Low-Complexity Iterative Soft-output Demodulation for Hierarchical Quadrature Amplitude Modulation

Daniel Kekrt and Zdenek Becvar

Dpt. of Telecommunication Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague
Prague, Czech Republic
{kekrtd1, zdenek.becvar}@fel.cvut.cz

Abstract—This paper proposes a novel design of low-complexity soft-output demodulation and soft-output demapping for multi-level iterative decoding of any double-binary code and high-order hierarchical quadrature amplitude modulation (HQAM) schemes. The proposed solution exploits two techniques of self-interference cancellation. The fist one, a blind successive self-interference cancellation, provides a coarse synchronization in an acquisition mode of a receiver. The second one, a hard decision directed parallel self-interference cancellation, is exploited in a tracking mode. The proposed solution is of a very low complexity corresponding only to QPSK demodulation even for modulations of higher orders. Such low complexity allows an efficient implementation of HQAM in mobile and wireless networks with no signaling or coordination between transmitter and receiver required for a selection of modulation. Thus, the proposed approach is suitable for many up-to-date solutions including communication via drones, transparent relaying, or device-to-device communication. The designed solution is verified via a reference implementation of 256-HQAM scheme in FPGA. The results confirm a suitability of the proposed scheme for HQAM demodulation and show that a low bit error rate is achieved by the proposed solution in a wide range of signal to noise ratio.

Index Terms—iterative detection, adaptive soft-output demodulation, soft-output demapping, hierarchical quadrature amplitude modulation, interference cancellation

I. INTRODUCTION

To improve a performance of future mobile networks with a flexible and dynamic architecture, various types of relaying are considered including conventional relay base stations, device-to-device (D2D) relaying [1] or flying base stations mounted on drones [2]. An integration of the D2D relaying or the flying base stations to mobile networks imposes constraints on an energy consumption and a cost for the relays. To reduce both the cost and the energy consumption, the relays should operate in a transparent mode with limited capabilities related to radio resource control and management. Due to the limited capabilities, the transparent relays do not transmit the reference signals (pilots) for a measurement of a received signal power at the receiver for a selection of a proper modulation and/or coding. However, a lack of such reference signal leads to an inability to facilitate an efficient communication with a certain guarantee of quality via the transparent relays (see more details about this issue, e.g., in [3]).

The selection of a suitable modulation even if the received signal power at the relaying link is not measured can be facilitated by a use of hierarchical modulations [4]. The hierarchical modulation is an efficient way of a superposition coding and allows to send a basic data stream with a specific modulation (e.g., QPSK) together with additional data streams with different modulating constellation (e.g., 16-QAM, 64-QAM) at the same channel without mutual impact of these on each other. The actual channel condition then determines which of the transmitted streams are successfully received and decoded and which are not. Such approach requires no interaction and coordination between the transmitter and the receiver and the maximum modulation order given by the channel quality is successfully demodulated on the receiver side.

In parallel, turbo codes [5] are widely used methods for forward error correction (FEC) in mobile networks due to their superior performance. Turbo code decoding in the receiver is typically based on an iterative design [6] with two or multiple soft-in soft-out (SISO) modules and a soft-output demodulator (SODEM). The SISO module is represented by soft decoding algorithms [7], soft (de)mappers, soft broadcasters, soft flow converters, etc. The output of each SISO module is considered as a priori input information for another SISO modules and the SISO modules iterates with each other to decode information.

The SODEM is a front-end of the iterative detector and computes a-posteriori metrics (log-APP) of channel symbols for the decoding. The SODEM closely cooperates with a soft-output (de)mapper (SOMAP). The SOMAP decomposes the output of SODEM to particular soft measures of code symbols. The decomposition is directed by a mapping technique on a side of the transmitter. In [8], [9], the authors focus on a design of the decoding algorithms. However, the work addresses the decoding only separately without respect to SODEM assuming the feed-forward injective interconnection between both functional blocks.

An essential mapping technique for high-order modulations is based on Gray-labeling, where a group of consequent bits of a single-binary code is mapped to particular channel symbols. In this case, the soft inversion of the mapping logic computes log-likelihood ratios (LLR) on the receiver side. A simplification of the Gray-labeled soft demodulation is addressed in, e.g., [10], [11], where the relation between the

(de)mapper and the decoder is assumed as non-iterative feed-forward. The Grey labeled soft demodulators [10] and [11] are affected by an implementation complexity due to recursive formulas and complex searching algorithms. Moreover, the Grey-labelled mapping reaches the maximum achievable data-rate only if the noise level is very low, otherwise, the Grey-labelled mapping does not work at all.

Another mapping technique for high-order modulations is built on multi-level platforms with layered encoders of many possible designs, such as single-binary codes with feed-forward and hierarchical successive SODEM [12] or a design based on a demapper aided by a subsequent re-encoding [13]. However, these solutions contain a consecutive processing, thus, suffer from a high latency.

In this paper, we introduce a novel iterative soft demodulation/(de)mapping for multi-level receiver to eliminate the drawbacks of the existing SODEM for hierarchical modulations. The proposed soft demodulation/(de)mapping principle is based on suboptimal, but numerically effective joint iterative synchronization and detection with two synchronization cores. These cores ensure a successive and parallel inter-layer interference suppression. The resulting SODEM-SOMAP cascade is easy to implement and it is sufficiently fast for a frequent re-computation of the forward a-posteriori metric during the iterative detection. The design of iterative decoding stages is matched to SODEM-SOMAP cascade. The interconnection between SODEM-SOMAP cascade and particular decoders is bidirectional and fully iterative. The proposed solution is robust against the channel noise and enables a successful decoding even at noisy channel.

