Blind Vector Parameter Estimation for Burst Type CPM Transmissions

ANDREAS LANG AND BERTHOLD LANKL

Institute of Information Technology, Bundeswehr University Munich, 85579 Neubiberg, Germany

CORRESPONDING AUTHOR: ANDREAS LANG (e-mail: andreas.lang@unibw.de)

ABSTRACT Short burst continuous phase modulation transmission is of practical relevance in e.g. frequency hopping systems applied in sensor or tactical networks. The channel conditions can be seen as mutually uncorrelated for each burst due to spectral and or temporal separation of those. Because of this time variant nature, a recurring acquisition of the impairment parameters is required for each burst. In this work, a blind joint estimation of several parameters in a flat fading environment for continuous phase modulation bursts is realized by the expectation maximization algorithm. The main contributions are first the formulation of the expectation and maximization steps to enable the joint computation of the maximum likelihood parameter estimates and second the analysis of the likelihood functions to obtain an optimized initialization for the algorithm. It is shown, that the joint estimator produces unbiased estimates and its performance in terms of the mean squared estimation error achieves the theoretical limits, i.e. the modified Cramér-Rao-Vector-Bound and slightly outperforms a state of the art pilot based estimator. Furthermore, the effective throughput is discussed and bit and frame error rates are compared to each other and to the perfectly synchronized estimator. Its computational complexity is analyzed and efficient computation steps and further approaches are outlined to decrease it.

INDEX TERMS CPM, joint estimation, expectation maximization, burst transmission, frequency hopping, VHF, UHF, frequency offset, phase offset, timing offset.

I. INTRODUCTION

Intentional or random interference can severely deteriorate the performance of communication systems. This is often tackled by frequency hopping (FH) systems, where transmission times in the order of $T_{\text{Burst}} = 1\,\text{ms}$ for single bursts are used. Continuous phase modulation (CPM) is a popular modulation choice in at least two application fields where FH is used. In Internet of things (IoT) networks, FH is mainly deployed to avoid random, bad channel conditions whereas in tactical networks also potential jamming devices shall be preempted [1], [2]. These networks can often be found in the very and ultra high (VHF and UHF) frequency bands, specifically in a range of 100 MHz to 600 MHz.

Such short burst durations pose difficulties in synchronizing the received signal, for which channel parameters have to be estimated and corrected. In this work, the transmission of CPM signals over a flat fading channel is considered with carrier and clock impairments, i.e. the fading factor (FF), carrier frequency offset (CFO), carrier phase offset (CPO) and

timing offset (TO) shall be estimated. The estimation problem can be tackled in numerous ways.

Non-Data Aided (NDA): NDA techniques usually utilize the signal statistics and thus are problematic for short bursts. Several NDA estimation techniques for the above channel parameters for CPM transmissions can be found in the literature, though many methods are limited to MSK-type transmissions [3], [4], [5] and thus are not applicable in general CPM configurations. There are also some methods that work with arbitrary configurations [6], [7], [8], [9], [10], though do not provide very accurate estimates in terms of the mean squared estimation error (MSEE) as shown in [11], [12]. Furthermore, these methods often only tackle the estimation of a single parameter.

Decision Directed: Similar to the NDA methods, no pilot sequences are utilized in the parameter estimation task. Instead initial estimates are improved by taking decision about data symbols into account. Though typical phase locked loop (PLL) methods [13] or Kalman filter approaches [14] use
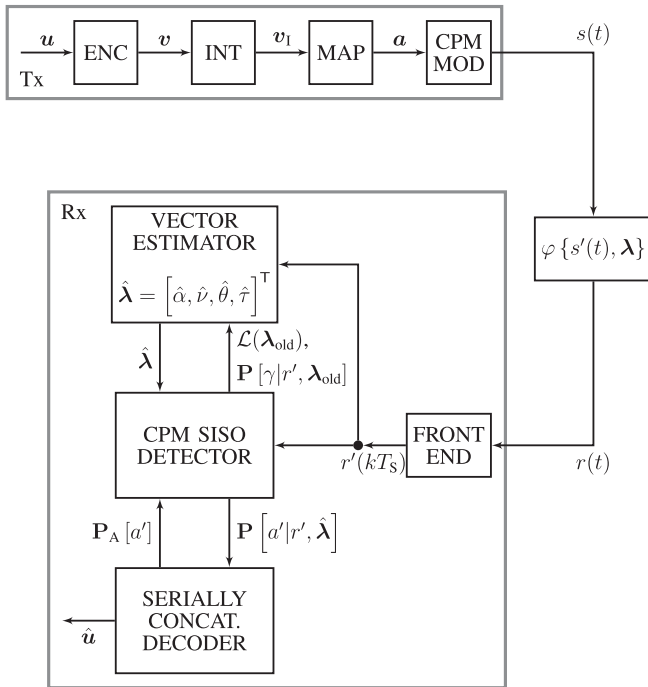
FIGURE 1. Overview of the whole transmission system. The transmitter encodes, interleaves and maps binary information symbols *u* to CPM symbols *a* that modulate the CPM signal *s(t)*. The channel output *r(t)* is tilted, convolved with the receive filter $H_R(f)$ and sampled in the front end block. Given the input $r'(kT_S)$, the CPM SISO detector is embedded in a doubly Turbo scheme. On the one hand, it exchanges likelihoods $\mathcal{L}(\boldsymbol{\lambda}_{old})$ and posterior probabilities $P[\gamma|r'(kT_S), \boldsymbol{\lambda}_{old}]$ for parameter estimates $\hat{\boldsymbol{\lambda}}$ with the vector estimator. On the other, it employs a serially concatenated CPM setup with a channel decoder, where a posteriori symbol probabilities $P[a'|r,'\hat{\lambda}]$ are exchanged for extrinsic information, i.e. the a priori probabilities $P_A[a']$.

pilots to obtain their initial estimates before tracking the errors over the course of a larger frame. For that reason they are not applicable in burst type transmission due to their acquisition constraints. For a subset of the channel parameters considered in this work, optimal DD estimators were presented by the authors in [11], [12].

Data Aided (DA): The most cited work on this subject in the recent past is a technique that utilizes a pilot sequence [15]. While the MSEE is very close to the theoretical minimum, there are some drawbacks to this method that are inherent to DA techniques. First, the spectral efficiency is degraded by the pilot sequence. Second, known pilot sequences can easily be detected and therefore disturbed, i.e. jammed. Jamming the pilot sequence for synchronization is a very efficient way to impede the complete transmission. This point is especially relevant in tactical networks.

In this work, a non-pilot, decision-directed estimator of FF, CFO, CPO and TO, which relies on the expectation maximization (EM) algorithm is proposed. The above referenced NDA and DD methods mostly cover scalar (i.e. single) parameter estimators, while a vector (i.e. joint) parameter estimator is required in case of the channel model considered in this work. For that reason, the vector pilot based (VPB) estimator

from [15] serves as comparison for the vector expectation maximization (VEM) parameter estimator that will be introduced in this paper.

The remainder of the paper is structured as follows: Section II reflects the system model including the CPM signal representation, the channel model and crucial elements of the receiver, which is in particular the typical serially concatenated CPM system involving the BCJR algorithm in the soft-input-soft-output (SISO) CPM detector. The two main contributions of deriving the EM steps for the case of channel parameter estimation in burst-type CPM and an analytical optimization of the algorithm's initialization are covered in Section III. Subsequently in Section IV, implementation aspects are detailed using an algorithmic description of the estimator and thoughts on efficiency are given. The performance analysis in Section V investigates the mean estimated value of both the VPB and VEM estimator as well as their MSEE compared to theoretical limits, the effective throughputs and the decoded bit and frame error rates. Before concluding the paper, the computational complexity of the proposed method is quantitatively discussed in Section VI.

II. SYSTEM MODEL
This section will step through the links of the transceiver chain and explain in detail the channel model. Fig. 1 summarizes this section visually. The deployment of a double Turbo setup is highlighted here, first the Turbo detection in the serially concatenated CPM scheme from [16] and second the Turbo estimation by means of the EM algorithm explained in detail in Section III.

A. EQUIVALENT LOWPASS DOMAIN CPM SIGNAL
In the equivalent lowpass domain, the burst CPM signal consisting of N_{CS} symbols is expressed as

$$s(t) = \begin{cases} \sqrt{\frac{E_S}{T}} \exp\left(j\phi(t, \boldsymbol{a})\right) & 0 \leq t < N_{CS}T \ , \\ 0 & \text{otherwise} \ . \end{cases} \quad (1)$$

The signal's amplitude is the square root ratio of the symbol's energy E_S and its duration T. The phase function $\phi(t, \boldsymbol{a})$ for $0 \leq t < N_{CS}T$ is defined by

$$\phi(t, \boldsymbol{a}) = 2\pi h \sum_{n=-(L-1)}^{N_{CS}-1} a_n q(t - nT) \quad (2)$$

in which $h = P/Q$ defines the modulation index as the fraction of two mutually prime integers P and Q. With N_{CS} symbols being transmitted, the negative subscripts of a only account for a phase initialization. The n'th CPM symbol a_n from an M-ary alphabet $\{-(M-1), \ldots, -1, +1, \ldots, +(M-1)\}$ modulates the phase pulse $q(t)$ with values equal to zero for $t \leq 0$ and 0.5 for $t \geq LT$. Its derivative $g(t) = dq(t)/dt$ is typically of a rectangular (REC), raised cosine (RC) or Gaussian (GA) shape. While the first two shapes are temporally restricted to LT with L being the pulse length, the latter is not. Usually Gaussian pulses are truncated after a few symbols
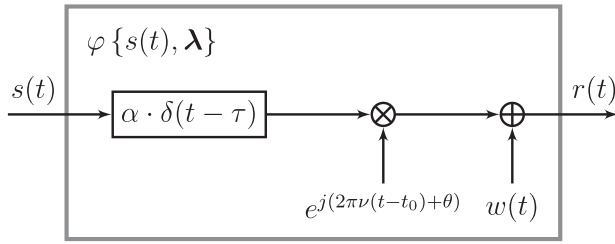
FIGURE 2. Block diagram of the static and frequency flat (SF) channel $\varphi\{s(t), \lambda\}$. First, an FF α and a TO τ scale and delay the input signal $s(t)$ before the CFO ν and CPO θ influence the carrier. Additive white Gaussian noise (AWGN) superposes the transmit signal. The time reference t_0 results in an additional CFO dependent phase offset.

when the resulting error becomes negligible. In this work, only partial response ($L > 1$) CPM schemes are considered.

B. CHANNEL MODEL

In this work, ground communication between two moving subscribers is considered. If both move with a velocity of $50\,\text{m/s}$ in opposite directions, this gives a maximum Doppler spread of $B_{\text{Dop}} = 200\,\text{Hz}$ at a carrier frequency $f_{\text{c}} = 300\,\text{MHz}$. The according coherence time $T_{\text{coh}} = B_{\text{Dop}}^{-1} = 5\,\text{ms}$ exceeds the burst duration of $T_{\text{Burst}} = 1\,\text{ms}$. In [17], the characteristics of channels in the VHF and UHF range have been investigated and for typical rural, urban and hilly scenarios delay spreads of up to $T_{\text{Delay}} = 4\,\mu\text{s}$ are reported. The more complex mountainous scenario from [17] is not considered here. Comparing the coherence bandwidth $B_{\text{coh}} = T_{\text{Delay}}^{-1} = 250\,\text{kHz}$ with the bandwidths of the waveforms considered in this work (cf. Table 1, the considered waveforms have both-sided bandwidths of $46\,\text{kHz}$ and $84\,\text{kHz}$), it shows to be greater than those. Based on these considerations on coherence time and bandwidth, the communication channel is assumed to be of static (i.e. time invariant) and frequency flat nature over the course of one burst. In order to evaluate this assumption, a time variant fading channel will be simulated in Section V given the actual Doppler and delay spreads. In any case, the channel conditions of every burst are assumed to be mutually independent and have to be repeatedly acquired.

Based on a static and flat fading (SF) channel model, only one propagation path with a scalar attenuation is taken into account and the effect of the Doppler spread f_{Dop} simplifies to a Doppler shift. Furthermore, a carrier phase mismatch and a time delay are considered. All necessary parameters (namely the FF α, CFO ν, CPO θ and TO τ) comprise the parameter vector

$$\lambda = [\alpha, \nu, \theta, \tau]^{\mathsf{T}} \tag{3}$$

and define the system operator alongside the additive white noise term $w(t)$

$$\varphi\{s(t), \lambda\} = \alpha \cdot s(t - \tau) \cdot e^{j(2\pi\nu(t-t_0)+\theta)} + w(t), \tag{4}$$

which is graphically presented in Fig. 2. This (SF) channel model serves as base to designing the synchronization algorithm in Section III.

