INVITED PAPER

Synchronization Techniques for Orthogonal Frequency Division Multiple Access (OFDMA): A Tutorial Review

One primary challenge in wireless communications today is the demand for multimedia services accessed by many simultaneous users. This paper provides a comprehensive survey of the latest results in the field of synchronization for OFDMA systems, and the authors have endeavored to include ample tutorial content.

By Michele Morelli, *Member IEEE*, C.-C. Jay Kuo, *Fellow IEEE*, and Man-On Pun, *Member IEEE*

ABSTRACT | Orthogonal frequency division multiple access (OFDMA) has recently attracted vast research attention from both academia and industry and has become part of new emerging standards for broadband wireless access. Even though the OFDMA concept is simple in its basic principle, the design of a practical OFDMA system is far from being a trivial task. Synchronization represents one of the most challenging issues and plays a major role in the physical layer design. The goal of this paper is to provide a comprehensive survey of the latest results in the field of synchronization for OFDMA systems, with tutorial objectives foremost. After quantifying the effects of synchronization errors on the system performance, we review some common methods to achieve timing and frequency alignment in a downlink transmission. We then consider the uplink case, where synchronization is made particularly difficult by the fact that each user's signal is characterized by different timing and frequency errors, and the base station has thus to estimate a relatively large number of unknown parameters. A second difficulty is related to how the estimated parameters must be employed to correct the uplink timing and frequency errors. The paper concludes with a comparison of the reviewed synchronization schemes in an OFDMA scenario inspired by the IEEE 802.16 standard for wireless metropolitan area networks.

KEYWORDS | Downlink synchronization; dynamic subcarrier assignment; frequency correction; frequency estimation; interblock interference (IBI); interchannel interference (ICI); interference cancellation; least square estimation; maximum likelihood estimation; multicarrier modulation; multiple access interference (MAI); multiuser detection; orthogonal frequency division multiple access (OFDMA); orthogonal frequency division multiplexing (OFDM); quasi-synchronous systems; subspace-based estimation; timing correction; timing estimation; uplink synchronization; wireless local area networks (WLANs); wireless metropolitan area networks (WMANs)

I. INTRODUCTION

The demand for multimedia wireless communications is growing today at an extremely rapid pace and this trend is expected to continue in the future. The common feature of many current wireless standards for high-rate multimedia transmission is the adoption of a multicarrier air interface based on orthogonal frequency division multiplexing (OFDM). The idea behind OFDM is to convert a frequency selective channel into a collection of frequency-flat subchannels with partially overlapping spectra. This goal

Manuscript received October 27, 2006; revised March 13, 2007.
M. Morelli is with the Department of Information Engineering, University of Pisa, 56126 Pisa, Italy (e-mail: michele.morelli@iet.unipi.it).
C.-C. J. Kuo is with the Ming Hsieh Department of Electrical Engineering and the Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: cckuo@sipi.usc.edu).
M.-O. Pun is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: mopun@princeton.edu).
Digital Object Identifier: 10.1109/JPROC.2007.897979

is achieved by splitting the input high-rate data stream into a number of substreams that are transmitted in parallel over orthogonal subcarriers [1], [2]. Compared to conventional single-carrier systems, OFDM offers increased robustness against multipath distortions as channel equalization can easily be performed in the frequency domain through a bank of one-tap multipliers [3]. Furthermore, it provides larger flexibility by allowing independent selection of the modulation parameters (like the constellation size and coding scheme) over each subcarrier [4]. Due to its favorable features, OFDM has been adopted in some commercial systems such as digital audio broadcasting (DAB) [5], terrestrial digital video broadcasting (DVB-T) [6], and the IEEE 802.11a wireless local area network (WLAN) [7].

There is currently a strong interest in extending the OFDM concept to multiuser communication scenarios. A prominent example of this trend is represented by the orthogonal frequency division multiple access (OFDMA) technology, which results from a combination of OFDM with a frequency division multiple access (FDMA) protocol. This scheme was originally suggested by Sari and Karam for cable TV (CATV) networks [8] and later adopted in the uplink of the Interaction Channel for Digital Terrestrial Television (DVB-RCT) [9]. More recently, it has become part of the emerging IEEE 802.16 standards for wireless metropolitan area networks (WMANs) [10] and is currently attracting vast research attention from both academia and industry as a promising candidate for next generation broadband wireless networks.

In OFDMA systems, the available subcarriers are divided into several mutually exclusive clusters (subchannels or subbands) that are assigned to distinct users for simultaneous transmission. The orthogonality among subcarriers guarantees intrinsic protection against multiple access interference (MAI) while the adoption of a dynamic subcarrier assignment strategy provides the system with high flexibility in resource management. Furthermore, OFDMA inherits from OFDM the ability to compensate channel distortions in the frequency domain without the need of computationally demanding time-domain equalizers.

Despite its appealing features, the design of an OFDMA system poses several technical challenges. One basic issue is related to the stringent requirement on frequency and timing synchronization [11]. Similarly to OFDM, OFDMA is extremely sensitive to timing errors and carrier frequency offsets between the incoming waveform and the local references used for signal demodulation. Inaccurate compensation of the frequency offset destroys orthogonality among subcarriers and produces interchannel interference (ICI) as well as MAI. Timing errors result in interblock interference (IBI) and must be counteracted to avoid severe error rate degradations. Using a sufficiently long guard interval between adjacent OFDMA blocks (in the form of a cyclic prefix) provides intrinsic protection against timing errors at the expense of some reduction in data throughput as a consequence of the extra overhead. However, timing accuracy becomes a stringent requirement in those applications where the cyclic prefix (CP) is made as short as possible to minimize the overhead.

In OFDMA systems, timing and frequency synchronization is usually accomplished by following a three-step procedure. The first step is taken during the downlink transmission, when each mobile terminal (MT) performs frequency and timing estimation by exploiting a pilot signal transmitted by the base station (BS). This operation reduces synchronization errors within a tolerable range and can easily be accomplished using the same methods available for OFDM systems since the users' signals appear at each MT with common frequency and timing errors. The estimated parameters are then exploited by each user not only to detect the downlink data stream, but also as synchronization references for the uplink transmission. However, due to Doppler shifts and propagation delays, the uplink signals arriving at the BS may still be plagued by residual synchronization errors. The second step of the synchronization process is thus represented by frequency and timing estimation in the uplink. This operation may represent a challenging task since, contrarily to the downlink situation, the uplink received waveform is a mixture of signals transmitted by different users, each affected by exclusive synchronization errors. Accordingly, frequency and timing recovery in the uplink can be categorized as a multiparameter estimation problem, where each user must be separated from the others before starting the synchronization procedure. As we shall see, the separation method is closely related to the particular carrier assignment scheme adopted in the system, i.e., the strategy according to which subcarriers are distributed among users.

Once the uplink timing and frequency offsets have been estimated, they must be employed in some way at the BS to restore orthogonality among subcarriers. This operation is commonly referred to as timing and frequency correction, and represents the final step of the synchronization process. A possible solution is that the BS returns the estimated offsets to the corresponding user, who exploits them to adjust his/her transmitted signal. In a time-varying scenario, however, users must periodically be provided with updated estimates of their synchronization parameters, which may result in excessive overhead and outdated information due to feedback delay. For these reasons, the current trend of OFDMA favors the use of advanced signal processing techniques to compensate for synchronization errors directly at the BS, i.e., without the need of returning timing and frequency estimates back to the subscriber terminals.

The aforementioned issues have been extensively studied in recent years and several practical solutions are currently available in the open literature. Nevertheless, they are scattered around in the form of various

conferences and journal publications and, therefore, they are hardly useful for giving a unified view of an otherwise seemingly heterogeneous field. This is the main reason that prompted us to work on a comprehensive survey of the latest state-of-the-art synchronization techniques for OFDMA transmissions as it stems out of our cultural and professional background. Particular emphasis is put on the uplink case, which represents the most challenging part of the overall synchronization process. Our goal is to make the unacquainted readers more familiar with the basic concepts of this rapidly growing research area without the pretence of showing new results or deriving novel schemes. The tutorial treatment is based on elementary knowledge of traditional multicarrier systems, as well as basic principles of synchronization techniques for digital transmissions [12].

A. Organization of This Paper

This paper has the following outline. Section II reviews basic concepts of an OFDMA downlink, with emphasis on synchronization functions and allocation strategies that are commonly adopted to distribute available subcarriers among active users. Section III quantifies the effects of timing and frequency synchronization errors on the performance of an OFDMA system. Section IV is devoted to some of the most popular synchronization schemes for OFDM transmissions that can directly be applied to an OFDMA downlink. The synchronization policy adopted in the OFDMA uplink is the subject of Section V together with the mathematical model of the uplink signals. Here, we introduce the concept of a quasi-synchronous system where the users' signals arriving at the BS are time-aligned within the CP. Section VI illustrates several advanced schemes to estimate the timing and frequency errors of all active users in an OFDMA uplink. In doing so, we distinguish among different subcarrier allocation strategies since the latter have a remarkable influence on the selection of the appropriate estimation method. Section VII deals with timing and frequency correction in the uplink. Again, we consider different subcarrier allocation schemes and show how estimates of the synchronization parameters can be exploited to restore orthogonality among the received users' signals. Our investigation proceeds in Section VIII with a comparison of some uplink synchronization techniques in an OFDMA scenario inspired by the recently standardized IEEE 802.16 Wireless MAN. The paper is ended by a short concluding section.

B. Notation

The following notation is used throughout the paper. Vectors and matrices are denoted by boldface letters. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ stand for transposition, conjugate transposition, and element-wise conjugation, respectively. $E\{\cdot\}$ denotes the expectation operator while $\|\cdot\|$ represents the Euclidean norm of the enclosed vector. $\mathbf{I}_m$ is the $m \times m$ identity matrix, $\mathbf{0}_m$ denotes an m-dimensional vector with all zero entries, and $\mathbf{A} = \mathrm{diag}\{a_1, a_2, \ldots, a_m\}$ is a diagonal matrix with $\{a_j : j = 1, 2, \ldots, m\}$ along its main diagonal. $[\mathbf{A}]_{i,j}$ stands for the element in the ith row and jth column of $\mathbf{A}$ while $\mathbf{A}^{-1}$ indicates the inverse of $\mathbf{A}$. Finally, $\Re e\{\cdot\}$ and $\Im m\{\cdot\}$ denote the real and imaginary components of a complex-valued quantity, respectively, while $\arg\{\cdot\}$ and $|\cdot|$ are used for the corresponding argument and amplitude.

II. FUNDAMENTALS OF OFDMA DOWNLINK

A. Subcarrier Allocation Strategies

From a physical layer perspective, the OFDMA downlink is essentially equivalent to an OFDM system. The only difference is that in OFDMA each transmitted block conveys simultaneous information for multiple subscribers while in OFDM it carries data for a single specific user. To fix the ideas, assume that the BS communicates with M users by exploiting N available subcarriers. The latter are evenly divided into R subchannels, each consisting of $P = N/R$ subcarriers. Without loss of generality, we consider the situation in which different subchannels are assigned to distinct users, even though in practice more subchannels may be allocated to the same user depending on its requested data rate. Since the maximum number of users that the system can simultaneously support is limited to R, in the ensuing discussion we assume $M \leq R$. The overall subcarriers are numbered from $n = 0$ to $n = N - 1$, while the subcarriers of the mth subchannel have indexes in the set $\mathcal{I}_m$. Clearly, to avoid that a given subcarrier is shared by different users, the sets $\{\mathcal{I}_m\}_{m=1}^{R}$ must be mutually exclusive, i.e., $\mathcal{I}_m \cap \mathcal{I}_j = \emptyset$ for $m \neq j$.

Three possible methods to distribute subcarriers among active users are illustrated in Fig. 1 [13]. For illustration, we adopt $N = 16$ and $M = R = 4$ as system parameters. In the subband carrier assignment scheme (CAS) of Fig. 1(a), each subchannel is composed by a group of P adjacent subcarriers. The main drawback of this approach is that it does not exploit the frequency diversity offered by the multipath channel since a deep fade might hit a substantial number of subcarriers of a given user. A viable solution to this problem is obtained by adopting the interleaved CAS shown in Fig. 1(b), where the subcarriers of each user are uniformly spaced over the signal bandwidth at a distance R from each other. Although this method can fully exploit the channel frequency diversity, the current trend in OFDMA favors a more flexible allocation strategy where users can select the best subcarriers [i.e., those with the highest signal-to-noise ratios (SNRs)] that are currently available. This scheme is called generalized CAS and its basic concept is shown in Fig. 1(c). Since there is no rigid association between subcarriers and users, the generalized CAS allows dynamic

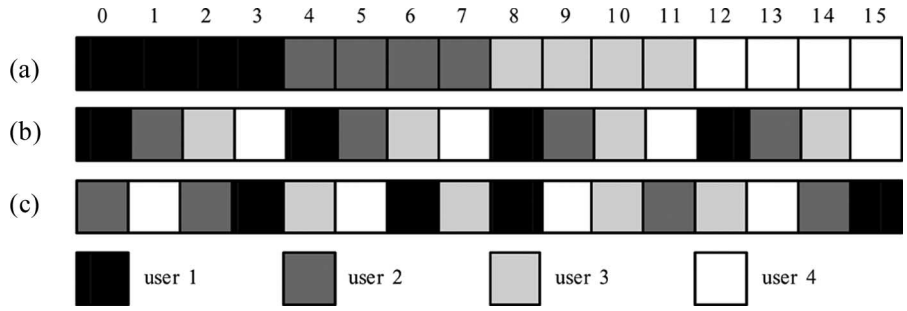

Fig. 1. *Examples of subcarrier allocation schemes: (a) subband CAS, (b) interleaved CAS, and (c) generalized CAS.*

resource allocation and provides more flexibility than subband or interleaved CAS.

B. OFDMA Transmitter

Fig. 2 illustrates the discrete-time block diagram of the OFDMA downlink transmitter. After channel coding and symbol mapping (not shown), the data stream of each user is divided into blocks of length P, with $\mathbf{b}_{m,i}$ denoting the ith block of the mth user. The CAS unit maps the P data symbols of each block onto the subcarriers assigned to the corresponding user. This operation is easily performed by extending $\mathbf{b}_{m,i}$ with the insertion of $N-P$ zeros and results in an N-dimensional vector $\mathbf{d}_{m,i}$ with entries

$$d_{m,i}(n) = \begin{cases} c_{m,i}(n), & \text{if } n \in \mathcal{I}_m, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $c_{m,i}(n)$ is the data symbol transmitted over the nth subcarrier. Vectors $\mathbf{d}_{m,i}$ are then summed up to produce the ith block of *frequency-domain* samples

$$\mathbf{d}_i = \sum_{m=1}^{M} \mathbf{d}_{m,i}. \tag{2}$$

The latter is fed to a conventional OFDM modulator that consists of an N-point inverse discrete Fourier transform (IDFT) unit followed by the insertion of an N_g-point CP to avoid interference between adjacent blocks. The IDFT output is collected into a vector $\mathbf{s}_i = [s_i(0), s_i(1), \ldots, s_i(N-1)]^T$ and the CP is appended to $\mathbf{s}_i$ such that $s_i(k) = s_i(k+N)$ for $-N_g \le k \le -1$. The ith block of *time-domain* samples can thus be written as

$$s_i(k) = \begin{cases} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} d_i(n) e^{j2\pi nk/N}, & \text{if } -N_g \le k \le N-1, \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\{d_i(n)\}$ are the entries of $\mathbf{d}_i$.

The discrete-time downlink signal is obtained as the serial concatenation of several blocks of time-domain samples and takes the form

$$s^{(T)}(k) = \sum_i s_i(k - iN_T) \tag{4}$$

where $N_T = N + N_g$ is the block length (with the CP included). In practice, sequence $s^{(T)}(k)$ is transmitted over the channel using a linear modulator with impulse

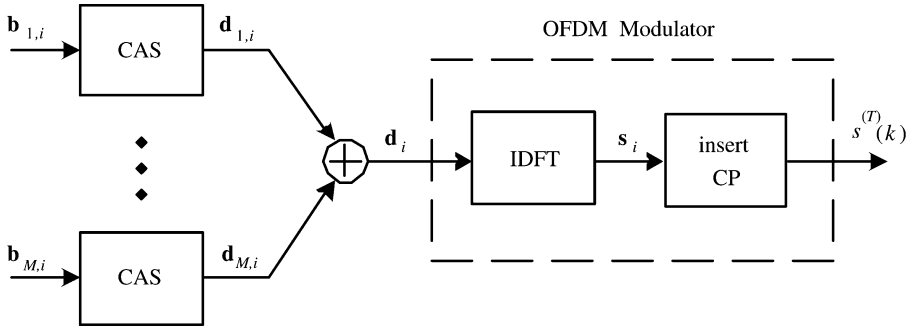

Fig. 2. *Block diagram of OFDMA downlink transmitter.*

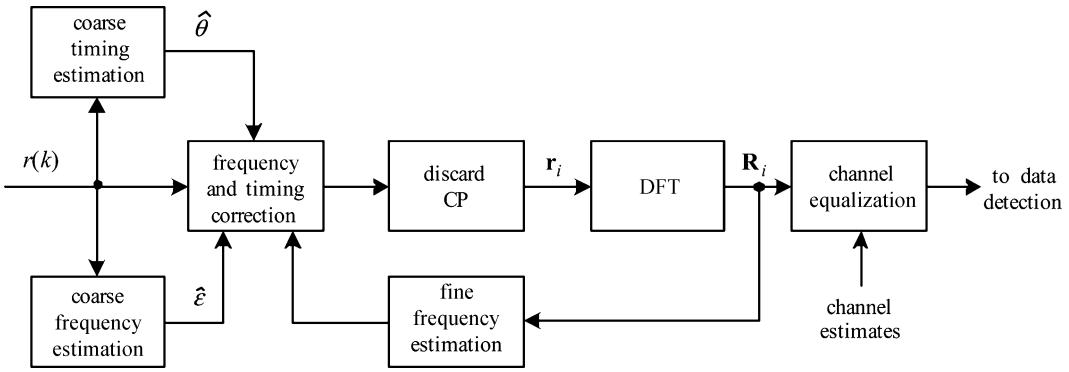

Fig. 3. ***Block diagram of OFDMA downlink receiver.***

response $g(t)$ and a signaling interval T_s. In this case, the distance in the frequency domain between two adjacent subcarriers (subcarrier spacing) is $1/(NT_s)$, which results in an overall signal bandwidth of approximately $1/T_s$. A root-raised-cosine filter is commonly used as a modulation pulse $g(t)$.

C. Channel Model

The transmitted signal propagates through a multipath channel that is assumed static over several OFDMA blocks. The main reason behind this assumption is that standardized systems employing OFDMA are mostly used to provide wireless connectivity to portable devices characterized by low mobility. In these applications, the channel coherence time is much larger than the block duration and the assumption of a static channel is reasonable.

We denote $\mathbf{h} = [h(0), h(1), \ldots, h(L-1)]^T$ the T_s-spaced samples of the overall channel impulse response (CIR) encompassing the physical channel as well as the transmit/receive filters. The channel taps $h(\ell)$ are modeled as statistically independent Gaussian random variables with zero mean (Rayleigh fading) and an exponentially decaying power delay profile

$$E\{|h(\ell)|^2\} = \beta e^{-\ell/L}, \quad 0 \le \ell \le L-1 \tag{5}$$

where β is a suitable factor chosen to normalize the average energy of the CIR to unity. Although signal shaping operated by the transmit and receive filters may result in some statistical correlation among channel taps, it is a common practice to neglect such a correlation when evaluating the performance of multicarrier systems. The channel order L is related to the duration of the impulse response of shaping filters as well as to the channel delay spread. Since the latter is usually unknown, in practice L is chosen according to the maximum expected delay spread.

D. OFDMA Receiver

The OFDMA downlink receiver is shown in Fig. 3. After down-conversion and low-pass filtering, the received signal is fed to an analog-to-digital (A/D) converter, where it is sampled with frequency $f_s = 1/T_s$. In a perfectly synchronized system, the A/D output is expressed by

$$r(k) = \sum_i \sum_{\ell=0}^{L-1} h(\ell) s_i(k - \ell - iN_T) + w(k) \tag{6}$$

where $w(k)$ is complex-valued additive white Gaussian noise (AWGN) with variance σ_w^2. Next, stream $r(k)$ is divided into adjacent segments of length N_T, each corresponding to a transmitted OFDMA block. Without loss of generality, we concentrate on the ith segment in the following. After removing the CP, the remaining N samples are collected into a vector $\mathbf{r}_i = [r_i(0), r_i(1), \ldots, r_i(N-1)]^T$ and passed to an N-point discrete Fourier transform (DFT) unit. Assuming that the CP is longer than the CIR duration, the DFT output $\mathbf{R}_i$ has entries

$$R_i(n) = H(n) d_i(n) + W_i(n), \quad 0 \le n \le N-1 \tag{7}$$

where $W_i(n)$ is the noise contribution with power σ_w^2 and

$$H(n) = \sum_{\ell=0}^{L-1} h(\ell) e^{-j2\pi n\ell/N} \tag{8}$$

is the channel frequency response over the nth subcarrier.