The rest of the paper is organized as follows. Next section describes a system model for the HQAM with the optimal multilevel iterative receiver that is, however, not friendly for a practical implementation due to a high complexity. Then, Section III outlines the novel design of a low-complexity SODEM-SOMAP cascade for the HQAM schemes of an arbitrary order. Then, in Section IV, we outline a reference implementation of the proposed design for 256-HQAM in FPGA and we verify its efficiency. The last section summarizes major findings and concludes the paper.

II. SYSTEM MODEL

In this section, we define a system model of the communication chain composed of the encoder and modulator on the transmitting side, the wireless communication channel, and the detector on the receiver side. We also describe a common activation schedule of the detector and the architectures of the encoding and decoding networks (structures) on the transmitter and receiver sides, respectively.

A. Transmitter model

The transmitting side is modeled as a layered encoder with HQAM modulator. The output of the HQAM, represented by modulated symbols, is defined as:

$$q[\ell] = \sum_{k=1}^{K} u_k[\ell], \qquad (1)$$

where $u_k[\ell] = 2^{k-1} f_{\text{QPSK}}(c_k[\ell])$ is the particular bit-shifted input QPSK modulated symbol of the code stream c_k with the constellation look-up table $f_{\text{QPSK}} : \{0, 1, 2, 3\} \mapsto \{1 + j; -1 - j; -1 + j; 1 - j\}$, index k denotes the layer in the hierarchical modulation, and ℓ denotes the sampling instant. The particular layers (levels) are of a different resistance to errors due to different energy that grows with k. The first code stream c_1 has the lowest resistance to errors and it carries the source data d_1, which has lowest importance. The last code stream c_K, in contrast, has the highest resistance and it is applicable for transmission of the data d_K with the highest importance. Combination of the individual layers ensures a proper operation and successful decoding of the hierarchical modulation in a wide range of signal to noise ratio (SNR).

B. Wireless channel model

The output symbols of the HQAM modulation is transmitted over a static channel with additive white Gaussian noise (AWGN) $w[\ell]$. Hence, the transmitted symbol $q[\ell]$ at the output of the channel is modeled as $r[\ell] = q[\ell] + w[\ell]$. Similarly as in related works, such as [14], we assume other nuisance parameters with long traces represented by highly correlated stochastic processes (i.e., fading, multipath propagation, symbol timing, frequency offsets, phase offsets, etc.) are removed by a joint soft-decision/hard-decision directed (SDD/HDD) synchronizers or independent data aided (DA) synchronizers. The log-likelihood function [6] for the standard Gaussian distribution is defined as: $M_{\text{AWGN}}(r, \check{q}) = (\Re(r) - \Re(\check{q}))^2 + (\Im(r) - \Im(\check{q}))^2$, where r is the received sample and q is the local replica (or testing estimator), i.e., some value from the alphabet $\{\check{q}\} = \{\check{q}^{(m')} = f(..., \check{u}_k^{(m)}, ...)\}_{m'}$ containing all possible values $\check{q}^{(m')}$ of the variable q that is random from the receiver perspective. Note that, in this paper, the check mark above any variable $\check{x}$ denotes a testing estimator of corresponding unchecked variable (i.e., true realization of a stochastic process).

C. Receiver model

The output of the channel enters an iterative detector on the receiver side. A general multi-level iterative detector is shown in Fig. 1. The first block is the SODEM that calculates the sequence (in time) of the forward a-posteriori metrics $\mathcal{M}_F^{(I)}(\check{q}) = \{\mathcal{M}_F^{(I)}(\check{q})[\ell]\}_\ell = \{M_F^{(I)}(\check{q}^{(m)})[\ell]\}_{m,\ell}$. The superset $\mathcal{M}_F^{(I)}(\check{q})$ of the subsets $\mathcal{M}_F^{(I)}(\check{q})[\ell]$ depends on the set $\hat{\mathcal{R}}^{(I)} = \{\hat{r}^{(I)}[\ell]\}_\ell$ of the synchronized received samples $\hat{r}^{(I)}[\ell]$. Note that the index I denotes the current system iteration. The elements $M_F^{(I)}(\check{q})[\ell]$ of the subsets $\mathcal{M}_F^{(I)}(\check{q})[\ell]$ are normalized metrics. These variables (depending on the testing estimator (.) and time [.]) arise from raw initial metrics $M_{FR}^{(I)}(\check{q})[\ell] = p M_{\text{AWGN}}(\hat{r}^{(I)}[\ell], \check{q})$ through the normalization $M_F^{(I)}(\check{q})[\ell] = M_{FR}^{(I)}(\check{q})[\ell] - \min_{\{\check{q}\}} M_{FR}^{(I)}(\check{q})[\ell]$ that shifts the smallest element to the zero. The normalization serves an arithmetic stabilization and also saves memory. The raw measurements $M_{FR}^{(I)}(\check{q})[\ell]$ depends on the precision factor $p = 1/2\sigma^2$ that is inversely proportional to the AWGN power denoted as a noise variance σ^2.