If $t_0 \neq 0$, the time origin of the CFO is shifted, which influences the CPO estimation [18]. In this work it is assumed that t_0 is chosen such that a so called symmetric observation window is created which is known to be optimal in terms of estimation performance, cf. Section III-A. The complex noise term $w(t)$ has a one-sided spectral noise power density N_0.

In flat fading channels, the FF is typically modeled according to a Rayleigh distribution [19]

$$\alpha \sim \mathcal{R}\left(\kappa = 1/\sqrt{2}\right), \tag{5}$$

where the scale parameter κ was chosen such that $\text{E}[\alpha^2] = 1$. In case of the CPO, typically there is no a priori information available and it is assumed to be uniformly distributed

$$\theta \sim \mathcal{U}(-\pi, +\pi). \tag{6}$$

Networks usually provide some form of frame synchronization, which restricts the range of the residual TO. Here it shall be uniformly distributed among one symbol interval

$$\tau/T \sim \mathcal{U}(-0.5, +0.5) \tag{7}$$

whereas it is not excluded that also a more informative distribution could be applicable. The CFO is subject to the aforementioned Doppler shift as well as opposing oscillator deviations. Setting a low-end requirement on the oscillator quality of 1 ppm at a carrier frequency $f_{\text{c}} = 300\,\text{MHz}$, the maximum CFO value would amount to $|\nu_{\text{max}}| = 700\,\text{Hz}$ and is without any further a priori knowledge again drawn from a uniform distribution

$$\nu T \sim \mathcal{U}(-0.0167, +0.0167) \tag{8}$$

in which the symbol rate of $T^{-1} = 42$ kBaud (cf. Table 1) was inserted to obtain normalized values.

C. PROCESSING IN THE RECEIVER

The received CPM signal is expressed as

$$\begin{aligned} r(t) &= \varphi\{s(t), \lambda\} \\ &= \alpha \cdot s(t-\tau) \cdot e^{j(2\pi\nu(t-t_0)+\theta)} + w(t) \end{aligned} \tag{9}$$

and is processed in in several stages that will be discussed in the following.

1) TILTING, SAMPLING AND NOISE ESTIMATION

Typical for CPM systems is inducing a phase tilt in the down mixing process to make the otherwise time variant CPM trellis time invariant [20]. This is realized by setting off the receiver carrier oscillator by $\xi = h\frac{M-1}{2T}$ with the tilted signal being described by

$$\begin{aligned} r'(t) &= r(t) \cdot e^{j2\pi\xi t} \\ &= \alpha e^{j(2\pi\nu(t-t_0)+\theta)} s(t-\tau) \cdot e^{j2\pi\xi t} + w'(t). \end{aligned} \tag{10}$$

Hereby the noise term $w'(t) = w(t) \cdot e^{j2\pi\xi t}$ is the tilted version of $w(t)$ with identical statistical properties. The following receive filter $H_R(f) = \text{rect}(\frac{f}{B_S})$ is given as the symmetric rectangular function of width B_S in the lowpass domain and shall let the CPM signal pass effectively without distortions (given the potential frequency offset due to ν and ξ). In the next step, the signal is sampled with a rate of T_S^{-1} and given as

$$r'(kT_S) = r'(t)|_{t=kT_s} \tag{11}$$

with k being an integer in the interval $0 \leq k < K \cdot N_{CS}$ and $K = T/T_S$ representing the oversampling factor. By choosing the receive filter width as the sample rate $T_S^{-1} = B_S$, the sampled noise process $w'(kT_S)$ keeps its whiteness due to the fitting spectral repetition and its presumed stationarity [21]. The noise power after the receive filter though is now limited to $\text{Var}[w'(kT_S)] = B_S N_0$ and thus the noise is modeled as a complex normal distribution

$$w'(kT_S) \sim \mathcal{CN}\left(0, \sigma_{w'}^2 = N_0 B_S\right) , \tag{12}$$

$$\Re\left\{w'(kT_S)\right\} \sim \mathcal{N}\left(0, \sigma_{\Re\{w'\}}^2 = \frac{N_0}{2}B_S\right) , \tag{13}$$

$$\Im\left\{w'(kT_S)\right\} \sim \mathcal{N}\left(0, \sigma_{\Im\{w'\}}^2 = \frac{N_0}{2}B_S\right) , \tag{14}$$

$$R_{\Re\{w'\},\Im\{w'\}} = 0 . \tag{15}$$

The operators $\Re\{\bullet\}$ and $\Im\{\bullet\}$ take the real and the imaginary parts, respectively and R stands for the cross correlation function (CCF) of its subscript's respective processes. The noise variance $\sigma_{w'}$ is generally of great interest if soft decisions are desired, as it also is the case in this work. It could e.g. be estimated directly by measuring the received power in an unoccupied time slot in the FH system. Another way in CPM would be to measure the amplitude variance that has to stem from the additive noise in an SF channel. The noise variance is estimated in this work by the formula

$$\hat{\sigma}_{w'}^2 = 2\sqrt{\text{Var}\left[|r'(kT_S)|^2\right]} , \tag{16}$$

which is derived in Appendix A.

2) SISO CPM DETECTOR

The tilting realized in (10) (which takes place before the sampling) is meant to transform the phase function $\phi(t, a)$ from (2) into

$$\phi_n'(t, a') = \phi(t, a) + \pi h(M-1)t/T$$

$$= 4\pi h \sum_{m=0}^{L-1} a'_{n-m} q(t - (n-m)T)$$

$$+ \left[2\pi h \sum_{l=-L+1}^{n-L} a'_l\right] \mod 2\pi \tag{17}$$

where the unipolar symbols $a'_n = \frac{a_n + M - 1}{2} \in \{0, \ldots, M-1\}$ are introduced. It is noted, that this transformation only works as intended if the tilting matches the transmitted signal's time origin, i.e. if the received signal's TO is synchronized (cf.
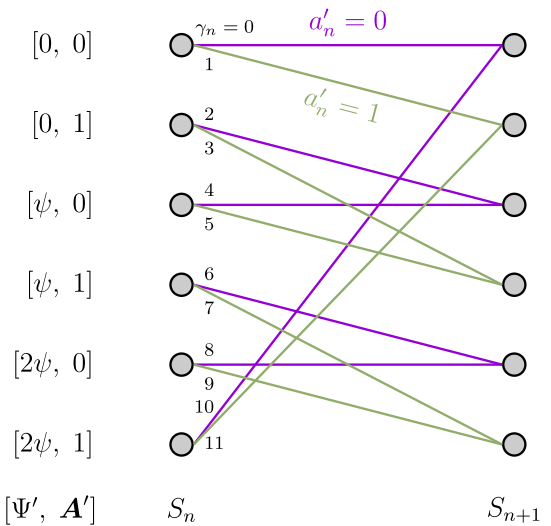

FIGURE 3. Trellis diagram of an $h = 1/3$, $M = 2$, $L = 2$ CPM waveform. It gives an overview about the states and transitions as well as data (a'_n) and pseudo (γ_n) symbols. The angle of one phase state equals $\psi = 2\pi/Q$. Violet and green lines indicate symbols $a'_n = 0$ and $a'_n = 1$, respectively. The pseudo symbols $\gamma = m(\tilde{s}, s)$ are given according to a mapping function that just consecutively numbers the trellis branches with ascending phase Ψ' and correlative state A' as wells as the current symbol a'.

(27). If the correct time origin is assumed, the n'th (tilted) correlative and phase state in the CPM trellis are given by

$$A'_n = \left[a'_{n-(L-1)}, \ldots, a'_{n-1}\right] , \tag{18}$$

$$\Psi'_n = \left[2\pi h \sum_{k=-(L-1)}^{n-L} a'_k\right] \mod 2\pi , \tag{19}$$

respectively. The number of phase states $N'_\Psi = Q$ in a tilted trellis is given by the denominator of the modulation index, while the number of correlative states $N'_A = M^{L-1}$ depends on the modulation order and pulse length. The set of all trellis states $S = [\Psi', {}'A']$ is given by the permutation of correlative and phase states. Any (inner) state is interconnected by M incoming and outgoing branches. Fig. 3 shows an exemplary trellis.

In CPM, all signals $s(t) \cdot e^{jz\psi}$ represent the same digital message as $s(t)$ [20] with z being an integer and $\psi = 2\pi/Q$ representing the phase state's angle. This property is called rotational invariance and will prove useful in Section III, because to synchronize a received signal with a CPO does not require the estimation of the exact CPO but rather only the remainder $\theta_0 = \theta \mod \psi$ of it. The reduced range makes the estimation easier (cf. Section III-F1) without deteriorating the signal's detection. For this to work, the CPM trellis must allow all states with a zero correlative state $A'_0 = \mathbf{0}$ as a start state opposed to only the all zero state $S_0 = [0, \mathbf{0}]$. This more lenient initialization is only necessary for the proposed EM estimation, as the pilot based alternative is able to estimate the actual CPO. The end state of the CPM trellis is for both cases not fixed since, as shown in [22], this will not lead to performance deterioration in coded systems as present in this work.

To obtain MAP probabilities of the data symbols, the BCJR algorithm can be used, as in any trellis, to compute the posterior probabilities $P(a_n'|r(kT_S), \hat{\lambda})$. Hereby it is assumed, that the CPM detector has the knowledge of a channel parameter estimate $\hat{\lambda}$. The posterior probabilities are expressed as

$$P(a_n'|r(kT_S), \hat{\lambda}) \propto p(a_n', r(kT_S)|\hat{\lambda}) \tag{20}$$

$$\propto \sum_{\substack{\tilde{s} \to s \\ a_n'}} p(S_n = \tilde{s}, S_{n+1} = s, r(kT_S)|\hat{\lambda}) \tag{21}$$

$$\propto \sum_{\substack{\tilde{s} \to s \\ a_n'}} f_n^F(\tilde{s}) \cdot f_n^P(\tilde{s}, s) \cdot f_{n+1}^B(s) \tag{22}$$

for $0 \le n < N_{CS}$. In each f_n, the knowledge of $\hat{\lambda}$ is assumed and the superscripts F, P and B stand for forward variable, path metric and backward variable, respectively. They are summed up over all state pairs $\tilde{s}$ and s that are connected by a trellis branch associated with a_n'. The forward and backward variables are calculated through recursions

$$f_n^F(s) = \sum_{\substack{\tilde{s} \\ \tilde{s} \to s}} f_{n-1}^F(\tilde{s}) \cdot f_{n-1}^P(\tilde{s}, s) , \tag{23}$$

$$f_n^B(\tilde{s}) = \sum_{\substack{s \\ \tilde{s} \leftarrow s}} f_{n+1}^B(s) \cdot f_n^P(\tilde{s}, s) \tag{24}$$

where it is summed over all states $\tilde{s}$ that are a predecessor and successor of the state s, respectively. The probability density $p[r(kT_S)|\gamma_n, \lambda]$ is proportional to the path metric and by omitting constant terms can be calculated as

$$f_n^P(\tilde{s}, s) \propto P_A[\gamma_n] \exp\left\{ \frac{2}{\hat{\sigma}_{w'}^2} \Re\{D_{\gamma_n}\} \right\} \tag{25}$$

with $\gamma = m(\tilde{s}, s)$ being a scalar representing the transition from state $\tilde{s}$ to state s by a mapping function m. It takes values between $0 \le \gamma_n \le \Gamma - 1$ with $\Gamma = QM^{L-1}$ being the number of trellis branches per stage. The a priori probability $P_A[\gamma_n]$ is provided by a potential channel decoder. If there is no channel decoder deployed in the system, $P_A[\gamma_n] = 1/\Gamma$ is constant. The scalars γ will in the remainder of this text be called pseudo symbols and their posterior probability is written as

$$P[\gamma_n|r(kT_S), \hat{\lambda}] \propto f_n^F(\tilde{s}) \cdot f_n^P(\tilde{s}, s) \cdot f_{n+1}^B(s) \tag{26}$$

with γ_n relating to the states by the mapping function mentioned above. The path metric exponential's argument D_{γ_n} writes as

$$D_{\gamma_n} = -\frac{\hat{\alpha}^2 E_S}{2} + \sum_{k=0}^{K-1} \hat{\alpha} e^{-j(2\pi[\hat{\nu}(kT_S+nT-t_0)+\xi\hat{\tau}]+\hat{\theta})}$$

$$\times c_\gamma^*(kT_S - \hat{\tau}) \cdot r'(kT_S + nT) \tag{27}$$

with $c_\gamma^*(kT_S)$ denoting the properly sampled complex conjugate of the corresponding pseudo symbol's signal representation, which is scaled, rotated and delayed appropriately to match the received signal. The first addend is the adjusted energy of one pseudo symbol and basically a constant in the path
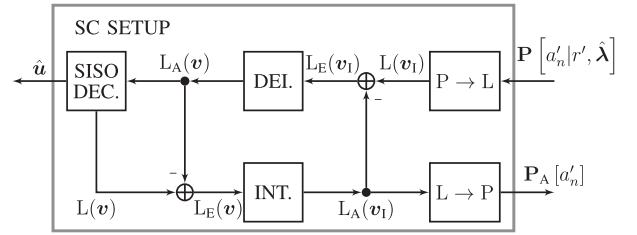

FIGURE 4. Block diagram about the serially concatenated coding setup. The CPM detector is excluded in this graphic and the interface to it is shown in Fig. 1. x denotes a vector of binary symbols and $L(x)$ denotes the vector of x's log likelihood ratios (LLRs). The subscripts A and E stand for a priori and extrinsic, respectively. While the input probabilities $P[a_n'|r'(kT_S), \hat{\lambda}]$ with $0 \le n < N_{CS}$ are converted to LLRs, the a priori LLRs $L_A(v_I)$ are converted to the output probabilities $P_A[a_n']$.

metric calculation. However, for (28) to produce comparable results for different FFs in λ, it must be included. The term $-j2\pi\xi\hat{\tau}$ in the exponential is a phase correction due to the offset introduced by tilting the signal ahead of a TO correction and is not subject to the CFO's time shift t_0. The sum over all end states constitutes the likelihood of the parameter vector

$$\mathcal{L}(\lambda) = p[r(t)|\lambda] = \sum_s f_{N_{CS}}^F(s) . \tag{28}$$

3) THE SERIALLY CONCATENATED CHANNEL DECODER

When deploying a serially concatenated CPM scheme, the CPM detector and the channel decoder in the CPM receiver can exchange information to achieve a Turbo detection decoding scheme [16]. Fig. 4 shows the decoding part of the scheme as well as the interface to the CPM detector.

