}, \mathbf{y}'_{m+2}, \dots, \mathbf{y}'_{m+L}]$  means the  $(m+1)$ th column to  $(m+L)$ th column of matrix  $\mathbf{Y}'$ . Then continue to execute step 3.2.

Step 3.2: The access point determines the collision users and estimate the CSI by the nonzero rows in  $\mathbf{Y}''$  except for the first row. And the algorithm ends.

Step 3.3: Calculating matrix  $\mathbf{Y}'''$  as follow

$$\mathbf{Y}''' = [\mathbf{y}'_{m+1}, \mathbf{y}'_{m+2}, \dots, \mathbf{y}'_{m+L}] - h\mathbf{F} \tag{17}$$

Then continue to execute the step 3.4.

Step 3.4: Checking how many columns with zero mean are there in matrix  $\mathbf{Y}'''$ , and assign the number to the variable  $m'$ ; Then back to step 3, and reassigned  $\mathbf{Y}' = \mathbf{Y}'''$  and  $m = m'$ .

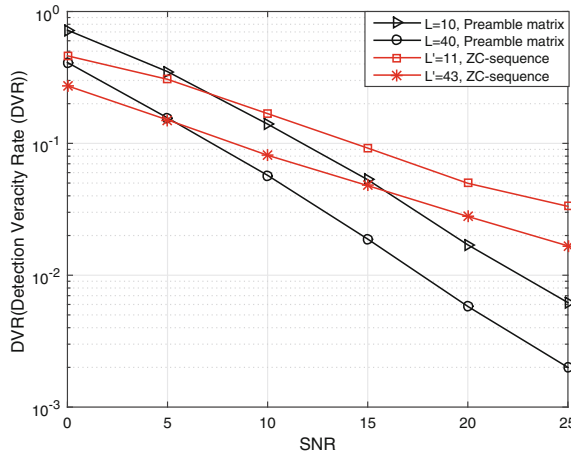
Step 4: Steps 3 and 4 are just the same, only when  $H = h$ , there are quite a few differences. If  $H = h$ , step 4 calculates the matrix  $\mathbf{Y}''$  as follow

$$\mathbf{Y}'' = [\mathbf{y}'_{m+1}, \mathbf{y}'_{m+2}, \dots, \mathbf{y}'_{m+L}] - H\mathbf{F}. \tag{18}$$

Then access point determines the collision users and estimate the CSI by the nonzero rows in matrix  $\mathbf{Y}''$  except for the first row. However, the algorithm does not end. The matrix  $\mathbf{Y} - E\mathbf{Y}'$  in step 2 needs to be expanded into a row vector by columns. After removing all the zero elements at the head and tail of the row vector (Zero-Suppression/ZS), start with the first element, and convert it into a new matrix  $[\mathbf{Y} - E\mathbf{Y}']_{ZS}$  by every  $M + 1$  elements as one column. Then return to step 1, reassign the matrix  $\mathbf{Y} = [\mathbf{Y} - E\mathbf{Y}']_{ZS}$  and start next loop.

### 5 Simulation Results and Analysis

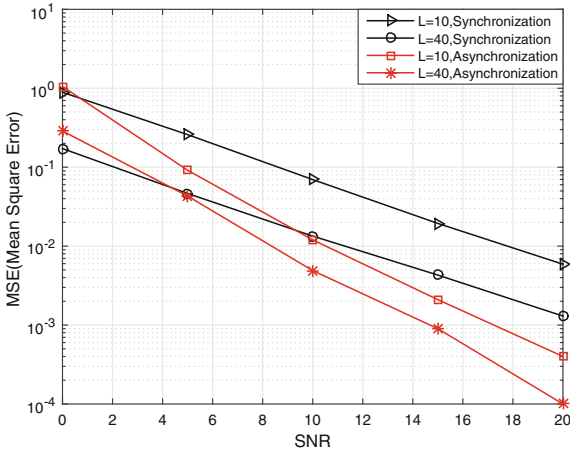
In this section, the numerical results are presented to validate the performance of collision detection and channel estimation. In the simulation, we evaluated the proposed scheme in a flat Rayleigh fading channel both in synchronous and asynchronous scenarios. Considered 10 candidate users with a single access point, the size of the preamble matrix is  $11 \times L$ . We also investigate the effect of variable  $L$  on the performance of collision detection. The detection veracity rate (DVR) represents the error probability of detection. For Channel Estimation, the mean square error (MSE) is used to show its performance. Besides, this paper also simulates the performance of ZC-sequence as a comparison, where  $L'$  (must be a prime number) is the length of ZC-sequence, so  $L'$  is 11 at least in theory.



