

Reliable OFDM Receiver with Ultra-Low Resolution ADC

Hanqing Wang, *Student Member, IEEE*, Wan-Ting Shih, Chao-Kai Wen, *Member, IEEE*,
and Shi Jin, *Senior Member, IEEE*

Abstract—The use of low-resolution analog-to-digital converters (ADCs) can significantly reduce power consumption and hardware cost. However, their resulting severe nonlinear distortion makes reliable data transmission challenging. For orthogonal frequency division multiplexing (OFDM) transmission, the orthogonality among subcarriers is destroyed. This invalidates conventional OFDM receivers relying heavily on this orthogonality. In this study, we move on to quantized OFDM (Q-OFDM) prototyping implementation based on our previous achievement in optimal Q-OFDM detection. First, we propose a novel Q-OFDM channel estimator by extending the generalized Turbo (GTurbo) framework formerly applied for optimal detection. Specifically, we integrate a type of robust linear OFDM channel estimator into the original GTurbo framework and derive its corresponding extrinsic information to guarantee its convergence. We also propose feasible schemes for automatic gain control, noise power estimation, and synchronization. Combined with the proposed inference algorithms, we develop an efficient Q-OFDM receiver architecture. Furthermore, we construct a proof-of-concept prototyping system and conduct over-the-air (OTA) experiments to examine its feasibility and reliability. This is the first work that focuses on both algorithm design and system implementation in the field of low-resolution quantization communication. The results of the numerical simulation and OTA experiment demonstrate that reliable data transmission can be achieved.

Index Terms—Low-precision ADC, OFDM, Turbo signal recovery, prototyping system.

I. INTRODUCTION

An essential pursuit of current research into future wireless communication systems is to achieve highly reliable data transmission using devices with ultra-low cost and power consumption. Low-resolution analog-to-digital converter (ADC) is the most representative one, whose application in communication receivers has been extensively explored by the academia and industry. This emphasis on low-resolution ADCs is motivated by their obvious advantages in reducing hardware cost and power consumption. Specifically, according to the power consumption expression of ADCs given by [1],

$$P_{\text{ADC}} \propto 2^b \times F_s,$$

H. Wang and S. Jin are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, P. R. China. P (e-mail: hqwanglyt@seu.edu.cn; jinshi@seu.edu.cn).

W.-T. Shih is with the Institute of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: sydney2317076@gmail.com).

C.-K. Wen is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw).

where b denotes its quantization precision and F_s denotes its sampling frequency, the reduction in the quantization precision b leads to an exponential decrease in the power consumption of ADCs. The hardware cost and implementation complexity of low-resolution ADCs are also relatively low [2]. Furthermore, storing and transferring data digitally represented by fewer bits are easier, which relaxes the quality requirement for radio frequency (RF) and interface circuits, and also reduces the power consumption of these components.

The feasibility of using low-resolution ADCs has been initially demonstrated by studies on the capacity or achievable rate degradation caused by such devices [3]–[9]. As indicated by the capacity derivation for additive white Gaussian noise (AWGN) channels in [4], only 10%-20% of capacity is lost under the quantization precision of as few as 2–3 bits. The research in [7] revealed that in multiple-input and multiple-output (MIMO) systems with mixed-precision ADCs, equipping only a small proportion of high-resolution ADCs can closely approximate the achievable rate of conventional architecture with all full-resolution ADCs. These results inspired subsequent studies into signal recovery techniques for dealing with strong nonlinear distortion caused by coarse quantization. Various methods have been proposed. They include linear receiver [10]–[12], projected gradient method based algorithms [13], [14], fast adaptive shrinkage/thresholding algorithm (FASTA) based method [15], coding theory-based approach [16], supervised learning based method [17], and message-passing approaches, including generalized approximate message passing (GAMP) [18]–[21] and its extension bilinear GAMP (BiGAMP) [22] and parametric BiGAMP [23], vector AMP [24], and generalized Turbo (GTurbo) [25].

In this study, we focus on orthogonal frequency division multiplexing (OFDM) systems with ultra-low resolution ADCs (i.e., 1–2 bits) at the receiver. We refer to these systems as quantized OFDM (Q-OFDM) systems. OFDM is a well-developed multicarrier technology for decomposing convolutive multipath channels into a bank of parallel flat-fading subchannels to facilitate wideband transmission with high spectral efficiency [26]. Hence, OFDM is applied as the physical layer waveform for variety of modern high-rate wireless systems, ranging from Long-Term Evolution and its evolutions to the standards such as IEEE 802.11ad [27]. It is also one of the strong waveform contenders for the 5th generation (5G) air interface [28], [29]. This research considers the most fundamental single-stream transmission case. Using ultra-low-resolution quantizers results in only minimal performance loss when many received streams, which average the quantization

noise, are available for combination [10]. However, in case of few received streams, in which diversity gain from a large number of received streams is unavailable, reliable reception is a challenging task in the presence of coarse quantization. Single stream is considered as an extreme case. In addition, this single stream input-output relation does not solely imply the single-antenna case. It can also be treated as an equivalent representation for multi-antenna systems with one transmitted and received data stream using analog beamforming when tackling data detection and channel estimation, as applied in [23], [30].

The conventional OFDM receiver proposed for the high-resolution quantization case performs by converting the time-domain received signal into the frequency-domain. The frequency-domain observations can be regarded as a bank of orthogonal flat-fading subchannels or subcarriers. Data detection and channel estimation are then implemented in parallel for each flat-fading subchannel. However, in the presence of coarse quantization, observations among subcarriers are not orthogonal anymore. Moreover, high correlations are exhibited in the frequency domain, and severe inter-carrier interference occurs. This conventional strategy, which is derived from orthogonality, can suffer from certain performance loss for Q-OFDM cases. Such occurrences have motivated us to propose advanced receiver strategies. Our pioneering work [25] proposed a powerful Q-OFDM detector based on the GTurbo signal recovery principle. This detector proved theoretically optimal in the sense of minimum mean-square error (MMSE) with ideal channel state information (CSI). In addition, an efficient pipelined and folding hardware architecture of the GTurbo algorithm was proposed in [31], which suggests that hardware friendly implementation of GTurbo-based algorithms is achievable.

In this work, we move on to the next step to realize the Q-OFDM prototyping system for reliable data transmission. To this end, several issues are waiting for being tackled. One important issue is channel estimation as frequency-domain channel responses are necessary for the Q-OFDM detector. Considering the algorithmic power of GTurbo in Q-OFDM detection, we develop a novel Q-OFDM channel estimator that is based on the GTurbo architecture. The main difference or challenge is that accurately characterizing the prior distribution of frequency-domain channel responses, which are time-variant and correlated, is difficult. Then the MMSE estimate of frequency-domain channel responses required in the original GTurbo framework is infeasible. To tackle this issue, we embed a type of robust OFDM channel estimator used for OFDM systems with high-resolution quantization into the GTurbo framework. In this way, our approach can be viewed as performing linear channel estimation followed by a certain linear approximation that is updated via iterations. As shown by the numerical results, the proposed scheme outperforms its counterparts, such as the conventional method, FASTA [15], expectation-maximization Gaussian-mixture GAMP (EM-GM-GAMP) [32], and Bussgang's theorem-based method [33]. In addition, the developments of detection and channel estimation based on the same algorithm architecture allow for sharing an identical circuit

module when designing specific circuits for the proposed algorithms.

For prototyping implementation, we also need to tackle several critical issues, including adjusting the input signal amplitude within a certain range (i.e., automatic gain control (AGC)), estimating the noise power, and identifying the starting point of each radio frame. AGC and noise estimation require the amplitude information of the received signal. As a low-resolution ADC eliminates its amplitude information, we introduce another high-resolution ADC with a considerably low sampling rate. With a power accumulator, signal and noise power can be measured. Moreover, we verify that correlation-based synchronization works well for coarsely quantized received signals. Combining the above issues with the proposed GTurbo-based algorithms, we put forward an efficient Q-OFDM receiver architecture and construct a prototyping system on the basis of such architecture. To the best of our knowledge, our current work is the first one which covers both theory (algorithm design) and practice (prototyping implementation and over-the-air (OTA) experiments) in the field of communication with low-resolution ADCs.

OTA experiments are conducted for justifying the reliability of the proposed architecture. Results reveal that reliable OFDM transmissions of 4-QAM symbols with 1-bit quantization per real dimension and 16-QAM symbols with 2-bit quantization per real dimension can be achieved. In addition, the well-known guideline for AWGN channels, which claims that the ADC precision per real dimension equal to the logarithm of half the modulation order is sufficient for reliable signal recovery [4], is also valid for OFDM transmissions. However, this guideline is not intuitively true for OFDM transmission, given that the inverse discrete Fourier transform (IDFT) operation produces continuous time-domain transmitted symbols instead of discrete constellation points. Thus, higher ADC precision is required. This interesting finding verifies the powerful capacity of the proposed algorithms and receiver framework.

This work, along with [25], forms a complete picture of the reliable Q-OFDM receiver for the one-stream transmission case. These studies help us clarify many essential technical issues, thereby establishing the basis of the future extension to multi-antenna and multi-user cases. Specifically, novel algorithm architectures potential for the single-user MIMO case can be developed by extending current GTurbo-based framework, such as [34]. The signal/noise power measurement, which is essential for algorithms, and the implementation of AGC, can be performed with the assistance of a high-resolution ADC. For MIMO systems, applying mixed ADC architectures [20], namely, equipping few high-resolution ADCs for improving detection and estimation accuracy, can also help signal/noise power measurement. Moreover, in the presence of coarse quantization, conventional correlation-based synchronization remains valid.

The rest of the paper is organized as follows. Section II models the Q-OFDM system, and highlights the inference problems and the system implementation issues considered in this study. Section III proposes algorithms based on GTurbo framework. Section IV evaluates and compares the performances of the proposed algorithms through the numerical

simulation results. Section V presents the proposed receiver architecture, the construction of the prototyping system and the OTA experiment results. Section VI discusses our ideas of future extensions. Finally, Section VII concludes this paper.