In the transmitter (cf. Fig. 1), the information bit vector u is first encoded and interleaved to reduce the influence of multi bit error events on the channel and to decorrelate the channel information from the decoder information, which is a crucial part of Turbo setups. The interleaved bits v_I are then Gray mapped onto CPM symbols a' that are fed into the CPM modulator. The symbol probabilities $P[a_n'|r'(t)]$ produced by the MAP detector are transformed into the according log likelihood ratios (LLRs) $L(v_I)$, which are deinterleaved and decoded. Before applying the Turbo principle and feeding the decoded information back into the CPM detector, the a priori LLRs must be subtracted to obtain extrinsic information (cf. Fig. 4), which is then interleaved again and converted to the a priori probabilities $P_A[a_n']$. By respecting which data symbol a_n' accounts for a trellis branch γ_n, these probabilities are used in the calculation of the path metric in (25). The setup is explained in detail in [16]. In this work, the Turbo setup is only carried out in the estimation of channel parameters (cf. IV-A), whereas only the upper branch in Fig. 4 is carried out for decoding the information data. This constitutes for such short burst lengths a reasonable trade-off between performance and computational complexity according to [23]. As a channel code, shortened Polar codes [24] with a soft decoder [25] are used.

III. VECTOR ESTIMATION OF CHANNEL PARAMETERS

In contrast to a scalar parameter estimation, a vector estimator's goal is to estimate a set of parameters jointly. Besides being bounded by slightly different theoretical performance limits, this requires in general more complex procedures. These aspects will be covered in this section.

A. THEORETICAL LIMITS ON THE ESTIMATION PERFORMANCE

For unbiased, joint estimators of unknown parameters (i.e. being treated without informative a priori knowledge) the Cramér-Rao-Vector-Bound (CRVB) is commonly used as a theoretical limit for the mean squared estimation error (MSEE). It is known that for CPM transmitting random data, there is no closed form of the true CRVB for the channel parameters considered in this paper. For that reason, the modified Cramér-Rao-Vector-Bound (MCRVB) [18], that principally lower bounds the true CRVB is used in this work as a benchmark tool. It was derived in [26], that for linear modulation of random data, the MCRVB coincides asymptotically with the true bounds at high E_S/N_0. Since CPM, which is non-linear, can be represented by a superposition of linearly modulated pulses [19], the above statement is assumed to be also valid for CPM. As shown below, the results in this paper support this claim.

In [27] the MCRVBs for the CFO, CPO and TO were derived under the assumption that a small TO (compared to the whole observation) does not influence the estimation of the CPO (see also [26] for this aspect). Here the influence will be taken into account because of the very short frame lengths considered. Furthermore, [27] derives the bound for the amplitude of the received signal, which has to be slightly adjusted to represent the FF instead. As described in all of the just mentioned works, a symmetric observation interval principally decouples the estimation of CFO and CPO. Decoupling in that sense means, that small estimation errors of one parameter does not impair the ability to estimate the other. Being generally a favorable property, such a symmetric observation window is applied in this work. For this, the time shift t_0 in (4) must be chosen to equal the half of the observation window that is relevant for the estimation, i.e. $N_{CS}/2$ for the EM and $N_{PS}/2$ for the pilot based method, where N_{PS} is the number of pilot symbols. The corresponding MCRVBs for each channel parameter are given as the diagonal elements of the inversed modified Fisher information matrix

$$\text{MCRVB}(\alpha) = \text{E}_\alpha \left[\frac{\alpha^2}{2N_{CS}} \cdot \frac{1}{E_S/N_0} \right]$$
$$= \frac{1}{2N_{CS}} \cdot \frac{1}{E_S/N_0} , \tag{29}$$

$$\frac{\text{MCRVB}(\nu)}{T^{-2}} = \frac{3}{2\pi^2 N_{CS}^3} \cdot \frac{1}{E_S/N_0} , \tag{30}$$

$$\text{MCRVB}(\theta) = \text{E}_\tau \left[\frac{N_{CS}^2 T^2 + 12\tau^2}{2N_{CS}^3 T^2} \cdot \frac{1}{E_S/N_0} \right]$$
$$= \left(\frac{1}{2N_{CS}} + \frac{1}{2N_{CS}^3} \right) \cdot \frac{1}{E_S/N_0} , \tag{31}$$

$$\frac{\text{MCRVB}(\tau)}{T^2} = \frac{3}{8\pi^2 h^2 \left(M^2 - 1 \right) G_2(0) T N_{CS}} \cdot \frac{1}{E_S/N_0} \tag{32}$$

with $G_2(0)T = T \int_{-\infty}^{+\infty} g^2(t)\mathrm{d}t$ denoting the normalized (unitless) energy of the respective frequency pulse. According to the definition of the MCRVB [18], the FF bound depends on the actual α, which is intuitive since the smaller the FF is, the smaller will the estimation errors get. To avoid this dependence, the bound is here defined as the average over all possible realizations of α. The FF was parametrized so that its second moment equals unity (cf. (5)), which leads to (29). The property of a decoupled CFO and CPO induced by the symmetric observation window is impaired by every TO $\tau \neq 0$, which obviously destroys the symmetry and affects the CPO bound by the appearance of the term $12\tau^2$ in (31). Similar to the FF case, an average for $\text{MCRVB}(\tau)$ is expressed by $\text{E}[\tau^2] = T^2/12$ (cf. (7)). The other two bounds do not differentiate from the mentioned literature.

B. MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood (ML) estimate of an unknown but constant parameter vector $\boldsymbol{\lambda}$ by utilizing an observation $r'(kT_S)$ is the solution of

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\mathrm{argmax}} \ \log \left(\mathrm{p} \left[r'(kT_S) | \boldsymbol{\lambda} \right] \right) \tag{33}$$

$$= \underset{\boldsymbol{\lambda}}{\mathrm{argmax}} \ \log \left(\sum_{\boldsymbol{\gamma}} \mathrm{p} \left[r'(kT_S), \boldsymbol{\gamma} | \boldsymbol{\lambda} \right] \right) . \tag{34}$$

The term to be maximized is called the log-likelihood function (LLF) with $\log(\bullet)$ taking the natural logarithm. While (33) is generally valid, (34) constitutes that the observed signal $r(kT_S)$ in a communication system is dependent on data symbols that are represented by the latent variable vector $\boldsymbol{\gamma}$

$$\boldsymbol{\gamma} = [\gamma_0, \ \ldots, \ \gamma_{N_{CS}-1}]^\mathsf{T}, \ 0 \leq \gamma_n \leq \Gamma - 1 , \tag{35}$$

which includes the N_{CS} pseudo symbols. By the use of Bayes' law, the logarithmic argument of (34) becomes

$$\sum_{\boldsymbol{\gamma}} \mathrm{p} \left[r'(kT_S), \boldsymbol{\gamma} | \boldsymbol{\lambda} \right] = \sum_{\boldsymbol{\gamma}} \mathrm{p}[r(kT_S) | \boldsymbol{\gamma}, \boldsymbol{\lambda}] \cdot \mathrm{P} \left[\boldsymbol{\gamma} \right] \tag{36}$$

$$\propto \sum_{\boldsymbol{\gamma}} \mathrm{p}[r(kT_S) | \boldsymbol{\gamma}, \boldsymbol{\lambda}] . \tag{37}$$

It was used that the pseudo symbols are independent of the channel parameters and that the a priori probabilities $\mathrm{P}[\boldsymbol{\gamma}] = 1/\Gamma$ are assumed to be constant, which is fulfilled when transmitting random data. The density terms are of a Gaussian form due to the additive noise character (cf. Section II-C1), which is in general a welcome property in LLFs due to the canceling of the logarithm and the exponential. In this case
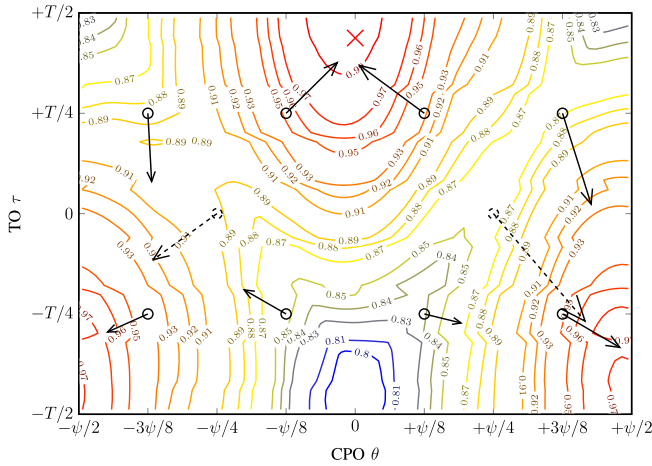
FIGURE 5. Contour plot of an exemplary LLF $\log(p[r(kT_S)|[\alpha = 1, \nu = 0, \theta, \tau]^{\mathsf{T}}])$ with fixed values for the FF and CFO at $10 \log_{10} E_S/N_0 = 10$ dB. Its normalized values are written into the contours and the colors indicate the height (ascending from blue to red). The waveform parameters are $h = 1/3$, $M = 2$, $L = 2$ and $g(t)$ has a rectangular form. The true values for the CPO and TO are $\theta = 0$ and $\tau = 7T/16$ with the LLF's global maximum (marked by a red cross) lying in the direct vicinity to that point. Two distinctive local maxima can be found at both bottom corners of the plot at around $(\pm\psi/2, -3T/8)$. The dashed circles indicate the CPO and TO starting point grid strictly according to the LLF analysis in Section III-F. Considering that in vector estimation CPO and TO can become somewhat ambiguous, the adjusted grid marked by solid circles was suggested in Section III-F5.

though, hey are summed over every possible realization of the pseudo symbol vector, which leads to a sum of exponentials inside the logarithm that cannot be reduced anymore. For this reason finding the ML solution becomes a non-trivial task in this case. Fig. 5 displays an exemplary LLF in the two dimensions of CPO and TO with the other parameters fixed.

C. EXPECTATION MAXIMIZATION (EM) FUNDAMENTALS

The EM algorithm is a method to obtain parameter estimates, when some random variables (e.g. $\boldsymbol{\gamma}$) cannot be observed directly as it is the case in the considered communication system. It provides a framework for iteratively approaching the ML solution with an acceptable computational effort in many use cases. Instead of solving (34) directly, which is non-trivial, a help function $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\text{old}})$ is defined and maximized instead

$$\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\text{old}}) = E_{\boldsymbol{\gamma}}\left\{\log p\left[\boldsymbol{\gamma}, r(kT_S)|\boldsymbol{\lambda}\right]|r(kT_S), \boldsymbol{\lambda}_{\text{old}}\right\}$$

$$= \sum_{\boldsymbol{\gamma}} P\left[\boldsymbol{\gamma}|r(kT_S), \boldsymbol{\lambda}_{\text{old}}\right] \cdot \log p\left[\boldsymbol{\gamma}, r(kT_S)|\boldsymbol{\lambda}\right],$$
(38)

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\arg\max}\; \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\text{old}}).$$
(39)

(38) is referred to as expectation step (E-Step), as it computes the expected value of the log-likelihood of the complete data (observed data $r(t)$ and the latent data $\boldsymbol{\gamma}$) with respect to the posterior distribution $P[\boldsymbol{\gamma}|r(kT_S), \boldsymbol{\lambda}_{\text{old}}]$ of the latent variable vector. Unlike to (34), the logarithm and density's exponential
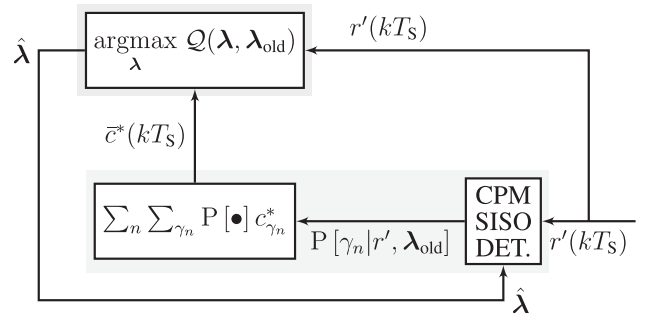
can be reduced, which makes the computation of the help function a more feasible task. In (39) the maximization step (M-Step) is realized by choosing the parameter set that maximizes the help function as the estimate. By updating the old estimate $\boldsymbol{\lambda}_{\text{old}} = \hat{\boldsymbol{\lambda}}$, a new iteration of E- and M-Steps can be carried out. The maximizing parameter set $\hat{\boldsymbol{\lambda}}$ is proved to be a better estimate than $\boldsymbol{\lambda}_{\text{old}}$ in the ML sense [28]. Fig. 6 visualizes the iterative EM procedure.