Fig. 1. Block scheme of multi-level iterative detector.

Fig. 2. Activation schedule of common HQAM detector.

The relationship between the SODEM and the iterative decoders is provided by the SOMAP. The activation of memory-less SOMAP leads to updating $\{\mathcal{M}_F^{(I)}(\check{q})[\ell], \{\mathcal{M}_B^{(I)}(\check{c}_k)[\ell]\}_k\} \mapsto \{\mathcal{M}_F^{(I+1)}(\check{c}_k)[\ell]\}_k$ and $\mathcal{M}_B^{(I+1)}(\check{q})[\ell]$, where $\mapsto$ denotes the symbol-oriented combination-marginalization process (factorizable in time) described in detail in following paragraphs. The resulting sequences of the forward a-posteriori metrics $\mathcal{M}_F^{(I+1)}(\check{c}_k) = \{\mathcal{M}_F^{(I+1)}(\check{c}_k)[\ell]\}_\ell = \{M_F^{(I+1)}(\check{c}_k^{(m)})[\ell]\}_{m,\ell}$ are used in the particular decoders while the sequence of the backward metrics $\mathcal{M}_B^{(I+1)}(\check{q}) = \{\mathcal{M}_B^{(I+1)}(\check{q})[\ell]\}_\ell = \{M_B^{(I+1)}(\check{q}^{(m)})[\ell]\}_{m,\ell}$ is used by external SDD synchronizers [15] to update the estimation $\hat{\mathcal{R}}^{(I+1)} = \{\hat{r}^{(I+1)}[\ell]\}_\ell$ based on the observed realization $\mathcal{R} = \{r[\ell]\}_\ell$ of the received signal.

The soft decoding process leads to a continuous updating of the sequences $\{\mathcal{M}_F^{(I)}(\check{c}_k), \mathcal{M}_B^{(I)}(\check{d}_k)\} \mapsto \mathcal{M}_F^{(I)}(\check{d}_k)$ and $\mathcal{M}_B^{(I)}(\check{c}_k)$, where $\mapsto$ refers to the general sequence-oriented combination-marginalization process. The apriori metric $\mathcal{M}_B^{(I)}(\check{d}_k)$ is constant and reflects stochastic properties of the k-th data source. The resulting backward metrics $\mathcal{M}_B^{(I)}(\check{c}_k)$ serves to the SOMAP for the next calculation of the forward metrics $\mathcal{M}_F^{(I+1)}(\check{q})$ in the system iteration $I + 1$. Through the SOMAP, the decoders interacts and iterate with each other.

The de-mapping of symbols is performed via an optimal max-log-APP based soft de-mapper. This de-mapper performs combinational $(+)$ and marginalization $\min(.)$ processing (see [6] for more details). The combinational processing combines the input metrics to the joint a-posteriori metric

$$M_{\text{SOMAP}}^{(I)}(\check{q})[\ell] = M_F^{(I)}(\check{q})[\ell] + \sum_{\{\check{c}_k\} \mapsto \check{q}} M_B^{(I)}(\check{c}_k)[\ell], \quad (2)$$

where the set $\{\check{c}_k\} \mapsto \check{q}$ contains all combinations of the testing estimators $\check{c}_k$ that leads to

$$\check{q}[\ell] = \sum_{k=1}^{K} 2^{k-1} f_{\text{QPSK}}(\check{c}_k[\ell]). \quad (3)$$

Then, the marginalization processing with an elimination of the apriori information [6] takes place so that

$$M_{BR}^{(I+1)}(\check{q})[\ell] = M_{\text{SOMAP}}^{(I)}(\check{q})[\ell] - M_F^{(I)}(\check{q})[\ell] \quad (4)$$

$$M_{FR}^{(I+1)}(\check{c}_k)[\ell] = \left(\min_{\{\check{q}\}:\check{c}_k} M_{\text{SOMAP}}^{(I)}(\check{q})[\ell] \right) - M_B^{(I)}(\check{c}_k)[\ell] \quad (5)$$

The set $\{\check{q}\} : \check{c}_k$ contains all testing estimators $\check{q}$ that may arise from the specific fixed $\check{c}_k$ and any ambiguous $\check{c}_{k' \neq k}$ through (3).

Finally, the both raw variables are normalized to the outputs $M_F^{(I+1)}(\check{c}_k)[\ell] = M_{FR}^{(I+1)}(\check{c}_k)[\ell] - \min_{\{\check{c}_k\}} M_{FR}^{(I+1)}(\check{c}_k)[\ell]$ and $M_B^{(I+1)}(\check{q})[\ell] = M_{BR}^{(I+1)}(\check{q})[\ell] - \min_{\{\check{q}\}} M_{BR}^{(I+1)}(\check{q})[\ell]$.

The quantity of the metrics $M_{\text{SOMAP}}^{(I)}(\check{q})[\ell]$ and $M_F^{(I)}(\check{q})[\ell]$ is 4^K for each symbol. Therefore the direct application of the optimal SODEM-SOMAP cascade in practice is not feasible and we propose a novel solution that eliminates this drawback.