Notations: This paper uses lowercase and uppercase bold-face letters to represent vectors and matrices, respectively. For vector \mathbf{a} , a_j denotes the j -th entry of \mathbf{a} , $\|\mathbf{a}\|$ denotes the l_2 norm of the vector \mathbf{a} , and the operator $\text{diag}(\mathbf{a})$ denotes the diagonal matrix with diagonal elements as the entries of \mathbf{a} . For matrix \mathbf{A} , the operator $\text{Diag}(\mathbf{A})$ denotes the vector composed of the diagonal elements of matrix \mathbf{A} . Moreover, the real and imaginary parts of a complex scalar a are represented by a^R and a^I , respectively. For random vectors \mathbf{a} and \mathbf{b} , the notation $\mathbf{R}_{\mathbf{ab}} = \mathbb{E}\{\mathbf{ab}^H\}$ represents their cross-correlation matrix. The distribution of a proper complex Gaussian random variable z with mean μ and variance ν is denoted by $\mathcal{CN}(z; \mu, \nu)$. Similarly, $\mathcal{N}(z; \mu, \nu)$ denotes the probability density function of a real Gaussian random variable z with mean μ and variance ν . The probability density function and cumulative distribution function for the standard Gaussian distribution are denoted by $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ and $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-\frac{t^2}{2}}dt$ respectively, and the Q function is defined as $Q(z) = 1 - \Phi(z)$.

II. SYSTEM MODEL

We consider an OFDM system that comprises N orthogonal subcarriers, where N_d of them corresponding to frequency $-\frac{N_d F_s}{2}, \dots, -\frac{F_s}{N}, \frac{F_s}{N}, \dots, \frac{N_d F_s}{2}$ (F_s denotes the sampling frequency) are assigned to carry data symbols. Meanwhile, the rest of the $N - N_d$ subcarriers are reserved as guard bands, which are padded by zero signals to reduce signal leakage into the adjacent frequency bands. Hereinafter, we use $\mathcal{X} = \{1, 2, \dots, N\}$ to denote the index set of all subcarriers, and $\mathcal{X}_d \subseteq \mathcal{X}$ to denote the index subset of the N_d subcarriers used for data transmission. In addition, the subscript “ d ” for the N -dimension vector and a matrix with N rows (e.g., \mathbf{a}_d and \mathbf{A}_d) represent the N_d -dimension subvector that comprises the elements of \mathbf{a} with index \mathcal{X}_d and the submatrix that comprises the rows of \mathbf{A} with index \mathcal{X}_d , respectively.

The signaling process at the OFDM transmitter is illustrated in Fig. 1.¹ The information bits to be transmitted, which were generated by the data source, are first encoded using a channel code technique, such as the turbo code. Then, the modulator maps the coded bits into the corresponding constellation symbols based on the selected modulation method, such as M -ary quadrature amplitude modulation (M -QAM), and each modulated symbol is assigned into one of the N_d subcarriers of the OFDM symbol. We denote the vector of the frequency-domain constellation symbols as $\mathbf{s} \in \mathbb{C}^{N \times 1}$, where $s_j = 0$ for $j \in \mathcal{X} \setminus \mathcal{X}_d$. Subsequently, the frequency-domain OFDM symbol \mathbf{s} is transformed into the time-domain using N -point IDFT. The CP with the length L_{cp} greater than the length of channel impulse response is appended at the beginning of each OFDM symbol. This contributes to convert inter-symbol

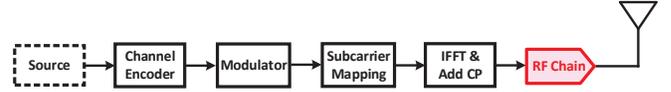


Fig. 1: Block diagram of an OFDM transmitter.

interference (ISI) channel caused by the multipath effect into a bank of flat-fading subchannels. The obtained sequence of the time-domain baseband samples is finally up-converted by the RF chain and transmitted over the wireless channel.

At the receiver, the RF signal received by the antenna is first down-converted into the analog baseband. Then the analog baseband signal is sampled with frequency F_s and quantized using an ultra-low resolution ADC. After CP removal and under the assumption of an ideal synchronization, the coarsely quantized time-domain samples of an OFDM symbol can be expressed as

$$\mathbf{q}_D = \mathcal{Q}_c[\sqrt{g_{AGC}}(\mathbf{F}^H \text{diag}(\mathbf{h})\mathbf{s} + \mathbf{n})], \quad (1)$$

where \mathbf{F} denotes the normalized DFT matrix with (m, n) -th entry being $\frac{1}{\sqrt{N}}e^{-2\pi j(n-1)(m-1)/N}$; $\mathbf{n} \in \mathbb{C}^{N \times 1}$ denotes the AWGN vector with distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I})$; \mathbf{h} is the frequency-domain channel response vector, which is obtained by performing DFT on the tapped delay line representation of the ISI channel denoted by $\mathbf{h}_t = [h_{t,1}, h_{t,2}, \dots, h_{t,N}]^T$ with $h_{t,j} = 0$ for $(L+1) \leq j \leq N$ and L being the maximum number of channel taps; g_{AGC} is the gain of the AGC unit used to adjust the amplitude of the ADC input within a certain range; $\mathcal{Q}_c(\cdot)$ denotes the elementwise mapping function of the complex-valued quantizer, which comprises two parallel real-valued B -bit quantizers $\mathcal{Q}(\cdot)$ that independently quantize the real and imaginary parts of each analog input sample. For example, the quantization of a scalar analog sample y will be

$$\mathcal{Q}_c(y) = \mathcal{Q}(y^R) + j\mathcal{Q}(y^I),$$

where quantizer $\mathcal{Q}(\cdot)$ outputs the discrete level c_b when the real-valued input y^R or y^I is within the interval $(r_{b-1}, r_b]$, with $-\infty = r_0 < r_1 < \dots < r_{2B-1} < r_{2B} = \infty$ being the thresholds of the quantizer. Notably, (1) can also be used as an equivalent input-output relation of OFDM systems with multi-antenna and analog beamforming when tackling data detection and channel estimation. Therefore, \mathbf{h} denotes the beamformed channel response in frequency domain, like the way in [30].

We design the baseband processing architecture at the receiver to recover the original information bits from each coarsely quantized received OFDM symbol \mathbf{q}_D with error rate given the channel estimate. Hereinafter we use the notation $\hat{\mathbf{h}}$ to denote the product of the estimate of \mathbf{h} and $\sqrt{g_{AGC}}$. For example, the conventional OFDM receiver first transforms the \mathbf{q}_D into the frequency-domain, which yields $\tilde{\mathbf{q}} = \mathbf{F}\mathbf{q}_D$, followed by the one-tap equalizer, which computes the estimate of s_j for $j \in \mathcal{X}_d$ by \tilde{q}_j/\hat{h}_j . However, this conventional strategy fails to work well in Q-OFDM systems because the orthogonality among subcarriers is destroyed. The iterative procedure for the MMSE estimation of \mathbf{s} was proposed in [25] under the assumption of a perfect CSI, noise

¹We only specify here the signaling process corresponding to the OFDM symbols used for data transmission. For OFDM symbols for other purposes, specific baseband symbols (e.g., pilot symbols and synchronization sequences) are directly assigned into their corresponding subcarriers without the need of channel encoding and constellation mapping.

power, and synchronization. Determining how to realize those assumptions comprises the content of this paper. In particular,

1) *Channel Estimation*: To obtain CSI, we periodically assign an OFDM symbol that contains pilot symbols at the subcarriers with index \mathcal{X}_d . The vector of the frequency-domain pilot symbols is denoted by $\mathbf{p} \in \mathbb{C}^N$, where $p_j = 0$ for $j \in \mathcal{X} \setminus \mathcal{X}_d$. The quantized received signal vector for each pilot-transmitting OFDM symbol can be expressed as

$$\mathbf{q}_P = \mathcal{Q}_c[\sqrt{g_{AGC}}(\mathbf{F}^H \text{diag}(\mathbf{p})\mathbf{h} + \mathbf{n})]. \quad (2)$$

For notation simplification, we define $\bar{\mathbf{p}} = \sqrt{g_{AGC}}\mathbf{p}$. The conventional OFDM channel estimator performs least-squares (LS) channel estimation to first yield $\mathbf{F}\mathbf{q}_P/\bar{\mathbf{p}}$, and then left-multiplies it by a certain weight matrix. This traditional approach relies heavily on orthogonality among subcarriers, which is already destroyed by the low-resolution ADC. This fact and the success of GTurbo in Q-OFDM detection have motivated us to develop a specialized channel estimator for Q-OFDM based on this advanced signal processing technique.

2) *Implementation Issues*: A number of issues should be considered in implementing the prototyping system, including adjusting the amplitude of the baseband signal within a certain range (i.e., AGC), measuring signal/noise power, and identifying the starting point of each frame, and subsequently, each OFDM symbol (i.e., synchronization). In conventional systems with high-resolution ADCs, the first two issues can be addressed by using a power accumulator, which is applied by taking the long-term time average of the amplitude of the corresponding received baseband samples. Meanwhile, synchronization is achieved through transmitting a pre-defined complex primary synchronization sequence (PSS) in each radio frame, computing the correlation between the received signal and the reference sequence, and finding the local maximum correlation to identify the starting point of each frame. We should clarify if these conventional solutions are still valid for the Q-OFDM receiver.

III. GTURBO-BASED ALGORITHMS

The GTurbo signal recovery principle was proposed in [35] for a quantized compressed sensing problem with an orthogonal sensing matrix. It consists of two concatenated modules. The first module produces a coarse estimate from the received quantized signal; the second module refines the estimate by considering the prior of the signal to be estimated. Each module computes its extrinsic information to be the input of the other module. The two modules are executed iteratively until convergence is achieved. This is why we use ‘‘turbo’’ to name it. GTurbo is a powerful tool for an inference problem formulated as a *generalized linear model* with an orthogonal linear transformation matrix. The Q-OFDM detection problem formulated in (1) belongs to this class of problem with a DFT matrix. However, the original GTurbo is only applicable to cases where the priori of the estimated signal is available. In this section, we extend the original GTurbo to the problem without a known priori and propose an efficient channel estimator for a Q-OFDM system.

A. Channel Estimator

The purpose of a channel estimator is to estimate the frequency-domain channel parameters of subcarriers dedicated for data transmission, namely \mathbf{h}_d , based on the given \mathbf{q}_P and $\bar{\mathbf{p}}$. The estimated channel parameter $\hat{\mathbf{h}}_d$ is subsequently utilized as the CSI for data detection to the remainder of the OFDM symbols in the current slot. Before formally introducing the proposed algorithm, we define the following two auxiliary vectors to facilitate our subsequent discussion:

$$\mathbf{x} := \text{diag}(\bar{\mathbf{p}})\mathbf{h}, \quad \mathbf{z} := \mathbf{F}^H \mathbf{x}, \quad (3)$$

where \mathbf{x} and \mathbf{z} represent the noiseless signals received in the frequency and time domains, respectively.