D. EXPECTATION STEP IN BURST TYPE CPM

By utilizing Bayes' law multiple times and respecting basic Markov chain properties (in connection with the CPM trellis), the help function (38) for the case of the joint estimation of FF, CFO, CPO and TO ($\boldsymbol{\lambda} = [\alpha, \nu, \theta, \tau]^{\mathsf{T}}$) in burst type CPM can be written as

$$\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\text{old}}) = -\frac{E_{r'} + N_{\text{CS}}\alpha^2 E_S}{\hat{\sigma}_{w'}^2}$$

$$+ \frac{2}{\hat{\sigma}_{w'}^2} \sum_{n=0}^{N_{\text{CS}}-1} \sum_{\gamma_n=0}^{\Gamma-1} P\left[\gamma_n|r'(kT_S), \boldsymbol{\lambda}_{\text{old}}\right]$$

$$\times \Re\left\{\sum_{k=0}^{K-1} r'_n(kT_S + nT) \cdot c_{\gamma_n}^*(kT_S, \boldsymbol{\lambda})\right\}$$
(40)

with $E_{r'}$ being the energy of the received signal, the posterior probabilities $P[\gamma_n|r'(kT_S), \boldsymbol{\lambda}_{\text{old}}]$ being output by the SISO CPM detector and

$$c_{\gamma_n}^*(kT_S, \boldsymbol{\lambda}) = \alpha e^{-j(2\pi\nu kT_S + \theta)} c_{\gamma_n}^*(kT_S - \tau)$$
(41)

describing the noise free signal elements under the influence of the actual (and unknown) channel parameters. It is pointed out, that in (40) there are only $N_{\text{CS}}\Gamma$ summations in contrast to $(N_{\text{CS}})^{\Gamma}$ in (38). The term in the real part operator is basically a zero lag CCF of a received symbol and a reference symbol (adjusted to the channel) which is then weighted by the probability of that reference symbol. Due to its independence to the other terms, the posterior probabilities and the according sum

FIGURE 6. Block diagram of the EM algorithm showing the expectation step (highlighted by the green box) and the maximization step (highlighted by the red box). The E-step consists of the calculation of the posterior trellis branch probabilities $P[\gamma_n|r'(kT_S), \boldsymbol{\lambda}_{\text{old}}]$ (cf. (26)) and the creation of the expected reference signal's complex conjugate $\bar{c}^*(kT_S)$ (cf. (42)). Equations (46)-(50) comprise the M-step.

can be drawn in front of the reference symbols to define

$$\bar{c}_n^*(kT_S, \boldsymbol{\lambda}) = \sum_{\gamma_n=0}^{\Gamma-1} \mathrm{P}\left[\gamma_n | r'(kT_S), \boldsymbol{\lambda}_{\mathrm{old}}\right] c_{\gamma_n}^*(kT_S, \boldsymbol{\lambda}) \quad (42)$$

as an expected reference symbol. By omitting constant terms and further shifting and merging sums, (40) is rewritten as

$$\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\mathrm{old}}) = -\alpha^2 \frac{N_{\mathrm{CS}} E_S}{\hat{\sigma}_{w'}^2} + \frac{2\alpha}{\hat{\sigma}_{w'}^2} \Re\left\{ e^{-j\theta} \cdot \chi_{r',\bar{c}}(\nu, \tau) \right\} \ . \quad (43)$$

Hereby, the product of received and reference signal is given as

$$\chi_{r',\bar{c}}(\nu, \tau) = \sum_{k=0}^{N_{\mathrm{CS}}K-1} r'(kT_S)\bar{c}^*(kT_S - \tau) \cdot e^{-j2\pi\nu kT_S} \quad (44)$$

with the reference signal $\bar{c}(kT_S) = \sum_{n=0}^{N_{\mathrm{CS}}-1} \bar{c}_n(\dot{k}T_S - nT)$, $k = \dot{k} + nK$ being the expected reference symbols strung together. $\bar{c}(kT_S)$ generally does not retain the constant amplitude property of a CPM signal anymore, but expresses the expected signal on basis of the CPM detector's knowledge.

Having brought the help function $\mathcal{Q}$ to a manageable form, it is noted, that the expectation step is principally associated with the posterior probabilities' calculation. In this case however, (42) is beneficial to compute in order to enable efficient maximization steps in Section III-E1. Thus the expectation step is associated with it in addition to the posterior probabilities' computation. The help function itself is not necessary to calculate in general and only used to formulate the maximization step.

E. MAXIMIZATION STEP IN BURST TYPE CPM
Having in mind the definition of the maximization step (39), it becomes clear, that this is not possible in an analytical way for (43). Instead the problem is split up in to three parts that are mutually decoupled.

1) MAXIMIZATION WITH REGARD TO ν AND τ
First, for (43) to be maximal the second addend has to be maximized. For this to happen the maximal value of $|\chi_{r',\bar{c}}(\nu, \tau)|$ has to be found, which boils down to the maximization with regard to the CFO ν and the TO τ. This problem is widely known in the field of radar signal processing under the maximization of the cross ambiguity function [29], but no straight forward joint solution exists for the case of arbitrary signal forms because both parameters are coupled in (43) (not in the sense from Section III-A). To determine the optimal order in which both parameters should be estimated, the normalized modified Fisher information (MFI) [18] is considered. The value of the MFI gives a measure of the steepness of the LLF around its maximum and thus the confidence of an ML solution under the influence of additive noise. The relation

$$T^{-2}\mathrm{MFI}(\nu) \gg T^2\mathrm{MFI}(\tau) \ , \quad (45)$$

which is derived in Appendix B, suggests estimating the CFO before the TO. While a slight TO mismatch should not influence its LLF too much, this is very well the case for the CFO due to its orders of magnitude larger MFI. Based on (45), it can be assumed that the estimation of CFO and TO are practically decoupled if carried out in the suggested order.

Since (44) is the definition of the discrete Fourier transform (DFT) of the received and reference signals' product, the cross ambiguity's absolute value can be maximized with regard to the CFO by computing this DFT and determining the maximizing ν

$$\hat{\nu} = \underset{\nu}{\mathrm{argmax}} \ \left| \chi_{r',\bar{c}}(\nu, \tau_{\mathrm{old}}) \right| \ . \quad (46)$$

To allow for the partial maximization, the TO is set to a fixed value τ_{old} stemming from the last EM iteration or the initialization. The next step of estimating the TO can be realized by the computing the absolute CCF of the frequency corrected receive signal and the reference

$$\hat{\tau} = \underset{\tau}{\mathrm{argmax}} \ \left| \chi_{r',\bar{c}}(\hat{\nu}, \tau) \right| \ . \quad (47)$$

These estimation results' accuracies are limited by the resolution of the respective operations, which are dependent on the sampling rate T_S in both cases and the number of channel symbols N_{CS} ((46)) for the CFO case. Unless the receiver's system is highly oversampled, theoretical limits (cf. Section III-A) cannot be achieved and thus some form of interpolation is required. This and an efficient implementation suggestion will be covered in Section IV-B.

2) MAXIMIZATION WITH REGARD TO θ
The estimation of the CPO is decoupled from the other parameters and can be realized by the closed form solution

$$\hat{\theta}_0 = \arg\left\{ \chi_{r',\bar{c}}(\hat{\nu}, \hat{\tau}) \right\} - 2\pi\xi\hat{\tau} \quad (48)$$

in which the angle of the cross ambiguity includes the phase offset due to a delayed tilting. The estimate exactly lets $\Re\{e^{-j\hat{\theta}_0} \cdot \chi_{r',\bar{c}}(\nu, \tau)\} = |\chi_{r',\bar{c}}(\hat{\nu}, \hat{\tau})|$, which is necessary for (43) to be maximized. The above operation computes an estimate for the fractional CPO $\hat{\theta}_0$, which is used in (27). The integer part $z\psi$ of the CPO θ is inherently handled by the trellis initialization as explained in Section II-C2.

3) MAXIMIZATION WITH REGARD TO α
The help function $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\mathrm{old}})$ is quadratic with regard to the FF α and due to its negative curvature it is maximized by setting the partial derivate $\frac{\partial \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\mathrm{old}})}{\partial \alpha} = 0$ and reordering with regard to α. The FF estimate is thus obtained through

$$\hat{\alpha} = \frac{\left| \chi_{r',\bar{c}}(\hat{\nu}, \hat{\tau}) \right|}{N_{\mathrm{CS}} E_S} \ . \quad (49)$$

Eventually the maximization step according to (39) is completed by

$$\hat{\boldsymbol{\lambda}} = \left[\hat{\alpha}, \hat{\nu}, \hat{\theta}_0, \hat{\tau} \right]^{\mathsf{T}} \ . \quad (50)$$
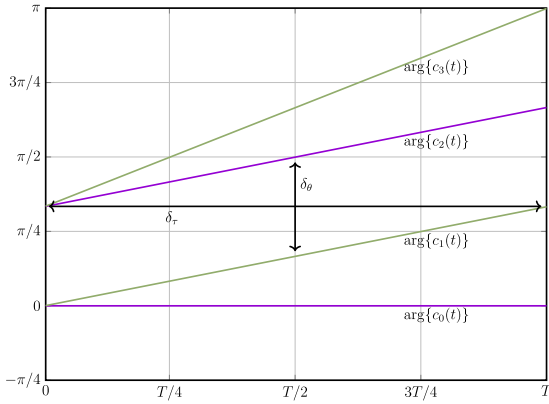
FIGURE 7. Phase trajectories of the first four time invariant trellis branches for a waveform with $h = 1/3$, $M = 2$, $L = 2$ and rectangular frequency pulse as parameters. The trajectories match the first four pseudo symbols (trellis branches) in Fig. 3. All shown trajectories share the same phase state $\Psi' = 0$, the correlative state is $A' = 0$ and $A' = 1$ for c_0, c_1 and c_2, c_3. The current symbol a' is zero for a violet and one for a green curve. The minimum phase offset δ_θ is shown as well as the minimum time offset δ_τ.

F. INITIALIZATION OF THE EM ALGORITHM

A common problem of iterative optimization algorithms and, as one of such, of EM is that different initializations can lead to different convergence results (cf. the local maxima in Fig. 5). The straight forward way to overcome this issue is to span a multi-dimensional grid with potential starting points. The choice of the grid density is naturally a trade off between effort and the certainty of choosing a start that leads to the ML solution. This issue is also described in detail in [13] for the joint time and phase estimation by means of a PLL and a similar approach is proposed to tackle it.

In this work the grid size shall be optimized analytically to match the distance between local maxima of the corresponding scalar LLFs. Generally speaking, a local maximum arises when a transmitted CPM symbol is changed (in phase, frequency or time) in such a way that the result is similar to another CPM symbol. The difference in height between the global and an adjacent local maximum can still be large, but since EM is a gradient ascent method it can get stuck in the local maximum and thus more than one initial starting point is needed in order to reliably find the global maximum and thus the ML solution. Since only the phase is modulated in CPM, the FF is not able to change symbols so that it appears similar to another one and thus no local maxima are expected.

In the following, optimal distances of starting points, i.e. the distance of local maxima in the LLF, are analytically derived under the assumption of a noise free CPM transmission with a rectangular pulse shape. While for the FF a coarse estimation is sufficient, for CFO, CPO and TO a grid with a minimum number of points shall be determined. Coarse estimators are not suitable here, as it was shown in [11], [12]. In these cases, the phase trajectories of some reference symbols are displayed in Fig. 7 on which the following derivations are based on. It shows the phase trajectories of a rectangular scheme in the

phase state $\Psi' = 0$. As explained before, due to CPM's rotational invariance, it is sufficient to estimate the phase offset θ_0 to lock into an arbitrary phase state and thus the consideration of only one phase state is sufficient.