The detection process is closed by the decision block (DEC), which selects the output so that $\hat{d}_k^{(I)}[\ell] = \arg\min_{\{\check{d}_k\}} M_F^{(I)}(\check{d}_k)[\ell] + M_B^{(I)}(\check{d}_k)[\ell]$. The sequence of estimates $\hat{\mathcal{D}}^{(I)} = \{\hat{d}_k^{(I)}[\ell]\}_{k,\ell}$ is stored in the memory (MEM) for a consequent link layer processing.

D. Activation schedule

The iterative detector in Fig. 1 is activated according to the schedule in Fig. 2. This activation schedule shows the order of an iterative activation of individual blocks to illustrate a spreading and updating of the soft information through the detector over time. First, the SODEM is activated and its input is inserted to SOMAP. The metrics $M_B^{(0)}(\check{c}_k) = 0$ are initially set to zero and the segment $\hat{\mathcal{R}}^{(0)}$ is initially synchronized and extracted by the data aided synchronizer. This activation progress provides an input soft information for all decoders. The decoders than perform one iteration according to their internal activation scheme. Then, the resulting metrics $M_B^{(1)}(\check{c}_k)$ are looped back to the SOMAP for the next iteration. All following iterations are performed in a similar manner. The number of performed iterations is upper bounded by hardware capabilities. The actual number of iteration is then dependent on an actual channel condition and required output data-rate.
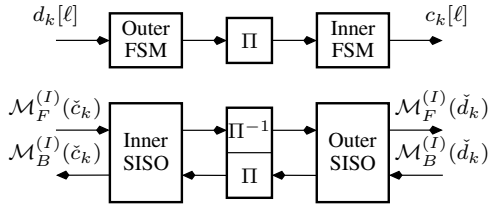
Fig. 3. Architecture of serial encoding (upper subplot) and iterative decoding (bottom subplot) networks.

Note that the tracking mode synchronization of residual nuisance parameters (remaining after their removal as explained in Section II.B) with a long trace is performed simultaneously with the iterative detection when the updated SOMAP output $\mathcal{M}_B^{(I)}(\check{q})$ passes through the SDD synchronizer(s). The returned and improved sequence $\hat{\mathcal{R}}^{(I+1)}$ is re-calculated by the SODEM to the new metrics $\mathcal{M}_F^{(I+1)}(\check{q})$ for the next iteration.

E. Architecture of encoding and iterative decoding networks

Encoding and decoding is done via convolutional coders. The coders and decoders are composed of finite state machines (FSMs) and SISO modules, respectively, concatenated in either serial or parallel way. Our proposed solution is suitable for both serial as well as parallel concatenations. For sake of clarity, in this paper, we focus on serial concatenation of the convolutional coders (SCCC) as shown in Fig. 3. The blocks Π and Π^{-1} represent forward and reverse interleavers, respectively.

III. PROPOSED LOW-COMPLEXITY SODEM FOR HQAM

The proposed low-complexity SODEM, shown in Fig. 4, is based on a principle of a joint iterative synchronization and detection. The proposed scheme, denoted as the low-complexity adaptive SODEM (LASODEM), exploits serial pipelined processing and contains two synchronization cores represented by HQAM demappers. The first core is non-data aided (NDA) and serves for a synchronization acquisition. The second core is hard decision directed (HDD) and ensures a tracking mode of the receiver. Both cores are complemented with a quality measurement logic deciding which synchronization core is used for the ℓ-th sample and with a simple QPSK SODEM shared by all layers. Individual parts are described in following subsections.

A. Non-data aided HQAM demapping

The blind successive self-interference cancellation (SIC) is applied in the NDA core. Note that for HQAM demapping purposes, the self-interference for the layer k is represented by QPSK modulation stream $u_{k' \neq k}[\ell]$ from any superior or inferior layers k'. The residual received signal is computed by the recurrent equation

$$\hat{r}_k^{(I)}[\ell] = \begin{cases} \hat{r}^{(I)}[\ell] & k = K \\ \hat{r}_{k+1}^{(I)}[\ell] - \hat{u}_{k+1}^{(I)}[\ell] & \text{otherwise} \end{cases}, \quad (6)$$
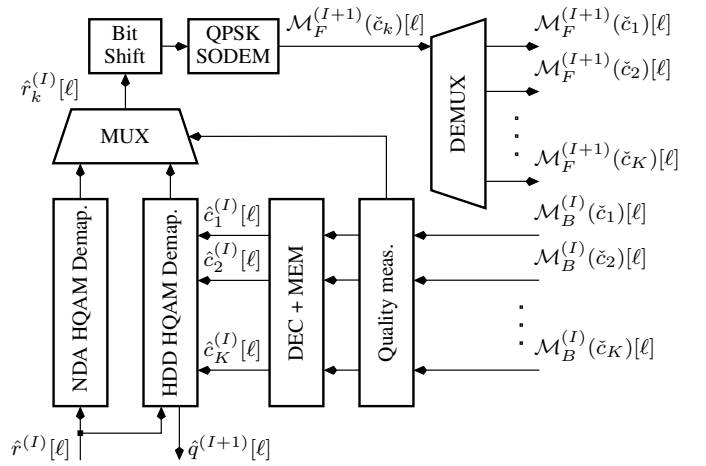
Fig. 4. Proposed LASODEM with low complexity for HQAM.

where $\hat{u}_k^{(I)}[\ell] = 2^{k-1}\text{sign}(\hat{r}_k^{(I)}[\ell])$ is the blind layer decision based on $\text{sign}(z) = |z|^{-1}z$. In the case of NDA demapping, the superior self-interferences are removed only and all inferior self-interferences are considered as an additional part of channel noise. The NDA core is activated at the beginning of the decoding process when a valid backward soft information $\mathcal{M}_B^{(I)}(\check{c}_k)[\ell]$ is not available yet.