Equation (7) indicates that OFDMA can be seen as a set of parallel transmissions over N Gaussian channels with different complex-valued attenuations $H(n)$. Thus, channel equalization can easily be performed through a bank of

one-tap multipliers, one for each subcarrier. The latter are designed according to either a zero-forcing (ZF) or minimum mean square error (MMSE) criterion using an estimate $\widehat{H}(n)$ of the channel response. The equalizer output is next passed to a threshold device, which delivers the final data decisions. In general, since each user is exclusively interested in the P data symbols transmitted by the BS over its assigned subchannel, only the DFT outputs with indexes in the set $\mathcal{I}_m$ are selected by the mth downlink receiver for channel equalization and data detection.

E. Downlink Synchronization Tasks

As mentioned previously, the signal model in (6) describes an ideal system where no frequency and timing errors are present. However, in practical situations, Doppler shifts and oscillator instabilities result in a carrier frequency offset (CFO) f_d between the received carrier and the local sinusoids used for signal demodulation. In addition, at the start-up the receiver does not know where the OFDMA blocks start and, accordingly, the DFT window may be placed in a wrong position. This results in a timing error, denoted by τ_d, which must properly be compensated to avoid severe performance degradation. Actually, since small (fractional) timing errors can be corrected through channel equalization (see Section III-A), it suffices to locate the beginning of each received OFDMA block within one sampling period. For this reason, it is a common practice to model the timing error as a multiple θ of the sampling period and consider the remaining fractional error as part of the CIR.

In summary, calling $\varepsilon = Nf_dT_s$ the frequency offset normalized to the subcarrier spacing $1/(NT_s)$, the received samples in the presence of synchronization errors take the form

$$r(k) = e^{j2\pi\varepsilon k/N} \sum_i \sum_{\ell=0}^{L-1} h(\ell) s_i(k - \theta - \ell - iN_T) + w(k). \tag{9}$$

As shown in Fig. 3, the coarse frequency and timing estimation units employ the received sequence $r(k)$ to compute estimates of ε and θ, denoted by $\widehat{\varepsilon}$ and $\widehat{\theta}$, respectively. The former is used to counter-rotate $r(k)$ at an angular speed $2\pi\widehat{\varepsilon}/N$ (coarse frequency correction), while the timing estimate is exploited to achieve the correct positioning of the receive DFT window (coarse timing correction). Specifically, the counter-rotated samples with indexes $iN_T + \widehat{\theta} \le k \le iN_T + \widehat{\theta} + N - 1$ are grouped together and passed to the DFT unit. Besides being used for data detection, the DFT output can also be exploited to track small short-term variations of the frequency offset (fine-frequency estimation). As mentioned earlier, the compensation of fractional timing errors (fine-timing synchronization) is usually accomplished by the channel equalizer.

III. SENSITIVITY TO TIMING AND FREQUENCY ERRORS

In this section, we assess the impact of uncompensated timing and frequency errors on the performance of an OFDMA system. To simplify the analysis, we concentrate on a downlink transmission, though similar results apply even to the uplink case.

A. Effect of Timing Offset

In multicarrier systems, the DFT window should include samples from only one single block in order to avoid IBI. As shown in Fig. 4, the tail of each received block extends over the first $L-1$ samples of the successive block as a consequence of multipath dispersion. Since the length of the CP must be greater than the CIR duration in a well designed system, a certain range of the guard interval is not affected by the previous block at the receiver. As long as the DFT window starts anywhere in this range, no IBI is present at the DFT output. This situation occurs whenever the timing error $\Delta\theta = \widehat{\theta} - \theta$ belongs to interval $-N_g + L - 1 \le \Delta\theta \le 0$ and only results in a cyclic shift of the received OFDMA block. Thus, recalling the time-shift property of the Fourier transform and assuming perfect frequency synchronization (i.e., $\varepsilon = 0$), the DFT output over the nth subcarrier takes the form

$$R_i(n) = e^{j2\pi n\Delta\theta/N} H(n) d_i(n) + W_i(n). \tag{10}$$

The equation indicates that timing error $\Delta\theta$ appears as a linear phase across subcarriers and, accordingly, it can be compensated for by the channel equalizer, which cannot distinguish between phase shifts introduced by the channel and those caused by timing misalignments.

On the other hand, if the timing error is outside interval $-N_g + L - 1 \le \Delta\theta \le 0$, samples at the DFT input will be contributed by two adjacent OFDMA blocks. In addition to IBI, this results in a loss of orthogonality among subcarriers which, in turn, generates ICI. In this case, the nth DFT output is given by

$$R_i(n) = e^{j2\pi n\Delta\theta/N} \alpha(\Delta\theta) H(n) d_i(n) + I_i(n, \Delta\theta) + W_i(n) \tag{11}$$

where $\alpha(\Delta\theta)$ is an attenuation factor while $I_i(n, \Delta\theta)$ accounts for IBI and ICI and can reasonably be modeled as a zero-mean random variable with power $\sigma_I^2(\Delta\theta)$. Both $\alpha(\Delta\theta)$ and $\sigma_I^2(\Delta\theta)$ depend on the timing error and the channel delay profile as discussed in [14].

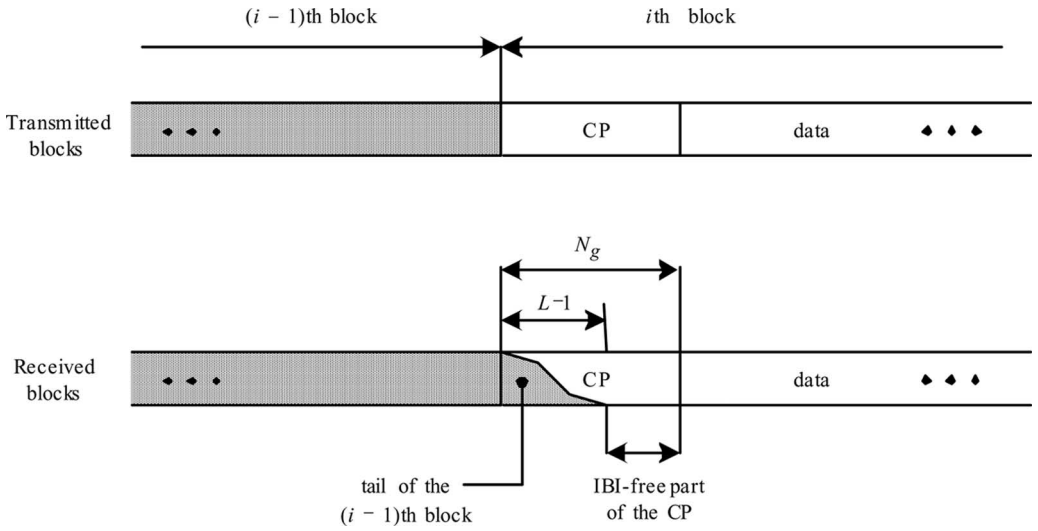

Fig. 4. *Partial overlapping between received blocks due to multipath dispersion.*

A useful indicator to evaluate the effect of timing errors on the system performance is the loss in the SNR. This quantity is defined as

$$\gamma(\Delta\theta) = \frac{\mathrm{SNR}^{(\mathrm{ideal})}}{\mathrm{SNR}^{(\mathrm{real})}} \tag{12}$$

where $\mathrm{SNR}^{(\mathrm{ideal})}$ is the SNR of a perfectly synchronized system and $\mathrm{SNR}^{(\mathrm{real})}$ is the SNR in the presence of a timing offset. Assuming a normalized channel response with unit average power (i.e., $E\{|H(n)|^2\} = 1$) and letting $D_2 = E\{|d_i(n)|^2\}$, we have $\mathrm{SNR}^{(\mathrm{ideal})} = D_2/\sigma_w^2$ from (7). On the other hand, since the three terms in the right-hand-side of (11) are statistically independent, it follows that $\mathrm{SNR}^{(\mathrm{real})} = D_2\alpha^2(\Delta\theta)/[\sigma_w^2 + \sigma_I^2(\Delta\theta)]$. Substituting these results into (12) yields

$$\gamma(\Delta\theta) = \frac{1}{\alpha^2(\Delta\theta)}\left[1 + \frac{\sigma_I^2(\Delta\theta)}{\sigma_w^2}\right]. \tag{13}$$

Fig. 5 illustrates $\gamma(\Delta\theta)$ (in decibels) versus timing error $\Delta\theta$ for $N = 256$ and some values of D_2/σ_w^2. The channel is Rayleigh fading with length $L = 8$ and an exponentially decaying power delay profile as given in (5). A CP of length $N_g = 16$ is appended to the transmitted blocks. At each simulation run, a new channel snapshot is generated and the results are numerically averaged with respect to the channel statistics. For a given timing error, we see that $\gamma(\Delta\theta)$ increases with D_2/σ_w^2. This can be explained by the fact that at low SNRs the main impairment is represented by thermal noise and the impact of synchronization errors becomes less evident. The results of Fig. 5 indicate that, to keep the SNR degradation to a tolerable level of less than 1.0 dB, the residual error $\Delta\theta$ after timing correction should be limited to a few percents of the block length. As discussed earlier, no IBI is present at the DFT output as long as $-N_g + L - 1 \le \Delta\theta \le 0$. In this case, we have $\mathrm{SNR}^{(\mathrm{real})} = \mathrm{SNR}^{(\mathrm{ideal})}$, so that $\gamma(\Delta\theta) = 0$ dB. The requirement of the timing synchronizer is therefore determined by the number of samples by which the CP exceeds the CIR duration.

B. Effect of Frequency Offset

A carrier frequency offset produces a shift of the received signal in the frequency domain and may result in a loss of mutual orthogonality among subcarriers. To better explain this concept, we assume ideal timing synchronization (i.e., $\widehat{\theta} = \theta$) and compute the DFT output corresponding to the ith OFDMA block in the presence of a frequency error ε. After performing standard manipulations, we obtain

$$R_i(n) = e^{j\varphi_i}\sum_{p=0}^{N-1} H(p)d_i(p)f_N(\varepsilon + p - n) + W_i(n) \tag{14}$$

where $\varphi_i = 2\pi i\varepsilon N_T/N$ and $f_N(x)$ is defined as

$$f_N(x) = \frac{\sin(\pi x)}{N\sin(\pi x/N)}e^{j\pi x(N-1)/N}. \tag{15}$$

At this stage, it is convenient to distinguish between two distinct situations depending on whether the frequency error is an integer multiple of the subcarrier spacing $1/NT_s$ or not. In the first case, ε is integer valued and (14) reduces to

$$R_i(n) = e^{j\varphi_i} H(|n-\varepsilon|_N) d_i(|n-\varepsilon|_N) + W_i(n) \tag{16}$$

where $|n-\varepsilon|_N$ is the value of $n-\varepsilon$ reduced to interval $[0, N-1]$. This equation indicates that an integer frequency offset only results in a shift of modulated subcarriers by ε positions. Orthogonality among subcarriers is thus preserved, even though the received symbols appear in a wrong position at the DFT output.

The situation is drastically different when ε is not integer valued. In this case, the subcarriers are no longer orthogonal and (14) can conveniently be rewritten as

$$R_i(n) = e^{j\varphi_i} H(n) d_i(n) f_N(\varepsilon) + I_i(n,\varepsilon) + W_i(n) \tag{17}$$

where $I_i(n,\varepsilon)$ is a zero-mean ICI term with power $\sigma_I^2(\varepsilon) = E\{|I_i(n,\varepsilon)|^2\}$. Letting $E\{|H(n)|^2\} = 1$, after some manipulations we find that

$$\sigma_I^2(\varepsilon) = D_2\left[1 - |f_N(\varepsilon)|^2\right]. \tag{18}$$

The impact of the frequency error on the system performance is assessed in terms of the SNR loss defined as

$$\gamma(\varepsilon) = \frac{\mathrm{SNR}^{(\mathrm{ideal})}}{\mathrm{SNR}^{(\mathrm{real})}} \tag{19}$$

where $\mathrm{SNR}^{(\mathrm{ideal})} = D_2/\sigma_w^2$ is the SNR of a perfectly synchronized system and $\mathrm{SNR}^{(\mathrm{real})} = D_2|f_N(\varepsilon)|^2/[\sigma_w^2 + \sigma_I^2(\varepsilon)]$ is the SNR in the presence of frequency offset ε. Substituting these results into (19) and bearing in mind (18), we obtain

$$\gamma(\varepsilon) = \frac{1}{|f_N(\varepsilon)|^2}\left\{1 + \frac{D_2}{\sigma_w^2}\left[1 - |f_N(\varepsilon)|^2\right]\right\}. \tag{20}$$

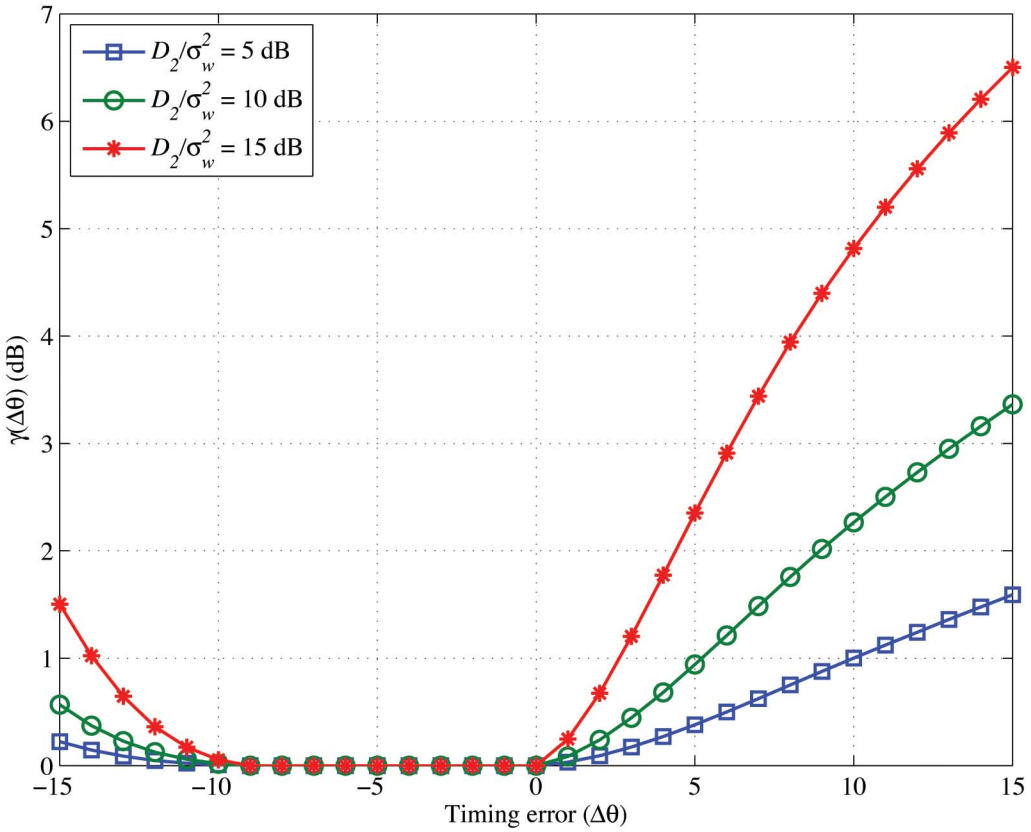

Fig. 5. ***SNR loss due to timing errors.***

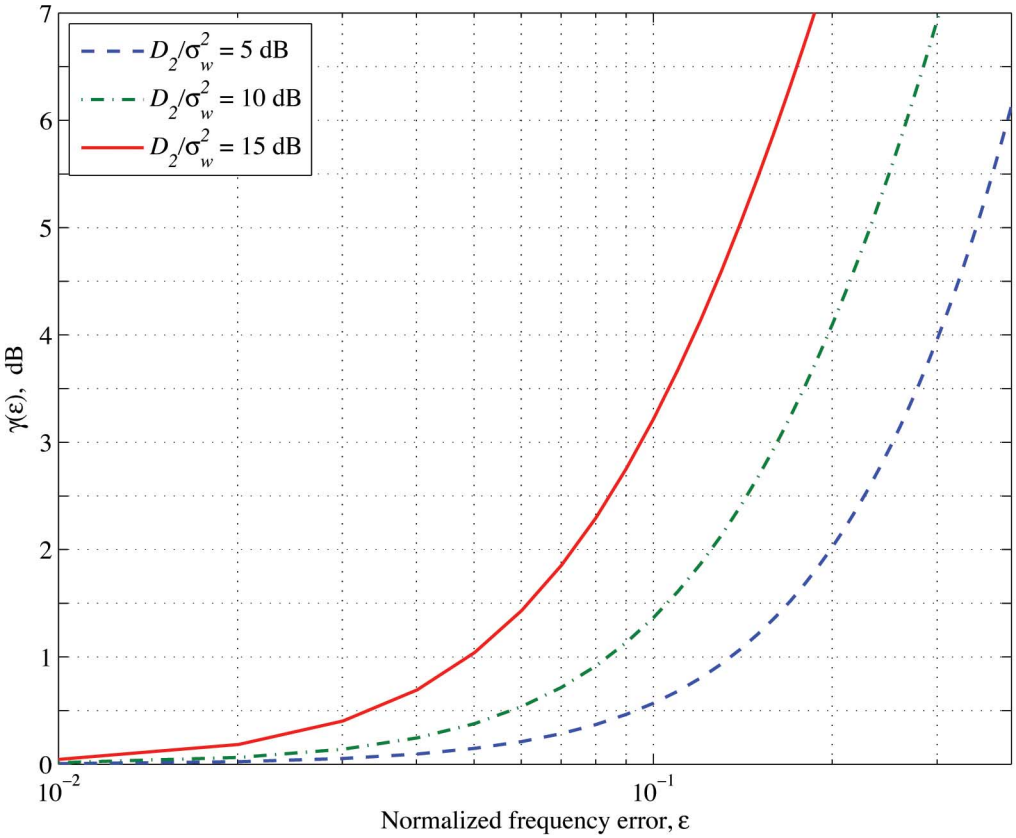

Fig. 6. ***SNR loss due to frequency errors.***

A simpler expression of $\gamma(\varepsilon)$ is obtained for small values of ε by resorting to the Taylor series expansion of $|f_N(\varepsilon)|^2$ around $\varepsilon = 0$. This produces

$$\gamma(\varepsilon) \approx 1 + \frac{1}{3}\frac{D_2}{\sigma_w^2}(\pi\varepsilon)^2 \tag{21}$$

from which it follows that the SNR loss is related to the square of the normalized frequency offset.

Equation (20) is shown in Fig. 6 as a function of ε for some values of D_2/σ_w^2. These results indicate that non-negligible performance degradations are incurred when the frequency error exceeds 4%–5% of the subcarrier distance.

IV. SYNCHRONIZATION ALGORITHMS FOR DOWNLINK TRANSMISSIONS

In OFDMA downlink transmissions, each terminal performs frequency and timing estimation by exploiting the broadcast signal transmitted by the BS in a way similar to single-user OFDM. As standardized in many commercial systems [5]–[10], the transmission is normally organized in frames. An example of the frame structure is illustrated in Fig. 7, where some reference blocks with a particular training pattern are appended in front of the payload segment to assist the synchronization process.

In this section, we review some popular schemes for timing and frequency estimation in an OFDMA downlink scenario. In doing so, we distinguish between an acquisition step and a tracking phase. During the acquisition, the reference blocks placed at the beginning of the frame are employed to get coarse estimates of the synchronization parameters [15]–[26]. These estimates are next refined during the tracking phase to counteract short-term variations produced by oscillator drifts and/or time-varying Doppler shifts. For this purpose, several techniques exploiting either the redundancy of the CP or pilot tones inserted in each block are available in the open literature [27]–[31]. Alternatively, blind methods operating over the DFT output can be resorted to [32], [33]. Blind methods can also be employed during the initial acquisition step. A good sample of the results obtained in this area is found in [34]–[44].

A. Timing Acquisition

In most multicarrier applications, timing acquisition represents the first step of the downlink synchronization

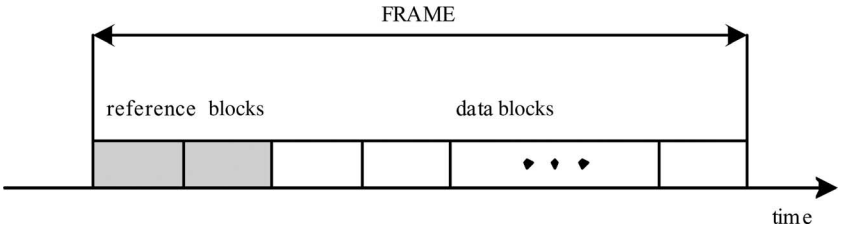

Fig. 7. *An example of the frame structure.*

process. This operation has two main objectives. First, it must detect the presence of a new frame in the received data stream. Second, once the frame has been detected, it must provide a coarse estimate of the timing error to find the correct position of the receive DFT window. Since the CFO is usually unknown at this stage, it is desirable that the timing recovery scheme be robust against possibly large frequency offsets.