1) CARRIER PHASE OFFSET
A grid set for the case of the CPO (and the TO) was given in [12] on a heuristic basis, which is replaced in this work by an optimized set. To do this, the minimum phase offset $\delta(\theta)$ of two otherwise (with respect to frequency and time) identical symbols is to be found. In Fig. 7, $\delta(\theta)$ changes e.g. $c_1(t)$ to $c_2(t)$ and calculates as the phase difference of identical trellis states except for the last symbol a'_{n-1} being minimally different (e.g. a 1 instead of a 0) in the correlative state

$$\delta(\theta) = 4\pi h \cdot (1 - 0) \cdot \frac{1}{2L}$$

$$= 2\pi \frac{h}{L} \, . \tag{51}$$

2) CARRIER FREQUENCY OFFSET
A CFO changes the frequency of a CPM symbol, so that (when referring to Fig. 7) e.g. c_0 can principally be changed to c_1. The required CFO therefore is generally very large and not realistic, i.e. out of bounds of (8). A CFO generally (apart from the time origin) also induces a phase offset which can lead to a similar effect as in the CPO case. The minimal CFO to cause a local LLF maximum would be, if in the beginning of the last (i.e. the $N_{CS} - 1$'th) symbol the induced phase offset matches $\delta(\theta)$. The resulting symbol would be identical to the corresponding reference symbol but for a very slight offset in the frequency. The required minimal CFO therefore calculates as

$$\delta(\nu) = \frac{\delta(\theta)}{2\pi(N_{CS} - 1)T}$$

$$= \frac{h}{(N_{CS} - 1)LT} \, . \tag{52}$$

Even though all symbols in between have a mismatch in phase and frequency, this still leads to a local maximum due to the falsely matching last symbol.

3) TIMING OFFSET
A timing offset that can lead to a local LLF maximum by changing one symbol to another is given by the fraction of that offset and the constant (for rectangular pulse shape) instantaneous frequency of said symbols. The minimum offset $\delta(\tau)$ therefore is determined by the maximum mutual frequency f_m of two symbols and is expressed by

$$f_m = 4\pi h \cdot \frac{1}{2LT} \left[(M - 2) + \sum_{l=1}^{L-1}(M - 1) \right]$$

$$= \frac{2\pi h}{LT} [L(M - 1) - 1] \, . \tag{53}$$

In Fig. 7, this is the frequency that $c_1(t)$ and $c_2(t)$ share. The highest instantaneous frequency possible is achieved by the current symbol $a'_n = M - 1$ and the correlative state $A'_n = [M - 1, \ldots, M - 1]$ and is exclusive to one symbol only. Eventually the minimum timing offset writes as

$$\delta(\tau) = \frac{\delta(\theta)}{f_{\mathrm{m}}}$$
$$= \frac{T}{L(M-1)-1} \ . \tag{54}$$

4) FADING FACTOR

The idea of this coarse estimation is, that CPM is a constant amplitude modulation and that the additive noise is naturally orthogonal to modulated signals. With the knowledge of the noise variance from (16), the FF α can be estimated by

$$\hat{\alpha} = \sqrt{\frac{\mathrm{E}\left[|r'(kT_{\mathrm{S}})|^2\right] - \hat{\sigma}_{w'}^2}{E_{\mathrm{S}}/T}} \tag{55}$$

in which the nominator calculates the sample energy of the received signal's wanted component and the denominator expresses the energy of one transmitted sample. The estimate $\hat{\alpha}$ is, under the assumed channel conditions, accurate enough for it to be serving as a single starting value and thus the grid set $\mathcal{I}_\alpha$ is simply given as

$$\mathcal{I}_\alpha = \hat{\alpha} \ . \tag{56}$$

5) JOINT LLF

The above derived minimal offsets $\delta(\lambda_i)$ for CFO, CPO and TO are for the case of a scalar parameter channel and are valid for these cases. However, in the case of vector (i.e. joint) estimation, things are generally more tricky. In case of the CFO, the number of grid points $N_\nu = \left\lceil \frac{\Delta(\nu)}{\delta(\nu)} \right\rceil$ is simply determined by dividing the potential parameter range by the minimal offset. For the joint case of CPO and TO, simulations suggested, that the number of grid points must be increased to reliably avoid local maxima. An intuitive justification could be, that both parameters can lead to temporally exact (not just similar as in the CFO case due to the slight frequency mismatch) copies of reference symbols, in which the causative offset cannot be reliably determined. Referring to Fig. 7, this could happen for a smaller h and or a larger L, thus a smaller $\delta(\tau)$. A time shifted $c_1(t)$ can become partly identical to $c_2(t)$. For that reason the number of grid points are chosen as $N_\theta = 2\left\lceil \frac{\psi}{\delta(\theta)} \right\rceil = 2\left\lceil \frac{L}{P} \right\rceil$ and $N_\tau = 2\left\lceil \frac{T}{\delta(\tau)} \right\rceil = 2(L(M-1)-1)$.

The grid points for CFO, CPO and TO are then distributed uniformly in the respective estimation range $\Delta(\lambda_i)$, ($2 \leq i \leq 4$)

$$\mathcal{I}_{\lambda_i} = \left\{ \pm \left(\frac{\Delta(\lambda_i)}{2} - \frac{\Delta(\lambda_i)}{2N_{\lambda_i}} \right), \right.$$
$$\left. \pm \left(\frac{\Delta(\lambda_i)}{2} - \frac{3\Delta(\lambda_i)}{2N_{\lambda_i}} \right), \ldots \right\} \tag{57}$$

Algorithm 1: Calculate ML Estimate $\hat{\lambda}$.

> **Input:** $r'(kT_{\mathrm{S}})$
> **Output:** $\hat{\lambda}$
> 1:   **for all** $\lambda_i \in \mathcal{I}$ **do**
> 2: $l_i = \mathcal{L}(\lambda_i)$ #(28)
> 3:   **end for**
> 4: $\mathcal{I}_{\mathrm{reduced}} = \bigcup_{\tau_i \in \mathcal{I}_\tau} \operatorname*{argmax}_{\lambda_i} \mathcal{L}(\lambda_i|\tau_i)$ #Select best λ for every τ of TO grid
> 5:   **for all** $\lambda_i \in \mathcal{I}_{\mathrm{reduced}}$ **do**
> 6: $\lambda_{\mathrm{old}} = \lambda_i$
> 7: $it = 1$
> 8:   **while** $it \leq N_{\mathrm{It}}$ **do**
> 9:   **if** $it < N_{\mathrm{It}}$ **then**
> 10:   E-Step CPM   #(26), (42)
> 11:   **else if** $it == N_{\mathrm{IT}}$ **then**
> 12:   E-Step SCCPM   #Section II-C3
> 13:   **end if**
> 14:   M-Step   #(46)-(50)
> 15: $\lambda_{\mathrm{old}} = \hat{\lambda}$
> 16: $it = it + 1$
> 17:   **end while**
> 18: $l_i = \mathcal{L}(\hat{\lambda})$ #(28)
> 19:   **end for**
> 20: $\hat{\lambda} = \operatorname*{argmax}_{\hat{\lambda}_i} l_i$

---

with λ_i specifying the i'th element of the parameter vector λ. Eventually the whole initialization grid with a cardinality of $|\mathcal{I}| = \prod_{i=1}^4 N_{\lambda_i}$ is given by the Cartesian product of the individual one-dimensional grids

$$\mathcal{I} = \{\mathcal{I}_\alpha \times \mathcal{I}_\nu \times \mathcal{I}_\theta \times \mathcal{I}_\tau\} \ . \tag{58}$$

In Fig. 5, the CPO and TO grid points are added to the LLF contour and also the LLF gradients (i.e. the direction in which the EM algorithm will converge) from these points are shown. The two points marked by dashed lines would be the starting points if the joint estimation would be treated as a scalar one. It can be seen that the EM converges from both of them to local maxima and thus is not able to find the ML solution. The eight points marked by solid lines indicate the adjusted grid points as reasoned above. The ML solution is reached by at least one of them.

IV. IMPLEMENTATION ASPECTS

A. ALGORITHMIC SYNCHRONIZATION DESCRIPTION

In Algorithm 1, the procedure of estimating $\hat{\lambda}$ is given. In the first loop (lines 1-3), for every starting point the respective likelihood is computed by the forward algorithm of the CPM detector. A starting point nearer to the true λ can have a lower likelihood than another point, e.g. when the other point lies in a high local maximum as can be seen in Fig. 5. For this reason more than the most likely starting point must be considered in the second loop (lines 5-19) where the EM is carried out.

The selection of starting points for the second loop happens in line 4. For every TO in $\mathcal{I}_\tau$ the most likely CFO and CPO grid points are chosen. In doing so, the CFO can be estimated accurately in (46) while using the fixed TO value. The number of EM iterations carried out determines the accuracy of the end result as well as the computational complexity, so a criterion to stop at some point must be defined (cf. line 8). This can be as e.g. the change in the estimate is small enough or a fixed number of iteration N_{It} is reached. The latter is used in this work with $N_{It} = 4$, which heuristically proved to provide satisfying results.

While it increases the computational complexity, the SC-CPM setup decreases the number of symbol errors. EM as a decision-directed estimator naturally gains from this and provides better estimates. As a compromise, the SCCPM is only applied in the last iteration, cf. lines 9 to 13. The maximization step is followed by the preparation of the new iteration. In line 20 the best estimate according to its likelihood is determined.

B. EFFICIENT IMPLEMENTATION

The most computationally complex parts in Algorithm 1 are the calculation of the path metrics, the expectation and the maximization steps for which some details are considered next.

1) CALCULATION OF PATH METRIC

The evaluation of $D_{\gamma,n}$ in (27) is initially an intensive task, but can be optimized in the following ways. First of all, by not shifting the received signal, but the reference in the opposite direction, the resampling operation on the already sampled received signal is saved and the result is kept the same. The shifted references can be computed offline and stored for every TO grid point, while the CFO compensation is precomputed beforehand in (27). The result $D_{\gamma,n}$ is just a complex number, every phase offset can be applied to it directly so the recomputation of the sum is unnecessary. The same holds true for the phase states of the CPM trellis, which are just phase shifts of the zero phase state $\Psi' = 0$. The path metrics are needed for the likelihood and E-Step computations.

2) E-STEP

The E-Step (line 10) consists of two parts. First, the pseudo symbols' posterior probabilities have to calculated, which is done by the BCJR algorithm. The weighted reference symbols are calculated in (42) in which the reference symbols are shifted to the right. The time shifted reference symbols can be reused since the old estimate τ_{old} from the last iteration (cf. (46)) is applied here as well as for the path metric calculation. For the E-Step under consideration of the SCCPM setup (line 12), the same optimization approaches are valid.

3) M-STEP

Four maximization operations comprise the M-Step in line 14. While two (CPO and FF) are closed form solutions and need no further consideration, the estimation for the CFO and TO in

(46) and (47) are basically grid searches and can be optimized. As mentioned in Section III-E1, the necessary operations are a DFT and a CCF, which can be both efficiently implemented by the use of the fast Fourier transform (FFT). The accuracy of both estimates is lower bounded by the oversampling factor K and in the case of CFO the number of channel symbols N_{CS}.

Both can be improved if the FFT operations are zero padded and thus the result's resolution is improved by ideal interpolation. This approach has the advantage over other techniques (such as e.g. parabolic interpolation) because it does not introduce an interpolation bias [30]. The DFT and CCF are extensive grid searches, i.e. they give results in a much larger range than the parameter ranges ($\Delta_\nu T \ll 1$ and $\Delta_\tau/T \ll 2N_{CS}$ for normalized CFO and TO, respectively). Spectral zoom methods can remedy the introduced complexity when extensive grid searches need a high resolution. The number of frequency (or time bins) determines the complexity and depends on the desired estimation range and resolution. As for the range, it is sufficient to search in the width of one LLF maximum $\delta(\nu)$ (or $\delta(\tau)$) (as others cannot be reached anyways due to getting stuck in local maxima). The resolution shall ensure that the introduced quantization error does not have any visible effect on the MSEE, i.e. it set to be $\frac{\varepsilon^2}{12} \overset{!}{=} 0.1 \cdot$ MCRVB (cf. (30), (32)). This gives the resolution of CFO $\varepsilon(\nu)$ and TO $\varepsilon(\tau)$ as

$$\varepsilon(\bullet) = \sqrt{\frac{6}{5} \cdot \text{MCRVB}(\bullet)} . \tag{59}$$

ε is directly proportional to $\sqrt{E_S/N_0}^{-1}$, i.e. in noisy channel conditions, a high resolution (i.e. a low ε) is unnecessary. The number of frequency (or time bins) is then given as

$$m(\bullet) = \left\lceil \frac{\delta(\bullet)}{\varepsilon(\bullet)} \right\rceil \tag{60}$$

and is even at a high signal to noise ratio (SNR) only in the double-digit range. An efficient way to compute these bins is e.g. the Goertzel algorithm [31]. In case of the TO estimation, basic DFT properties have to be applied, so that time bins are calculated through this methods.