B. Hard decision directed HQAM demapping

The parallel self-interference cancellation (PIC) is applied in the HDD core. The residual received signal in the HDD core is calculated as

$$\hat{r}_k^{(I)}[\ell] = \hat{r}^{(I)}[\ell] - \hat{q}^{(I)}[\ell] + \hat{u}_k^{(I)}[\ell], \quad (7)$$

where $\hat{q}^{(I)}[\ell]$ is the output of HQAM mapper (as defined in (1)) if and only if the inputs are $\hat{u}_k^{(I)}[\ell] = 2^{k-1}f_{\text{QPSK}}(\hat{c}_k^{(I)}[\ell])$ and $\hat{c}_k^{(I)}[\ell] = \arg\min_{\{\check{c}_k\}} \text{M}_B^{(I)}(\check{c}_k)[\ell]$. In this case, the self-interferences are mutually removed across all layers and only the AWGN remains.

The HDD core is activated when the backward metrics $\mathcal{M}_B^{(I)}(\check{c}_k)[\ell]$ reach the sufficient quality. The quality is determined by the quality measurement block (see Fig. 4). The set of output metrics of the demapping cores $\mathcal{M}_B^{(I)}(\check{c}_k)[\ell]$ is valid if only one element in the set is equal to zero and all other elements are non-zero and higher than the chosen quality threshold QTH_k. If the metrics are valid at the current layer k or at the arbitrary inferior layer, the output $\hat{r}_k^{(I)}[\ell]$ of the HDD core is used for further processing. Otherwise the output of the NDA core is used.

C. QPSK soft-output demodulation

One QPSK SODEM block with a fixed structure is reused by all layers of the demodulator. Thus, the complexity is reduced and corresponds only to the complexity of QPSK SODEM even if the demodulation of higher order modulations is possible. The residual signal $\hat{r}_k^{(I)}[\ell]$ of the demmaper is first scaled (bit shifted) as the level of the signal decreases
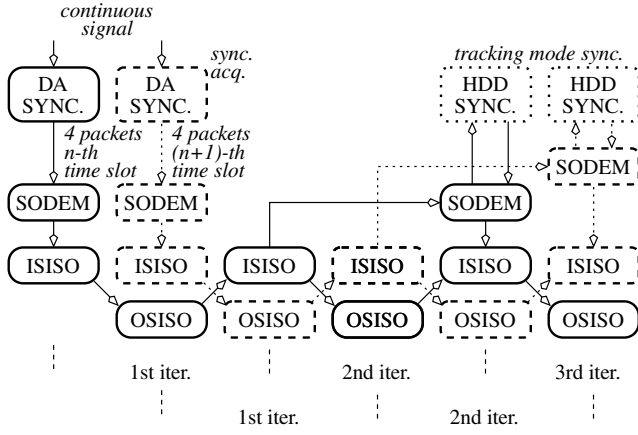
Fig. 5. Activation schedule of reference 256-HQAM detector.

Fig. 6. Reference implementation of 256-HQAM detector with LASODEM.

with the layer (i.e., with the index k). Then, the soft demodulation and demapping process starts with $k = K$. The metric $\mathrm{M_{AWGN}}(2^{K-k}\hat{r}_k^{(I)}[\ell], \check{u})$ of the shifted residual signal $2^{K-k}\hat{r}_k^{(I)}[\ell]$ is computed for all $\check{u} \equiv \check{u}_K$ and stored. Then, we find the argument of the smallest metric so that $\hat{v}_k^{(I)}[\ell] = \arg\min_{\{\check{u}\}} \mathrm{M_{AWGN}}(2^{K-k}\hat{r}_k^{(I)}[\ell], \check{u})$.

Now, we approximate, with respect to the definition of channel model (see Section II.B), the instantaneous noise power $(\hat{\sigma}_k^{(I)}[\ell])^2$ of the k-th layer by the smallest metric. The approximation assumes $|w[\ell]| < 2^{K-k}$ and the modulation components from the inferior layers are considered as the part of the additive noise together with $w[\ell]$ during the synchronization acquisition phase consider. Accordingly, the mean precision factor p is:

$$\hat{p}_k^{(I)} = 2^P L \left(\sum_{\ell=1}^{L} (\hat{\sigma}_k^{(I)}[\ell])^2 \right)^{-1}, \tag{8}$$

where $(\hat{\sigma}_k^{(I)}[\ell])^2 \approx \mathrm{M_{AWGN}}(2^{K-k}\hat{r}_k^{(I)}[\ell], \hat{v}_k^{(I)}[\ell])$ is the stored smallest metric, the positive integer P is set so that the estimation $\hat{p}_k^{(I)}$ fits into a suitable dynamic range, L is the integration length set to allow a usage of a slow and simple serial divider for calculating of (8). The precision factor is updated once per L symbols and its initial value is determined as $\hat{p}_k^{(I)} = 2^P$ at the beginning of each block of data.