The channel estimation problem (2) has been formulated as a generalized linear model of the frequency-domain channel vector \mathbf{h} , suggesting that the GTurbo principle [35] can be applied for estimating \mathbf{h}_d . The proposed channel estimation technique is illustrated in Fig. 2. The GTurbo-based algorithm comprises two processing modules. Module A is used to directly produce a coarse estimate of \mathbf{x} from the quantized measurement \mathbf{q}_P , which is equivalent to producing a coarse estimate of \mathbf{h} following the linear transform in (3). Module B is used to produce a refinement of the estimate of \mathbf{h}_d . The extrinsic mean and variance of each module are passed to the other module as input. In each iteration, Modules A and B are implemented successively to refine the estimate of \mathbf{h}_d until convergence is achieved.

The proposed channel estimation algorithm is presented in Algorithm 1. We use t to denote the iteration index. We then provide several intuitive explanations to allow readers to improve their understanding of the proposed channel estimator. In Module A, (6a) and (6b) compute the MMSE estimate of z_j under the given quantized observation q_j , and its corresponding MSE, where $\mathbb{E}[z_j^R | q_{j,P}^R]$ and $\text{var}[z_j^R | q_{j,P}^R]$ denote the posteriori expectation and variance of z_j^R given $q_{j,P}^R$ under the assumption of $P(z_j^R) = \mathcal{N}(z_j^R; z_{j,A}^{\text{pri},R}(t), \frac{1}{2}v_A^{\text{pri}}(t))$. Their explicit expressions can be derived from [22] as

$$\mathbb{E}[z_j^R | q_{j,P}^R] = z_{j,A}^{\text{pri},R} + \frac{v_A^{\text{pri}}(t)}{\sqrt{2(v_A^{\text{pri}}(t) + \bar{\sigma}^2)}} \left(\frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right), \quad (4a)$$

$$\text{var}[z_j^R | q_{j,P}^R] = \frac{v_A^{\text{pri}}(t)}{2} - \frac{(v_A^{\text{pri}}(t))^2}{2(v_A^{\text{pri}}(t) + \bar{\sigma}^2)} \times \left[\left(\frac{\phi(\eta_1) - \phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right)^2 + \frac{\eta_1\phi(\eta_1) - \eta_2\phi(\eta_2)}{\Phi(\eta_1) - \Phi(\eta_2)} \right], \quad (4b)$$

where $\bar{\sigma}^2 = g_{AGC}\hat{\sigma}^2$ with $\hat{\sigma}^2$ being the estimated noise power, and

$$\eta_1 = \frac{z_{j,A}^{\text{pri},R}(t) - l(q_{j,P}^R)}{\sqrt{(v_A^{\text{pri}}(t) + \bar{\sigma}^2)/2}}, \quad \eta_2 = \frac{z_{j,A}^{\text{pri},R}(t) - u(q_{j,P}^R)}{\sqrt{(v_A^{\text{pri}}(t) + \bar{\sigma}^2)/2}} \quad (5)$$

where $l(q_{j,P}^R)$ and $u(q_{j,P}^R)$ are the lower and upper bounds that correspond to the quantizer output value $q_{j,P}^R$. Moreover, $\mathbb{E}[z_j^I | q_{j,P}^I]$ and $\text{var}[z_j^I | q_{j,P}^I]$ can be similarly computed

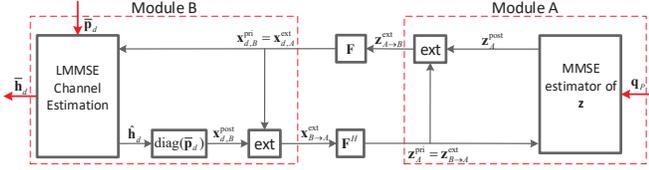


Fig. 2: Block diagram of the GTurbo-LMMSE channel estimator. The “ext” block represents the extrinsic information computation. The block of a certain matrix represents the left-multiplying operation of the input vector by the matrix in the block.

by simply replacing $z_{j,A}^{\text{pri},R}(t)$ with $z_{j,A}^{\text{pri},I}(t)$ when computing η_1 and η_2 in (5). Subsequently, the extrinsic information of Module A can be computed using (7a)–(7d). In particular, (7a) and (7b) compute the extrinsic variance of Module A, whereas (7c) and (7d) compute the extrinsic mean of Module A, which is then used as an input of Module B.

Algorithm 1: GTurbo-LMMSE Channel Estimator

Initialization: $\mathbf{z}_A^{\text{pri}}(1) = \mathbf{0}_{N \times 1}$, $v_A^{\text{pri}}(1) = 1 - g_{AGC}\widehat{\sigma}^2$,
 $\mathbf{x}_B^{\text{pri}} = \mathbf{0}_{N \times 1}$, $\mathbf{x}_B^{\text{post}} = \mathbf{0}_{N \times 1}$;

for $t = 1 : T_{\max}$ **do**

Module A:

(1) Compute the a posteriori mean/variance of \mathbf{z} :

$$z_{j,A}^{\text{post}} = \text{E} \left[z_j^R | q_{j,P}^R \right] + j \text{E} \left[z_j^I | q_{j,P}^I \right], \quad (6a)$$

$$v_{j,A}^{\text{post}} = \text{var} \left[z_j^R | q_{j,P}^R \right] + \text{var} \left[z_j^I | q_{j,P}^I \right], \quad (6b)$$

(2) Compute the extrinsic information of Module A:

$$v_A^{\text{post}} = \frac{1}{N} \sum_{j=1}^N v_{j,A}^{\text{post}}, \quad (7a)$$

$$v_B^{\text{pri}} = v_{A \rightarrow B}^{\text{ext}} = \left(\frac{1}{v_A^{\text{post}}} - \frac{1}{v_A^{\text{pri}}(t)} \right)^{-1}, \quad (7b)$$

$$\mathbf{x}_A^{\text{ext}} = v_{A \rightarrow B}^{\text{ext}} \left(\frac{\mathbf{F} \mathbf{z}_A^{\text{post}}}{v_A^{\text{post}}} - \frac{\mathbf{F} \mathbf{z}_A^{\text{pri}}(t)}{v_A^{\text{pri}}(t)} \right), \quad (7c)$$

$$\mathbf{x}_{d,B}^{\text{pri}} = \mathbf{x}_{d,A \rightarrow B}^{\text{ext}}, \quad (7d)$$

Module B:

(3) LMMSE channel estimation:

$$\hat{\mathbf{h}}_d = \mathbf{W}_{\text{LMMSE}} \text{diag}^{-1}(\bar{\mathbf{p}}_d) \mathbf{x}_{d,B}^{\text{pri}}, \quad (8)$$

(4) Compute the extrinsic information of Module B:

$$\mathbf{x}_{d,B}^{\text{post}} = \text{diag}(\bar{\mathbf{p}}_d) \hat{\mathbf{h}}_d, \quad (9a)$$

$$\mathbf{x}_A^{\text{pri}} = \mathbf{x}_{B \rightarrow A}^{\text{ext}} = c \left(\mathbf{x}_B^{\text{post}} - \alpha \mathbf{x}_B^{\text{pri}} \right), \quad (9b)$$

$$\mathbf{z}_A^{\text{pri}}(t+1) = \mathbf{F}^H \mathbf{x}_A^{\text{pri}}, \quad (9c)$$

$$v_A^{\text{pri}}(t+1) = v_{B \rightarrow A}^{\text{ext}} = \frac{\left\| \mathbf{x}_{d,B \rightarrow A}^{\text{ext}} - \mathbf{x}_{d,B}^{\text{post}} \right\|^2}{N_d}. \quad (9d)$$

end

We then consider the operations of Module B for extracting the desired channel parameters. The input of Module B $\mathbf{x}_B^{\text{pri}}$ can be approximately represented by the AWGN observation

of $\mathbf{x} = \text{diag}(\bar{\mathbf{p}}) \mathbf{h}$ as follows:

$$\mathbf{x}_B^{\text{pri}} = \text{diag}(\bar{\mathbf{p}}) \mathbf{h} + \boldsymbol{\omega}_B, \quad (10)$$

where $\boldsymbol{\omega}_B \sim \mathcal{CN}(\mathbf{0}, v_B^{\text{pri}} \mathbf{I})$. To be continuous, the following challenges should be addressed:

- The MMSE estimate of \mathbf{h}_d from (10) as done in the original GTurbo framework requires the prior distribution of \mathbf{h}_d , which is typically unavailable in practical systems. Hence the operation of Module B needs to be re-designed.
- The derivation of the extrinsic mean and variance that correspond to the preceding operations guarantees the convergence of the proposed algorithm.

The first issue can resort to the well-developed “denoiser” in the image processing field. Various types of denoisers, such as SURE-LET [36], BM3D [37], dictionary learning [38], and SVT [39], are available. However, these denoisers are specialized for i.i.d. sparse signal, low-rank matrix, or image. In our earlier trial computer simulations, the mechanical application of these denoisers cannot achieve satisfactory performance in estimating \mathbf{h}_d with strongly correlated entries. Therefore, we must select a new type of denoiser for this channel estimation problem.