V. PERFORMANCE ANALYSIS

This section will carry out a comprehensive evaluation of the vector expectation maximization (VEM) estimator and compares it to a widely referenced vector pilot based (VPB) method [15].

A. NOTES ON THE VPB METHOD

- *FF estimation:* Since [15] does not include a method for estimating the FF, the following estimation rule is introduced for the VPB method (which proved to be optimal in the simulations)

$$\hat{\alpha} = \frac{|\Lambda_1(\hat{\nu})| + |\Lambda_2(\hat{\nu})|}{N_{PS} E_S} . \tag{61}$$

To avoid confusion with this paper's notation, the functions λ_i from [15] were capitalized to Λ_i.

- *CFO estimation:* Due to the nature of the pilot in [15] being a preamble, the (in the presence of noise always uncertainly) estimated and corrected CFO introduces a phase offset at the beginning of the data block. To combat this effect, [15] suggests the use of a DD phase and timing tracking loop and shows, that the residual CFO estimation error has virtually no effect on error rates with this approach. It is noted that [15] is able to estimate a CFO range of $\Delta_\nu T = [-0.5, +0.5]$ which would induce a high computational complexity in the VEM method due to the need of many CFO grid points. Though a range this large is arguably not of practical relevance (cf. Section II-B).

- *TO estimation:* Contrary to the statement in [15], that a timing offset in the range of $[-0.5T, +0.5T]$ (also cf. (7)) can be estimated by a closed form expression, this is not the case for arbitrary CPM parameters, as it is shown in Appendix C. The maximum TO that can theoretically be estimated by the VPB method is

$$\left| \frac{\tau_{\max}}{T} \right| = \frac{1}{2h(M-1)} \tag{62}$$

which can be lower than 0.5 for waveforms with high modulation indexes and orders. In practice, the range should be even lower because of noise influences and the range $\Delta_\tau = [-0.75\tau_{\max}, +0.75\tau_{\max}]$ is set accordingly with headroom. It is noted that in [15], the simulated CPM parameters provide enough headroom to be able to estimate the range in (7). Fig. 9 visualizes the effect explained above and waiving the TO range reduction would lead to a useless 4/7-Q2RC VPB curve in Fig. 12. The VEM method is not limited with regard to the TO range, but profits of the lower range by having potentially less grid points and thus a lower computational complexity.

B. SIMULATION PARAMETERS

Two CPM waveform were chosen for the performance analysis, that are energy and spectrally efficient in terms of Euclidean distance and occupied bandwidth [20]. All relevant parameters are summarized in Table 1.

The first part of the table summarizes the waveform parameters including the bandwidth time product of $B_{GA}T = 0.5$ of 2/3-B3GA's frequency pulse and the normalized pulse energies used in (32). The symbol rate T^{-1} and oversampling factor K as well as the normalized two-sided bandwidth $2B_{99}T$ are given in the second part.

Several comparisons will be carried out in the following subsections that will require different setups in terms of the numbers of transmitted symbols and included information symbols. In particular, the pilot based VPB system is adjusted to enable a fair comparison to the blind VEM system. The last three parts in Table 1 account for these adjustments. In Sections V-C and V-D, the theoretical estimator performance is evaluated and $N_{CS} = 32$ is chosen as a multiple of 4 which is recommended for the VPB system [15]. The burst shall only contain pilot symbols in that case, whereas the pilot less VEM

TABLE 1. Waveform Parameters and System Setups for Several Simulated Comparisons

	2/3-B3GA	4/7-Q2RC
h	2/3	4/7
M	2	4
L	3	2
$g(t)$	GA ($B_{GA}T = 0.5$)	RC
$G_2(0)T$	0.1755	0.1875
T^{-1}	42 kBaud	
K	6	
$2B_{99}T$	1.10	2.32
Sec. V-C, V-D		
N_{CS}	32	
N_{PS} (VPB)	32	
k_C	21 (VEM), 0 (VPB)	42, 0
R_{VEM}	0.5	
Sec. V-E		
N_{CS}	42	
N_{PS} (VPB)	16	
k_C	21 (VEM), 13 (VPB)	42, 26
R_{VEM}	0.5	
R_{VPB}	0.5	
Sec. V-F		
N_{CS}	42	
N_{PS} (VPB)	16	
k_C	21 (VEM), 21 (VPB)	42, 42
R_{VEM}	0.5	
R_{VPB}	0.81	

system deploys a code rate of $R_{VEM} = 0.5$. To obtain more practical metrics, the VPB system must also contain information bits in the subsequent sections. The number of transmitted symbols is here increased to 42 to enable bursts of length $T_{Burst} = 1$ ms. The code rate for VEM is kept the same, while $N_{PS} = 16$ symbols are spent on pilots in the VPB system. In Section V-E, the effective throughput is investigated for which the code rate shall be fixed to 0.5 across all systems with the number of information bits k_C differing as consequence. The systems' according spectral efficiencies are defined by

$$\eta = \frac{k_C}{2B_{99}T_{Burst}} . \tag{63}$$

To enable the fair (in terms of spectral efficiency) error rate comparison in Section V-F, the same number of information
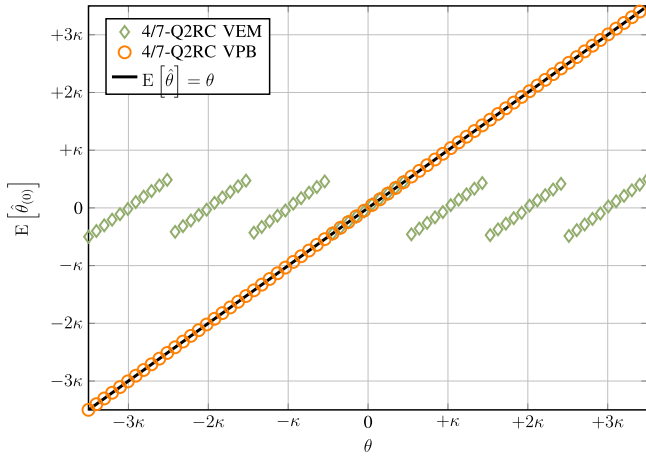
FIGURE 8. CPO MEV E[$\hat{\theta}_0$] and E[$\hat{\theta}$] of VEM and VPB, respectively, in the SF channel for the 4/7-Q2RC waveform at $10 \log_{10}(\alpha^2 E_S/N_0) = 15$ dB compared to the bisector line corresponding to an unbiased estimator.
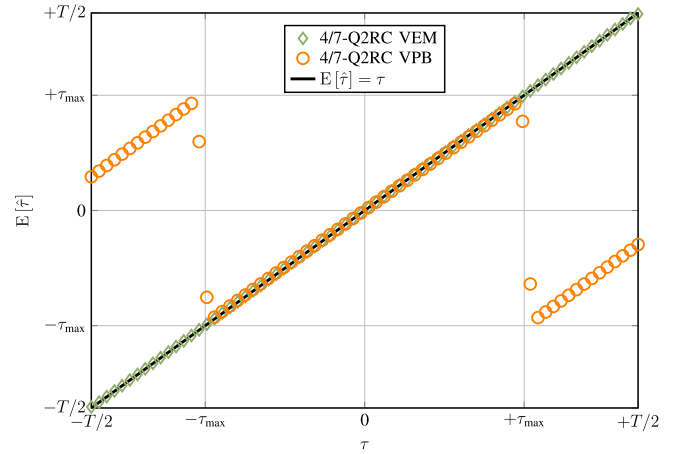

FIGURE 9. TO MEV E[$\hat{\tau}$] of VEM and VPB in the SF channel for the 4/7-Q2RC waveform at $10 \log_{10}(\alpha^2 E_S/N_0) = 15$ dB compared to the bisector line corresponding to an unbiased estimator.

bits k_C is transmitted using the same transmission time T_{Burst} and occupied two-sided bandwidth $2B_{99}$ with the same energy E_b spent on each information bit. Since the VPB method has to rely on the burst containing a certain amount of non-informative (in the sense of information theoretic entropy) symbols. For this burst to carry the same amount of information bits as in the VEM case, different code rates have to be applied, i.e. the VEM scheme bursts can carry more code redundancy and thus profit from a lower code rate, which has higher error correction capabilities. The relation between the VEM code rate R_{VEM} and the VPB code rate R_{VPB} in this section is given as

$$R_{\text{VPB}} = R_{\text{VEM}} \cdot \frac{N_{\text{CS}}}{N_{\text{CS}} - N_{\text{PS}}} \tag{64}$$

with N_{CS} and N_{PS} giving the number of transmitted and pilot symbols, respectively. It is noted, that the VEM and VPB systems are set to have the same spectral efficiency only for one waveform, but not across both waveforms.

To emphasize the performance of the estimators, an adjusted SNR metric $\alpha^2 E_{\text{S/b}}/N_0$ is used in the SF channel from (4), where the energy is scaled by the FF's square. By this, the fading effect is diminished and a comparison with the AWGN channel is made reasonable.

C. MEAN ESTIMATION VALUE

This section analyzes the mean estimation value (MEV) of the estimators in the SF channel. Both VPB and VEM follow ML oriented approaches and thus should be unbiased which is the case if the MEV asymptotically equals the true parameter, i.e. E[λ_i] − λ_i = 0. The cases of the CPO θ and TO τ for the 4/7Q2RC waveform are shown in Figs. 8 and 9.

The VPB estimator provides unbiased estimates over the whole CPO range $\theta \in [-\pi, +\pi)$ without any signs of unfavorable values. In case of the VEM estimator it becomes clear, that the MEV is periodic with the phase state angle ψ. This

means that only the fine offset within a phase state is estimated and not the actual CPO that occurred in the channel. For the reason outlined in Section II-C2, this is an intended behavior and for that the VEM is still considered to provide unbiased CPO estimates. It is noted that both estimators are prone to handling CPO values in the vicinity of their corresponding range's edges in a wrong way, i.e. that $-\pi$ is handled as $+\pi$ or $-\psi/2$ is handled as $+\psi/2$, respectively. This is completely unproblematic in terms of synchronizing because of the equality $e^{-j\pi} = e^{+j\pi}$ and CPM's rotational invariance. For the displaying of the results in this section, the estimates were mapped in the correct interval to avoid meaningless ripples in the curves. The same applies to Fig. 11 in Section V-D.

The TO MEV of the VEM estimator matches perfectly with the angle bisector line that indicates unbiasedness. The VPB estimator's curve also gives unbiased estimates within the range stated in (62). Outside of it, the estimates are shifted precisely by $\pm 2\tau_{\text{max}}$ due to the estimator's ambiguity explained in appendix C. It is emphasized that in contrast to the CPO case, this cannot be considered as intentional behavior, since there lies no natural periodicity or the like in time offsets. To avoid the resulting large deteriorations in estimation performance, the TO range in these cases is limited as mentioned in Section V-A.

Both estimators show unbiasedness in the estimation of the FF α and the CFO ν without bringing any further insights and thus their corresponding plots are left out for the sake of saving space.

D. MEAN SQUARED ESTIMATION ERROR

The mean squared estimation error (MSEE) is arguably the most important metric for parameter estimators' performance. In case of unbiased estimators (as given in this work) the theoretical minimum is given as true CRVB. While these exact bounds are available for pilot based estimation [32], the MCRVB is useful because of its relatively simple closed form
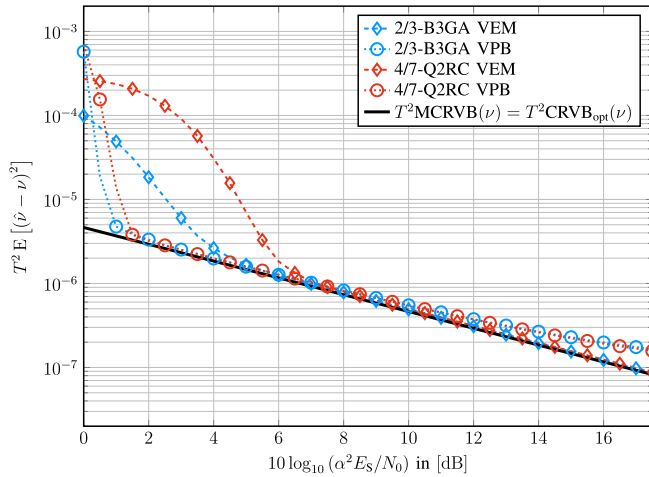
FIGURE 10. Normalized CFO MSEE $T^2\mathrm{E}[(\hat{\nu} - \nu)^2]$ over $\alpha^2 E_\mathrm{S}/N_0$ of VEM and VPB in the SF channel for both waveforms compared to the theoretical limit.