One of the first timing acquisition algorithms for OFDM transmissions was proposed by Nogami and Nagashima [45] and was based on the use of a null reference block where nothing is transmitted (no signal power). The drop of the received power corresponding to the null block was exploited to reveal the arrival of a new frame. Unfortunately, this method provides highly inaccurate timing estimates. Also, it is not suited for a burst-mode transmission since the null block cannot be distinguished by the idle period between successive bursts. A popular approach to overcome these difficulties is the use of some reference blocks exhibiting a repetitive structure in the time domain. In this case, a robust timing estimator can be designed by searching for the peak of the correlation among repetitive parts. This idea was originally employed by Schmidl and Cox (S&C) in [16], where a reference block composed by two identical halves of length $N/2$ is transmitted at the beginning of each frame. Note that a block with such a structure can easily be generated in the frequency domain by modulating subcarriers with even indexes by a pseudonoise (PN) sequence while forcing to zero the remaining subcarriers with odd indexes.

To explain the rationale behind the S&C algorithm, one can verify that, as long as the CP is longer than the CIR duration, the two halves of the reference block will remain identical after passing through the transmission channel except for a phase shift induced by the CFO. In other words, if we model the received samples corresponding to the first half as

$$r(k) = s^{(R)}(k)e^{j2\pi\varepsilon k/N} + w(k), \qquad \theta \le k \le \theta + N/2 - 1 \quad (22)$$

with $s^{(R)}(k)$ denoting the useful signal and $w(k)$ the noise contribution, then the samples in the second half take the following form:

$$r(k + N/2) = s^{(R)}(k)e^{j2\pi\varepsilon k/N}e^{j\pi\varepsilon} + w(k + N/2), \qquad \theta \le k \le \theta + N/2 - 1. \quad (23)$$

In this case, timing acquisition can be performed by feeding the time-domain samples to a sliding window correlator of lag $N/2$, which is expected to exhibit a peak when the sliding window is perfectly aligned with the received reference block. The resulting timing estimate is thus given by [16]

$$\widehat{\theta} = \arg\max_{\tilde{\theta}} \left\{ \left| \Gamma(\tilde{\theta}) \right| \right\} \quad (24)$$

where $\Gamma(\tilde{\theta})$ is the following normalized autocorrelation function:

$$\Gamma(\tilde{\theta}) = \frac{\sum_{q=\tilde{\theta}}^{\tilde{\theta}+N/2-1} r(q + N/2)r^*(q)}{\sum_{q=\tilde{\theta}}^{\tilde{\theta}+N/2-1} |r(q + N/2)|^2}. \quad (25)$$

Fig. 8 shows an example of the timing metric, $|\Gamma(\tilde{\theta})|$, as a function of the difference $\delta_\theta = \tilde{\theta} - \theta$. The results are obtained numerically over a Rayleigh multipath channel with $L = 8$ taps. The number of subcarriers is $N = 256$ and the CP has length $N_g = 16$. The SNR over received samples is defined as SNR $= \sigma_s^2/\sigma_w^2$ with $\sigma_s^2 = E\{|s^{(R)}(k)|^2\}$ and is set to 20 dB.

As mentioned before, the first step of the timing acquisition process is represented by the detection of a new frame in the received data stream. For this purpose, $|\Gamma(\tilde{\theta})|$ is continuously monitored and the start of a frame is declared whenever it overcomes a given threshold λ. The latter must be designed by properly taking into account the statistics of the timing metric so as to achieve a reasonable

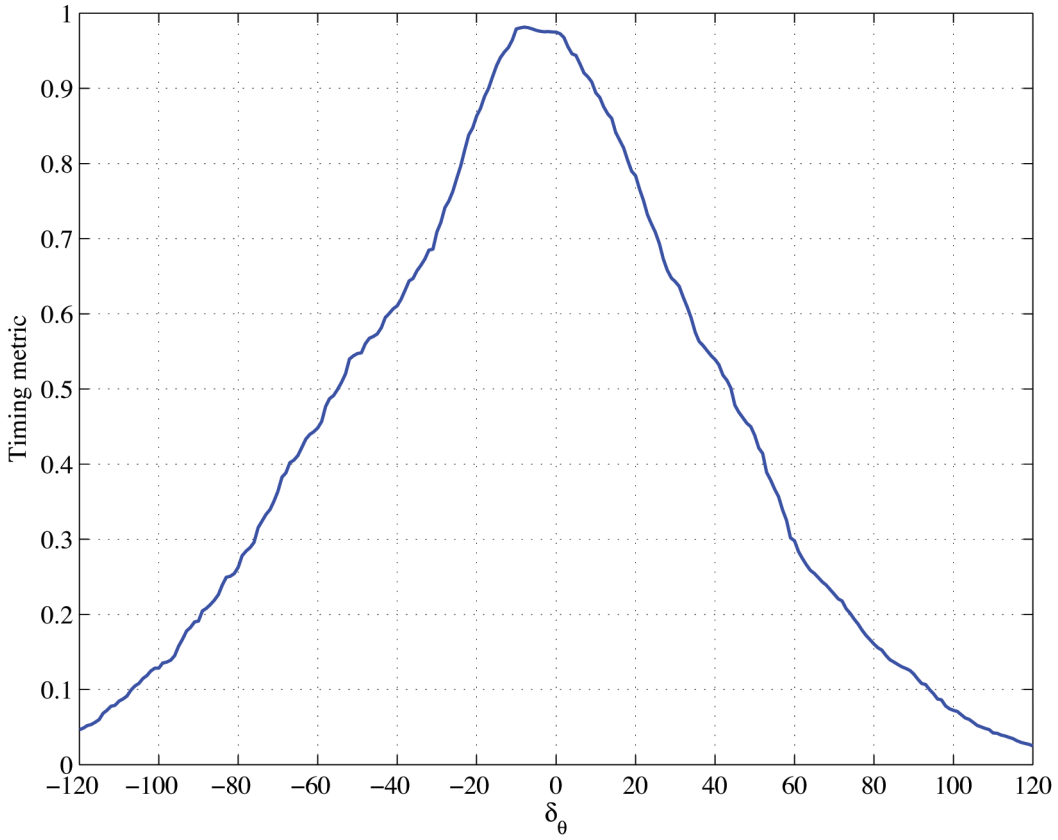

Fig. 8. *Example of timing metric for S&C algorithm.*

tradeoff between the false alarm and misdetection probabilities. Once the presence of a new frame has been detected, timing estimate $\widehat{\theta}$ is computed by searching for the maximum of $|\Gamma(\widetilde{\theta})|$ as indicated in (24).

Unfortunately, we see from Fig. 8 that the timing metric of the S&C algorithm exhibits a large "plateau" that may greatly reduce the estimation accuracy. Solutions to this problem are proposed in some recent works, where reference blocks with suitably designed patterns are exploited to obtain sharper timing metric trajectories [25], [26]. For instance, Shi and Serpedin (S&S) used a training block composed of four repetitive parts $[+B\ +B\ -B\ +B]$ with a sign inversion in the third segment [26]. As depicted in Fig. 9, a sliding window of length N spans received time-domain samples with indexes $\widetilde{\theta} \le k \le \widetilde{\theta} + N - 1$ and collects them into four vectors $\mathbf{r}_j(\widetilde{\theta}) = \{r(\ell + jN/4 + \widetilde{\theta}); 0 \le \ell \le N/4 - 1\}$ with $j = 0, 1, 2, 3$. The timing metric is then computed as

$$\Gamma(\widetilde{\theta}) = \frac{\left|\Lambda_1(\widetilde{\theta})\right| + \left|\Lambda_2(\widetilde{\theta})\right| + \left|\Lambda_3(\widetilde{\theta})\right|}{\frac{3}{2}\sum_{j=0}^{3}\left\|\mathbf{r}_j(\widetilde{\theta})\right\|^2} \tag{26}$$

where

$$\begin{aligned}
\Lambda_1(\widetilde{\theta}) &= \mathbf{r}_0^H(\widetilde{\theta})\mathbf{r}_1(\widetilde{\theta}) - \mathbf{r}_1^H(\widetilde{\theta})\mathbf{r}_2(\widetilde{\theta}) - \mathbf{r}_2^H(\widetilde{\theta})\mathbf{r}_3(\widetilde{\theta}) \\
\Lambda_2(\widetilde{\theta}) &= \mathbf{r}_1^H(\widetilde{\theta})\mathbf{r}_3(\widetilde{\theta}) - \mathbf{r}_0^H(\widetilde{\theta})\mathbf{r}_2(\widetilde{\theta}) \\
\Lambda_3(\widetilde{\theta}) &= \mathbf{r}_0^H(\widetilde{\theta})\mathbf{r}_3(\widetilde{\theta}).
\end{aligned} \tag{27}$$

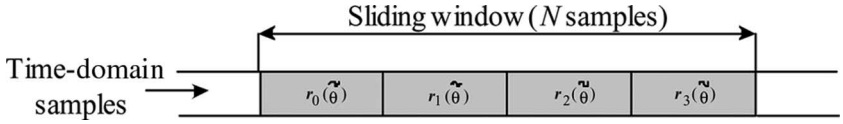

Fig. 9. *Sliding window used in S&S timing acquisition scheme.*

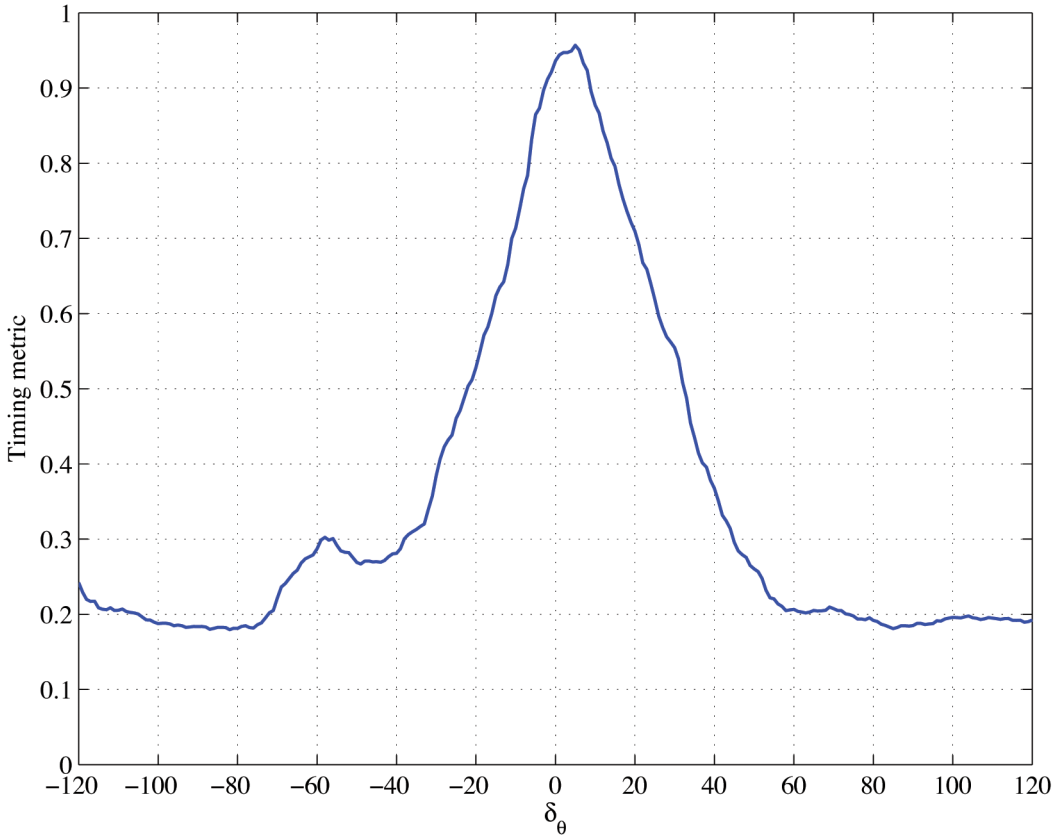

Fig. 10. *Timing metric for S&S algorithm.*

Fig. 10 illustrates $\Gamma(\widetilde{\theta})$ for the S&S algorithm in the same operating conditions of Fig. 8. Since the plateau region present in the S&C metric is now significantly reduced, more accurate timing estimates are expected. As indicated in [25], reference blocks with more than four repetitive segments can be designed to further increase the sharpness of the timing trajectory.

Although the use of reference blocks is the most popular approach for timing acquisition in burst-mode applications, blind methods seem preferable in a continuous-mode transmission since any unnecessary overhead can be avoided. An example of a blind scheme can be found in [29], where autocorrelation properties induced by the CP on the time-domain samples are exploited for timing estimation. Specifically, the following N-lag autocorrelation function is used as the timing metric:

$$\Gamma(k) = \sum_{q=0}^{N_g-1} r(k-q)r^*(k-q-N) \qquad (28)$$

where k is the time index of the last received sample. Since the CP is just a duplication of the last N_g samples of each OFDMA block, we expect that the magnitude of $\Gamma(k)$ may exhibit periodic peaks whenever samples $r(k-q-N)$ with $0 \leq q \leq N_g - 1$ belong to the CP. The peak locations indicate the beginning of the received blocks and are exploited to control the position of the DFT window.

B. Fine Timing Tracking

If the transmit and the receive clock oscillators are adequately stable, the estimate $\widehat{\theta}$ provided by the timing acquisition unit can be used to detect data symbols over the entire frame. In certain applications, however, the presence of non-negligible errors in the sampling clock frequency may result in a long-term variation of timing error $\Delta\theta$, which must be tracked in some way. From a mathematical viewpoint, we may consider $\Delta\theta$ as introduced by the *physical channel* rather than by the oscillator drift. This amounts to absorbing $\Delta\theta$ into the CIR vector or, equivalently, replacing $\mathbf{h} = [h(0), h(1), \ldots, h(L-1)]^T$ with its time-shifted version $\mathbf{h}'(\Delta\theta) = [h(\Delta\theta), h(1+\Delta\theta), \ldots, h(L-1+\Delta\theta)]^T$. Therefore, in the presence of small sampling frequency offsets, channel estimates computed over different OFDMA blocks are differently delayed as a consequence of long-term fluctuations of $\Delta\theta$. A possible method to track these fluctuations is to look for

the delay of the first significative taps in the estimated CIR vector. This approach is adopted in [20], where the timing estimate is updated at each new received block and used for a fine adjustment of the DFT window. As mentioned before, fractional timing errors are inherently compensated for by the channel equalization unit since they only appear as phase shifts at the DFT output.

C. Frequency Acquisition

After frame detection and timing acquisition, each terminal must compute a coarse frequency estimate to align its local oscillator to the received carrier frequency. This operation is referred to as frequency acquisition and is normally accomplished at each new received frame by exploiting the same reference blocks used for timing acquisition in addition to possibly other dedicated blocks. One common approach is to employ a training pattern composed by some repetitive parts which remain identical after passing through the transmission channel except for a phase shift produced by the frequency error [15]–[26]. The latter is thus estimated by measuring the induced phase shift. This method was originally employed by Moose in [15], where the phase shift between two successive identical blocks is measured in the frequency domain at the DFT output. More precisely, assume that timing acquisition has already been achieved and let $R_1(n)$ and $R_2(n)$ be the nth DFT output corresponding to the two reference blocks. Then, we may write

$$R_1(n) = S(n) + W_1(n) \tag{29}$$

$$R_2(n) = S(n)e^{j2\pi\varepsilon N_T/N} + W_2(n) \tag{30}$$

where $S(n)$ is the signal component (the same over each block as long as the channel is static) while $W_1(n)$ and $W_2(n)$ are noise terms. The above equations indicate that an estimate of ε can be computed as

$$\widehat{\varepsilon} = \frac{1}{2\pi(N_T/N)} \arg\left\{ \sum_{n=0}^{N-1} R_2(n) R_1^*(n) \right\}. \tag{31}$$

The main drawback of this scheme is the relatively short acquisition range. Actually, since the $\arg\{\cdot\}$ function returns values in the range $[-\pi, \pi)$, we see from (31) that $|\widehat{\varepsilon}| \leq N/(2N_T)$, which is less than one-half of the subcarrier spacing. A viable method to enlarge the acquisition range is proposed by Schmidl & Cox in [16]. They decompose the frequency error into a *fractional part*, less than $1/(NT_s)$ in magnitude, plus an integer part which is a multiple of $2/(NT_s)$. The normalized frequency error is thus rewritten as

$$\varepsilon = \nu + 2\eta \tag{32}$$

where $\nu \in (-1, 1]$ and η is an integer. The S&C estimator relies on the transmission of two suitably designed reference blocks as depicted in Fig. 11. The first block is the same as that used for timing acquisition. In particular, it is composed by two identical halves of length $N/2$ that are generated by modulating only subcarriers with even indices while setting the others to zero. The second block contains a differentially encoded pseudo-noise sequence PN1 on even subcarriers and another pseudo-noise sequence PN2 on odd subcarriers.

The samples in the two halves of the first block are obtained after substituting (32) into (22) and (23). This yields

$$r(k) = s'(k) + w(k), \quad \theta \leq k \leq \theta + N/2 - 1 \tag{33}$$

$$r(k + N/2) = s'(k)e^{j\pi\nu} + w(k + N/2), \quad \theta \leq k \leq \theta + N/2 - 1 \tag{34}$$

where $s'(k) = s^{(R)}(k)e^{j2\pi(\nu+2\eta)k/N}$ for notational simplicity and the identity $e^{j2\pi\eta} = 1$ is used in the derivation. The above equations indicate that, apart from thermal noise, the two halves are identical except for a phase shift of $\pi\nu$. Thus, an estimate of ν is obtained as

$$\widehat{\nu} = \frac{1}{\pi} \arg\left\{ \sum_{k=\theta}^{\theta+N/2-1} r(k + N/2) r^*(k) \right\}. \tag{35}$$

As shown above, timing information is necessary to compute $\widehat{\nu}$. In practice, quantity θ in (35) is replaced by its corresponding estimate $\widehat{\theta}$ given in (24).

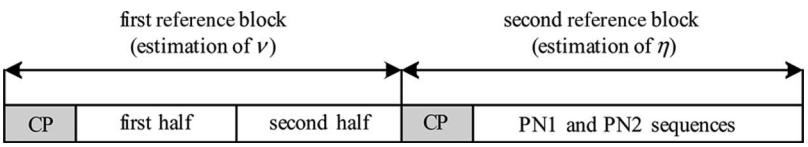

Fig. 11. *Reference blocks of S&C algorithm for frequency acquisition.*

The next step is the estimation of the integer frequency offset η. For this purpose, the samples belonging to the two reference blocks of Fig. 11 are first counter-rotated at an angular speed $2\pi\widehat{\nu}/N$ to compensate for the fractional offset ν. Next, they are fed to the DFT unit, which produces quantities $R_1(n)$ and $R_2(n)$ for $0 \le n \le N-1$. As explained in Section III-B, in case of perfect compensation of ν the DFT outputs are not affected by ICI. However, they will be shifted from their correct position by a quantity 2η due to the uncompensated integer frequency offset. Actually, from (16) we have

$$R_1(n) = e^{j\varphi_1} H(|n-2\eta|_N) d_1(|n-2\eta|_N) + W_1(n) \quad (36)$$

$$R_2(n) = e^{j\varphi_2} H(|n-2\eta|_N) d_2(|n-2\eta|_N) + W_2(n) \quad (37)$$

where $|n-2\eta|_N$ is the value $n-2\eta$ reduced to interval $[0, N-1]$. Neglecting for simplicity the noise terms and calling $p(n) = d_2(n)/d_1(n)$ the differentially encoded PN sequence on the even subcarriers of the second block, we see from (36) and (37) that

$$R_2(n) = e^{j(\varphi_2-\varphi_1)} p(|n-2\eta|_N) R_1(n)$$

for even n. Therefore, an estimate of η can be calculated by finding integer $\widehat{\eta}$ that maximizes the following metric:

$$B(\widetilde{\eta}) = \frac{\left| \sum_{n\ even} R_2(n) R_1^*(n) p^*(|n-2\widetilde{\eta}|_N) \right|}{\sum_{n\ even} |R_2(n)|^2} \quad (38)$$

where $\widetilde{\eta}$ varies over the range of possible integer frequency offsets. With (32), the estimated CFO is finally obtained in the form $\widehat{\varepsilon} = \widehat{\nu} + 2\widehat{\eta}$.