Extracting \mathbf{h}_d from its approximate AWGN observation (10) is similar to the channel estimation problem in conventional full-resolution OFDM systems. The LMMSE denoiser for extracting \mathbf{h}_d from $\mathbf{x}_B^{\text{pri}}$ is constructed using (8), where $\mathbf{W}_{\text{LMMSE}}$ is the corresponding linear weight matrix with the form

$$\mathbf{W}_{\text{LMMSE}} = \mathbf{R}_{\mathbf{h}\mathbf{h}_d} (\mathbf{R}_{\mathbf{h}_d\mathbf{h}_d} + \gamma^2 \mathbf{I})^{-1}, \quad (11)$$

where $\mathbf{R}_{\mathbf{h}\mathbf{h}_d} = \text{E}[\mathbf{h}\mathbf{h}_d^H]$ and $\mathbf{R}_{\mathbf{h}_d\mathbf{h}_d} = \text{E}[\mathbf{h}_d\mathbf{h}_d^H]$ are the channel cross-correlation and auto-correlation matrices, respectively; and γ^2 is the variance of additive noise in (10). However, to implement a strict-sense LMMSE denoiser, $\mathbf{R}_{\mathbf{h}\mathbf{h}_d}$ and $\mathbf{R}_{\mathbf{h}_d\mathbf{h}_d}$ should be known. Therefore, we assume the time-domain profile of the channel to be a uniformly tapped delay line, with the maximum number of channel taps as \hat{L} . Given this assumption, the (m, n) -th entry of $\mathbf{R}_{\mathbf{h}_d\mathbf{h}_d}$ can be computed using [40, Appendix A]

$$[\mathbf{R}_{\mathbf{h}_d\mathbf{h}_d}]_{(m,n)} = \begin{cases} 1, & \mathcal{X}_d(m) = \mathcal{X}_d(n) \\ \frac{1 - e^{-j2\pi\hat{L} \frac{\mathcal{X}_d(m) - \mathcal{X}_d(n)}{N}}}{j2\pi\hat{L} \frac{\mathcal{X}_d(m) - \mathcal{X}_d(n)}{N}}, & \mathcal{X}_d(m) \neq \mathcal{X}_d(n) \end{cases}$$

where $\mathcal{X}_d(n)$ is the n -th element of set \mathcal{X}_d . And entries of $\mathbf{R}_{\mathbf{h}\mathbf{h}_d}$ can be computed similarly. In addition, we let γ^2 be an extremely small number compared with the variance of additive noise. The preceding assumptions have been proven by [41] to perform close to the strict-sense LMMSE estimation. Furthermore, the reliability and hardware efficiency of the approximate LMMSE channel estimator are confirmed by our previously developed massive MIMO-OFDM prototyping system [42].

The remaining task is to derive the extrinsic mean and variance of Module B that are associated with the approximate LMMSE denoiser. Following the divergence-free property of the series of algorithms extended from the turbo principle [43], we first construct the extrinsic mean $\mathbf{x}_{B \rightarrow A}^{\text{ext}}$ as the linear combination of $\mathbf{x}_B^{\text{post}}$ and $\mathbf{x}_B^{\text{pri}}$ with the form of (9b), and then

determine coefficients α and c based on the following two conditions:

- The errors of the input vector and extrinsic mean of Module B are orthogonal, namely,

$$E[(\mathbf{x}_B^{\text{pri}} - \mathbf{x})^H (\mathbf{x}_{B \rightarrow A}^{\text{ext}} - \mathbf{x})] = 0. \quad (12)$$

- The error of the extrinsic mean of Module B, i.e., $E[\|\mathbf{x}_{B \rightarrow A}^{\text{ext}} - \mathbf{x}\|^2]$, is minimized.

Coefficients α and c that satisfy the aforementioned conditions can be obtained by following the derivations in [44], which is given by

$$\alpha = \frac{\mathbf{1}_{1 \times N_d} \text{diag}^{-1}(\bar{\mathbf{p}}_d) \text{Diag}(\mathbf{W}_{d, \text{LMMSE}})}{N_d}, \quad (13a)$$

$$c = \frac{(\mathbf{x}_B^{\text{pri}})^H (\mathbf{x}_B^{\text{post}} - \alpha \mathbf{x}_B^{\text{pri}})}{\|\mathbf{x}_B^{\text{post}} - \alpha \mathbf{x}_B^{\text{pri}}\|^2}, \quad (13b)$$

where $\mathbf{1}_{1 \times N_d}$ denotes the N_d -dimension row vector whose elements are all one. In addition, we can approximate the extrinsic variance $v_{B \rightarrow A}^{\text{ext}} = \frac{1}{N_d} \|\mathbf{x}_{d, B \rightarrow A}^{\text{ext}} - \mathbf{x}_d\|^2$ by relaxing the exact \mathbf{x}_d as its estimate $\mathbf{x}_{d, B}^{\text{post}}$.

Notably, nearly all amplitude information is lost irretrievably under 1-bit quantization. We must recover amplitude information, particularly the average channel gain $P_h = \frac{1}{N_d} \sum_{j \in \mathcal{X}_d} |h_j|^2$, via additional means to overcome this intrinsic defect of the 1-bit ADC and further improve performance. We refine the channel estimate $\hat{\mathbf{h}}_d$ with the additional amplitude information by operating the channel normalization expressed by

$$\tilde{\mathbf{h}}_d = \sqrt{\hat{P}_h} \frac{\hat{\mathbf{h}}_d}{\|\hat{\mathbf{h}}_d\|}, \quad (14)$$

which is the final output of the channel estimator. In our system, the estimated P_h can be obtained from the average power of the received signal P_r , which is defined by

$$\hat{P}_h = (P_r - \hat{\sigma}^2) \times \frac{N}{N_d}. \quad (15)$$

We will specify how to measure P_r in Section V. Then, $\bar{\mathbf{h}}_d = \sqrt{g_{\text{AGC}}} \tilde{\mathbf{h}}_d$ will be used as the CSI for the data detector after obtaining the normalized channel estimate $\hat{\mathbf{h}}_d$. Channel normalization is unnecessary when ADC precision is more than 1 bit because the majority of amplitude information can be recovered using the proposed GTurbo-LMMSE channel estimator. This result can be verified by the numerical results presented in Section IV. Let $\bar{\mathbf{h}}_d = \sqrt{g_{\text{AGC}}} \hat{\mathbf{h}}_d$ be the CSI for the data detector under the case without channel normalization.

B. Data Detection and Channel Decoding

The data detector is used to compute the MMSE estimate of \mathbf{s} and its corresponding MSE based on the quantized received signal \mathbf{q}_D modeled by (1). The GTurbo-based detector proposed in [25, Algorithm 1] can be directly used in data detection with a slight modification by replacing the known CSI with the estimated channel parameters $\bar{\mathbf{h}}$. In this subsection, we briefly describe the processing procedure for the data detector. For the details of the algorithm, refer to [25].

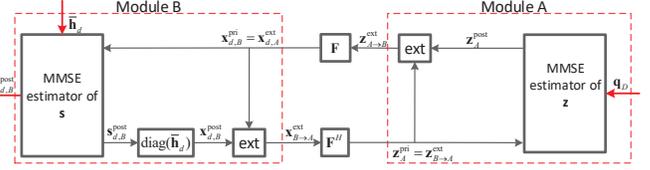


Fig. 3: Block diagram of the GTurbo-based data detector with an estimated CSI $\bar{\mathbf{h}}$. The “ext” block represents the extrinsic information computation. The block of a certain matrix represents the left-multiplying operation of the input vector by the matrix in the block.

We also define two auxiliary vectors similar to (3) as follows:

$$\mathbf{x} := \text{diag}(\bar{\mathbf{h}})\mathbf{s}, \quad \mathbf{z} := \mathbf{F}^H \mathbf{x}.$$

The computation procedures are demonstrated in Algorithm 2, and the corresponding flow diagram of this algorithm is presented in Fig. 3. The two modules of this detector are as follows. Module A calculates the coarse estimate of \mathbf{x} without considering the prior distribution of \mathbf{s} . Module B refines the estimate by considering the prior distribution of \mathbf{s} . In each iteration, Modules A and B are performed successively to refine the estimate of \mathbf{s} until convergence is achieved.

In Module A, (18a) and (18b) compute the posteriori mean and variance of z_j with explicit expressions provided by (4a) and (4b), respectively. The extrinsic variance of Module A can be computed using (19a) and (19b), whereas its extrinsic mean can be computed using (19c) and (19d), respectively. In Module B, we compute the estimate of \mathbf{s} and its MSE for $j \in \mathcal{X}_d$ in (20a) and (20b). Their explicit expressions are given by

$$s_{j,B}^{\text{post}} = \frac{\sum_{s \in \mathcal{S}} s \mathcal{CN}\left(s; \frac{x_{j,B}^{\text{pri}}}{h_j}, \frac{v_B^{\text{pri}}}{|h_j|^2}\right)}{\sum_{s \in \mathcal{S}} \mathcal{CN}\left(s; \frac{x_{j,B}^{\text{pri}}}{h_j}, \frac{v_B^{\text{pri}}}{|h_j|^2}\right)}, \quad (16a)$$

$$v_{j,B}^{\text{post}} = \frac{\sum_{s \in \mathcal{S}} |s|^2 \mathcal{CN}\left(s; \frac{x_{j,B}^{\text{pri}}}{h_j}, \frac{v_B^{\text{pri}}}{|h_j|^2}\right)}{\sum_{s \in \mathcal{S}} \mathcal{CN}\left(s; \frac{x_{j,B}^{\text{pri}}}{h_j}, \frac{v_B^{\text{pri}}}{|h_j|^2}\right)} - \left|s_{j,B}^{\text{post}}\right|^2. \quad (16b)$$

where \mathcal{S} denotes the set of constellation points. Subsequently, we compute the extrinsic variance of Module B using (21a) and (21b) and its extrinsic mean using (21c) and (21d). (21d) is identical to (9b) by letting $c = \frac{v_{B \rightarrow A}^{\text{ext}}}{v_B^{\text{post}}}$ and $\alpha = \frac{v_B^{\text{post}}}{v_B^{\text{pri}}}$, which also satisfy conditions (a) and (b).

In the practical applications, we can set an additional termination criterion, namely, the iteration should be terminated when v_B^{pri} or v_A^{pri} is less than the predefined thresholds. This modification is reasonable because that the excessively small v_B^{pri} and v_A^{pri} , along with errors in the estimated channel responses, possibly trap the estimates (18a) and (20a) into wrong values, which deteriorates detection performance. Finally, $\mathbf{x}_{d,B}^{\text{pri}}$ and v_B^{pri} are used for the computation of the log-likelihood ratio (LLR) that corresponds to each coded bit, which is used as the input of the channel decoder. We denote

the i -th bit associated with the modulated symbol s_j by b_{ji} . The LLR of b_{ji} can be computed by

$$\text{LLR}(b_{ji}) = \frac{1}{\beta_j} \left(\left| \frac{x_{j,B}^{\text{pri}}}{\bar{h}_j} - s_{j,i}^- \right|^2 - \left| \frac{x_{j,B}^{\text{pri}}}{\bar{h}_j} - s_{j,i}^+ \right|^2 \right), \quad (17)$$

where $\beta_j = v_B^{\text{pri}} / |\bar{h}_j|^2$, $s_{j,i}^+$ and $s_{j,i}^-$ are defined as follows:

$$s_{j,i}^- = \underset{s \in \mathcal{S}_i^-}{\text{argmin}} \left| s - \frac{x_{j,B}^{\text{pri}}}{\bar{h}_j} \right|^2,$$

$$s_{j,i}^+ = \underset{s \in \mathcal{S}_i^+}{\text{argmin}} \left| s - \frac{x_{j,B}^{\text{pri}}}{\bar{h}_j} \right|^2,$$

where \mathcal{S}_i^+ and \mathcal{S}_i^- denote the subsets of the constellation symbols of the adopted modulation scheme with i -th bit being 1 and 0, respectively.