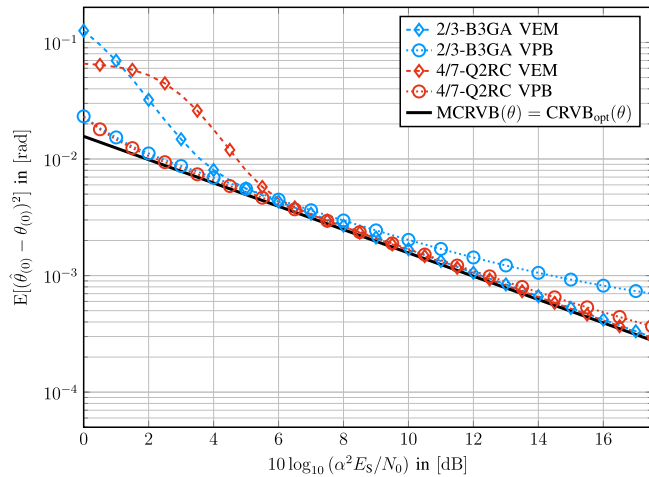**FIGURE 12.** Normalized TO MSEE $T^{-2}\mathrm{E}[(\hat{\tau} - \tau)^2]$ over $\alpha^2 E_\mathrm{S}/N_0$ of VEM and VPB in the SF channel for both waveforms compared to the theoretical limit. The subscript 'opt' at the CRVB indicates the use of the optimal synchronization sequence from [32].

FIGURE 11. CPO MSEE $\mathrm{E}[(\hat{\theta}_0 - \theta_0)^2]$ and $\mathrm{E}[(\hat{\theta} - \theta)^2]$ of VEM and VPB, respectively in the SF channel over $\alpha^2 E_\mathrm{S}/N_0$ for both waveforms compared to the theoretical limit.

The VPB estimator gives accurate estimates for every parameter covering all relevant SNRs. Hereby the MSEE is slightly (or in the case of CFO immensely) increased at very low and significantly increased at very high SNRs. While the former originates from a natural uncertainty regarding the LLF's global maximum even when a pilot sequence is present, the latter is the result of the assumptions made in [15] to enable an efficient and decoupled computation method. In between, the estimates' MSEE lie very close to the theoretical limits.

Every curve corresponding to a VEM estimate shows a high MSEE at low SNRs and begins to match the MCRVB at some SNR threshold. The curves do not bend up at high SNRs as the VPB ones do, since no simplifying assumptions were necessary in the EM algorithm's derivation and the interpolations for CFO and TO were optimal in the sense of interpolation errors. The SNR threshold value is around 4 dB for the 2/3-B3GA and around 6 dB for the 4/7-Q2RC waveform. In both waveforms, this relates to an average of 0.024 (hard) pseudo symbol errors after one SCCPM iteration. It is expected that the VEM estimator produces useful estimates as soon as few mistakes are made on received symbols' decisions, since VEM is basically a decision-directed estimation method.

E. EFFECTIVE THROUGHPUT
The effective throughput is defined as the number of binary information symbols that are transmitted in a correctly decoded burst. Moreover, it is normalized by the occupied bandwidth and the burst duration, which gives the formal definition

$$\mathcal{T}_\mathrm{eff} = \frac{(1 - P_\mathrm{Frame}) \cdot k_\mathrm{C}}{2B_{99}T_\mathrm{Burst}} \tag{65}$$

$$= (1 - P_\mathrm{Frame}) \cdot \eta \,, \tag{66}$$

derivation. It is further noted that for unbiased estimators, the variance of the estimation error equals the MSEE. Figs. 10–12 show the CFO, CPO and TO MSEEs of both estimators in the SF channel for both waveform setups over increasing signal to noise ratio (SNR) given as relation of symbol energy to the scaled noise power density in dB. The FF's performance plot is omitted for the lack of insights and the hereby justified sake of briefness.

The theoretical performance limit is given in each plot for every curve. While the CRVB for pilot based [32] and the MCRVB [18] for non-pilot based estimation coincide for the FF, CPO and CFO [26] and are also CPM parameter independent, four bounds are given in the TO plot for each combination of estimator and waveform. The analysis covers all four parameters at once because of their similar behavior.
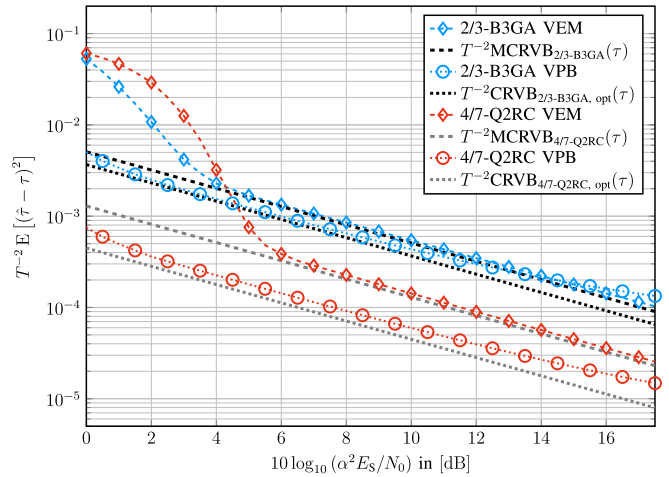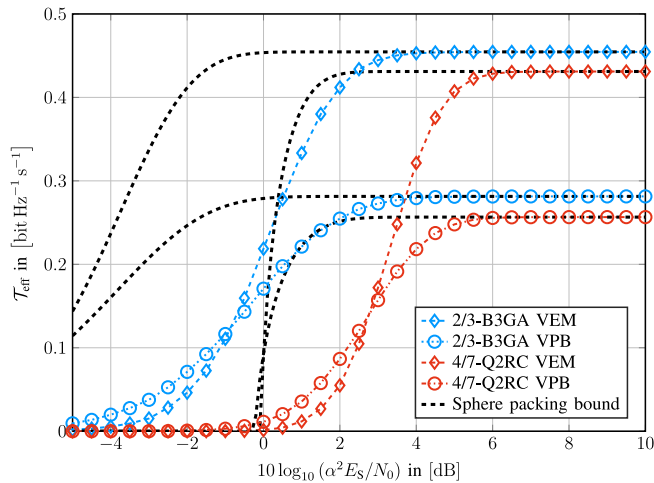
FIGURE 13. Normalized effective Throughput $\mathcal{T}_{\text{eff}} = (1 - P_{\text{Frame}}) \cdot \eta$ with code rates being fixed to $R_{\text{VEM}} = R_{\text{VPB}} = 0.5$ over $\alpha^2 E_b/N_0$ in the SF channel. The throughput, when the sphere packing bound is applied, upper bounds the performance. The respective corresponding sphere bound curve can be deducted by the asymptotic values.

FIGURE 14. Coded bit error rate (BER) over $\alpha^2 E_b/N_0$ in the SF channel.

FIGURE 15. Coded frame error rate (FER) over $\alpha^2 E_b/N_0$ in the SF channel.

where the second line follows from (63). The probability of the event of a wrongly decoded frame P_{Frame} is obtained as the frame error rate (FER) through Monte-Carlo simulations. An upper bound for the throughput can be obtained by applying Shannon's sphere packing bound [33] to above equation. This bound constitutes the performance in terms of frame errors of an optimal spherical code for a continuous input channel, which itself is a lower bound to the FER. Fig. 13 shows the normalized effective throughput $\mathcal{T}_{\text{eff}}$ for both waveforms and estimation systems for the fixed code rate of $R_{\text{VEM}} = R_{\text{VPB}} = 0.5$ as stated in Table 1 and the respective upper bounds obtained through the sphere packing bound.

All throughput curves have their spectral efficiency according to (63) as asymptote at high SNRs, i.e. when the FER approaches zero. The VPB has generally a lower probability of a wrongly decoded frame throughout all SNR regions because of their known pilot structure compared to the VEM setups. Though it can only convert this advantage into a slightly higher throughput for very low SNRs, which is due to VEM systems' lack of pilot symbols and hence their higher spectral efficiency. Comparing the curves to their respective upper bound, it can be observed, that the VPB performs slightly closer to the theoretic limit.

F. CODED ERROR RATES

The performance in terms of error rates are investigated through the coded bit and frame error rates in Figs. 14 and 15 for the SF channel. The SNR definition differs from the last sections and is now with regard to the information bit energy, that relates to the former definition by $E_b/N_0 = E_S/N_0/(\log_2(M) \cdot R)$.

Both in terms of BER and FER, the VEM curves lie very close to the AWGN optimum, i.e. about 0.2 dB and 0.4 dB for the binary and quaternary waveform, respectively. However
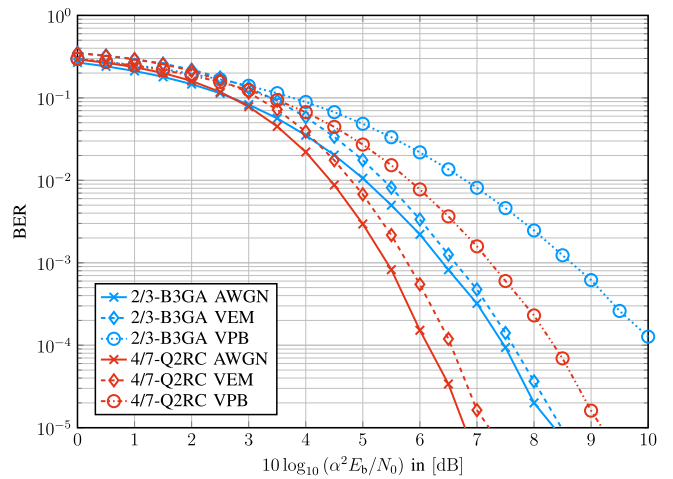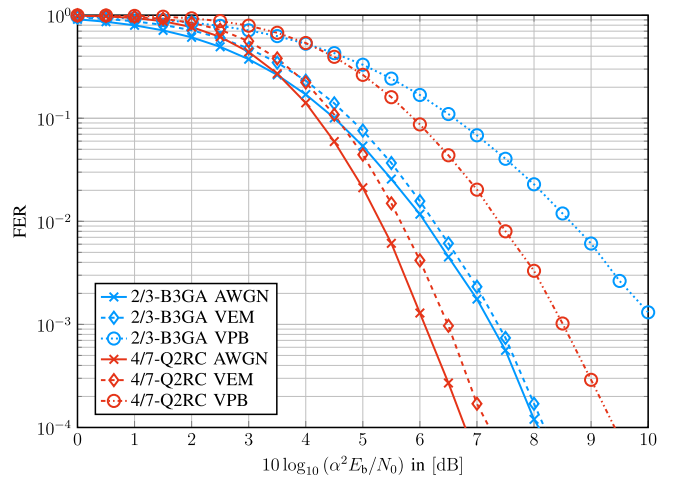
the VPB performance is about 2 dB and little short of 3 dB worse than its contender. The reasons hereby lie foremost in the SNR penalty to ensure equal spectral efficiency for both estimation methods. Another reason is that the VPB only uses $N_{\text{PS}} = 16$ symbols for the parameter estimation, while VEM utilizes the whole burst length of $N_{\text{CS}} = 42$ symbols and the channel code for it. Using the above relation between symbol and information bit energy and by investigating Fig. 14, the SNR threshold from the last section translates to a (coded) bit error rate of $5 \cdot 10^{-4}$ for both waveforms.

To evaluate the effect of the assumptions made on the channel characteristics in Section II-B, a time variant, frequency selective channel is simulated. Hereby, the channel parameters are simulated according to the Urban profile from [17] under a maximum Doppler spread of $f_{\text{Dop}}T = 0.024$. A receiver with genie-aided perfect channel state information (PCSI) using it in a non-adaptive minimum mean square error (MMSE) equalizer [19] serves as a comparison candidate. Deploying an adaptive equalizer would pose difficulties in the frame design
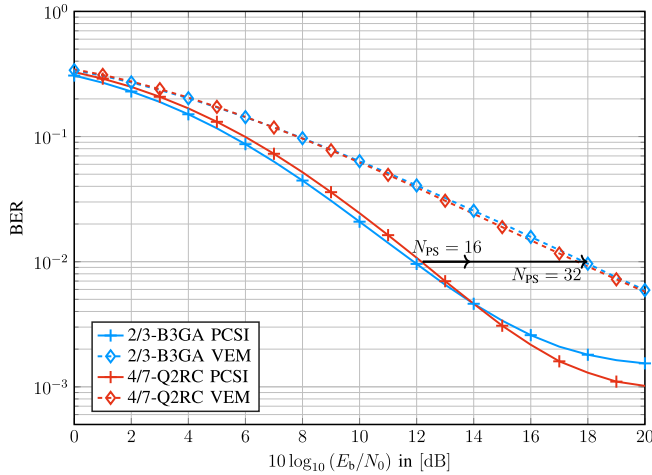
FIGURE 16. Coded bit error rate (BER) over E_b/N_0 in the Urban environment [17] with a maximum Doppler spread of 100 Hz at a symbol rate of 42 kBaud. The VEM receiver is compared to a coherent receiver following a non-adaptive MMSE equalizer having perfect channel state information. The SNR losses for two pilot sequence lengths are shown by arrows.