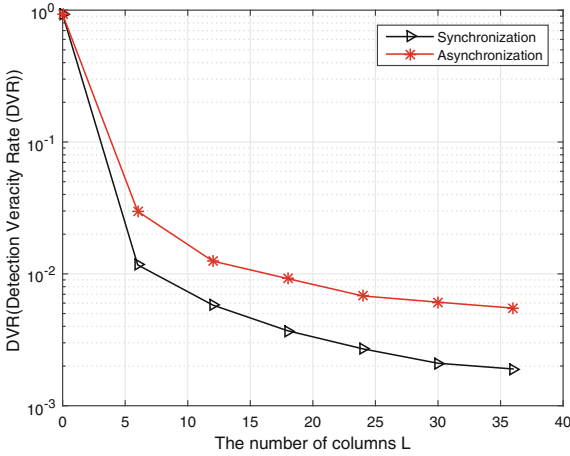
**Fig. 2.** Detection veracity rate of the proposed scheme and ZC-sequence of different lengths in asynchronous Rayleigh channel.

Figure 2 shows the DVR in asynchronous scenario. With the increase of SNR, the performance of preamble matrix becomes better and better, and exceeds ZC-sequence gradually. In fact, both ZC-sequence and preamble matrix are very sensitive to the synchronization problem. However, we optimize the preamble matrix by adding phase, so it performs better than ZC-sequence in the Asynchronous scenario. Figure 3 shows the Mean Square Error (MSE) of channel estimation in the Rayleigh channel when the communication is synchronous and asynchronous. In this figure, we can find the performance of channel estimation in an synchronous condition is worse. This is mainly because when the communication is asynchronous and the  $d_i$  is not integral multiple of  $M + 1$ , access point can separate the signals mixing together completely. If ignore the noise temporarily, the access point can get the original preamble matrix which carries the channel information, and CSI can be estimated with the whole matrix.





**Fig. 3.** MSE of channel estimation in synchronous and asynchronous Rayleigh channel.



**Fig. 4.** Effects of  $L$  on DVR in synchronous and asynchronous Rayleigh channel (SNR = 20 dB).

However, if the communication is synchronous, the access point can only use the nonzero rows to estimate channel. So the performance in an asynchronous condition is better after SNR > 8 dB. Here we can see asynchronous scenario also has some advantages.

In Figs. 2 and 3, the performance of collision detection and channel estimation is affected by  $L$ , the number of columns of preamble matrix. As shown in Fig. 4, the simulation environment is in Rayleigh channel with SNR = 20 dB, we can see the synchronous performance is better than asynchronous performance as a whole. With the increase of  $L$ , the performance of collision detection overall is becoming better and better, but when the variable  $L$  grows to a certain

degree, the performance does not change any more. The reason is the method of averaging or weighted average to suppress the noise is limited. Increasing the dimension of the preamble matrix can reduce the interference of noise. However, this trend will stop changing and keep smooth when  $L$  is large enough.

## 6 Conclusion

We proposed a new structure of preamble for contention-based communications. With the well-designed preamble structure, we can detect the number of collided users, which users are in the collision, and the channel state information of each user in both synchronous and asynchronous scenarios. In addition, it involves low amount of storage, which is the size of DFT matrix, and low computational complexity. The numerical results show that the proposed scheme exhibits better performance than using ZC-sequence as preamble.

## References

1. Liang, T.-J., Fettweis, G.: MIMO preamble design with a subset of subcarriers in OFDM-based WLAN. In: Vehicular Technology Conference, vol. 2, pp. 1032–1036 (2005)
2. Hao, B., Wang, J., Wang, Z.: Low-complexity signaling-embedded preamble design based on relative subcarrier position. *IEEE Commun. Lett.* **18**(9), 1657–1660 (2014)
3. Mavrokefalidis, C., Kofidis, E., Rontogiannis, A., Theodoridis, S.: Preamble design for channel estimation in OFDM/OQAM cooperative systems. In: Proceedings of the Tenth International Symposium on Wireless Communication Systems, Immenau, Germany, pp. 1–5 (2013)
4. Liu, J., Kang, G., Lu, S., Zhang, P.: Preamble design based on complete complementary sets for random access in MIMO-OFDM systems. In: Wireless Communications and Networking Conference, Kowloon, pp. 858–862 (2007)
5. Zihan, E., Choi, K.W., Kim, D.I.: Distributed random access scheme for collision avoidance in cellular device-to-device communication. *IEEE Trans. Wirel. Commun.* **14**(7), 3571–3585 (2015)
6. Kim, D.-H., Park, C.-K.: CSMA/CD with reservations in wireless communication: a novel approach to resolve collisions. In: International Conference on Control, Automation and Systems, Seoul, pp. 2858–2864 (2007)
7. Piret, P.: A stochastic model for burst-correcting convolutional decoders. *IEEE Trans. Inf. Theory* **18**(4), 532–535 (1972)
8. Kim, T., Choudhury, S., Jin, Z.Y., Doppler, K., Ghosh, C.: Simultaneous polling mechanism for low power sensor networks using ZC sequences. In: 2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Sydney, NSW, pp. 2024–2029 (2012)
9. Hua, M., Wang, M., Yang, K.W., Zou, K.J.: Analysis of the frequency offset effect on Zadoff-Chu sequence timing performance. *IEEE Trans. Commun.* **62**(11), 4024–4039 (2014)