Algorithm 2: GTurbo-based Data Detector

Initialization: $\mathbf{z}_A^{\text{pri}}(1) = \mathbf{0}_{N \times 1}$, $v_A^{\text{pri}}(1) = \frac{1}{N} \sum_{j \in \mathcal{X}_d} |\bar{h}_j|^2$,
 $\mathbf{x}_B^{\text{pri}} = \mathbf{0}_{N \times 1}$, $\mathbf{x}_B^{\text{post}} = \mathbf{0}_{N \times 1}$;

for $t = 1 : T_{\text{max}}$ **do**

Module A:

(1) Compute the posteriori mean/variance of \mathbf{z} :

$$z_{j,A}^{\text{post}} = \text{E} \left[z_j^R \mid q_{j,D}^R \right] + j \text{E} \left[z_j^I \mid q_{j,D}^I \right], \quad (18a)$$

$$v_{j,A}^{\text{post}} = \text{var} \left[z_j^R \mid q_{j,D}^R \right] + \text{var} \left[z_j^I \mid q_{j,D}^I \right], \quad (18b)$$

(2) Compute the extrinsic mean/variance of \mathbf{x} :

$$v_A^{\text{post}} = \frac{1}{N} \sum_{j=1}^N v_{j,A}^{\text{post}}, \quad (19a)$$

$$v_B^{\text{pri}} = v_{A \rightarrow B}^{\text{ext}} = \left(\frac{1}{v_A^{\text{post}}} - \frac{1}{v_A^{\text{pri}}(t)} \right)^{-1}, \quad (19b)$$

$$\mathbf{x}_{A \rightarrow B}^{\text{ext}} = v_{A \rightarrow B}^{\text{ext}} \left(\frac{\mathbf{F} \mathbf{z}_A^{\text{post}}}{v_A^{\text{post}}} - \frac{\mathbf{F} \mathbf{z}_A^{\text{pri}}}{v_A^{\text{pri}}(t)} \right), \quad (19c)$$

$$\mathbf{x}_{d,B}^{\text{pri}} = \mathbf{x}_{d,A \rightarrow B}^{\text{ext}}, \quad (19d)$$

Module B:

(3) Compute the posteriori mean/variance of \mathbf{s}_d :

$$s_{j,B}^{\text{post}} = \text{E} \left[s_j \mid \bar{h}_j, x_{j,B}^{\text{pri}} \right], \quad j \in \mathcal{X}_d, \quad (20a)$$

$$v_{j,B}^{\text{post}} = \text{var} \left[s_j \mid \bar{h}_j, x_{j,B}^{\text{pri}} \right], \quad j \in \mathcal{X}_d, \quad (20b)$$

(4) Compute the extrinsic mean/variance of \mathbf{z} :

$$v_B^{\text{post}} = \frac{1}{N} \sum_{j \in \mathcal{X}_d} |\bar{h}_j|^2 v_{j,B}^{\text{post}}, \quad (21a)$$

$$v_A^{\text{pri}}(t+1) = v_{B \rightarrow A}^{\text{ext}} = \left(\frac{1}{v_B^{\text{post}}} - \frac{1}{v_B^{\text{pri}}(t)} \right)^{-1}, \quad (21b)$$

$$x_{j,B}^{\text{post}} = \bar{h}_j s_{j,B}^{\text{post}}, \quad j \in \mathcal{X}_d, \quad (21c)$$

$$\mathbf{z}_A^{\text{pri}}(t+1) = \mathbf{z}_{B \rightarrow A}^{\text{ext}} = v_{B \rightarrow A}^{\text{ext}} \left(\frac{\mathbf{F}^H \mathbf{x}_B^{\text{post}}}{v_B^{\text{post}}} - \frac{\mathbf{F}^H \mathbf{x}_B^{\text{pri}}}{v_B^{\text{pri}}(t)} \right). \quad (21d)$$

end

IV. NUMERICAL RESULTS

Before proceeding to the OTA experiments, we present the numerical simulation results of the proposed GTurbo-based algorithms. The system parameters are specified as follows. Notably, the following parameter setups will be followed in the OTA experiments in Section V-B.

- The number of orthogonal subcarriers N is 2048, and the number of subcarriers used for data transmission N_d is 1186.²
- The turbo code is adopted for the channel coding technique, where the interleaver rule set based on [45], and the recursive systematic convolutional encoders represented by [13, 15] with the constraint length of 4.
- Uniform mid-rise quantization is adopted by each real-value quantizer $\mathcal{Q}(\cdot)$. The thresholds are set as $r_0 = -\infty$, $r_b = (-2^{B-1} + b)\Delta$ for $b = 1, 2, \dots, 2^B - 1$, and $r_{2^B} = \infty$, whereas the quantized outputs are set as $c_b = \left(-\frac{2^B+1}{2} + b \right) \Delta$ for $b = 1, 2, \dots, 2^B$, where quantization step size $\Delta = \sqrt{P_q} \Delta_B$ and $P_q = g_{\text{AGC}} P_r / 2$ denotes the input power of $\mathcal{Q}(\cdot)$. The values of Δ_B are typically selected to minimize the quantization MSEs, which are given by [46] and listed in Table I.

TABLE I: Δ_B for different quantization bits B

B	1	2	3	4	5
Δ_B	$\sqrt{8/\pi}$	0.9957	0.5860	0.3352	0.1881

- The entries of the pilot vector \mathbf{p}_d are drawn independently from the equiprobable 4-QAM constellation. In addition, the entries of \mathbf{p}_d and \mathbf{s}_d are normalized, such that $\text{E}[|p_j|^2] = 1$ and $\text{E}[|s_j|^2] = 1$ for $j \in \mathcal{X}_d$.

In this section, we set the number of channel taps $L = 4$, with power profiles 0, -7, -12, and -18 dB for each tap. We conduct Monte Carlo simulations of 1,000 independent channel realizations to obtain the performance metrics to be evaluated. Imperfections in parameters $1/P_r$, σ^2 and L will be simulated by imposing certain mismatches on these parameters. In particular, we artificially impose an error that is uniformly distributed in [-30%, 30%] on parameters $1/P_r$ and σ^2 to obtain \hat{g}_{AGC} and $\hat{\sigma}^2$, in addition, we let $\hat{L} = 6$. These mismatches are set to simulate the practical situation in which exact parameters $1/P_r$, σ^2 and L are unavailable. The influence of channel normalization in (14) will also be discussed in addition to the imperfect parameters. The performance under the perfect parameters, i.e., $\hat{g}_{\text{AGC}} = 1/P_r$, $\hat{\sigma}^2 = \sigma^2$, and $\hat{L} = 4$, serves as the benchmark. The following methods will be used later as benchmarks for performance comparison:

- Conventional approach: the procedures described in section II that ignore quantization effect.
- Bussgang based method: the procedures described in section II based on Bussgang linearization for (1) and (2) in the same way of [23, section IV-A].

²For the transmission of 16-QAM symbols, only 1185 of them are actually used.

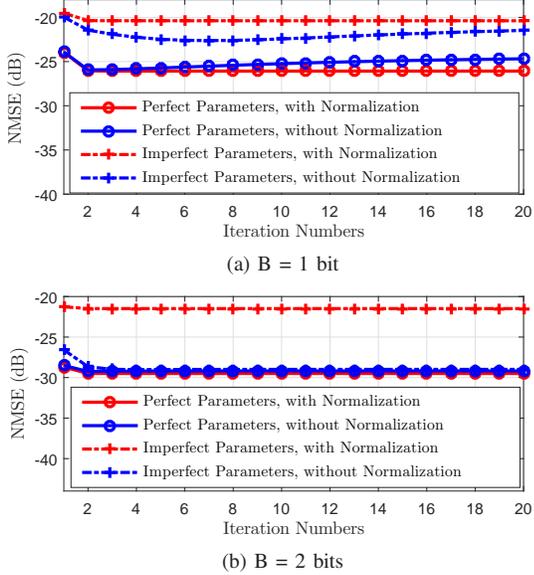


Fig. 4: NMSEs versus iteration numbers of the proposed GTurbo-LMMSE channel estimator (i.e., Algorithm 1) for an SNR of 12 dB under four conditions.

- GAMP-based methods: EM-GM-GAMP [32] for channel estimation and GAMP [47, Algorithm 2] for detection.
- FASTA-based³ maximum likelihood channel estimator and approximated MMSE detector [15].

In this section, we set the maximum iteration number as the stopping criteria for iterative methods. We also mark the maximum iteration numbers we set in each figure, which provides clues about computational delay required by each iterative algorithm.

We first present the normalized MSEs (NMSEs) of $\tilde{\mathbf{h}}_d$, which are defined as

$$\text{NMSE} = \frac{\|\tilde{\mathbf{h}}_d - \mathbf{h}_d\|^2}{\|\mathbf{h}_d\|^2},$$

versus the iteration numbers of Algorithm 1. Performances are compared under four conditions produced by combining perfect or imperfect parameters, and whether channel normalization is performed. Fig. 4 shows the results under an SNR⁴ of 12dB and B of 1 bit and 2 bits. For $B = 1$ bit without channel normalization, the NMSEs are initially reduced and then increase again after a few iterations. This is expected because the 1-bit ADCs implemented by signal zero-crossing comparators eliminate the information regarding average channel gain P_h , which is difficult to recover using the proposed iterative procedure. After phase information is recovered, the error in amplitude information leads to divergence. Therefore, channel normalization is recommended when $B = 1$ bit to guarantee the numerical stability of the proposed algorithm. Different observations can be made when $B = 2$ bits, under

³In our simulation, we utilize part of the code downloaded from https://github.com/quantizedmassivemimo/mu_mimo_ofdm/ and also FASTA toolbox [48], [49].

⁴We refer to SNR as the ratio of the average power of the noiseless received time-domain samples to the average noise power.