As mentioned before, the main advantage of the S&C method over the Moose scheme is the enlarged frequency acquisition range. An interesting question is whether a similar result can be obtained even with a smaller overhead than that required by S&C. A solution in this sense was proposed by Morelli and Mengali (M&M) in [19]. They consider a *single* reference block composed by $Q > 2$ identical parts, each containing N/Q samples. The estimated CFO is obtained as

$$\widehat{\varepsilon} = \frac{1}{2\pi/Q} \sum_{q=1}^{Q/2} \chi(q) \arg\{\Psi(q)\Psi^*(q-1)\} \quad (39)$$

where $\chi(q)$ are suitably designed coefficients given by

$$\chi(q) = \frac{12(Q-q)(Q-q+1) - Q^2}{2Q(Q^2-1)} \quad (40)$$

and $\Psi(q)$ is the following qN/Q-lag autocorrelation:

$$\Psi(q) = \sum_{k=\widehat{\theta}}^{\widehat{\theta}+N-1-qN/Q} r(k+qN/Q) r^*(k), \quad q = 1, 2, \ldots, Q/2. \quad (41)$$

As shown in [19], the estimation range of this scheme is $|\varepsilon| \le Q/2$. Hence, if Q is designed such that the possible frequency offsets lie in interval $[-Q/2, Q/2]$, the CFO is estimated by means of a single reference block so that the training overhead is reduced by a factor two with respect to the S&C method.

Fig. 12 compares S&C and M&M in terms of the MSE values of frequency estimates versus SNR $= \sigma_s^2/\sigma_w^2$, where σ_w^2 is the noise power and $\sigma_s^2 = E\{|s^{(R)}(k)t|^2\}$. We assume a total of $N = 256$ subcarriers and a Rayleigh multipath channel with $L = 8$ taps. Parameter Q in the M&M algorithm is fixed to eight. For comparison, we also show the Cramer–Rao bound (CRB) for the frequency estimation of a complex sinusoid embedded in AWGN [46]. Although the accuracy of both schemes is relatively close to the bound, M&M exhibits improved performance and achieves a gain of approximately 1.0 dB over S&C.

D. Frequency Tracking

The CFO estimate $\widehat{\varepsilon}$ obtained during the acquisition phase is used to counter-rotate the received samples $r(k)$ at an angular speed $2\pi\widehat{\varepsilon}/N$ to produce the new sequence $r'(k) = r(k)e^{-j2\pi k\widehat{\varepsilon}/N}$. In the presence of time-varying Doppler shifts, the residual frequency error $\Delta\varepsilon = \varepsilon - \widehat{\varepsilon}$ must continuously be tracked and compensated for in order to avoid ICI at the DFT output. This operation is usually accomplished on a block-by-block basis by resorting to a closed-loop structure as depicted in Fig. 13. Here, $r_i'(m)$ $(-N_g \le m \le N-1)$ are the rotated time-domain samples of the ith received OFDMA block (included the CP) and e_i is an error signal that provides information on $\Delta\varepsilon$. This signal is employed by the loop filter to update the frequency estimate according to the following recursion:

$$\Delta\widehat{\varepsilon}_{i+1} = \Delta\widehat{\varepsilon}_i + \alpha \cdot e_i \quad (42)$$

where $\Delta\widehat{\varepsilon}_i$ is the estimated frequency error over the ith block and α is a parameter (step-size) that controls the dynamic behavior of the loop. In practice, α must properly be designed so as to achieve a reasonable tradeoff between

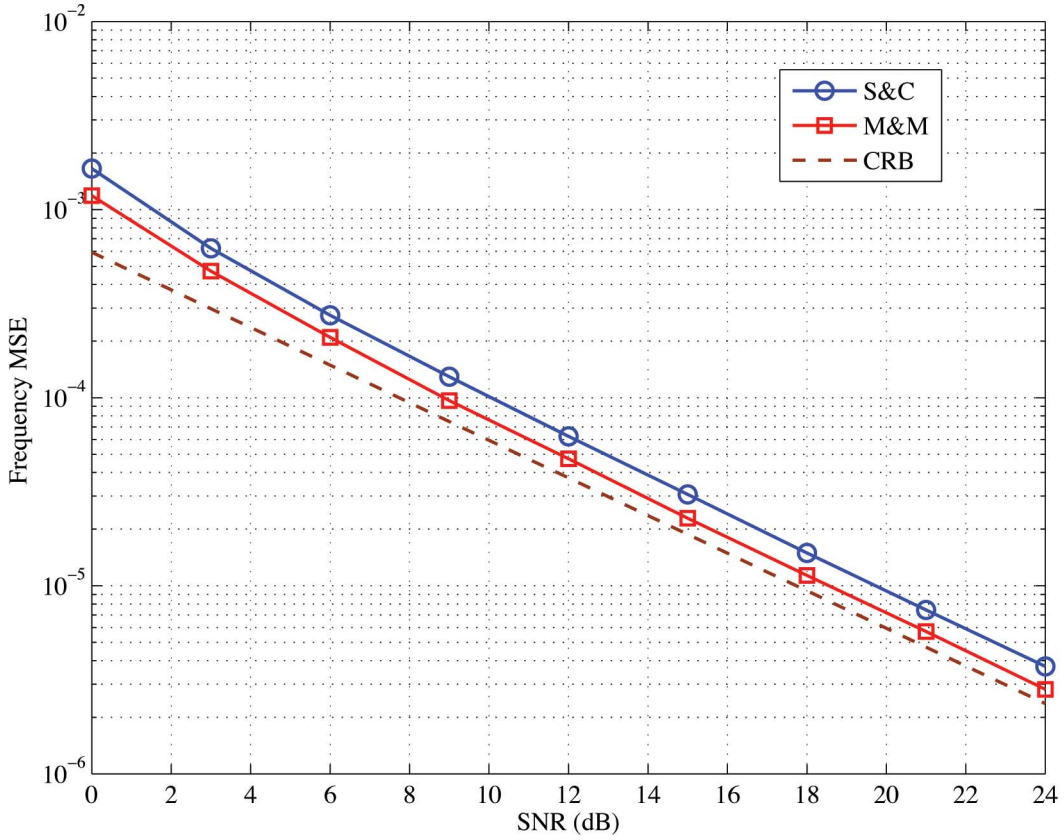

Fig. 12. *Accuracy of frequency estimates versus SNR for S&C and M&M algorithms.*

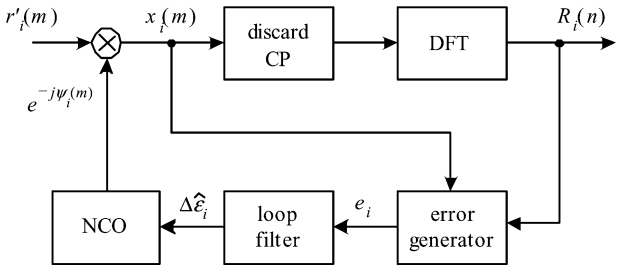

Fig. 13. *Closed-loop structure to track residual CFO.*

the convergence speed and accuracy in the steady state. The estimate $\Delta\widehat{\varepsilon}_i$ is then fed to a numerically controlled oscillator (NCO), which generates an exponential term $e^{-j\psi_i(m)}$. The phase $\psi_i(m)$ varies linearly in time with a slope proportional to $\Delta\widehat{\varepsilon}_i$ and is recursively computed as

$$\psi_i(m) = \psi_i(m-1) + 2\pi\Delta\widehat{\varepsilon}_i/N, \quad -N_g \le m \le N-1 \tag{43}$$

where $\psi_i(-N_g-1)$ is set equal to $\psi_{i-1}(N-1)$ in order to avoid any phase jump between two adjacent blocks. The exponential term is used to obtain the frequency-corrected samples $x_i(m) = r'_i(m)e^{-j\psi_i(m)}$ for $-N_g \le m \le N-1$. After discarding the CP, the latter are finally fed to the DFT unit which generates the frequency-domain samples $R_i(n)$ $(0 \le n \le N-1)$.

Several frequency tracking schemes available in the literature are based on the closed-loop structure in Fig. 13 and only differ in their specific error signal e_i [28]–[33]. In particular, we distinguish between frequency-domain and time-domain algorithms, depending on whether e_i is computed from the DFT output $R_i(n)$ or from rotated samples $x_i(m)$. An example of frequency-domain scheme is given in [32], where e_i is derived using a maximum likelihood (ML) approach and reads

$$e_i = \Re e\left\{\sum_{n=1}^{N-1} R_i^*(n)[R_i(n+1) - R_i(n-1)]\right\}. \tag{44}$$

A similar method with improved performance is proposed in [33], which employs the following error signal:

$$e_i = \Re e\left\{\sum_{n=1}^{N-1} \frac{R_i^*(n)[R_i(n+1) - R_i(n-1)]}{1+\rho|R_i(n)|^2}\right\} \quad (45)$$

where ρ is a design parameter related to the operating SNR.

Examples of time-domain methods are provided in [28] and [29], where the phase shift between the CP and the last N_g samples of each block is used as an indicator of the residual frequency offset. The resulting error signal is given by

$$e_i = \frac{1}{N_g}\Im m\left\{\sum_{m=-N_g}^{-1} x_i(m+N)x_i^*(m)\right\} \quad (46)$$

where $x_i(m)$ $(-N_g \le m \le -1)$ are samples taken from the CP of the ith received block. It is worthwhile to point out that error signals in (44)–(46) result in *blind* schemes that do not exploit any pilot symbols embedded into the transmitted data stream.

V. FUNDAMENTALS OF OFDMA UPLINK

A. Uplink Transmitter

We now consider an OFDMA uplink scenario where M users transmit their signals to a central BS. The discrete-time block diagram of the mth transmitter is sketched in Fig. 14 and is basically similar to the downlink transmitter of Fig. 1. The main difference is that, while the downlink signal conveys information for all active users, in the uplink each terminal only transmits its own data. Transmission occurs in a block-wise fashion, where each block $\mathbf{b}_{m,i}$ of P information symbols is fed to the CAS unit and mapped over the P subcarriers assigned to the considered user. This produces the N-dimensional vector $\mathbf{d}_{m,i}$ of frequency-domain samples, whose entries $d_{m,i}(n)$ are the same given in (1). After an IDFT operation and the insertion of an N_g-point CP, each $\mathbf{d}_{m,i}$ is transformed into the corresponding block of time-domain samples, which are expressed by

$$s_{m,i}(k) = \begin{cases} \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} d_{m,i}(n) \\ \quad \times e^{j2\pi nk/N}, & \text{if } -N_g \le k \le N-1, \\ 0, & \text{otherwise.} \end{cases} \quad (47)$$

The mth discrete-time uplink signal results from the concatenation of several blocks of time-domain samples and may be written as

$$s_m^{(T)}(k) = \sum_i s_{m,i}(k - iN_T). \quad (48)$$

B. Synchronization Policy

A peculiar feature of the uplink system is that each transmitted signal propagates through a different multipath channel and is received at the BS with distinct timing and frequency offsets. This makes uplink synchronization a rather difficult task. To alleviate this problem, it is a common practice to adopt a particular synchronization policy where timing and frequency estimates computed by each terminal during the downlink phase are used not only to detect the downlink data stream, but also as synchronization references for uplink transmission. Even with this approach, however, residual timing and frequency errors may still be present on the uplink signals arriving at the BS. To see how this happens, we use $T_B = N_T T_s$ to denote the length of the cyclically extended OFDMA blocks and

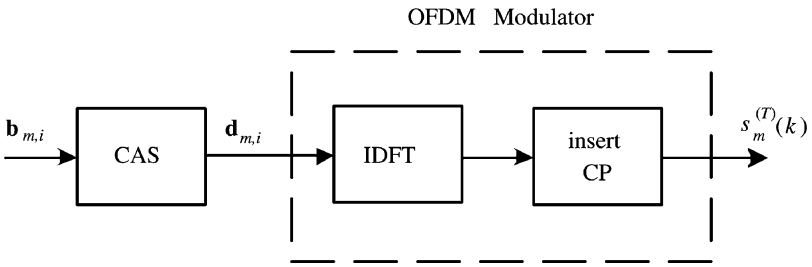

Fig. 14. *Block diagram of mth OFDMA uplink transmitter.*

assume that the BS starts to transmit the ℓth downlink block at time $t = \ell T_B$ $(\ell = 0, 1, 2, \ldots)$ and carrier frequency f_0. The block is then received by user m at time $t = \ell T_B + \tau_m$ and frequency $f_0 + \Delta f_m$, where τ_m and Δf_m are the propagation delay and Doppler shift of the considered user, respectively. They can be expressed by

$$\tau_m = \frac{D_m}{c} \tag{49}$$

$$\Delta f_m = \frac{f_0 v_m}{c} \tag{50}$$

where c is the speed of light, v_m is the speed of the mth mobile terminal, and D_m is the distance between the considered terminal and the BS. As mentioned previously, each user tries to align its uplink signal to the BS time and frequency scales by exploiting synchronization information obtained during the downlink phase. In particular, the mth user transmits its uplink blocks at time $t = iT_B + \tau_m$ $(i = 0, 1, 2, \ldots)$ and frequency $f_0 + \Delta f_m + F$, where F is the nominal separation between the uplink and downlink frequency bands (clearly, $F = 0$ in time-division-duplex systems). On the other hand, as a consequence of the propagation delay and Doppler shift, the BS receives blocks at time $iT_B + 2\tau_m$ and frequency $f_0 + 2\Delta f_m + F$, which results in uplink timing and frequency errors of $2\tau_m$ and $2\Delta f_m$, respectively.

Intuitively speaking, the adoption of the policy described above may greatly alleviate the uplink synchronization problem since frequency and timing offsets are expected to be much smaller than those encountered in a completely asynchronous system, where uplink signals are not locked to any downlink control channel. Interestingly, this policy makes uplink frequency synchronization even unnecessary as long as the Doppler shift is sufficiently smaller than the subcarrier spacing since the resulting ICI at the receive side is negligible in this case. Moreover, in the next section we show how timing synchronization can be skipped as well if the CP is designed to accommodate both the CIR duration and the *two-way* propagation delay $2\tau_m$. The reason is that in the latter case timing errors only result in phase shifts at the DFT outputs and are compensated for by the channel equalizer.

C. Uplink Signal Model

At the BS, the M incoming waveforms are implicitly combined by the receive antenna and down-converted to the baseband before passing to the A/D unit. The resulting T_s-spaced samples are modeled as

$$r(k) = \sum_{m=1}^{M} s_m^{(R)}(k) + w(k) \tag{51}$$

where $w(k)$ is the noise contribution while $s_m^{(R)}(k)$ is the signal component from the mth user and reads

$$s_m^{(R)}(k) = e^{j2\pi\varepsilon_m k/N} \sum_{\ell=0}^{L_m-1} h_m(\ell) s_m^{(T)}(k - \theta_m - \ell) \tag{52}$$

where $\mathbf{h}_m = [h_m(0), h_m(1), \ldots, h_m(L_m - 1)]^T$ denotes the discrete-time CIR of the mth user, ε_m represents the frequency error (normalized to the subcarrier spacing), and $\theta_m = \mathrm{int}(\tau_m / T_s)$ is the integer timing offset expressed in sampling periods. Note that the residual fractional timing error $\tau_m - \theta_m T_s$ is asborbed into the CIR and compensated for by the channel equalizer.

In order to restore orthogonality among uplink signals, the BS must compute estimates of θ_m and ε_m for each active user. A simple way to counteract the effects of the uplink timing errors is to design the CP length to accommodate both the channel delay spread and the two-way propagation delay. This leads to a *quasi-synchronous* system [47] in which timing errors are incorporated as part of the channel responses. The received samples can thus be rewritten as

$$s_m^{(R)}(k) = e^{j2\pi\varepsilon_m k/N} \sum_{\ell=0}^{L-1} h'_m(\ell) s_m^{(T)}(k - \ell) \tag{53}$$

where $\mathbf{h}'_m = [h'_m(0), h'_m(1), \ldots, h'_m(L-1)]^T$ is the mth *extended* channel vector with entries $h'_m(\ell) = h_m(\ell - \theta_m)$ and length $L = \max_m\{L_m + \theta_m\}$.

In practice, a quasi-synchronous system is equivalent to a perfectly time-synchronized network in which the duration of the mth CIR is *artificially* extended from L_m to L. Thus, if the CP is longer than L, samples $s_{m,i}^{(R)}(k) = s_m^{(R)}(k + iN_T)(0 \le k \le N-1)$ falling within the ith DFT window are free from IBI and take the form

$$s_{m,i}^{(R)}(k) = \frac{1}{\sqrt{N}} e^{j(2\pi\varepsilon_m k/N + \varphi_{m,i})} \sum_{n \in \mathcal{I}_m} H'_m(n) d_{m,i}(n) e^{j2\pi nk/N}, \quad 0 \le k \le N-1 \tag{54}$$

where $\varphi_{m,i} = 2\pi\varepsilon_m iN_T/N$, while $H'_m(n)$ is the N-point DFT of $\mathbf{h}'_m$ and reads

$$H'_m(n) = e^{-j2\pi n\theta_m/N} H_m(n), \quad 0 \le n \le N-1 \tag{55}$$

with

$$H_m(n) = \sum_{\ell=0}^{L_m-1} h_m(\ell) e^{-j2\pi n\ell/N}. \tag{56}$$

The fact that θ_m only appears in the phase of $H'_m(n)$ makes quasi-synchronous systems extremely appealing since timing errors can be compensated for by the channel equalizer. Timing recovery is thus unnecessary and the BS has only to estimate frequency offsets ε_m. Clearly, the price for this simplification is a certain loss of efficiency due to the extended CP length that accommodates both the two-way propagation delay and the CIR duration. To keep the loss to a tolerable level, the CP must be maintained within a short fraction of the block length. This poses an upper limit to the maximum tolerable propagation delay and, ultimately, to the maximum distance between users' terminals and the BS receiver.

VI. TIMING AND FREQUENCY ESTIMATION FOR UPLINK TRANSMISSIONS

Multiuser uplink synchronization is, in general, a difficult task that presents a drastically different outlook from the corresponding downlink situation. The main reason is that uplink signals transmitted by distinct users are characterized by different timing and frequency offsets and, accordingly, the BS has to estimate much more parameters than in the downlink. A second difficulty is related to how timing and frequency estimates are used to restore orthogonality among uplink signals. In downlink transmissions, frequency correction is easily achieved by counter-rotating time-domain samples at an angular speed $2\pi\widehat{\varepsilon}_m/N$, while timing adjustment is accomplished by shifting the DFT window by $\widehat{\theta}_m$ sampling intervals. Unfortunately, these methods do not apply to an uplink scenario because uplink signals are affected by different synchronization errors and the correction of one user's time and frequency offset would misalign other initially aligned users.

In this section, we concentrate on the problem of timing and frequency estimation for OFDMA uplink transmissions. Since the resulting algorithms are closely related to the adopted carrier assignment strategy, in what follows we distinguish between systems employing subband, interleaved, or generalized CAS. The use of estimated synchronization parameters to restore orthogonality among subcarriers is the subject of Section VII.

A. Timing and Frequency Estimation With Subband CAS

In OFDMA systems with subband CAS, groups of P contiguous subcarriers (subbands) are exclusively assigned to active users as illustrated in Fig. 1(a). In the presence of frequency errors, subbands of distinct users are differently shifted in the frequency domain from their nominal positions and, in consequence, subcarriers located at the edges of a given group may experience significant ICI. To mitigate this problem, a specified number of unmodulated subcarriers is typically inserted among subbands to provide adequately large guard intervals.