The required estimations of the raw forward a-posteriori metrics are the bit-shifted products

$$\mathrm{M}_{FR}^{(I)}(\check{c}_k)[\ell] \approx \frac{\hat{p}_k^{(I)} \mathrm{M_{AWGN}}(2^{K-k}\hat{r}_k^{(I)}[\ell], \check{u})}{2^P} \tag{9}$$

of the stored squared Euclidean distances (as a function of $\mathrm{M_{AWGN}}$ defined in Section II.B) and the estimated precision factor.

Finally, we reuse the stored argument of the smallest metric $\hat{v}_k^{(I)}[\ell]$ once more for the normalization $\mathrm{M}_F^{(I)}(\check{c}_k)[\ell] = \mathrm{M}_{FR}^{(I)}(\check{c}_k)[\ell] - \mathrm{M}_{FR}^{(I)}(\hat{c}_k[\ell])[\ell]$, where the pointer $\hat{c}_k[\ell] = f_{\mathrm{QPSK}}^{-1}(2^{1-K}\hat{v}_k^{(I)}[\ell])$ is obtained through the inverse unambiguous mapping $f_{\mathrm{QPSK}}^{-1} : \{1 + j; -1 - j; -1 + j; 1 - j\} \mapsto \{0, 1, 2, 3\}$.
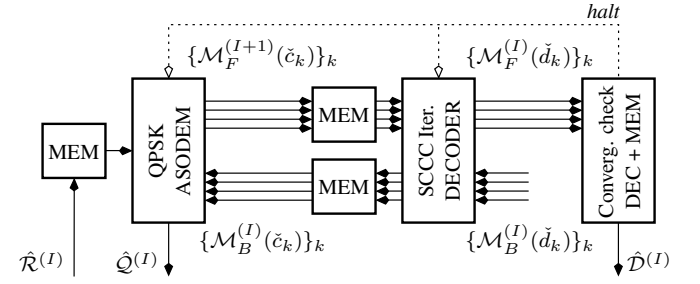
Then, the layer index k is decremented by "1" and demodulation on the next (inferior) level starts. This process is repeated for each received sample and each layer.

D. Activation schedule of 256-HQAM detector

Now, we outline the activation schedule for the 256-HQAM detector in Fig. 5. Note that this activation schedule corresponds to the reference implementation outlined in Fig. 6. The maximum bit-rate is scalable and depends on available memory capacity. We can distinguish low- and high-end variants. For the low-end variant with reduced cost and memory, the activation schedule works in a single time-slot mode with four packets decoded together. In this case, only the branch depicted by solid line in Fig. 5 is used and the detector consumes four clock ticks per 1 bit and 1 iteration and the processing is purely serial with the FBA cores working on 25% of capacity only. Another option of the low-end variant uses a double time-slot mode and eight packets are decoded together. In this case, only a half of the activation schedule is performed again, but the forward and backward state recursions in both SISO modules are calculated simultaneously. Hence, the detector consumes 2 clock ticks per 1 bit and 1 iteration and the FBA cores work on 50% of capacity. Note that the outer SISO (OSISO) and inner SISO (ISISO) alternate their operation.

The high-end variant allocates four time-slots to enable decoding of 16 packets together. The both twisted branches in Fig. 5 are activated and internal state recursions are performed simultaneously. The detector spends only 1 clock tick per 1 bit and 1 iteration. The FBA cores are fully exploited if the length of packets is same in all consequent time-slots.

The detection process is terminated by the convergence checking logic if two consecutive sets of metrics $\mathcal{M}_F^{(I-1)}(\check{d}_k)$ and $\mathcal{M}_F^{(I)}(\check{d}_k)$ are the same for all K layers. In such case, the hard decisions $\hat{\mathcal{D}}^{(I)}$ are forwarded to the link layer. Otherwise, the detection process continues until the maximum number of iterations is performed. Then, only the hard decisions $\hat{\mathcal{D}}_k^{(I)}$ from the stable layers are forwarded to the link layer. The hard decisions from other layers at which the stable state is not reached are discarded.

IV. PERFORMANCE EVALUATION

In this section, we first describe a reference implementation of HQAM demodulator in FPGA. Then, we present and

analyze the results obtained from FPGA implementation to demonstrate an efficiency of the proposed LASODEM.

A. Reference implementation of 256-HQAM detector in FPGA

The reference implementation in Intel Cyclone V FPGA to prove the proposed concept is shown in Fig. 6. We validate the proposal for detection of 256-HQAM scheme containing 4-layer pipelined SCCC in line with the system model. In the transmitter, the outer FSM with binary convolutional recursive code with the rate of 1/2 is considered and we implement 8-state redundant outer code adopted in 4G LTE-A standard with the generator polynomials $G(D) = (1 + D + D^3)/(1 + D^2 + D^3)$. The inner FSM contains a modulo 4 integrator with code rate 1/1. The integrator ensures a resistance to the nuisance phase rotations of 0, $\pi/2$, π and $3\pi/2$. The interleaver in the SCCC are based on quadratic permutation polynomials [16]. In the receiver, each SISO module in decoder contains two processing units for forward and backward state recursion respectively. The processing units have single (2-radix) pipelined core able to decode four packets simultaneously in the chop mode. The results are obtained using an automated testbench with an AWGN emulator combining the Box-Muller transform and the central limit theorem [17].