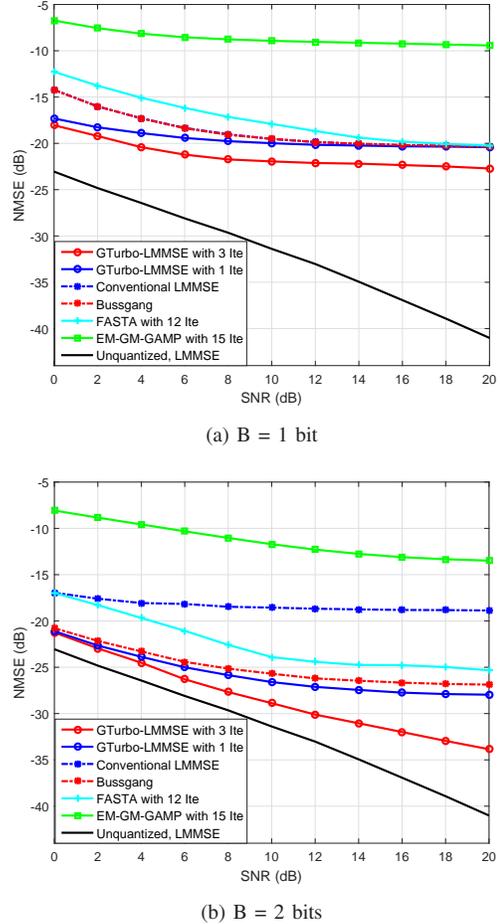


Fig. 5: Comparison of the NMSE performances of various channel estimators when $B = 1$ bit and 2 bits under imperfect parameters.

which NMSEs remains stable with only a minor fluctuation under all four conditions after a few iterations. This result indicates that the phase information and amplitude information of \mathbf{h}_d can be properly recovered using the proposed method. When we compare the NMSE performances, better result is observed in the case of 2-bit quantization without channel normalization with imperfect parameters, which can also be observed when $B = 1$ bit, whereas the opposite result is observed with perfect parameters. This finding implies the inaccuracy of \hat{P}_h estimated by (15) from incorrect g_{AGC} and $\hat{\sigma}^2$, while the proposed algorithm exhibits strong signal recovery capability. Hence, with $B = 2$ bits or more, channel normalization is not preferable in the application of Algorithm 1 to practical systems. Fig. 4 also shows that a satisfactory channel estimate can be obtained with less than five iterations in all cases.

Fig. 5 compares the NMSE performance of different algorithms under imperfect parameters. The performance with full-resolution quantization is also included as a lower bound. Evidently, the proposed algorithm outperforms its counterparts, which can be attributed to its capacity to update the linear approximation of the Q-OFDM system in (10) through

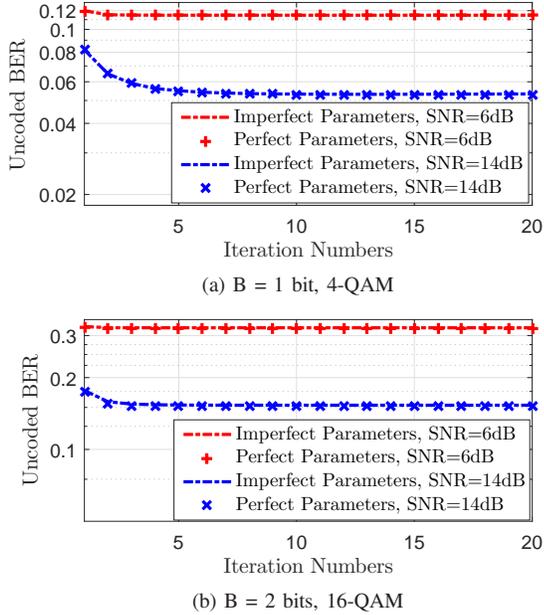


Fig. 6: Uncoded BERs versus iteration numbers of the GTurbo-based detector (i.e., Algorithm 2) under estimated channel responses.

iterations. A channel estimate is produced (and can also be updated) based on this approximation. We find from Fig. 5 that the Bussgang-based method achieves a similar NMSE performance (with minimal performance degradation) to that of the proposed GTurbo-LMMSE channel estimator with one iteration. EM-GM-GAMP does not work efficiently because it applies i.i.d. Gaussian mixture model to approximate the a priori distribution of each element of frequency-domain channel response, but the real channel responses are correlated. The results in Fig. 5 also indicate that when $B = 2$ bits, the gap between the NMSE performance of the proposed channel estimator and that of the LMMSE channel estimator under full-resolution quantization is minimal.

We next evaluate and compare the information recovery performances of various algorithms under the estimated channel responses. Fig. 6 shows the uncoded bit error rate (BER) of Algorithm 2 versus the iteration numbers under $\hat{\mathbf{h}}_d$ estimated by Algorithm 1. At lower SNR (6 dB) and higher SNR (14 dB), and with perfect or imperfect parameters, the uncoded BERs are reduced firstly and then remain stable after a few iterations. At most five iterations (typically 2 or 3 iterations are enough) are adequate for the algorithm to converge. In addition, performance loss from imperfect parameters is negligible. This result justifies the robust signal recovery capability of the GTurbo-based algorithms. In Figs. 7a and 7b, coded BER performances are compared. We set the code rate to be $1/2$ in this section to prevent the information recovery capabilities of the algorithms from being hidden by the strong channel code. However, we set the code rate to be $1/3$ to enable reliable communication in our prototyping system shown in section V-B. From Figs. 7a and 7b, we have a similar finding that the Bussgang-based method has a similar coded BER performance

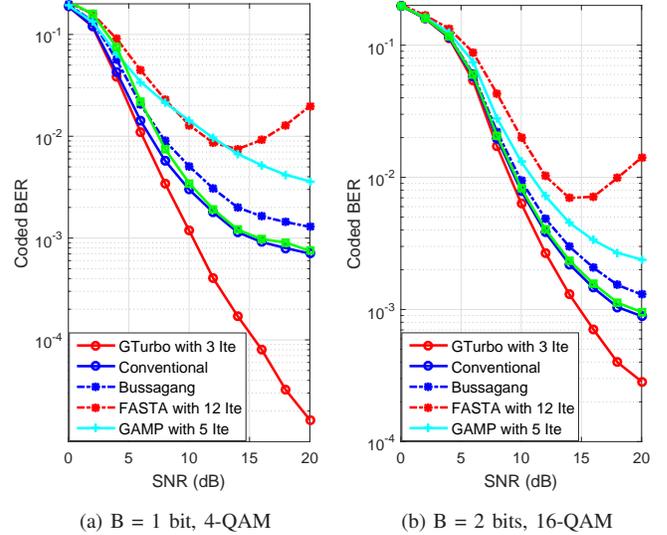


Fig. 7: Comparison of data detection performances of various detectors when $B = 1$ bit and 2 bits under imperfect parameters and estimated channel responses.

(with minimal performance degradation) to that of the GTurbo-based detector with one iteration. In addition, we find that GTurbo is superior to GAMP because GAMP is derived for problems with linear transformation matrices that comprise i.i.d. Gaussian elements rather than orthogonal ones. The proposed schemes also outperforms FASTA-based algorithms, given that the latter one assumes that the entries of \mathbf{s} follow i.i.d. circularly symmetric complex Gaussian distribution instead of their true ones. When computing LLR, the FASTA-based method cannot produce the variance information, i.e., β_j in (17), and approximate it by 1. This approximation is inaccurate, especially under high SNR, which causes the coded BER curves to go up after a certain SNR for the cases of $B = 1$ bit and 2 bits.

V. RECEIVER DESIGN AND IMPLEMENTATION

In this section, we first introduce the proposed Q-OFDM receiver architecture and then conduct OTA experiments to verify the reliability of our proposed architecture.

A. Receiver Architecture

Additional processing modules are required to guarantee that the inference algorithms proposed in Section III can be successfully implemented in practice, including AGC, signal/noise power estimator, and synchronization module. Correspondingly, we propose the Q-OFDM receiver architecture shown in Fig. 8. Once the RF signal is received by the antenna from the wireless channel, the signal is down-converted into the analog baseband for subsequent processing. The down-converter module typically consists of low-noise amplifiers, local oscillators, filters, and mixers. The received analog baseband signal is divided into two streams to be sampled and quantized in parallel with different sampling rates and quantization

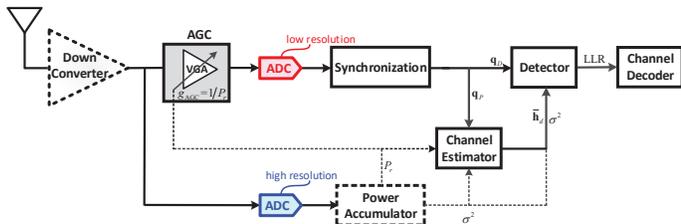


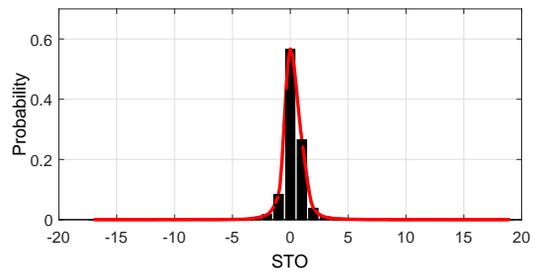
Fig. 8: Q-OFDM receiver architecture.

precisions for different purposes. One stream is quantized with ultra-low resolution (e.g., 1–2 bits) and sampled with frequency F_s , from which synchronization searching, channel estimation, data detection and channel decoding are performed according to the adopted frame structure. Another stream is quantized with high precision but sampled with a considerably lower sampling frequency (tens or hundreds times lower than F_s), which is used by the power accumulator to assist in AGC, noise signal/noise estimation, and channel amplitude information recovery under 1-bit quantization. We then elaborate how the power accumulator and synchronization module work.

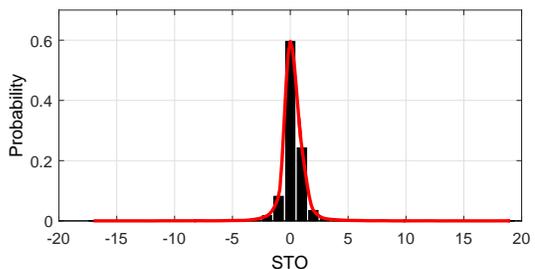
1) *Power Accumulator*: The main purpose of this accumulator is to measure the average power of the high-resolution quantized samples within specific intervals. The average power of the received signal P_r is measured as a necessary parameter in the receiver. In particular, we take the reciprocal of P_r as the gain of the variable gain amplifier in the AGC unit. Hence, $g_{AGC} = 1/P_r$. This step enables us to adjust the amplitude of the baseband signal within a certain range to suit the fixed thresholds of the ADC. For 1-bit quantization, the AGC unit becomes unnecessary. However, P_r is still a requisite parameter for the channel norm recovery shown in (15), which helps improve channel estimation performance as previously discussed and demonstrated. In addition, noise power σ^2 is another necessary parameter of the proposed algorithms. Noise power estimation is conducted by assigning a few zero-padding OFDM symbols to each frame and measuring the average power of these OFDM symbols at the receiver. The power accumulator measures noise power by using the high-resolution quantized samples of these OFDM symbols.