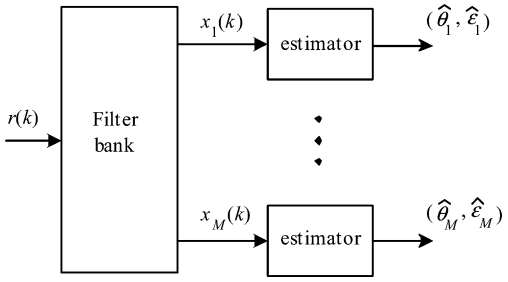

Fig. 15. *Uplink timing and frequency estimation in OFDMA system with subband CAS.*

Assigning groups of adjacent subcarriers to each user simplifies the synchronization task to a large extent. Actually, if the frequency offsets are adequately smaller than the guard intervals, users' signals can easily be separated at the BS by passing the received samples through a bank of digital band-pass filters, each selecting one subband. As shown in Fig. 15, the filtering operation allows the BS to perform timing and frequency estimation independently for each active user. Clearly, perfect users' separation cannot be achieved in practice since this would require ideal brickwall filters and/or large guard intervals. Hence, the output from the filter tuned on the mth subband can be written as

$$x_m(k) = s_m^{(R)}(k) + I_m(k) + w_m(k) \tag{57}$$

where $s_m^{(R)}(k)$ is the mth received uplink signal as given in (52), $w_m(k)$ is thermal noise, and $I_m(k)$ accounts for interference arising from imperfect users' separation. Intuitively, estimates of θ_m and ε_m can be obtained from $x_m(k)$ by applying the same methods used in downlink transmissions. One possibility is to adopt the method discussed in [29], which exploits the correlation induced on $x_m(k)$ by the use of the CP. This leads to the following timing and frequency estimates:

$$\widehat{\theta}_m = \arg\max_{\tilde{\theta}} \left\{ \left| \Gamma_m(\tilde{\theta}) \right| \right\} \tag{58}$$

$$\widehat{\varepsilon}_m = \frac{1}{2\pi} \arg\left\{ \Gamma_m(\widehat{\theta}_m) \right\} \tag{59}$$

where

$$\Gamma_m(\widetilde{\theta}) = \sum_{k=\tilde{\theta}-N_g}^{\tilde{\theta}-1} x_m(k+N)x_m^*(k) \tag{60}$$

is the N-lag autocorrelation of sequence $x_m(k)$. A slightly modified version of this algorithm is presented in [48], where it is shown that the estimator's performance depends heavily on the number of subcarriers in one subband and deteriorates as this number becomes smaller due to the increased correlation among samples $x_m(k)$. The accuracy of the timing and frequency estimates can be improved by averaging $\Gamma_m(\widetilde{\theta})$ over Q successive blocks. This yields a new metric

$$\overline{\Gamma}_m(\widetilde{\theta}) = \sum_{q=0}^{Q-1} \Gamma_m(\widetilde{\theta} + qN_T) \tag{61}$$

which is used in (58) and (59) in place of $\Gamma_m(\theta)$.

An alternative scheme to obtain estimates of θ_m and ε_m from $x_m(k)$ is discussed in [47]. This method exploits unmodulated (virtual) subcarriers inserted in each subband and updates timing and frequency estimates until the energy of the DFT outputs corresponding to virtual carriers achieves a minimum. Mathematically, we have

$$(\widehat{\theta}_m, \widehat{\varepsilon}_m) = \arg \min_{(\widetilde{\theta}_m, \widetilde{\varepsilon}_m)} \left\{ J(\widetilde{\theta}_m, \widetilde{\varepsilon}_m) \right\} \tag{62}$$

where $\widetilde{\theta}_m$ and $\widetilde{\varepsilon}_m$ represent trial values of θ_m and ε_m, respectively, and the cost function $J(\widetilde{\theta}_m, \widetilde{\varepsilon}_m)$ is proportional to the average energy of the time- and frequency-corrected samples $x_m(k+\widetilde{\theta}_m)e^{-j2\pi\widetilde{\varepsilon}_m k/N}$ falling across virtual carriers. The main drawback of this method is that the minimization problem in (62) requires a complicated 2-D grid search over the set spanned by $\widetilde{\theta}_m$ and $\widetilde{\varepsilon}_m$.

As mentioned before, the subband CAS offers the possibility of separating signals from different users through a simple filter bank even in a completely asynchronous scenario with arbitrarily large timing errors. On the other hand, grouping subcarriers together prevents the possibility of optimally exploiting the inherent diversity offered by the multipath channel. The adoption of an interleaved CAS can provide users with some form of frequency diversity. However, this strategy greatly complicates the synchronization task.

B. Timing and Frequency Estimation With Interleaved CAS

In OFDMA systems with interleaved CAS, each user modulates an exclusive set of P subcarriers which are uniformly spaced in the frequency domain at distance R as shown in Fig. 1(b). This makes users' separation a challenging task as compared to subband transmissions. The reason is that, in the presence of frequency errors, users' signals partially overlap in the frequency domain and cannot simply be isolated through a filter bank. However, the interleaved CAS provides uplink signals with a special periodic structure that can effectively be exploited for the synchronization purpose. Since the joint estimation of all synchronization parameters in an interleaved OFDMA system appears a formidable problem, we consider a quasi-synchronous scenario and limit our attention to the frequency estimation problem for the time being. A method for estimating timing offsets of all users will be discussed later.

We denote $\mathcal{I}_m = \{i_m + pR; 0 \le p \le P-1\}$ the indexes of subcarriers assigned to the mth user, where i_m may be any integer in the interval $[0, R-1]$. Then, we see from (54) that samples $s_{m,i}^{(R)}(k)$ of the mth uplink signal within the ith DFT window can be rewritten as

$$s_{m,i}^{(R)}(k) = \frac{1}{\sqrt{N}} e^{j2\pi\xi_m k/P} \sum_{p=0}^{P-1} \Phi_{m,i}(p) e^{j2\pi pk/P} \tag{63}$$

where

$$\Phi_{m,i}(p) = e^{j\varphi_{m,i}} H'_m(i_m + pR) d_{m,i}(i_m + pR) \tag{64}$$

is an amplitude- and phase-distorted version of the pth data symbol and ξ_m is defined as

$$\xi_m = \frac{i_m + \varepsilon_m}{R}. \tag{65}$$

Equation (63) shows that

$$s_{m,i}^{(R)}(k) = e^{j2\pi\xi_m \ell} s_{m,i}^{(R)}(k + \ell P) \tag{66}$$

which indicates that any OFDMA block has a periodic structure that repeats every P samples. As anticipated, this property reveals useful for the synchronization purpose. In particular, it has been used in [49] to derive a frequency estimation scheme based on the principle of

multiple signal classification (MUSIC) [50]. The resulting procedure can be summarized as follows.

1) Let $r_i(k) = r(k + iN_T)$ $(k = 0, 1, \ldots, N-1)$ be the received samples within the ith DFT window and arrange them into the following $R \times P$ matrix:

$$\mathbf{G}_i = \begin{bmatrix} r_i(0) & \cdots & r_i(P-1) \\ r_i(P) & \cdots & r_i(2P-1) \\ \vdots & \ddots & \vdots \\ r_i(N-P) & \cdots & r_i(N-1) \end{bmatrix}. \quad (67)$$

2) Compute the $R \times R$ sample-correlation matrix

$$\mathbf{Q}_i = \frac{1}{P}\mathbf{G}_i\mathbf{G}_i^H. \quad (68)$$

3) Determine the noise subspace by finding the $R - M$ smallest eigenvalues of $\mathbf{Q}_i$ and arrange the corresponding eigenvectors into an $R \times (R - M)$ matrix denoted by $\mathbf{U}_i$.
4) Compute estimates $\{\widehat{\xi}_m\}_{m=1}^M$ of quantities ξ_m by locating the M highest peaks of the following metric:

$$\Gamma(\widetilde{\xi}) = \frac{1}{\left\|\mathbf{U}_i^H\mathbf{a}(\widetilde{\xi})\right\|^2} \quad (69)$$

where $\mathbf{a}(\widetilde{\xi}) = [1, e^{j2\pi\tilde{\xi}}, e^{j4\pi\tilde{\xi}}, \ldots, e^{j2\pi(R-1)\tilde{\xi}}]^T$.
5) Use quantities $\{\widehat{\xi}_m\}_{m=1}^M$ to obtain frequency estimates by

$$\widehat{\varepsilon}_m = R \cdot \widehat{\xi}_m - i_m, \quad 0 \le m \le M-1. \quad (70)$$

In the following discussion, this structure-based algorithm is referred to as the Cao–Tureli–Yao (CTY) estimator. Its main advantage is that it provides estimates of users' CFOs in the absence of any training block or pilot sequence. The only requirement is that CFOs must be confined within half the subcarrier spacing since otherwise the intervals spanned by quantities ξ_m partially overlap and in such a case it would be difficult to match each $\widehat{\xi}_m$ with the corresponding user.

It is shown in [49] that the accuracy of CTY degrades as the number of active users approaches the number of available subchannels. A simple way to improve the system performance is to enlarge the observation window to comprehend a specified number I of adjacent OFDMA blocks. In this case, the CTY proceeds as indicated earlier except that the sample correlation matrix $\mathbf{Q}_i$ is now computed as

$$\mathbf{Q}_i = \frac{1}{PI}\sum_{k=i}^{i+I-1}\mathbf{G}_k\mathbf{G}_k^H. \quad (71)$$

A major assumption for the application of CTY is that the OFDMA uplink signals are quasi-synchronous. As mentioned earlier, this poses an upper limit on the maximum distance between the BS and mobile terminals and may prevent the use of CTY in a number of applications. A possible solution to this problem relies on the transmission of suitable training blocks placed at the beginning of the uplink frame, which are exploited for synchronization as well as channel estimation purposes. These blocks are preceded by relatively long CPs comprising both the channel delay spread and the two-way propagation delay. In this way, during the training period, uplink signals are quasi-synchronous and CTY can be used for frequency estimation. To reduce unnecessary overhead, however, it is desirable that data blocks have a shorter prefix (only a bit longer than the channel response duration). In this case, accurate knowledge of the timing offsets is necessary to align all users in time and avoid IBI over the data section of the frame. A simple way to get timing estimates is based on knowledge of users' channel responses. This method is now explained by reconsidering the mth extended channel vector $\mathbf{h}'_m$ defined in Section V-C. Recall that

$$\mathbf{h}'_m = \begin{bmatrix} \mathbf{0}_{\theta_m}^T & \mathbf{h}_m^T & \mathbf{0}_{L-\theta_m-L_m}^T \end{bmatrix}^T \quad (72)$$

where $\mathbf{h}_m = [h_m(0), h_m(1), \ldots, h_m(L_m - 1)]^T$ is the mth channel response. Assume that an estimate of $\mathbf{h}'_m$, say $\widehat{\mathbf{h}}'_m$, has been computed by exploiting training blocks placed at the beginning of the uplink frame. Then, it is intuitively clear from (72) that an estimate of θ_m can be obtained by looking for the maximum of the energy of $\widehat{\mathbf{h}}'_m$ over a sliding window of length L_m. This is tantamount to setting

$$\widehat{\theta}_m = \arg\max_{\tilde{\theta}}\left\{\sum_{\ell=\tilde{\theta}}^{\tilde{\theta}+L_m-1}\left|\widehat{h}'_m(\ell)\right|^2\right\} \quad (73)$$

where $\widehat{h}'_m(\ell)$ is the ℓth entry of $\widehat{\mathbf{h}}'_m$.

C. Frequency Estimation With Generalized CAS

The generalized CAS is a dynamic resource allocation scheme in which the BS exploits knowledge of users' channel responses to assign the best subcarriers that are currently available. This allocation strategy is more flexible

than the subband or interleaved CAS and provides the system with some form of *multiuser diversity* [51] since a subcarrier that is in a deep fade for one user may exhibit a relatively large gain for another. On the other hand, the absence of any rigid structure in the allocation policy makes the synchronization task even more challenging than with interleaved CAS.

One of the first synchronization schemes for OFDMA systems with generalized CAS was proposed by Morelli in [52]. This method employs the ML principle to compute estimates of the timing and frequency offsets of a new user entering the network. Unfortunately, it relies on the fact that other users have already been acquired and aligned to the BS references, an assumption that may be too stringent in practical applications. Alternative ML-based synchronization schemes for generalized CAS are described in [53] and [54]. Here, timing and frequency estimates are obtained on the basis of a training block which is transmitted by each user at the beginning of the uplink frame. These methods are now reviewed over a quasi-synchronous scenario wherein the CP of the training block is sufficiently long to comprehend both the channel delay spread and propagation delays incurred by users' signals. Since in this case timing information can be achieved as indicated in (73), in the following we only concentrate on the joint ML estimation of the channel responses and frequency errors. For notational simplicity, the block index i is omitted in all subsequent derivations.

We denote $p_m(n)$ $(n \in \mathcal{I}_m)$ the pilot symbols transmitted by the mth user during the training block. Then, from (47), we may write the corresponding time-domain samples as

$$s_m(k) = \frac{1}{\sqrt{N}} \sum_{n \in \mathcal{I}_m} p_m(n) e^{j2\pi nk/N}, \quad -N_g \le k \le N-1. \tag{74}$$

At the BS receiver, the CP is removed and the remaining samples are given by

$$r(k) = \sum_{m=1}^{M} e^{j2\pi\varepsilon_m k/N} \sum_{\ell=0}^{L-1} h'_m(\ell) s_m(k-\ell) + w(k), \quad 0 \le k \le N-1 \tag{75}$$

where $\mathbf{h}'_m = [h'_m(0), h'_m(1), \ldots, h'_m(L-1)]^T$ is the mth extended channel vector and $w(k)$ is a noise term.

Collecting the received samples into an N-dimensional vector $\mathbf{r} = [r(0), r(1), \ldots, r(N-1)]^T$, we can rewrite (75) into the equivalent form

$$\mathbf{r} = \sum_{m=1}^{M} \mathbf{r}_m + \mathbf{w} \tag{76}$$

where $\mathbf{w} = [w(0), w(1), \ldots, w(N-1)]^T$ is the noise vector and

$$\mathbf{r}_m = \mathbf{D}(\varepsilon_m)\mathbf{S}_m\mathbf{h}'_m \tag{77}$$

is the signal component from the mth user. In (77),

$$\mathbf{D}(\varepsilon_m) = \operatorname{diag}\left\{1,\ e^{j2\pi\varepsilon_m/N}, \ldots, e^{j2\pi(N-1)\varepsilon_m/N}\right\} \tag{78}$$

is a diagonal matrix and $\mathbf{S}_m$ is an $N \times L$ matrix with known entries $[\mathbf{S}_m]_{k,\ell} = s_m(k-\ell)$ for $0 \le k \le N-1$ and $0 \le \ell \le L-1$.

In principle, the received vector $\mathbf{r}$ could be exploited to get joint ML estimates of frequency offsets and channel responses of all active users. As shown in [53], however, the exact solution to this problem turns out to be too complex for practical purposes as it involves a grid-search over a multidimensional domain. An efficient method to recursively approach the ML solution is the space-alternating projection expectation-maximization (SAGE) algorithm [55]. Similarly to the well known EM algorithm [56], this technique operates in an iterative fashion where the original measurements are replaced with some *complete data set* from which the original measurements can be obtained through a many-to-one mapping. The SAGE algorithm alternates between an E-step, calculating the log-likelihood function of the complete data, and an M-step, maximizing that expectation with respect to unknown parameters. At any iteration, parameter estimates are updated and the process continues until no significant changes in the updates are observed.

In [53] and [54], the SAGE algorithm is applied to the signal model (76), (77) to compute estimates of ε_m and $\mathbf{h}'_m$ for $m = 1, 2, \ldots, M$. The resulting procedure consists of *iterations* and *cycles*. An iteration is made of M cycles, where each cycle updates the parameters of a single user while keeping those of the others at their latest updated values. To show the idea, we call $\widehat{\varepsilon}_m^{(j)}$ and $\widehat{\mathbf{h}}_m'^{(j)}$ the estimates of ε_m and $\mathbf{h}'_m$ after the jth iteration, respectively, and assume that some initial estimates $\widehat{\varepsilon}_m^{(0)}$ and $\widehat{\mathbf{h}}_m'^{(0)}$ are available at the BS. The latter are exploited to compute the following M vectors, one for each user:

$$\widehat{\mathbf{r}}_m^{(0)} = \mathbf{D}\left(\widehat{\varepsilon}_m^{(0)}\right)\mathbf{S}_m\widehat{\mathbf{h}}_m'^{(0)}, \quad 1 \le m \le M. \tag{79}$$

Then, during the mth cycle of the jth iteration, the SAGE algorithm proceeds as follows.

1) **E-step**: Compute

$$\mathbf{y}_m^{(j)} = \mathbf{r} - \sum_{k=1}^{m-1} \widehat{\mathbf{r}}_k^{(j)} - \sum_{k=m+1}^{M} \widehat{\mathbf{r}}_k^{(j-1)} \tag{80}$$

where a notation of the type $\sum_{\ell}^{u}$ is zero whenever $u < \ell$.

2) **M-step**: Compute estimates of ε_m and $\mathbf{h}'_m$ by locating the minimum of the following cost function:

$$\Lambda\left(\widetilde{\varepsilon}_m, \widetilde{\mathbf{h}}'_m\right) = \left\| \mathbf{y}_m^{(j)} - \mathbf{D}(\widetilde{\varepsilon}_m)\mathbf{S}_m \widetilde{\mathbf{h}}'_m \right\|^2 \tag{81}$$

with respect to $\widetilde{\varepsilon}_m$ and $\widetilde{\mathbf{h}}'_m$. This yields

$$\widehat{\varepsilon}_m^{(j)} = \arg\max_{\widetilde{\varepsilon}_m} \left\{ \left\| \mathbf{S}_m \mathbf{S}_m^{\dagger} \mathbf{D}^H(\widetilde{\varepsilon}_m) \mathbf{y}_m^{(j)} \right\|^2 \right\} \tag{82}$$

$$\widehat{\mathbf{h}}_m'^{(j)} = \mathbf{S}_m^{\dagger} \mathbf{D}^H\left(\widehat{\varepsilon}_m^{(j)}\right) \mathbf{y}_m^{(j)} \tag{83}$$

where $\mathbf{S}_m^{\dagger} = (\mathbf{S}_m^H \mathbf{S}_m)^{-1} \mathbf{S}_m^H$ is the Moore–Penrose generalized inverse of $\mathbf{S}_m$. The estimated parameters are used to obtain the following vector:

$$\widehat{\mathbf{r}}_m^{(j)} = \mathbf{D}\left(\widehat{\varepsilon}_m^{(j)}\right) \mathbf{S}_m \widehat{\mathbf{h}}_m'^{(j)} \tag{84}$$

which is then exploited in the E-step of the next cycle or iteration.

A physical interpretation of (81) is as follows. From (76) and (77), we see that the signal component in $\mathbf{r}$ results from the contributions $\mathbf{r}_k$ of several users ($1 \le k \le M$), each depending on a specific set of parameters $(\varepsilon_k, \mathbf{h}'_k)$. If all these sets were known except for $(\varepsilon_m, \mathbf{h}'_m)$, the contributions of users with indexes $k \neq m$ could be subtracted from $\mathbf{r}$, yielding a MAI-free vector

$$\mathbf{y}_m = \mathbf{r} - \sum_{k \neq m} \mathbf{r}_k. \tag{85}$$

Then, we should estimate $(\varepsilon_m, \mathbf{h}'_m)$ based on the observation of $\mathbf{y}_m$. In practice, $\mathbf{y}_m$ is not available at the receiver so that we are compelled to replace it with a suitable estimate $\mathbf{y}_m^{(j)}$ as indicated in (80), which can be rewritten as

$$\mathbf{y}_m^{(j)} = \mathbf{D}(\varepsilon_m)\mathbf{S}_m \mathbf{h}'_m + \mathbf{i}_m + \mathbf{w} \tag{86}$$

where $\mathbf{i}_m$ is a disturbance term that accounts for imperfect cancellation of interfering signals. Vector $\mathbf{y}_m^{(j)}$ is used in place of true $\mathbf{y}_m$ to get estimates of unknown quantities $(\varepsilon_m, \mathbf{h}'_m)$. By ignoring the interference term in (86), the ML estimator of $(\varepsilon_m, \mathbf{h}'_m)$ aims at minimizing the right-hand-side of (81). Based on the above arguments, the SAGE algorithm can be viewed as a recursive approximation to the ML estimator in which previous estimates of synchronization parameters are exploited to cancel out the MAI.

A possible risk associated with the use of EM-type algorithms is that local peaks present in the likelihood function might attract the solution toward spurious locks. In practice, the SAGE algorithm has a higher chance to converge to the global maximum of the likelihood function if accurate frequency and channel estimates are used in the initialization setp. A viable method to obtain these initial estimates can be stated as follows. First, we compute the N-point DFT of $\mathbf{r}$ and select those outputs that correspond to the subcarriers of the mth user while putting the others to zero. Next, we return to the time domain through an IDFT operation and exploit the resulting samples to compute $\widehat{\varepsilon}_m^{(0)}$ and $\widehat{\mathbf{h}}_m'^{(0)}$. In doing so, we can resort to the single-user scheme discussed in [57]. Intuitively, computing the DFT of $\mathbf{r}$ and forcing subcarriers allocated to interfering users to zero helps mitigate the MAI effect.

D. Computational Complexity

We now assess the complexity of the timing and frequency synchronization schemes discussed in previous sections. We begin with the method proposed in [48] for subband CAS, which evaluates the timing metrics $\{\Gamma_m(\widetilde{\theta})\}_{m=1}^{M}$ in (60) over a number N_θ of possible timing errors $\widetilde{\theta}$. Since computing each $\Gamma_m(\widetilde{\theta})$ needs approximately N_g operations, the overall computational requirement is $\mathcal{O}(N_g M N_\theta)$. As indicated by (59), frequency estimation comes for free once the quantities $\{\Gamma_m(\widetilde{\theta})\}$ have been computed over all candidate values $\widetilde{\theta}$.