B. Results and discussions

In Fig. 7, we show the impact of SNR on bit error rate (BER) of the proposed LASODEM for each particular turbo-code layers (TCL). When the proposed scheme operates without the proposed QTH and the convergence check, BER is relatively high even for a high SNR. An improvement of up to 8 dB (observed for the 2nd layer) is accomplished by enabling QTH. Note that we set QTH to 2, 3, 5, and 9 for 1st, 2nd, 3rd, and 4th layer, respectively, as these values lead to the lowest BER as tested experimentally. Additional noteworthy reduction in BER is reached by enabling the QTH together with the convergence check. This complete version of the proposed LASODEM leads to an outstanding suppression of BER below 10^{-4} for all levels of SNR. Moreover, for SNR>5 dB and SNR>7 dB, the complete LASODEM pushes BER further down below 10^{-5} and 10^{-6}, respectively. Such a low BER makes the LASODEM suitable for a wide range of applications in mobile and wireless networks, as this BER is largely below requirements of common communications standards. Note that the proposed solution requires, in average, only four iterations to reach the correct estimation for SNR over 30 dB.

Note that existing SODEMs for HQAM are of a huge complexity and their implementation is not possible. Thus, for benchmarking purposes, we present a performance of uncoded layers (UCLs), when data bits d_k are directly applied in as indicated in (1) through BPSK and estimated by the successive demapping core (6) on the receiver side. The performance of UCL is notably worse comparing to our proposed LASODEM and BER is alwas above 10^{-4} even for SNR of 30 dB.

In Fig. 8, we show a relative throughput (RTP) achieved by the proposed solution. The RTP is understood as the ratio of
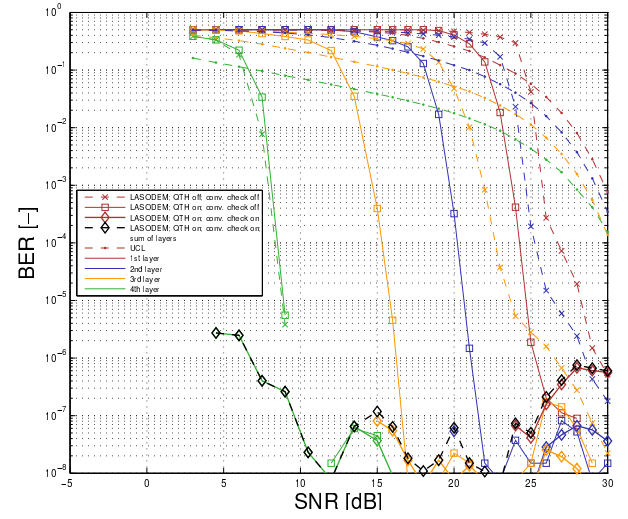

Fig. 7. Impact of SNR on BER for individual layers (TCL) and impact of convergence check and QTH on BER of proposed LASODEM.
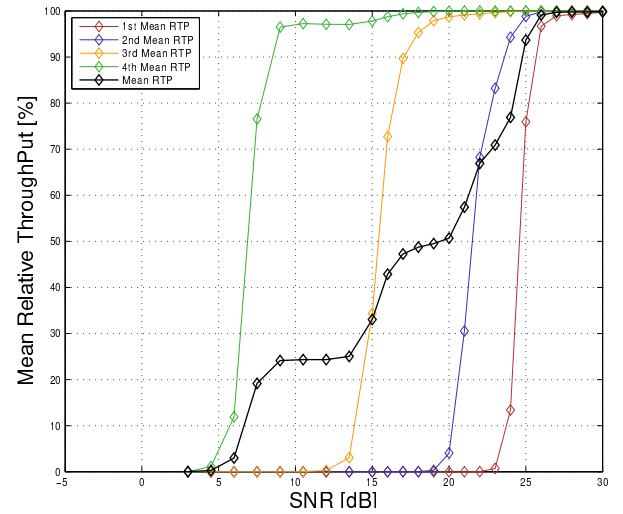

Fig. 8. Impact of SNR on relative throughput (RTP) achieved by proposed LASODEM for 256-HQAM. Colors represent individual layers TCL.

the approved and rejected packets by the convergence checking logic. Note that we assume QTH set to 2, 3, 5, 9, respectively, for individual layers and the results are obtained for same random payload of 4×2^{33} bits segmented into packets of random length out of 128/256/512/1024/2048 symbols. The figure shows that individual layers reach more than 95% of their individual maximum RTP at 9, 18, 25, and 27 dB, respectively.

Note that the existing high-order QAM modulations using Gray-labeled LDPC codes [18] or polar codes [19] can reach higher data-rates in comparison to the introduced LASODEM. However, these are strictly limited for a high SNR regions only and the performance for a low SNR is not sufficient. Moreover, for the low SNR, the modulation scheme should be changed (modulation order decreased) for both Gray-labeled LDPC codes [18] as well as polar codes [19] and, thus, a coordination

TABLE I
OVERVIEW OF THE NUMBER OF CONSUMED UNITS IN FPGA TO
ILLUSTRATE COMPLEXITY OF THE LASODEM.