Remark 1: The power consumption of the assistant high-resolution ADC will be considerably reduced by decreasing its sampling frequency as much as possible, thereby leading to a reduction in the precision of signal/noise power estimation. The variations of channel gain and noise power are relatively slow. Therefore, we can compensate for the precision reduction by reasonably prolonging the time of the power accumulator. ■

2) *Synchronization*: We perform the symbol-timing synchronization by following the correlation-based synchronization utilized in many extant communication systems to precisely identify the starting point of each frame. We select the Zadoff Chu (ZC) sequence as the PSS because of its low cross-correlation and high auto-correlation. For the OFDM symbols in each frame arranged to transmit PSS, the ZC sequence is placed at $N_s = 62$ subcarriers corresponding to frequency $-\frac{N_s}{2} \frac{F_s}{N}, \dots, -\frac{F_s}{N}, \frac{F_s}{N}, \dots, \frac{N_s}{2} \frac{F_s}{N}$, whereas other subcarriers



(a) B = 1 bit



(b) B = 2 bits

Fig. 9: Distribution of the STO that corresponds to the exact synchronization position.

are padded with zero to facilitate accurate synchronization. We denote the vector of the frequency-domain OFDM symbol that contains the ZC sequence as $\mathbf{t} \in \mathbb{C}^{N \times 1}$ with $t_j = 0$ for $j \in \mathcal{X} \setminus \mathcal{X}_s$, where \mathcal{X}_s denotes the index subset of the N_s subcarriers used for transmitting PSS. We construct the reference sequence \mathbf{t}_r by first transferring the frequency-domain OFDM symbol that transmits PSS into the time-domain which yields $\tilde{\mathbf{t}} = \mathbf{F}^H \mathbf{t}$, and then normalizing the power of the resulting sequence which yields $\mathbf{t}_r = \tilde{\mathbf{t}} / \|\tilde{\mathbf{t}}\|$. Synchronization search is conducted by calculating the cross-correlation between the received quantized sequence (denoted by \mathbf{q}_r) and a reference sequence, namely, $\mathbf{q}_r^H \mathbf{t}_r$. The starting point of each frame is identified by determining the highest cross-correlation among a few adjacent symbol periods, which follows the same mechanism as that in full-resolution OFDM systems. Once the starting point of each frame is identified, the CP of each OFDM symbol is removed, and the quantized time-domain OFDM symbols are inputted into their corresponding processing modules, namely, the channel estimator or data detector, by using the algorithms presented in Section III.

We conduct numerical simulation to examine whether the correlation-based synchronization is still valid for the coarsely quantized received signal. Then, we display the probability distribution of the symbol time offsets (STOs), which are measured by the relative offset in the number of time-domain samples of the synchronization position that corresponds to the exact starting point of each frame under an SNR of 15 dB and a quantization precision of 1–2 bits, as shown in Fig. 9. We draw the red curve to approximately depict the profile of the STO distribution. The profile shows that the exact synchronization position can be identified with a large probability. In particular, the probability of the exact synchronization is approximately 0.55 and 0.60 under the quantization precision



Fig. 10: Picture of the proposed prototyping system.

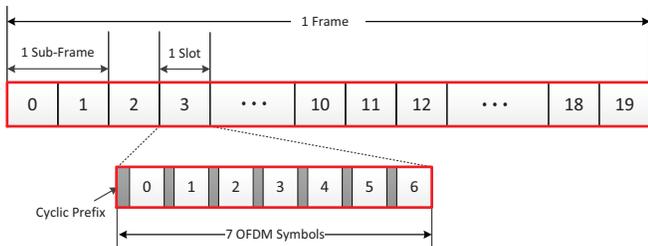


Fig. 11: Frame structure

of 1 bit and 2 bits, respectively. Moreover, STO is retained within a range of -4 to 4 with the highest probability, which is an acceptable synchronization error range. We can conclude from the above preceding results that the correlation-based synchronization technique is still reliable for the received samples with low-resolution quantization.

B. OTA Experiments

We then establish a proof-of-concept prototyping system based on the proposed architecture presented in the previous subsection (Fig. 10). The transmitter and the receiver are set up with a separate baseband and RF processing unit. Baseband processing is performed through the software programmed in the host computer. The corresponding procedures are specified in Sections III and V-A. RF processing is conducted using the Rohde & Schwarz equipments. In particular, we use SGT100A for the transmitter and RTO2044 for the receiver. The output baseband samples of RTO2044 are quantized by a high-resolution ADC. Thus, coarse quantization is performed virtually through the software before further processing. This setup enables us to obtain certain reference parameters for subsequent performance evaluations. Data exchange between the baseband processing software and RF processing components is performed through off-the-shelf high-speed interfaces.

The frame structure used in our prototyping system is illustrated in Fig. 11. This structure is designed based on the frame structure of the LTE system, which can be flexibly reconfigured for future extensions. An entire radio frame consists of 10 subframes, and each subframe has 2 consecutive slots. In each slot, 7 OFDM symbols are transmitted. In our system, L_{cp} is set to 144 for OFDM symbols 1–6, and 160 for OFDM symbol 0 in each slot. For each frame, slots 2–19 are dedicated for data transmission, in which the predefined pilot sequence is transmitted via the first OFDM symbol of these slots for channel estimation, whereas the remaining 6 OFDM symbols are used for transmitting data sequences. To establish synchronization between the transmitter and the receiver, OFDM symbol 6 of slots 0 to 10 are reserved for transmitting PSS. The first OFDM symbol of slot 1 is padded by null symbols for noise power estimation.

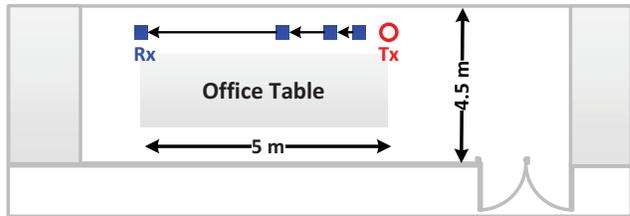


Fig. 12: Experiment environment and equipment deployment.

TABLE II: System Parameters

Parameter	Value
System Bandwidth	100 MHz
Sampling Rate F_s	153.6 MHz
Subcarrier Interval	75 KHz
Frame Duration	2 ms
Subframe Duration	0.2 ms
OFDM symbol Duration (without CP)	13.33 μ s

The experiments are conducted in an indoor environment displayed in Fig. 12. The transmitter and the receiver are placed in a corridor beside an office table. We adjust their distance by fixing the transmitter and placing the receiver in different positions. In our experiments, the carrier frequency is set to 3.5 GHz, and we set the bandwidth to be 100 MHz to support high data transmission with a constant frame duration. The resulting system parameters are listed in Table II.

Experiments are conducted to evaluate the error rate performance of the proposed prototyping system under different quantization precisions and modulation schemes, including 1 bit for 4-QAM and 2 bits for 16-QAM. Various values of average SNRs are generated by changing the transmission power and distance between the transmitter and the receiver. The performance metrics to be evaluated include the MSE of the estimated channel parameter, coded and uncoded BER, and PER of the information bits in each code word. The average SNRs and these metrics are measured from 10,000 received frames and obtained by calculating the average of these measured results. We use the LMMSE channel estimate from the high-resolution received signal as the reference when calculating the MSE of the estimated channel parameters. The experiment results are presented in Tables III and IV, which show the measured performance metrics when transmitting 4-QAM with 1-bit quantization and 16-QAM with 2-bit quantization, respectively. We can conclude that reliable data transmission can be guaranteed when SNR is higher than approximately 6 dB for both cases.

VI. DISCUSSIONS

A. Implementation Issues

In this subsection, we discuss the implementation issues of GTurbo-based algorithms. Compared with linear receivers, which involve only one matrix-vector multiplication, GTurbo-based algorithms involve many challenges in their hardware-friendly implementation. These challenges include various nonlinear computations in Module A, several matrix-vector multiplications, and the need for iteration. The rest of this

TABLE III: B = 1 bit, 4-QAM

Transmitted Power (dBm)	Distance (cm)	SNR (dB)	MSE (dB)	Uncoded BER	Coded BER	PER
-18	50	12.6768	-21.1096	0.0385	0	0
	100	7.8791	-17.1139	0.0762	0	0
	200	5.4760	-15.2942	0.1610	7.6436×10^{-5}	0.0230
	250	3.9743	-14.0242	0.1820	0.0017	0.2030
-15	200	5.8339	-15.5368	0.1574	5.6588×10^{-5}	0.0080
	250	5.5803	-15.4472	0.1644	$1.0093e \times 10^{-4}$	0.0310

TABLE IV: B = 2 bits, 16-QAM

Transmitted Power (dBm)	Distance (cm)	SNR (dB)	MSE (dB)	Uncoded BER	Coded BER	PER
-18	50	12.6867	-22.8611	0.1128	0	0
	100	7.9481	-18.5790	0.1633	0	0
	200	6.3335	-17.2023	0.2022	0.0108	0.5160
	250	4.0372	-15.1170	0.2455	0.1791	1
-15	200	6.7478	-17.6764	0.1879	1.4032×10^{-4}	0.0269
	250	6.5489	-17.3654	0.1968	0.0020	0.2405

subsection shows that these GTurbo-based algorithms can be hardware-friendly and efficiently implemented.

First, we elaborate on the implementation of those nonlinear computations and matrix-vector multiplications. The computations of the posteriori expectation and variance in (4a) and (4b) require deriving the integral of the Gaussian distribution function $\phi(x)$, which poses the largest bottleneck in algorithm implementation. Hardware-friendly approximations for $\phi(x)$ and $Q(x)$ were proposed in [31], which also facilitate the nonlinear operations in (16a) and (16b). The matrix multiplications in (7c), (9d), (19c) and (21d) can be implemented using fast Fourier transform processors with computational complexity $\mathcal{O}(N \log_2 N)$. Our approximation for the weight matrix of the LMMSE channel estimator in (11) enables us to compute this matrix off-line and store the result in the memory, thus we can avoid the online computation of high-dimensional matrix inversion. We can also observe that the two algorithms demonstrate the same operations in Module A, which indicates that they can share the identical circuit module when implementing the proposed architecture in practice. Interestingly, from Algorithms 1 and 2, we find that the proposed GTurbo-based methods contain the operation of the conventional method, specifically, (8) with $\mathbf{x}_B^{\text{pri}} = \mathbf{F}\mathbf{q}_P$ in Algorithm 1, and (17) with $\mathbf{x}_B^{\text{pri}} = \mathbf{F}\mathbf{q}_D$ and v_B^{pri} being the estimate of noise power (and quantization error) in algorithm 2. This feature provides the flexibility to reduce the iterative procedures to the conventional operations by disabling other operations and allowing only one iteration in the condition that iteration cannot yield sufficient performance improvement.