Next, we consider the CTY frequency estimator for interleaved CAS. The eigen-decomposition of the sample-correlation matrix $\mathbf{Q}_i$ involves $\mathcal{O}(R^3)$ operations, whereas it takes approximately $\mathcal{O}(R^2)$ operations to obtain the metric $\Gamma(\widetilde{\xi})$ given in (69). Denoting by N_ε the number of candidate frequency offsets (which equals the number of candidate values $\widetilde{\xi}$) and observing that, in general, we have $N_\varepsilon > R$, the overall complexity of CTY turns out to be $\mathcal{O}(R^2 N_\varepsilon)$.

We finally move our attention to the SAGE-based method for joint frequency and channel estimation with a

generalized CAS. For presentational simplicity, we omit the calculations required to get the initial estimates $\{\widehat{\mathbf{r}}_m^{(0)}, \hat{\varepsilon}_m^{(0)}, \widehat{\mathbf{h}}_m'^{(0)}\}$ and assume that $\mathbf{S}_m^{\dagger}$ has been precomputed and stored in the receiver. Thus, according to (82) and (83), we see that each cycle needs $\mathcal{O}(2LNN_\varepsilon)$ and $\mathcal{O}(LN)$ operations to update $\widehat{\varepsilon}_m^{(j)}$ and $\widehat{\mathbf{h}}_m'^{(j)}$, respectively. This results into an overall complexity of $\mathcal{O}(LNM(2N_\varepsilon+1))$ per SAGE iteration. These figures indicate that the SAGE estimator is by far the most computationally demanding among all the considered schemes.

VII. TIMING AND FREQUENCY OFFSET COMPENSATION IN UPLINK TRANSMISSIONS

Once the uplink timing and frequency offsets have properly been estimated, they are employed by the BS to restore orthogonality among users' signals. This operation is known as timing and frequency correction, which is the final stage of the overall synchronization process. As mentioned before, timing and frequency correction in the uplink cannot be achieved with the same methods employed in the downlink since the alignment of one specific user would misalign all the others. A first solution to this problem was presented in [48], where estimates of users' offsets are returned to active terminals via a downlink control channel and exploited by each user to properly adjust its transmitted signal. More recently, advanced signal processing techniques have been proposed to compensate for synchronization errors *directly* at the BS. These schemes are largely inherited from the multiuser detection area and avoid the need for any exchange of side information between the BS and active terminals.

For the rest of this section, we first concentrate on the problem of timing and frequency correction for OFDMA systems with subband CAS. Then, we show how linear multiuser and interference cancellation schemes may be employed for frequency correction with a generalized CAS.

A. Timing and Frequency Compensation With Subband CAS

An appealing feature of subband CAS is that signals' separation at the BS is easily accomplished through a bank of bandpass filters. The receiver can thus estimate and correct synchronization errors independently for each user as shown in the block diagram of Fig. 16. Here, the output $x_m(k)$ from the filter tuned on the mth subband is exploited to get estimates $\widehat{\theta}_m$ and $\widehat{\varepsilon}_m$ of the timing and frequency offsets using one of the methods described in Section VI-A. The estimated parameters are then employed to correct synchronization errors in a way similar to conventional OFDM systems. In particular, the samples $x_m(k)$ are multiplied by the exponential term $e^{-j2\pi k\hat{\varepsilon}_m/N}$ to compensate for the CFO while the timing estimate $\widehat{\theta}_m$ is used to select N samples that are next processed by the DFT unit.

A possible shortcoming of the receiver architecture of Fig. 16 is that perfect signals' separation is not possible even with ideal brick-wall filters due to the frequency leakage among adjacent subbands caused by synchronization errors. In such a case, some residual MAI will be present at the output of each DFT unit. In addition, a distinct N-point DFT operation is required for each active user. Since the computational load of the DFT represents a major concern for system implementation, the receiver structure of Fig. 16 may be too complex in practical situations where the number M of users and/or the number N of subcarriers are relatively large.

A certain reduction of complexity can be obtained by employing the frequency correction scheme proposed by

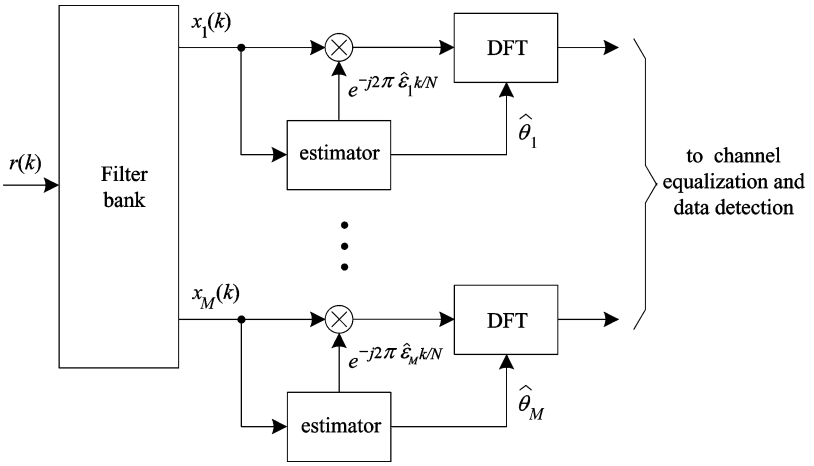

Fig. 16. *Timing and frequency synchronization in OFDMA uplink receiver with subband CAS.*

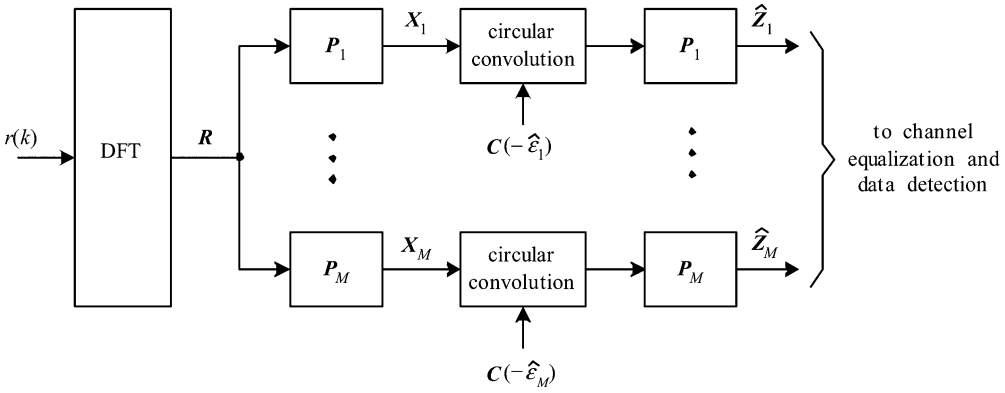

Fig. 17. ***Frequency correction by means of circular convolutions at DFT output.***

Choi–Lee–Jung–Lee (CLJL) in [58] as depicted in Fig. 17. This solution avoids the need for multiple DFT operations but cannot perform timing adjustment. Thus, it can only be applied to a quasi-synchronous scenario where timing correction is unnecessary.

To explain the rationale behind CLJL, we reconsider the N received samples $r_i(k)$ $(0 \le k \le N-1)$ belonging to the ith DFT window, which can be written as

$$r_i(k) = \sum_{m=1}^{M} z_{m,i}(k) e^{j2\pi\varepsilon_m k/N} + w_i(k), \quad 0 \le k \le N-1 \tag{87}$$

where $w_i(k)$ is thermal noise and $z_{m,i}(k)$ is the multipath distorted version of the mth user signal of the following form:

$$z_{m,i}(k) = \frac{1}{\sqrt{N}} e^{j\varphi_{m,i}} \sum_{n\in\mathcal{I}_m} H'_m(n) d_{m,i}(n) e^{j2\pi nk/N}, \quad 0 \le k \le N-1 \tag{88}$$

with $\varphi_{m,i} = 2\pi\varepsilon_m i N_T/N$.

For convenience, we collect the N-point DFT of sequences $r_i(k)$, $z_{m,i}(k)$, and $w_i(k)$ into three N-dimensional vectors $\mathbf{R}_i$, $\mathbf{Z}_{m,i}$, and $\mathbf{W}_i$, respectively. In the following, index i is omitted for notational simplicity. Since a multiplication in the time domain corresponds to a circular convolution in the frequency domain, we have from (87) that

$$\mathbf{R} = \sum_{m=1}^{M} \mathbf{Z}_m \otimes \mathbf{C}(\varepsilon_m) + \mathbf{W} \tag{89}$$

where $\mathbf{C}(\varepsilon_m)$ is the N-point DFT of $\{e^{j2\pi\varepsilon_m k/N}; 0 \le k \le N-1\}$ and $\otimes$ denotes the N-point circular convolution. The entries of $\mathbf{Z}_m$ can be derived from (88) as

$$Z_m(n) = \begin{cases} e^{j\varphi_m} H'_m(n) d_m(n), & \text{if } n \in \mathcal{I}_m \\ 0, & \text{otherwise.} \end{cases} \tag{90}$$

We see that $\mathbf{Z}_m$ represents the contribution of the mth uplink signal to the DFT output in the absence of any interference caused by frequency offsets. Thus, the purpose of the frequency synchronization scheme is to recover $\{\mathbf{Z}_m\}_{m=1}^{M}$ from the observation of $\mathbf{R}$. The CLJL scheme achieves this goal in two steps. First, it computes the N-dimensional vectors $\{\mathbf{X}_m\}_{m=1}^{M}$, where each $\mathbf{X}_m$ is obtained by selecting the entries of $\mathbf{R}$ with indexes $n \in \mathcal{I}_m$ while forcing the others to zero. Mathematically, this is equivalent to setting $\mathbf{X}_m = \mathbf{P}_m\mathbf{R}$, where $\mathbf{P}_m$ is a diagonal matrix whose (n,n)th entry is unitary if $n \in \mathcal{I}_m$ and is zero otherwise. In practice, $\mathbf{P}_m$ acts as a bandpass filter which serves to isolate the contribution of the mth uplink signal at the DFT output. Clearly, perfect signal separation is not possible since the uncompensated CFOs produce some frequency leakage among users' subbands. By ignoring for simplicity the effect of the frequency leakage and based on (89), we have

$$\mathbf{X}_m \approx \mathbf{Z}_m \otimes \mathbf{C}(\varepsilon_m) + \mathbf{W}_m \tag{91}$$

where $\mathbf{W}_m = \mathbf{P}_m\mathbf{W}$ is the noise contribution. The second step of CLJL is to get an estimate of $\mathbf{Z}_m$ from $\mathbf{X}_m$. For this purpose, we observe that $\mathbf{Z}_m \otimes \mathbf{C}(\varepsilon_m) \otimes \mathbf{C}(-\varepsilon_m) = \mathbf{Z}_m$ and recall that the nth entry of $\mathbf{Z}_m$ is zero for $n \notin \mathcal{I}_m$ as

indicated in (90). Then, from (91), it follows that an estimate of $\mathbf{Z}_m$ can be obtained in the form

$$\widehat{\mathbf{Z}}_m = \mathbf{P}_m[\mathbf{X}_m \otimes \mathbf{C}(-\widehat{\varepsilon}_m)] \tag{92}$$

where $\mathbf{C}(-\widehat{\varepsilon}_m)$ collects the N-point DFT of sequence $\{e^{-j2\pi\widehat{\varepsilon}_m k/N}; 0 \le k \le N-1\}$. Substituting (91) into (92) and assuming ideal frequency estimation (i.e., $\widehat{\varepsilon}_m = \varepsilon_m$), we get

$$\widehat{\mathbf{Z}}_m = \mathbf{Z}_m + \mathbf{P}_m[\mathbf{W}_m \otimes \mathbf{C}(-\widehat{\varepsilon}_m)] \tag{93}$$

which indicates that $\widehat{\mathbf{Z}}_m$ is unbiased and, apart from thermal noise, is free from interference. In practice, however, nonideal frequency compensation and imperfect users' separation give rise to some residual ICI and MAI on $\widehat{\mathbf{Z}}_m$, thereby leading to some performance degradation with respect to the ideal setting described by (93).

B. Frequency Compensation Through Interference Cancellation

The CLJL scheme discussed in the previous section is only suited for an OFDMA system with subband CAS. The reason is that the bank of matrices $\mathbf{P}_m$ $(1 \le m \le M)$ in Fig. 17 can provide accurate users' separation as long as the subcarriers of a given user are grouped together and sufficiently large guard intervals are inserted among adjacent subchannels. When used in conjunction with an interleaved or a generalized CAS, however, the CLJL cannot significantly reduce the MAI induced by frequency errors. In this case, alternative approaches must be resorted to. One possibility is offered by the concept of multiuser detection [59]. The latter was originally developed for the joint demodulation of mutually interfering data streams and includes linear as well as interference cancellation (IC) architectures. In this section, we show how the IC concept can be applied to CLJL to reduce any residual interference present on $\widehat{\mathbf{Z}}_m$. The resulting scheme has been derived by Huang and Letaief (HL) in [60] and operates in the following iterative fashion, where $\widehat{\mathbf{Z}}_m^{(j)}$ is the mth restored signal after the jth iteration.

1) **Initialization**: Use the CLJL vectors defined in (92) as initial estimates of $\{\mathbf{Z}_m\}_{m=1}^{M}$, i.e.,

$$\widehat{\mathbf{Z}}_m^{(0)} = \mathbf{P}_m[(\mathbf{P}_m\mathbf{R}) \otimes \mathbf{C}(-\widehat{\varepsilon}_m)], \quad 1 \le m \le M. \tag{94}$$

2) **jth iteration** $(j = 1, 2, \ldots)$: For each active user $(m = 1, 2, \ldots, M)$, perform interference cancellation in form of

$$\widetilde{\mathbf{Z}}_m^{(j)} = \mathbf{R} - \sum_{k=1, k\ne m}^{M} \widehat{\mathbf{Z}}_k^{(j-1)} \otimes \mathbf{C}(\widehat{\varepsilon}_k), \quad 1 \le m \le M \tag{95}$$

and then compensate for ε_m in a way similar to CLJL

$$\widehat{\mathbf{Z}}_m^{(j)} = \mathbf{P}_m\left[\left(\mathbf{P}_m\widetilde{\mathbf{Z}}_m^{(j)}\right) \otimes \mathbf{C}(-\widehat{\varepsilon}_m)\right], \quad 1 \le m \le M. \tag{96}$$

As indicated in (95), at each iteration the contribution of interfering users is regenerated and subtracted from the original DFT output $\mathbf{R}$. The expurgated vectors $\widetilde{\mathbf{Z}}_m^{(j)}$ are then used to obtain the restored signals $\widehat{\mathbf{Z}}_m^{(j)}$ according to (96). In this respect, the HL can be regarded as a parallel interference cancellation (PIC) scheme. However, in contrast to other IC-based frequency compensation methods existing in the literature [61], HL does not exploit any data decision and, in consequence, is not plagued by the error propagation phenomenon. Simulation results reported in [60] indicate that HL exhibits a significant advantage over CLJL after just a few iterations. In particular, its increased robustness against ICI and MAI makes it suited for any CAS.

C. Frequency Compensation Through Linear Multiuser Detection

Linear multiuser detection can be used as an alternative to the IC concept to mitigate interference arising from uplink CFOs. An example in this sense is provided by Cao–Tureli–Yao–Honan (CTYH) in [62]. Their method is suitable for any CAS but can only be applied to a quasi-synchronous scenario where signals are time aligned within the CP and no IBI is present.

The CTYH aims at restoring orthogonality among users by applying a linear transformation to the DFT output $\mathbf{R}$. To see this, we return to (89) and observe that the nth entry of $\mathbf{R}$ is given by

$$R(n) = \sum_{m=1}^{M}\sum_{\ell=0}^{N-1} Z_m(\ell) f_N(\varepsilon_m + \ell - n) + W(n), \quad 0 \le n \le N-1 \tag{97}$$

where $f_N(x)$ was defined in (15) and $Z_m(\ell)$ is a channel distorted version of the symbol transmitted by the mth user

over the ℓth subcarrier as given in (90). Equation (97) indicates that MAI is present at the DFT output when $\varepsilon_m \neq 0$ since $R(n)$ is contributed by all active users in this case. To proceed further, we use $\mathcal{I}_m = \{q_m(p); 0 \leq p \leq P-1\}$ to denote the indexes of P subcarriers assigned to the mth user. Then, bearing in mind (90), we may rewrite (97) in matrix notation as

$$\mathbf{R} = \sum_{m=1}^{M} \mathbf{F}_m(\varepsilon_m)\mathbf{V}_m + \mathbf{W} \tag{98}$$

where $\mathbf{W}$ is the noise vector, $\mathbf{F}_m(\varepsilon_m)$ is an $N \times P$ matrix with entries

$$[\mathbf{F}_m(\varepsilon_m)]_{n,p} = f_N[\varepsilon_m + q_m(p) - n], \quad 0 \leq n \leq N-1, 0 \leq p \leq P-1 \tag{99}$$

and $\mathbf{V}_m$ is a P-dimensional vector collecting the mth user's data multiplied by the corresponding channel frequency response, i.e.,

$$[\mathbf{V}_m]_p = e^{j\varphi_m} H'_m[q_m(p)] d_m[q_m(p)], \quad 0 \leq p \leq P-1. \tag{100}$$

From (98), it turns out that $\mathbf{R}$ can also be rewritten as

$$\mathbf{R} = \mathbf{F}(\boldsymbol{\varepsilon})\mathbf{V} + \mathbf{W} \tag{101}$$

where $\mathbf{V} = [\mathbf{V}_1^T \ \mathbf{V}_2^T \ \cdots \ \mathbf{V}_M^T]^T$ and $\mathbf{F}(\boldsymbol{\varepsilon}) = [\mathbf{F}_1(\varepsilon_1) \ \mathbf{F}_2(\varepsilon_2) \ \cdots \ \mathbf{F}_M(\varepsilon_M)]^T$ is an $N \times MP$ matrix whose elements are related to users' CFOs $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M]^T$. As shown in (100), each entry of $\mathbf{V}$ is a channel distorted version of a single data symbol. In this respect, $\mathbf{V}$ is the vector that would be ideally present at the DFT output in the absence of any interference and thermal noise. Orthogonality among users' signals is thus restored after providing the receiver with an estimate of $\mathbf{V}$. As illustrated in Fig. 18, the CTYH scheme achieves this goal by means of a linear transformation applied to $\mathbf{R}$. The estimated vector $\widehat{\mathbf{V}}$ is then fed to the channel equalizer and data detection unit, which provides decisions on transmitted data symbols.

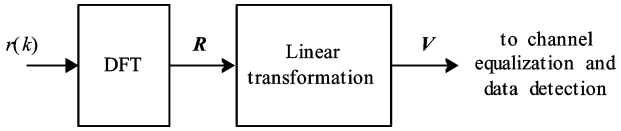

Fig. 18. ***Frequency correction by means of linear transformation applied to DFT output.***

Two possible methods for computing $\widehat{\mathbf{V}}$ are illustrated in [62]. Based on an LS approach, the first one is equivalent to the well-known *linear decorrelating detector* (LDD) [59]

$$\widehat{\mathbf{V}}_{LDD} = \mathbf{F}^{\dagger}(\boldsymbol{\varepsilon})\mathbf{R} \tag{102}$$

where $\mathbf{F}^{\dagger}(\boldsymbol{\varepsilon}) = [\mathbf{F}^H(\boldsymbol{\varepsilon})\mathbf{F}(\boldsymbol{\varepsilon})]^{-1}\mathbf{F}^H(\boldsymbol{\varepsilon})$ is the generalized inverse of $\mathbf{F}(\boldsymbol{\varepsilon})$. Substituting (101) into (102) yields

$$\widehat{\mathbf{V}}_{LDD} = \mathbf{V} + \mathbf{F}^{\dagger}(\boldsymbol{\varepsilon})\mathbf{W} \tag{103}$$

from which it is seen that LDD can totally suppress any interference produced by frequency errors. However, this result is obtained at the price of a certain enhancement of the output noise level.

The second solution is based on the MMSE criterion and aims at minimizing the overall effect of interference plus ambient noise. The resulting scheme is known as the linear MMSE detector [59] and can be written as

$$\widehat{\mathbf{V}}_{\mathrm{MMSE}} = \left[\mathbf{F}^H(\boldsymbol{\varepsilon})\mathbf{F}(\boldsymbol{\varepsilon}) + \sigma_w^2 \mathbf{I}_{MP}\right]^{-1}\mathbf{F}^H(\boldsymbol{\varepsilon})\mathbf{R}. \tag{104}$$

Although some residual MAI may be present on $\widehat{\mathbf{V}}_{\mathrm{MMSE}}$, the noise enhancement phenomenon is greatly reduced as compared to LDD.