Entity / Sub-entity	AML	DSP	M10k
Whole 256-HQAM LASODEM detector	**2480**	**12**	**98**
QPSK demodulator for LASODEM	380	8	2
SCCC decoder	1780	4	84
Inner SISO (max-log-APP)	940	0	40
Outer SISO (max-log-APP)	720	0	44

between the transmitter and the receiver is required.

The compilation results, representing the complexity of the proposed LASODEM in FPGA, are reported in Table I. The table presents the number of consumed adaptive logic modules (AML), digital signal processing blocks (DSP) representing embedded multipliers, and memory blocks (M10k). The detector is compiled in the double time-slot variant and for the maximum packet length of 1024 symbols. The table illustrates that only 2480 AMLs, 12 DSPs and 98 memory blocks M10k. Most of these (1780, 4, and 84 respectively) are consumed by the SCCC decoder. Such low numbers of individual elements of the FPGA confirm the efficiency of the designed LASODEM.

V. CONCLUSION

In this paper, we have proposed and verified the novel design of low-complexity adaptive soft-output demodulator for high-order HQAM schemes. The solution is based on a joint iterative synchronization and detection technique and the complexity is linearly proportional to the number of modulated double-binary streams. All existing solutions for HQAM demodulation are of a high complexity that does not allow their practical implementation. However, our proposed scheme is of a very low complexity corresponding only to QPSK demodulation even for a high order modulation. Thus, our concept enables an exploitation of HQAM in real world applications, such as flying base station acting as transparent relay or D2D communication while preserving benefits of HQAM no need for apriori modulation format setting for communication. The proposed concept is tested in the Intel Cyclone V FPGA. The results confirm that the proposed LASODEM reaches very low BER (below 10^{-6}) already at SNR = 7 dB. This makes it suitable for all common mobile and wireless networks.

In the future, the LASODEM should be evaluated and validated for different coding and optimized also for another types of hierarchical modulations.

REFERENCES

[1] A. Asadi, Q. Wang, V. Mancuso, A Survey on Device-to-Device Communication in Cellular Networks IEEE Communications Surveys & Tutorials, vol. 16, no. 4, 2014.
[2] M. Mozaffari, et al, A tutorial on FlyBSs for wireless networks: Applications, challenges, and open problems, IEEE Communications Surveys & Tutorials, 2019.
[3] Motorola, Discussion of Type II (Transparent) Relays for LTE, 3GPP TSG RAN WG1 Meeting #57, R1 091941, San Francisco, USA, 2009.
[4] H. Jiang, P.A. Wilford, "A Hierarchical Modulation for Upgrading Digital Broadcast Systems," IEEE Transactions on Broadcasting, 2005.
[5] K. Gracie and M.H. Hamon, "Turbo and turbo-like codes: Principles and applications in telecommunications," *Proc. IEEE*, vol. 95, 2007.
[6] K. Chugg, A. Anastasopoulos, X. Chen, *Iterative detection: Adaptivity, Complexity reduction and Applications*. Kluwer Academic Pub., 2001.
[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284-287, Mar. 1974.
[8] R. Shrestha and R. P. Paily, "High-throughput turbo decoder with parallel architecture for LTE wireless communication standards," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 9, Sept. 2014.
[9] R. Shrestha, A. Sharma, "VLSI-Architecture of Radix-2/4/8 SISO Decoder for Turbo Decoding at Multiple Data-rates," IFIP/IEEE VLSI-SoC, 2018.
[10] Q. Wang, Q. Xie, Z. Wang, S. Chen, and L. Hanzo, "A universal low-complexity symbol-to-bit soft demapper," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 119130, Jan. 2014.
[11] C.W. Chang, P.N. Chen, and Y.S. Han, "A systematic bit-wise decomposition of M-ary symbol metric," *IEEE Trans. Wireless Commun.*, 2006.
[12] J. Kim, S. Lee, and J. Seo, "Successive MAP detection with soft interference cancellation for iterative receivers in hierarchical M-ary QAM systems," *IEEE Vehicular Technology Conference*, 2009.
[13] G. Gül, et al, "Low complexity demapping algorithms for multilevel codes," *IEEE Trans. Commun.*, 2011.
[14] T. Cui, F. Gao, A. Nallanathan, H. Lin and C. Tellambura, "Iterative Demodulation and Decoding Algorithm for 3GPP/LTE-A MIMO-OFDM Using Distribution Approximation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1331-1342, Feb. 2018.
[15] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise and L. Vandendorpe, "Turbo synchronization: an EM algorithm interpretation," *IEEE ICC*, May 2003.
[16] O. Y. Takeshita, "Permutation polynomial interleavers: An algebraic-geometric perspective," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, 2007.
[17] J. L Danger, A. Ghazel, E. Boutillon, and H. Laamari, "Efficient FPGA implementation of Gaussian noise generator for communication channel emulation," *in Proc. Int. Conf. Elec. Circuits Syst.*, Dec. 2000, pp. 366-369.
[18] B. Zheng, L. Deng, M. Sawahashi and N. Kamiya, "High-order circular QAM constellation with high LDPC coding rate for phase noise channels," *20th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2017, pp. 196-201.
[19] C. Cao, T. Koike-Akino, Y. Wang and S. C. Draper, "Irregular Polar Coding for Massive MIMO Channels," *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1-7.