The hardware-friendly implementation of the entire iterative procedure can be achieved by applying the pipelined and folding hardware architecture proposed in [31]. Moreover, Section IV presents that the iterative process can converge after only two or three iterations, thereby preventing high computational delay. These facts make efficient and real-time

implementation of algorithms feasible. The simulation results in [31] demonstrate that the fixed-point setting combined with Q-function approximation causes only a minimal performance degeneration compared with the original floating-point simulation. Furthermore, our MIMO-OFDM prototyping [42] verifies the capacity of multi-core processors in implementing high-speed signal processing and complicated operations, which implies the potential of using multi-core processors to effectively implement advanced signal-processing techniques (like GTurbo) in practice.

B. Future Works

In this subsection, we discuss our initial ideas about the extensions to multi-user and multi-antenna cases. The following are the two potential directions for these extensions. The first direction involves aligning the proposed algorithms with advanced beamforming techniques. The second direction involves developing novel techniques that are specialized for scenarios associated with multiple received streams. The first direction is promising for the systems in which we can generate high-directional beams between transmitters and receivers. Here, we use the uplink transmission as an example. Narrow beams are steered via a specific beamforming technique for the formation of high-directional spatial links between different users and the base station (BS). Through such beamforming techniques aligned with proper user selection, the multi-antenna and multi-user systems can be regarded as a bank of parallel single-stream transmissions. Therefore, the proposed techniques can be directly used at the BS side for each individual spatial link. Moreover, the key point, which requires our future works on this direction, is the design of reliable beamforming techniques.

However, for scenarios with a wide angle spread or when beams steered for different links overlap, the previous idea is ineffective. Novel algorithm frameworks dedicated to mitigating nonlinear distortion and multi-user interference must

therefore be developed. The inference problems in Q-OFDM systems with multiple users and multiple received streams involve similar formulation as (1) with a generic linear transformation matrix. The extended algorithm framework is proposed in [34] by constraining the estimation into the linear space characterized by a linear transformation matrix, which is reduced to GTurbo when the linear transformation matrix is an orthogonal matrix. This work exhibits potential for the detection of the single-user MIMO case. Subsequent researches on this new framework is currently ongoing on the following two aspects. First, the use of algorithm in [34] requires unitary precoding among all transmitted streams at the transmitter, which restricts its use to multi-user cases. Second, the efficient channel estimation method based on this new framework is being developed.

VII. CONCLUSION

This study focuses on the development of a reliable Q-OFDM receiver architecture and the prototyping system implementation. A novel Q-OFDM channel estimator was developed by integrating a type of robust linear OFDM channel estimator into the original GTurbo framework with extrinsic information derived to guarantee its convergence. In addition, synchronization, AGC, and signal/noise power measurement were also properly considered. We proposed an efficient Q-OFDM receiver architecture by combining the above issues with the proposed GTurbo-based algorithms, based on which we constructed a Q-OFDM prototyping system. OTA experiments are conducted to justify the reliability of the proposed architecture. The following conclusions can be drawn from the results.

- The proposed GTurbo-LMMSE channel estimator outperforms its counterparts in terms of NMSE performance. Performance under 2-bit quantization is very close to that of full-resolution quantization. In addition, additional recovery of amplitude information is required in the case of 1-bit quantization.
- The GTurbo-based detector, with support of the proposed GTurbo-LMMSE channel estimator, obtains performance gain in terms of coded BER compared with its counterparts.
- Correlation-based synchronization still works well in the coarsely quantized cases.
- A communication system that supports reliable OFDM transmission with ultra-low resolution ADCs is achievable.

ACKNOWLEDGMENT

We would like to thank Prof. Yonina Eldar and Dr. Nir Shlezinger of Technion-Israel Institute of Technology for their great efforts and valuable suggestions on this work.

REFERENCES

- [1] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [2] B. Murmann. ADC performance survey 1997–2018. [Online]. Available: <http://web.stanford.edu/~murmann/adcsurvey.html>
- [3] A. Mezghani and J. Nosssek, "Analysis of rayleigh-fading channels with 1-bit quantized output," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, Canada, Jul. 2008, pp. 260–264.
- [4] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [5] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *Proc. IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, 2010, pp. 1–5.
- [6] J. Mo and R. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
- [7] N. Liang and W. Zhang, "Mixed-ADC massive MIMO uplink in frequency-selective channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.
- [8] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.
- [9] S. Rini, L. Barletta, Y. Eldar, and E. Erkip. (2017) A general framework for low-resolution receivers for MIMO channels. [Online]. Available: <https://arxiv.org/abs/1702.08133v1>
- [10] C. Risi, D. Persson, and E. G. Larsson. (2014) Massive MIMO with 1-bit ADC. [Online]. Available: <http://arxiv.org/abs/1404.7736>
- [11] A. Mezghani, M.-S. Khoufi, and J. Nosssek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. ITG Workshop Smart Antennas (WSA)*, 2007.
- [12] C. Mollén, J. Choi, E. G. Larsson, and R. Heath, "One-bit ADCs in wideband massive MIMO systems with OFDM transmission," in *Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3386–3390.
- [13] J. Choi, J. Mo, and R. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
- [14] P. N. Alevizos, "Low-complexity multiuser QAM detection for uplink 1-bit massive MIMO," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1592–1595, Aug. 2018.
- [15] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, June 2016.
- [16] Y.-S. Jeon, S.-N. Hong, and N. Lee. (2017) Uplink multiuser massive MIMO systems with low-resolution ADCs: A coding-theoretic approach. [Online]. Available: <https://arxiv.org/abs/1704.03287v1>
- [17] ——. (2017) Supervised-learning-aided communication framework for massive MIMO systems with low-resolution ADCs. [Online]. Available: <https://arxiv.org/abs/1610.07693v3>
- [18] S. Wang, Y. Li, and J. Wang, "Multiuser detection for uplink large-scale MIMO under one-bit quantization," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 4460–4465.
- [19] A. Mezghani and J. Nosssek, "Belief propagation based MIMO detection operating on quantized channel output," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, 13–18 June 2010, pp. 2113–2117.
- [20] T. Zhang, C.-K. Wen, S. Jin, and T. Jiang, "Mixed-ADC massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.
- [21] J. García, J. Munir, K. Roth, and J. A. Nosssek. Channel estimation and data equalization in frequency-selective MIMO systems with one-bit quantization. 2016. [Online]. Available: <http://arxiv.org/abs/1504.04799v1>.preprint.
- [22] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
- [23] P. Sun, Z. Wang, R. W. Heath, and P. Schniter. (2018) Joint channel-estimation/decoding with frequency-selective channels and few-bit ADCs. [Online]. Available: <https://arxiv.org/abs/1807.02494>
- [24] P. Schniter, S. Rangan, and A. K. Fletcher. (2016) Vector approximate message passing for the generalized linear model. [Online]. Available: <https://arxiv.org/abs/1612.01186>
- [25] H. Wang, C. K. Wen, and S. Jin, "Bayesian optimal data detector for mmwave OFDM system with low-resolution ADC," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1962–1979, Sept. 2017.
- [26] J. G. Proakis, *Digital Communications*. Boston, USA: McGraw-Hill Companies, 2007.
- [27] *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012)*, Std., 2012.

- [28] F. Schaich and T. Wildinproceedings, "Waveform contenders for 5G - OFDM vs. FBMC vs. UPMC," in *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, May 2014, pp. 457–460.
- [29] A. A. Zaidi, R. Baldemair, H. Tullberg *et al.*, "Waveform and numerology to support 5g services and requirements," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 90–98, Nov. 2016.
- [30] B. Yin, S. Abu-Surra *et al.*, "High-throughput beamforming receiver for millimeter wave mobile communication," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Atlanta, GA, Dec. 9–13, 2013.
- [31] J. Yang, C. Zhang, S. Jin, C.-K. Wen, and X. You, "Efficient hardware architecture for compressed sensing with DFT sensing matrix," in *Proc. IEEE International Workshop on Signal Processing Systems (SiPS)*, Oct. 26–28, 2016, pp. 207–212.
- [32] J. P. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, Oct. 2013.
- [33] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," *Res. Lab. Electron.*, vol. 216, Mar. 1952.
- [34] H. He, C. Wen, and S. Jin, "Generalized expectation consistent signal recovery for nonlinear measurements," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2333–2337.
- [35] T. Liu, C.-K. Wen, S. Jin, and X. You, "Generalized Turbo signal recovery for nonlinear measurements and orthogonal sensing matrices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 10–15, 2016, pp. 2883–2887.
- [36] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.
- [37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," vol. 6064, 2006.
- [38] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [39] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [40] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, July 1998.
- [41] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.
- [42] X. Gao, Z. Huang, X. Lu, S. Zhang, C. K. Wen, and S. Jin, "Implementation of massive MIMO uplink receiver on rapro prototyping platform," *IEEE Access*, vol. 6, pp. 8133–8147, 2018.
- [43] J. Ma and P. Li, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
- [44] Z. Xue, J. Ma, and X. Yuan, "Denoising-based Turbo compressed sensing," *IEEE Access*, vol. 5, pp. 7193–7204, 2017.
- [45] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (3GPP TS 36.212 version 9.0.0 Release 9)*, Std., 2010.
- [46] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, 1960.
- [47] S. Rangan. (2010) Generalized approximate message passing for estimation with random linear mixing. [Online]. Available: <http://arxiv.org/abs/1010.5141v2>
- [48] T. Goldstein, C. Studer, and R. Baraniuk. (2014) A field guide to forward-backward splitting with a FASTA implementation. [Online]. Available: <http://arxiv.org/abs/1411.3406>
- [49] ——. (2015) FASTA: A generalized implementation of forward-backward splitting. [Online]. Available: <http://arxiv.org/abs/1501.04979>