D. Computational Complexity

The timing and frequency compensation methods discussed in the previous sections are now compared in terms of computational complexity. We begin by considering the receiver architecture of Fig. 16, where frequency correction is accomplished in the time domain at the output of the filter bank. Clearly, the crux in the calculations is represented by the need for M distinct DFT operations (one for each active terminal), which results into an overall complexity $\mathcal{O}(MN \log_2 N)$.

The CLJL scheme operates in the frequency domain and performs CFO compensation by means of M circular convolutions as indicated in (92). Since vectors $\{\widehat{\mathbf{Z}}_m\}_{m=1}^{M}$ must only be computed over the P subcarriers assigned to the considered user, $\mathcal{O}(PN)$ operations are required to obtain each $\widehat{\mathbf{Z}}_m$. Hence, the overall computational requirement is approximately $\mathcal{O}(PNM)$. The HL scheme improves upon CLJL by iteratively subtracting the regenerated interference from the received signal. Inspection of (95) and (96) reveals that $\mathcal{O}(PNM)$ operations are required at each iteration for every user. The resulting

Table 1 Physical Layer Parameters of IEEE 802.16 Wireless MAN

Parameters	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
System bandwidth (MHz)	1.25	2.5	5.0	10.0	20.0
Sampling frequency (MHz)	1.429	2.857	5.714	11.429	22.857
Sampling period (ns)	700	350	175	88	44
FFT size	128	256	512	1024	2048
Subcarrier spacing	11.16 kHz				
CP length	11.2 μs				
Block duration	100.8 μs				

complexity is thus $\mathcal{O}(PNM^2N_i)$, where N_i is the number of iterations.

We finally consider the CTYH scheme, which obtains the restored signal vector after applying a linear transformation to the received frequency-domain samples. In principle, this method requires $\mathcal{O}(N^3)$ operations, even though some computational saving is possible by taking into account that most of the interference afflicting a given subcarrier is caused by data symbols transmitted over neighboring subcarriers [62].

VIII. APPLICATIONS

Although originally suggested for cable TV networks [8], OFDMA made its first appearance in the uplink channel of the DVB-RCT system [9]. However, the great success of OFDMA started in 2002 when the IEEE 802.16 Working Group recommended it as an air-interface for WMAN applications [10]. The purpose of these systems is to provide broadband network access to business offices as well as residential customer premises through rooftop antennas communicating with a central radio BS, thereby

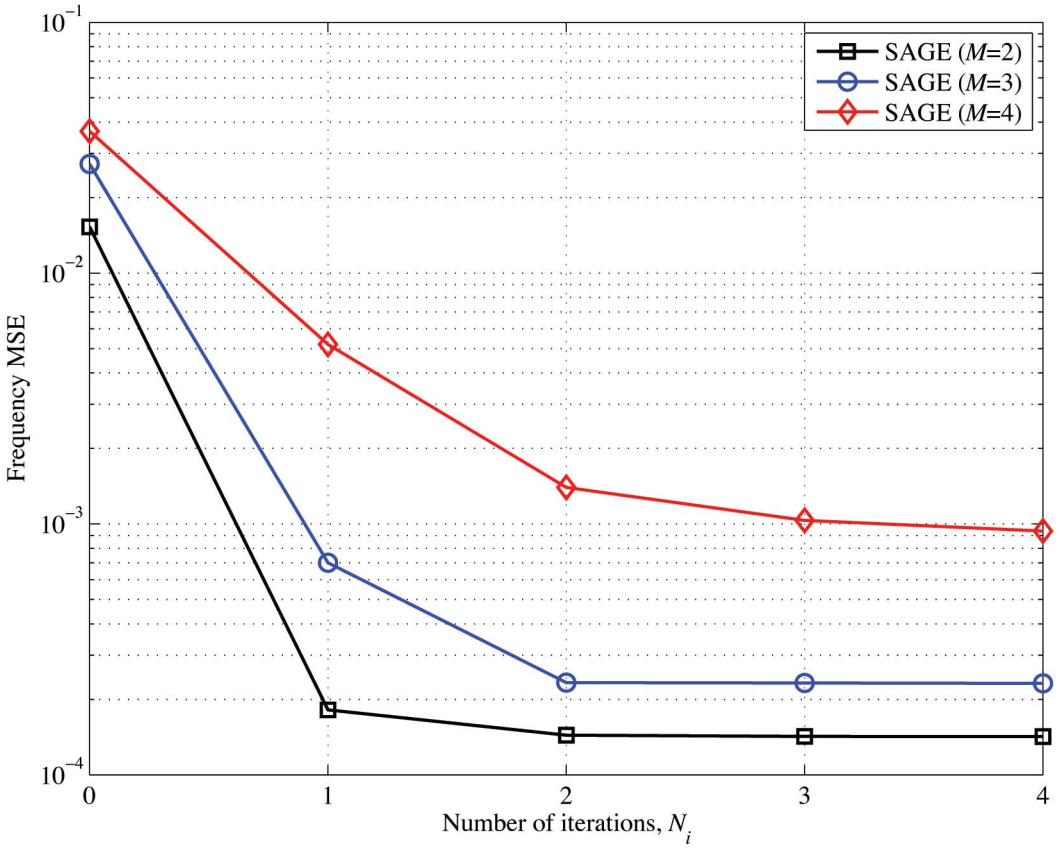

Fig. 19. ***Convergence performance of SAGE algorithm for frequency estimation.***

replacing the "last-mile" connection by a wireless link. This approach offers an appealing alternative to cabled access networks or digital subscriber line (DSL) links and promises ubiquitous broadband access to rural or developing areas where broadband is currently unavoidable for lack of a cabled infrastructure [63]. The typical application of WMANs is to deliver high data rates (up to 50 Mb/s) to fixed or portable subscriber stations (SSs) over metropolitan areas with cell radii up to 25–30 km. A more challenging situation occurs if the SS is mounted on a moving vehicle to provide passengers with continuous Internet connectivity.

In this section, the synchronization algorithms discussed throughout the paper are applied to a WMAN uplink scenario inspired by the IEEE 802.16 family of standards. The system parameters are summarized in Table 1.

As shown in the table, five transmission modes are available where the DFT size and signal bandwidth are suitably scaled to keep the subcarrier spacing fixed. This scalable architecture enables large flexibility without significant impact on the product cost [64]. Since we expect qualitatively similar performance from each of the transmission modes, we concentrate on the first mode to reduce the computational effort and speed up the simulations. Accordingly, the considered system has a DFT size of 128 and a signal bandwidth of 1.25 MHz, which corresponds to 112 modulated subcarriers. Unless otherwise specified, we assume an interleaved CAS, where each user is provided with a set of 28 subcarriers uniformly spaced over the signal bandwidth. In this way, up to four users can simultaneously be active on a given OFDMA block.

The channel has an exponential power delay profile with a delay spread of 5.2 μs, as specified by the Stanford University Interim (SUI) hilly terrain model [65]. This corresponds to a CIR length of approximately eight sampling periods. Observing that the CP has duration 11.2 μs, the considered system turns out to be quasi-synchronous as long as the two-way propagation delay is limited to 6.0 μs. As explained previously, this situation is extremely appealing as it allows one to incorporate timing errors into the channel response, thereby avoiding the need for any timing adjustment. Unfortunately, the above condition poses a limit of 900 m to the maximum distance between users' terminals and BS receiver, which is too stringent for WMAN applications with a coverage range of several kilometers. In this case, timing synchronization becomes unavoidable and may represent a critical problem. With interleaved or generalized CAS,

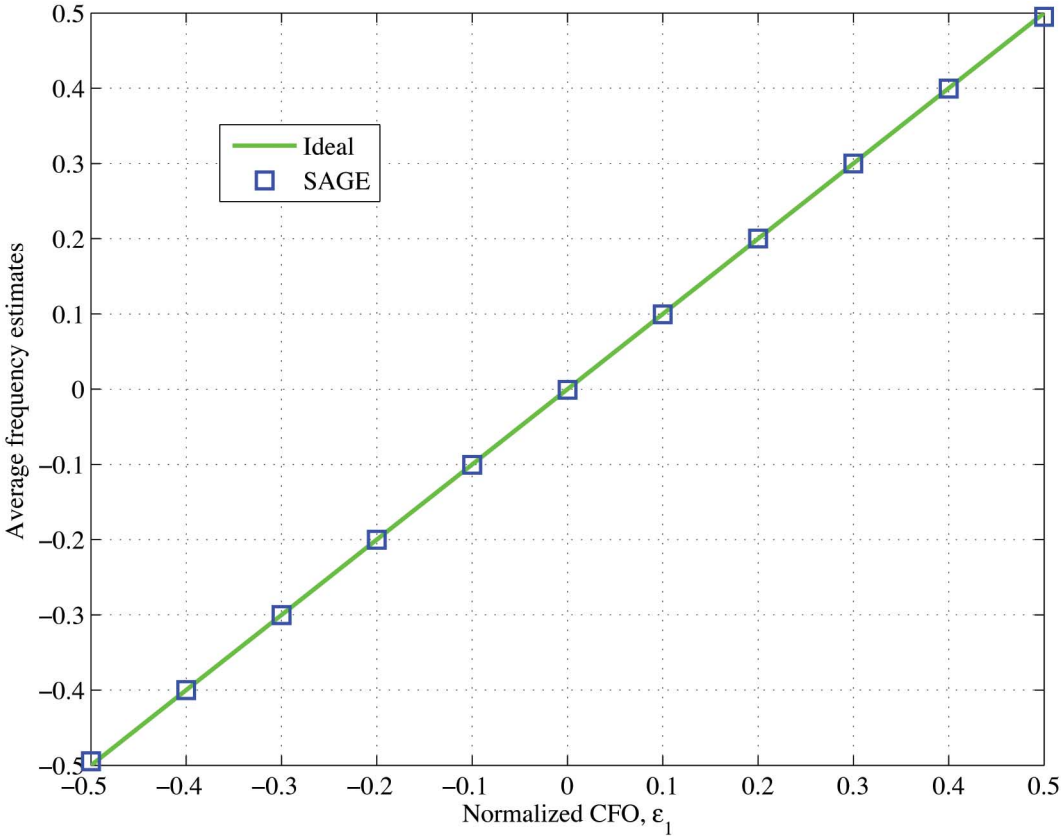

Fig. 20. *Average frequency estimates versus ε_1 for SAGE algorithm.*

the timing estimates can only be computed as indicated in (73) on the basis of suitable channel estimates. The latter are derived from dedicated training blocks equipped with a long CP comprising both the channel delay spread and the propagation delay. The timing estimates are next returned to users on a downlink control channel and exploited for timing adjustment at each mobile terminal. The reason for adopting this strategy is that no method has been devised so far to compensate timing errors directly at the BS except for a system with subband CAS (see Section VII-A). For simplicity, in the following discussion we assume a quasi-synchronous scenario and only concentrate on the performance of the frequency synchronization schemes when applied to an uncoded QPSK transmission.

Assuming a carrier frequency of 5 GHz and users' speeds up to 50 km/h, the maximum Doppler shift is limited to $\Delta f_{\max} = 230$ Hz, and the coherence time is given by [66]

$$T_{coh} = \sqrt{\frac{9}{16\pi \cdot \Delta f_{\max}^2}} = 1.84 \text{ ms}. \qquad (105)$$

Since the latter is much larger than the block duration, the channel is essentially static over several blocks. For this reason, a new channel snapshot is generated at each simulation run and kept fixed over an entire frame.

A. Performance of Frequency Estimation Schemes

Fig. 19 illustrates the MSE of frequency estimates obtained with the SAGE algorithm as a function of the number of iterations N_i in case of two, three, and four users. The latter have equal power with $E_b/N_0 = 20$ dB, where E_b is the average received energy per bit and N_0 represents the two-sided noise power spectral density. The SAGE algorithm is initialized as indicated in [53] and the normalized CFOs are uniformly distributed over interval $[-0.3, 0.3]$ and vary at each new simulation run. Without loss of generality, we only provide results for the first user. We see that the algorithm achieves convergence in only two iterations and no significant gains are observed with $N_i > 2$. The average frequency estimates provided by the SAGE algorithm are shown in Fig. 20 versus ε_1, assuming that three users are present in the systems. Here, ε_1 is kept fixed while ε_2 and ε_3 vary independently over the range $[-0.3, 0.3]$. The ideal line $E\{\widehat{\varepsilon}_1\} = \varepsilon_1$ is also drawn for

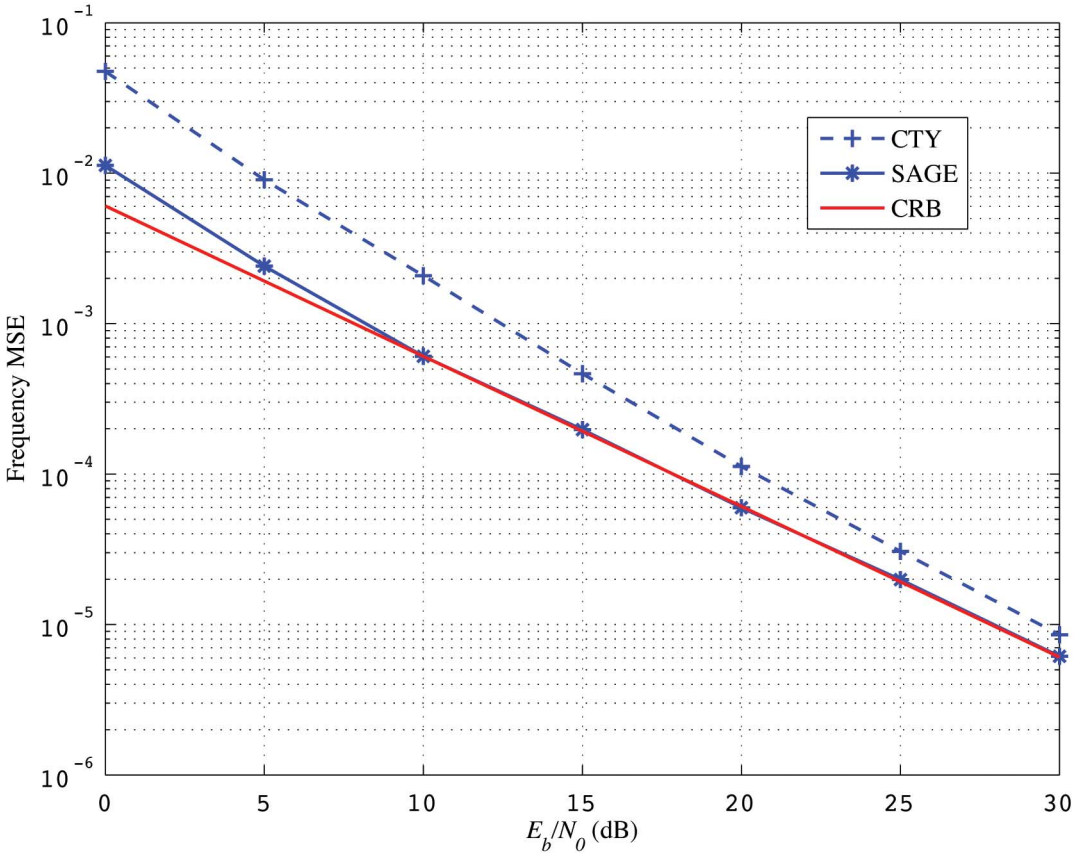

Fig. 21. ***Accuracy of CTY and SAGE as function of*** E_b/N_0.

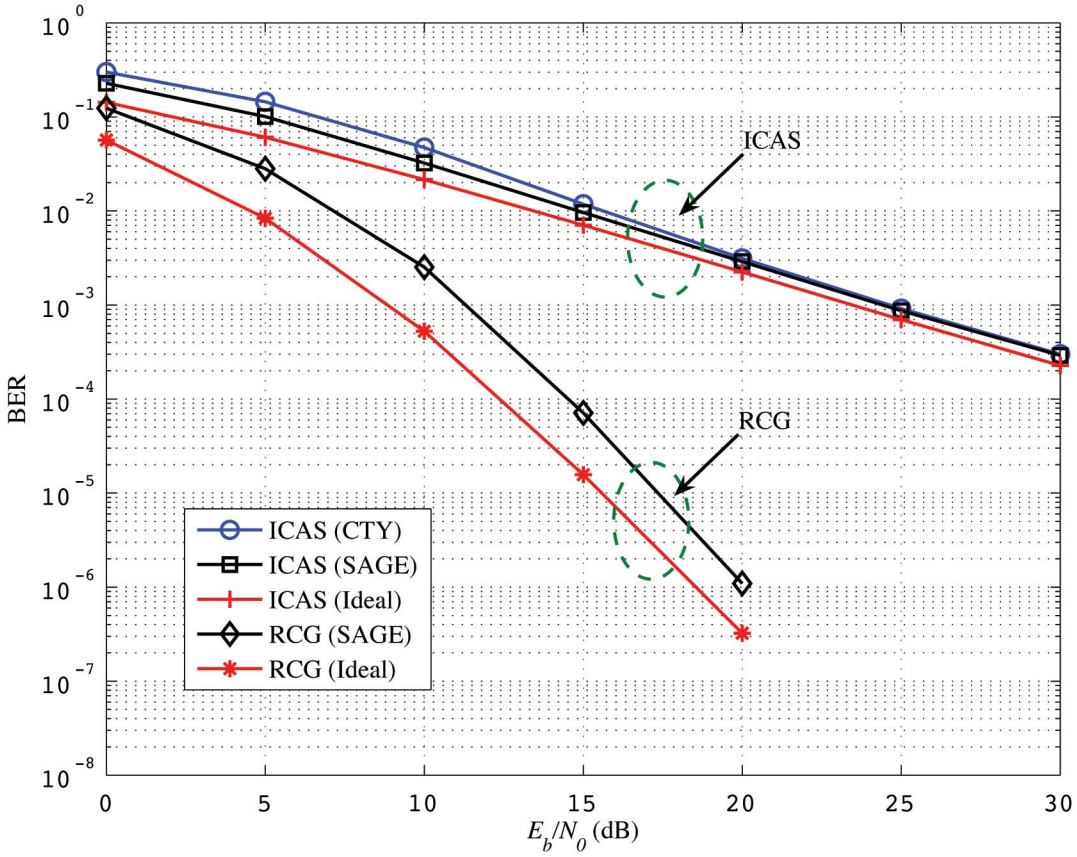

Fig. 22. ***BER performance with frequency estimates provided by CTY and SAGE.***

comparison. These results indicate that the SAGE algorithm provides unbiased estimate over interval $|\varepsilon_1| < 0.5$.

Fig. 21 compares SAGE and CTY in terms of frequency MSE as a function of E_b/N_0 in the case of three active users. The lowest line represents the CRB for frequency estimation in a quasi-synchronous OFDMA uplink transmission as derived in [53] and is shown as a benchmark. We see that SAGE achieves the CRB for $E_b/N_0 > 10$ dB. The CTY exhibits good performance at high SNR values, yet some degradations are observed with respect to SAGE.

Fig. 22 shows the impact of residual frequency estimation errors on the bit-error rate (BER) performance in case of three active users. Here, we assume that the BS performs frequency estimation whereas frequency correction is made at each user's side based on instructions transmitted via a downlink control channel. To take this into account, during the data section of the frame we replace ε_m by the corresponding residual error $\varepsilon_m - \widehat{\varepsilon}_m$. This approach is useful to highlight the BER degradation produced by residual synchronization errors while avoiding any loss that might derive from compensating CFOs directly at the BS. Two different scenarios are considered. The first one adopts an interleaved CAS (ICAS) with frequency estimates provided by either SAGE or CTY. In the second scenario, subcarriers are dynamically assigned to active users according to the rate-craving greedy (RCG) algorithm, which was proposed in [67] to maximize the transmission rate of an OFDMA system. In the latter case, only the SAGE algorithm is considered since CTY can exclusively be used in conjunction with an interleaved CAS. Channel estimates are computed as indicated in [53] and exploited for ZF equalization. The ideal curves have been obtained assuming perfect knowledge of channel responses and frequency offsets and are shown as benchmarks. We see that CTY is only marginally worse than SAGE. As expected, RCG leads to a dramatic improvement of the error rate performance as compared to ICAS. The reason is that dynamic subcarrier assignment provides the system with multiuser diversity, which increases the asymptotic slope of the BER curves [51].

B. Performance of Frequency Correction Schemes

It was shown in Section VII that frequency correction in the uplink can be accomplished directly at the BS using advanced signal processing techniques like HL or CTYH. The performance of these schemes are now assessed in a fully loaded system where four users with equal average

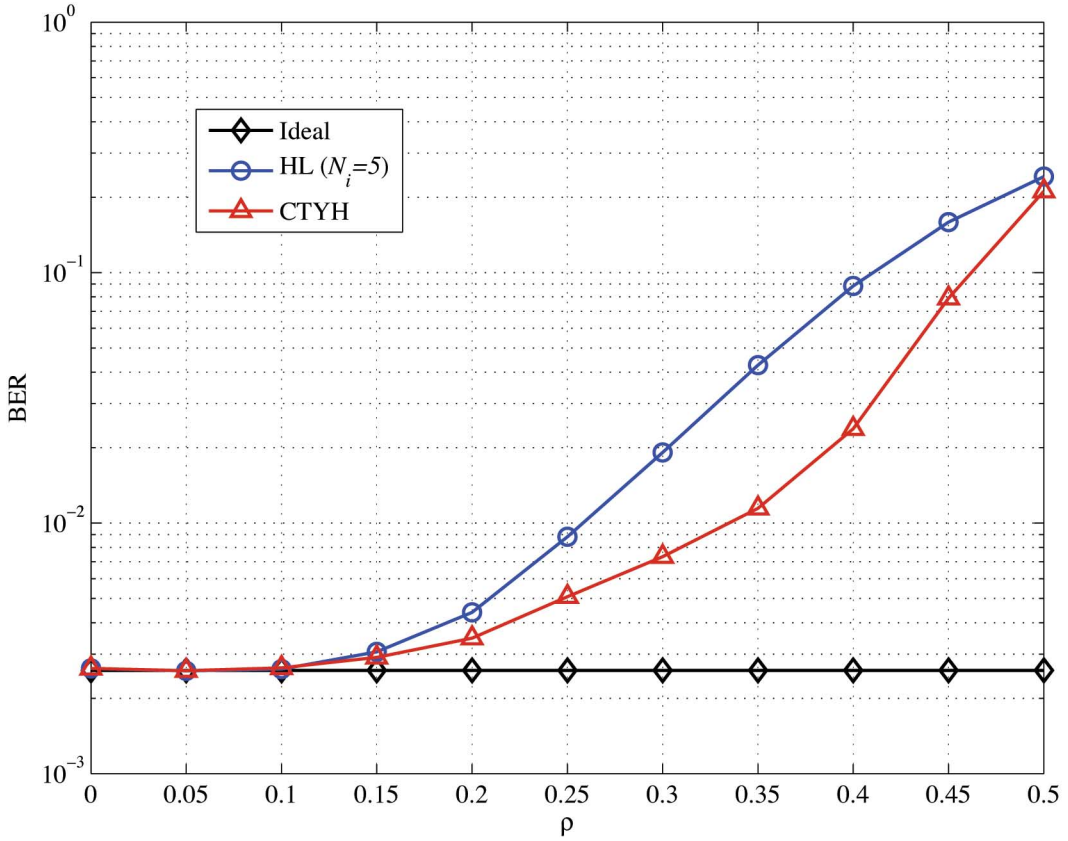

Fig. 23. *BER performance of HL and CTYH as function of ρ with $E_b/N_0 = 20$ dB in case of perfect frequency and channel state information.*

power are simultaneously active. Their CFOs are modeled as $\boldsymbol{\varepsilon} = \rho[1, -1, 1, -1]^T$, where ρ is a deterministic parameter belonging to interval [0, 0.5] and known as *frequency attenuation factor* [60]. Five iterations are performed by HL, while CTYH employs the decorrelating matrix $\widehat{\mathbf{V}}_{LDD}$ given in (102).

Fig. 23 illustrates the BER versus ρ for $E_b/N_0 = 20$ dB. The results have been obtained under the assumption of perfect frequency and channel state information at the BS. The reason is that in this way we can highlight the distinct impact of imperfect frequency correction on the system performance while avoiding the loss induced by inaccuracy in the CFO estimates. The curve labeled "ideal" has been obtained in the absence of any synchronization error and refers to a system where uplink signals at the DFT output are perfectly orthogonal and no interference is thus present. We see that the BER degrades with ρ due to the increased amount of MAI. As anticipated, CTYH outperforms HL as it can totally suppress interference at the expense of some noise enhancement.

Finally, Fig. 24 shows the BER of the considered schemes versus E_b/N_0 with $\rho = 0.2$. In contrast with Fig. 23, frequency correction is accomplished here using channel estimates provided by SAGE. Again, we see that CTYH outperforms HL. In particular, at an error rate of 10^{-2}, the loss of CTYH with respect to the ideally synchronized system is approximately 4 dB. As for HL, it exhibits an error floor at high SNR values.

IX. CONCLUSION AND FUTURE WORK

An overview of the current status in the field of timing and frequency synchronization for OFDMA applications, from the very basics to up-to-date advances, was provided. This research area is attracting considerable attention and significant efforts are underway to develop new schemes with improved performance and affordable complexity.

While synchronization techniques for downlink transmissions can directly be taken from the vast literature devoted to single-user OFDM, only a few solutions are currently available for the uplink case. The main difficulty is that the uplink signals arriving at the BS are characterized by different synchronization parameters. The latter cannot be estimated with the same methods employed in the downlink because each user must be separated from the others before the synchronization process can be started. The separation method is closely related to the particular carrier assignment scheme adopted in the system. With

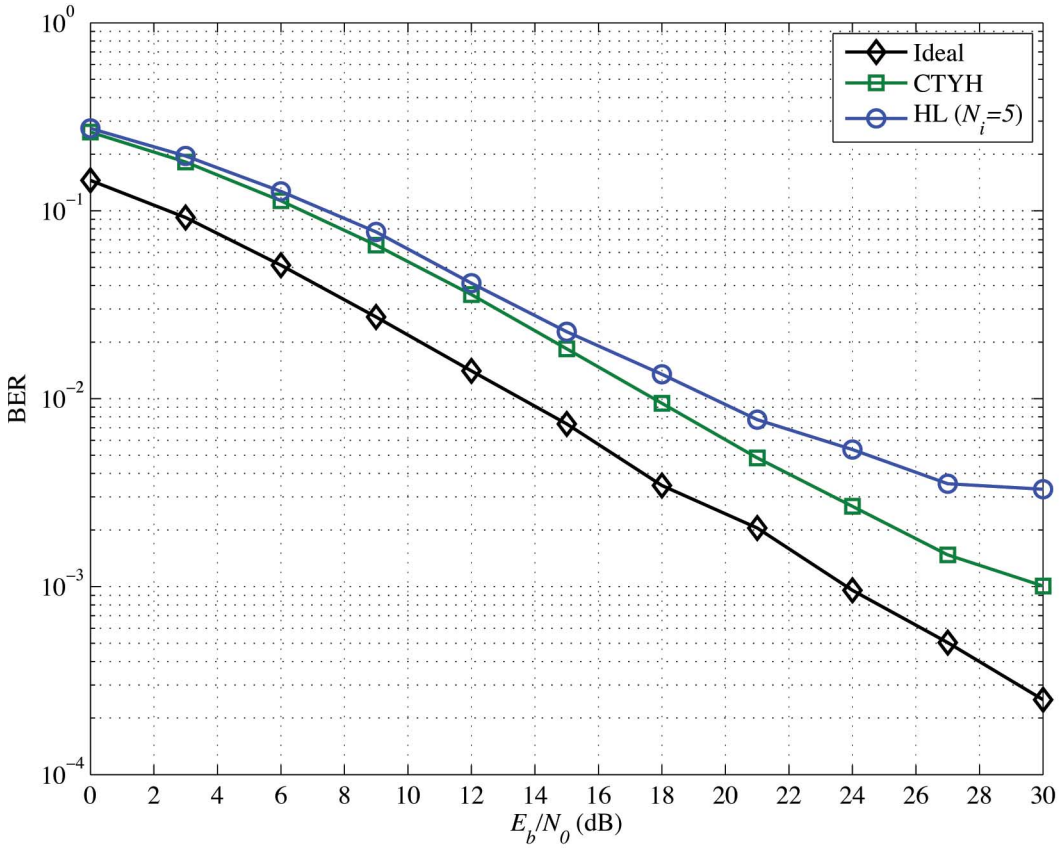

Fig. 24. ***BER performance of HL and CTYH as function of E_b/N_0 with $\rho = 0.2$ and frequency estimates provided by SAGE algorithm.***

subband CAS, the uplink signals are easily separated through a bank of bandpass filters and synchronization is next performed independently over each separated signal. The situation is more challenging in the presence of an interleaved or generalized CAS. In both cases, we showed how CFO estimates can be obtained using the SAGE algorithm, while CTYH reveals a promising approach to compensate the frequency errors directly at the BS.

The most challenging task is represented by timing synchronization, for which no effective solution is currently available in the open literature. Extending the length of the CP to comprehend both the channel delay spread and the two-way propagation delay results in a quasi-synchronous system and provides intrinsic protection against timing errors. Unfortunately, this approach cannot be used in recently standardized wireless metropolitan area networks. The reason is that in these applications the cell radius may be as large as 25 km and the propagation delay is thus comparable with the length of the OFDMA blocks. Novel ideas must be found to solve the timing synchronization problem in these scenarios. We hope that the material presented in this paper may stimulate those who are already working in the field, while providing the necessary background to those who are approaching this active research area. ■

REFERENCES

[1] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COM-19, no. 10, pp. 628–634, Oct. 1971.

[2] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[3] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, no. 7, pp. 665–675, Jul. 1985.

[4] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, no. 5, pp. 611–640, May 2000.

[5] *Radio Broadcasting Systems: Digital Audio Broadcasting to Mobile, Portable and Fixed Receivers*, ETS 300 401, Eur. Telecommun. Standard, 1995, ETSI.

[6] *Digital Video Broadcasting (DVB-T); Frame Structure, Channel Coding, Modulation for Digital Terrestrial Television*, ETS 300 744, Eur. Telecommun. Standard, 1997, ETSI.

[7] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Higher-Speed Physical Layer Extension in the 5 GHz Band*, IEEE802.11a, 1999.

[8] H. Sari and G. Karam, "Orthogonal frequency-division multiple access and its application to CATV networks," *Eur.*

Trans. Commun., vol. 45, pp. 507–516, Nov.–Dec. 1998.

[9] *Interaction Channel for Digital Terrestrial Television (RCT) Incorporating Multiple Access OFDM*, ETSI DVB RCT, Mar. 2001.

[10] *Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems-Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2–11 GHz*, IEEE P802.16a/D3-2001, Mar. 2002.

[11] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun*, vol. 43, no. 2, pp. 191–193, Feb. 1995.

[12] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1997.

[13] H. Wang and B. Chen, "A comparison of the subcarrier allocation schemes for uplink OFDMA systems," in *Proc. CISS'2004*, Princeton, NJ, Mar. 2004.

[14] M. Speth, S. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broadband systems using OFDM, Part I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.

[15] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.

[16] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.

[17] U. Lambrette, M. Speth, and H. Meyr, "OFDM burst frequency synchronization by single carrier training data," *IEEE Commun. Lett.*, vol. 1, no. 2, pp. 46–48, Mar. 1997.

[18] F. Tufvesson, O. Edfors, and M. Faulkner, "Time and frequency synchronization for OFDM using PN-sequence preamble," in *Proc. IEEE Vehicular Technology Conf.*, Sep. 1999, vol. 4, pp. 2202–2207.

[19] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Commun. Lett.*, vol. 3, no. 3, pp. 75–77, Mar. 1999.

[20] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Timing recovery for OFDM transmission," *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, pp. 2278–2291, Nov. 2000.

[21] H. K. Song, Y. H. You, J. H. Park, and Y. S. Cho, "Frequency-offset synchronization and channel estimation for OFDM-based transmission," *IEEE Commun. Lett.*, vol. 4, no. 3, pp. 95–97, Mar. 2000.

[22] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: Algorithms," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2486–2494, Dec. 2001.

[23] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: Analysis," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2495–2503, Dec. 2001.

[24] T. Keller, L. Piazzo, P. Mandarini, and L. Hanzo, "Orthogonal frequency division multiplex synchronization techniques for frequency-selective fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 6, pp. 999–1008, Jun. 2001.

[25] H. Minn, V. K. Bhargava, and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 822–839, Jul. 2003.

[26] K. Shi and E. Serpedin, "Coarse frame and carrier synchronization of OFDM systems: A new metric and comparison," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1271–1284, Jul. 2004.

[27] F. Classen and H. Meyr, "Frequency synchronization algorithms for OFDM systems suitable for communication over frequency selective fading channels," in *Proc. IEEE Vehicular Technology Conf.*, Aug. 1994, vol. 3, pp. 1655–1659.

[28] F. Daffara and O. Adami, "A novel carrier recovery technique for orthogonal multicarrier systems," *Eur. Trans. Telecommun.*, vol. 7, pp. 323–334, Jul./Aug. 1996.

[29] J. J. van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of timing and frequency offset in OFDM systems," *IEEE Trans. Signal Proc.*, vol. 45, no. 7, pp. 1800–1805, Jul. 1997.

[30] N. Lashkarian and S. Kiaei, "Class of cyclic-based estimators for frequency-offset estimation of OFDM systems," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2139–2149, Dec. 2000.

[31] J. Lei and T. Ng, "A consistent OFDM carrier frequency offset estimator based on distinctively spaced pilot tones," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 588–599, Mar. 2004.

[32] F. Daffara and A. Chouly, "Maximum likelihood frequency detectors for orthogonal multicarrier systems," in *Proc. ICC'93*, Jun. 1993, pp. 766–771.

[33] M. Morelli, A. N. D'Andrea, and U. Mengali, "Feedback frequency synchronization for OFDM applications," *IEEE Commun. Lett.*, vol. 5, no. 1, pp. 28–30, Jan. 2001.

[34] T. M. Schmidl and D. C. Cox, "Blind synchronization for OFDM," *Elect. Lett.*, vol. 33, no. 2, pp. 113–114, Feb. 1997.

[35] H. Liu and U. Tureli, "A high-efficiency carrier estimator for OFDM communications," *IEEE Commun. Lett.*, vol. 2, no. 4, pp. 104–106, Apr. 1998.

[36] U. Tureli, H. Liu, and M. D. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1459–1461, Sep. 2000.

[37] G. Santella, "A frequency and symbol synchronization system for OFDM signals: Architecture and simulation results," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 254–275, Jan. 2000.

[38] X. Ma, C. Tepedelenlioglu, G. B. Giannakis, and S. Barbarossa, "Non-data-aided carrier offset estimator for OFDM with null subcarriers," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2504–2515, Dec. 2001.

[39] M. Ghogho, A. Swami, and G. B. Giannakis, "Optimized null-subcarrier selection for CFO estimation in OFDM over frequency selective fading channels," in *Proc. IEEE Globecom 2001*, Nov. 2001, vol. 48, pp. 202–206.

[40] H. Bolcskei, "Blind estimation of symbol timing and carrier frequency offset in wireless OFDM systems," *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 988–999, Jun. 2001.

[41] U. Tureli, D. Kivanc, and H. Liu, "Experimental and analytical studies on high-resolution OFDM carrier frequency offset estimator," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 629–643, Mar. 2001.

[42] M. Ghogho and A. Swami, "Semi-blind frequency offset synchronization for OFDM," in *Proc. ICASSP 2002*, May 2002, vol. 3, pp. 2333–2336.

[43] M. Ghogho and A. Swami, "Blind frequency-offset estimator for OFDM systems transmitting constant-modulus symbols," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 343–345, Aug. 2002.

[44] F. Yang, K. H. Li, and K. C. Teh, "A carrier frequency offset estimator with minimum output variance for OFDM systems," *IEEE Commun. Lett.*, vol. 8, no. 11, pp. 677–679, Nov. 2004.

[45] H. Nogami and T. Nagashima, "A frequency and timing period acquisition technique for OFDM systems," in *Proc. Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 1995, vol. 3, pp. 1010–1015.

[46] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591–598, Sep. 1974.

[47] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE J. Select. Areas Commun.*, vol. 20, no. 2, pp. 474–486, Feb. 2002.

[48] J. J. van de Beek, P. O. Börjesson, M. L. Boucheret, D. Landström, J. M. Arenas, O. Ödling, C. Östberg, M. Wahlqvist, and S. K. Wilson, "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Select. Areas Commun.*, vol. 17, no. 11, pp. 1900–1914, Nov. 1999.

[49] Z. Cao, U. Tureli, and Y. D. Yao, "Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1585–1594, Sep. 2004.

[50] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Mar. 1986, vol. 34, no. 3, pp. 243–258.

[51] J. Li, H. Kim, Y. Lee, and Y. Kim, "A novel broadband wireless OFDMA scheme for downlink in cellular communications," in *Proc. IEEE WCNC 2003*, Mar. 2003, pp. 1907–1911.

[52] M. Morelli, "Timing and frequency synchronization for the uplink of an OFDMA system," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 296–306, Feb. 2004.

[53] M. O. Pun, M. Morelli, and C.-C. J. Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.

[54] M. O. Pun, M. Morelli, and C.-C. J. Kuo, "Iterative detection and frequency synchronization for OFDMA uplink transmissions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 629–639, Feb. 2007.

[55] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[56] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, vol. 39, pp. 1–38, 1977.

[57] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sep. 2000.

[58] J. Choi, C. Lee, H. W. Jung, and Y. H. Lee, "Carrier frequency offset compensation for

uplink of OFDM-FDMA systems," *IEEE Commun. Lett.*, vol. 4, no. 12, pp. 414–416, Dec. 2000.

[59] S. Verdù, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[60] D. Huang and K. B. Letaief, "An interference-cancellation scheme for carrier frequency offsets correction in OFDMA systems," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1155–1165, Jul. 2005.

[61] A. M. Tonello, "Multiuser detection and turbo multiuser decoding for asynchronous multitone multiple access systems," in *Proc. IEEE Vehicular Technology Conf.*, Vancouver, BC, Canada, Sep. 2002, pp. 970–974.

[62] Z. Cao, U. Tureli, Y. D. Yao, and P. Honan, "Frequency synchronization for generalized OFDMA uplink," in *Proc. Globecom 2004*, Dallas, TX, 2004, pp. 1071–1075.

[63] A. Ghosh, D. R. Wolter, J. G. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: Current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, pp. 129–136, Feb. 2005.

[64] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 WirelessMAN," *Intel Technol. J.*, vol. 8, pp. 201–212, Aug. 2004.

[65] *Channel Models for Fixed Wireless Applications*, IEEE 802.16.3c 01/29r4.

[66] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed. Upper Saddle River, NJ: Prentice–Hall, 2002.

[67] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150–1158, Nov. 2003.

ABOUT THE AUTHORS

Michele Morelli (Member, IEEE) received the Laurea (*cum laude*) degree in electrical engineering, the "Premio di Laurea SIP," and the Ph.D. degree from the University of Pisa, Pisa, Italy, in 1991, 1992, and 1995, respectively.

From 1992 to 1995, he was with the Department of Information Engineering, University of Pisa. In September 1996, he joined the Centro Studi Metodi e Dispositivi per Radiotrasmissioni (CSMDR) of the Italian National Research Council (CNR), Pisa, where he held the position of Research Assistant. Since 2001, he has been with the Department of Information Engineering, University of Pisa, where he is currently an Associate Professor of Telecommunications. His research interests are in wireless communication theory, with emphasis on synchronization algorithms and channel estimation in multiple-access communication systems.

C.-C. Jay Kuo (Fellow, IEEE) received the B.S. degree from National Taiwan University, Taipei, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is Director of the Signal and Image Processing Institute and Professor of electrical engineering, computer science, and mathematics at the University of Southern California (USC), Los Angeles. His research interests are in the areas of digital image/video analysis and modeling, multimedia data compression, communication and networking, and biological signal/image processing. He has guided about 80 students to their Ph.D. degrees and supervised 20 postdoctoral research fellows. He is coauthor of about 130 journal papers, 700 conference papers, and seven books. He delivered more than 300 invited lectures in conferences, research institutes, universities, and companies.

Dr. Kuo is Editor-in-Chief for the *Journal of Visual Communication and Image Representation* and has served as Editorial Board member for about ten international journals. He received the National Science Foundation Young Investigator Award and Presidential Faculty Fellow Award in 1992 and 1993, respectively. He is a Fellow of SPIE.

Man-On Pun (Member, IEEE) received the B.Eng. (Hon.) degree in electronic engineering from the Chinese University of Hong Kong in 1996, the M.Eng. degree in information sciences from University of Tsukuba, Japan, in 1999, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2006, respectively.

Since August 2006, he has been a Postdoctoral Research Fellow with the Department of Electrical Engineering, Princeton University, Princeton, NJ. From 1999 to 2001, he was with the Sony Corporation, Tokyo, Japan. His current research interests are in the area of statistical signal processing for wireless communications.

Dr. Pun received the best student paper award from the IEEE Vehicular Technology Fall Conference (VTC-Fall), Montreal, Canada, in 2006, and a number of scholarships including the Japanese Government (Monbusho) Scholarship, the Sir Edward Youde Memorial fellowship for Overseas Studies, and the Croucher postdoctoral fellowship.