TABLE 2. Relevant Parameters for Computational Complexity

	2/3-B3GA	4/7-Q2RC
$\|\mathcal{I}_\nu\|$	7	5
$\|\mathcal{I}_\theta\|$	4	2
$\|\mathcal{I}_\tau\|$	4	4
$\|\mathcal{I}\|$	112	40
$10\log_{10}\left(\alpha^2 E_S/N_0\right)$	4 dB	6 dB
$m(\nu)$	6	9
$m(\tau)$	12	12
n_{MC}	64	128
N_{ZP}	512	512

because of the need of enough pilot symbols distributed within one burst of only 1 ms duration. For that reason, the PCSI receiver ignores the time variant nature of the channel. No scaling of N_0 happens in the SNR metric for the comparison in Fig. 16 to reflect reality's conditions best possible.

While the waveform parameters do not seem to have a large impact on the system performance, even the PCSI receiver does not reach BERs of under 10^{-3} for high SNRs. Thus it is argued that in a fading case like this, an outer channel code with larger interleavers should be deployed on an aggregation of single bursts. This would create diversity in the channel and enable smaller error rates at lower SNR points. A burst BER of around 10^{-2} would arguably serve as a good basis for an outer high rate code, at which the PCSI receiver has an advantage of around 6 dB compared to the VEM scheme. Considering that genie-aided PCSI is an ideal not achievable in reality, this gap would shrink by giving up spectral efficiency in exchange for pilot symbols (that would gain PCSI only for $N_{\mathrm{PS}} \to \infty$). Spending $N_{\mathrm{PS}} = 16$ symbols on a pilot (as above in the VPB schemes) would result in an SNR loss of 2.1 dB, whereas $N_{\mathrm{PS}} = 32$ (as recommended in [15]) would completely close the SNR gap by inducing an SNR loss of 6.2 dB. While the simplifying assumptions made on the channel characteristics did not hold as shown in Fig. 16, the VEM performance can be reasoned to be satisfying under moderate Doppler and delay spread. Other use cases of short burst transmissions include e.g. an indoor sensor network, that suffers less from both Doppler and delay spread and hence are arguably suited for the simplified channel model.

VI. COMPLEXITY ANALYSIS

In order to evaluate practical relevance, the computational complexity of the results of Section V-D is considered. This section will point out the main drivers of complexity in the VEM receiver as well as compare it with its contender VPB. For the sake of simplicity, only a subset of the complete receiver that is dominant with respect to the complexity is considered and furthermore, each operation (additions, multiplications, etc.) is treated equally. These inaccuracies are accepted to allow for a simple and still informative comparison of one frame's synchronization complexity. For the analysis the main elements from Algorithm 1 are taken into account and relevant parameters are listed in Table 2.

Calculation of likelihoods in line 2: The calculation of $|\mathcal{I}|$ likelihoods, the according path metrics of every grid point must be computed (cf. Section IV-B1 for the necessary steps), which amounts to

$$
\begin{aligned}
\mathcal{C}_{\mathcal{L}, 1} = N_{\mathrm{CS}} M^L \big[&|\mathcal{I}_\tau| \, |\mathcal{I}_\nu| \cdot (8K - 2) \\
&+ (Q - 1)|\mathcal{I}| \cdot 3 + Q\,|\mathcal{I}| \cdot 2 \big] \\
&+ |\mathcal{I}| \, N_{\mathrm{CS}} Q M^{L-1} \cdot (2M - 1) .
\end{aligned} \tag{67}
$$

The first three addends represent the grid points' path metric computation that is comprised by the multiplication of $r'(kT_S + nT) \cdot c_\gamma^*(kT_S - \hat{\tau})$ and the product's summation, the metrics' rotations to include every phase state and CPO combination, the division by the noise power and the exponential in (25), respectively. The multiplication with a priori information is ignored since none is available yet. The execution of BCJR's forward part, which computes the likelihood according to (28) is considered in the last term. The complexity $\mathcal{C}_{\mathcal{L}, 2}$ of line 18 is derived similarly.

E-Step in line 10: The computation of the posterior probabilities for multiple coarse estimates and iterations is done by the execution of the CPM detector (including the calculation of the path metrics for $\boldsymbol{\lambda}_{\mathrm{old}}$) and costs

$$
\begin{aligned}
\mathcal{C}_{\mathrm{E}, 1} = |\mathcal{I}_\tau| \, (N_{\mathrm{It}} - 1) N_{\mathrm{CS}} \big[&Q M^{L-1} \cdot (6M - 2) \\
&+ M^L (8K - 2 + 3(Q-1) + 2Q) \big]
\end{aligned} \tag{68}
$$

operations. The complexity evaluation of (42) is ignored, since most of the posterior probabilities are zero or very close to zero and thus its overall complexity is negligible.

E-Step in line 12: To include the channel decoder in the SCCPM setup, two executions of the CPM detector (the path metrics must only be calculated once, because the appliance of the decoder information is just an additional multiplication) and one of the decoder are necessary. Summarized, this line is responsible for

$$
\begin{aligned}
\mathcal{C}_{\text{E, 2}} = &\ |\mathcal{I}_\tau| N_{\text{CS}} \left[2QM^{L-1} \cdot (6M - 2) \right.\\
&+ M^L \left(8K - 2 + 3(Q - 1) + 3Q \right) \big]\\
&+ |\mathcal{I}_\tau| \left(40 \log_2(n_{\text{MC}}) \cdot n_{\text{MC}} \right)
\end{aligned}
\tag{69}
$$

operations (cf. [25] for the complexity on the decoder). $n_{\text{MC}} = 2^{\lceil \log_2(n_{\text{C}}) \rceil}$ is the length in bits for the mother Polar code, that is shortened in the process of providing flexible code rates. Of course, it is noted that the specific decoder complexity is solely dependent on its choice.

M-Step in line 14 In the maximization step, only the maximizations with regard to CFO and TO are considered, since the other parameters are estimated with simple closed form expressions. As suggested in Section IV-B3, the Goertzel algorithm is used to evaluate the cross ambiguity function at $m(\nu)$ CFO and $m(\tau)$ values (cf. (60)). The SNR determines the number of bins to be calculated and is chosen according to the threshold values from Section V-D (cf. Table 2). For the CFO case this amounts to

$$
\mathcal{C}_{\text{M, 1}} = 6N_{\text{CS}}K + m(\nu) \cdot (3N_{\text{CS}}K + 4)
\tag{70}
$$

operations with the first term accounting for the multiplication of $r'(kT_{\text{S}})$ and $\bar{c}^*(kT_{\text{S}} - \tau)$ and the second for the Goertzel executions. To compute a partially interpolated cross ambiguity function, the same basic approach as in [12] is followed. First, the received and the reference signal are transformed by means of an FFT. Hereby both signals are zero padded to ensure a linear (in contrast to circular) correlation and to utilize a radix-2 FFT implementation. While [12] computes a spectrally zero padded inverse FFT of this product, this work again suggests the Goertzel algorithm to compute the few necessary time bins. This takes

$$
\begin{aligned}
\mathcal{C}_{\text{M, 2}} = &\ 2 \left(4N_{\text{ZP}} \log_2(N_{\text{ZP}}) - 6N_{\text{ZP}} + 8 \right)\\
&+ 6N_{\text{ZP}} + m(\tau) \cdot (3N_{\text{ZP}} + 4)
\end{aligned}
\tag{71}
$$

operations with the complexity of an FFT taken from [34] and N_{ZP} being the properly zero padded length in samples.

VPB method The main complexity part in the VPB method is comprised by two zero padded FFTs of signal products ((22) and (23) in [15]) that cost

$$
\begin{aligned}
\mathcal{C}_{\text{VPB}} = &\ 2 \cdot (6N_{\text{PB}}KK_{\text{f}}\\
&+ 4N_{\text{PB}}KK_{\text{f}} \log_2(N_{\text{PB}}KK_{\text{f}}) - 6N_{\text{PB}}KK_{\text{f}} + 8)
\end{aligned}
\tag{72}
$$

with $K_{\text{f}} = 4$ as an interpolation factor and $N_{\text{PB}} = 16$.

Table 3 lists the number of operations per frame (considering the SF model in the error rate simulation's section) for

TABLE 3. Listing of Complexities in Terms of Number of Operations Per Burst ($T_{\text{Burst}} = 1$ ms)

	2/3-B3GA	4/7-Q2RC
$\mathcal{C}_{\mathcal{L}, 1}$	1053696	1807680
$\mathcal{C}_{\text{E}, 1}$	294336	939456
$\mathcal{C}_{\text{E}, 2}$	183744	578816
$\mathcal{C}_{\text{M}, 1}$	6072	8352
$\mathcal{C}_{\text{M},2}$	52288	52288
$\mathcal{C}_{\mathcal{L}, 2}$	84000	242592
$\mathcal{C}_{\text{VEM}}$	1674136	3629184
$\mathcal{C}_{\text{VEM}}/\mathcal{C}_{\text{AWGN}}$	42	32
$\mathcal{C}_{\text{VEM}}/\mathcal{C}_{\text{VPB}}$	45	98

every crucial part of the VEM estimator and sets that metric in relation to the Genie aided synchronized system, i.e. a coherent CPM detector with decoder in an AWGN channel and the VPB estimator. Relevant parameters can be gathered from Tables 1 and 2.

It is obvious that the gross share of complexity falls in the calculation of likelihoods for the coarse estimation. Any way to shrink the parameter ranges would be helpful to reduce that part, e.g. by inserting few pilot symbols. Furthermore, typical trellis reductions and efficient logarithmic implementations of the BCJR were not considered in this work. Anyway, the coarse estimation and also the $|\mathcal{I}_\tau|$ EM instances are perfectly suited to be parallelized. Even considering the low-cost field programmable array Xilinx Spartan-7 XC7S6 operating at 741 MHz and incorporating ten DSP units (i.e. 7.41 million possible operation per T_{Burst}), multiple parallel instances of an VEM receiver can be run by it in real time, roughly speaking.

Clearly the VPB complexity is magnitudes smaller than VEM's. Though it is ignored to the benefit of the VPB that a phase and timing tracking loop has to be carried out in the following detector, while a simple coherent detector can be deployed in the VEM case. In case the oscillator hardware is not chosen to be low-end but at a quality of 0.1 ppm, which would be a sensible decision considering the complexity savings, the ratio $\mathcal{C}_{\text{VEM}}/\mathcal{C}_{\text{AWGN}}$ would be significantly lower. Having also in mind, that the systems considered in [2] are optimally detected in the SCCPM setup with up to four iterations, this ratio is only around 5.5 for both waveforms.

VII. CONCLUSION

This paper introduced a double Turbo estimation and decoding scheme in order to enable the use of the EM algorithm to jointly estimate four channel parameters. The first main contribution of this work was to derive the expectation and maximization steps. For these, efficient implementations were introduced as well. The second main contribution was the analysis of the channel parameter LLFs to optimize the choice of a starting point grid, which is necessary for gradient ascent

methods to which EM belongs, in order to ensure finding the joint LLF's global maximum.

In the performance analysis, the decision-directed VEM estimator was benchmarked against the VPB from [15] in a non-data (i.e. bursts consist of only pilot symbols) setup and was shown to provide more accurate estimates that are theoretically optimal in terms of MSEE for relevant SNR regions. By fixing the code rates, the effective throughputs of the systems were compared with the result, that VEM provides a higher throughput than VPB due to the lack of pilot symbols while having about the same gap from the theoretical optimum. Furthermore, the VEM receiver performed significantly better than its VPB contender regarding bit and frame error rates when fixing the spectral efficiency by adjusting the systems' code rates.

The drawback of the VEM receiver is its higher computational complexity compared to the VPB solution. An unquantifiable aspect is the easily detectable (since known) preamble of the VPB method and thus an inherent vulnerability against jamming which is not existent in the presented VEM method. Further complexity reduction approaches as well as the adjustment and evaluation of this technique to more challenging fading environments is left for future work to further improve the VEM estimator.

APPENDIX
A. VARIANCE ESTIMATION
The presented noise estimation in a constant amplitude waveform relies on a flat fading channel and the received signal amplitude's variation is utilized for which additive noise must be the sole cause. Letting a and b represent the real and imaginary part of the received signal $r' = a + jb$, the variance of the squared absolute value writes as

$$
\begin{aligned}
\mathrm{Var}[|a + jb|^2] &= \mathrm{Var}[a^2 + b^2] \\
&= \mathrm{Var}[a^2] + \mathrm{Var}[b^2] + 2\mathrm{Cov}[ab] \\
&= \mathrm{E}[a]^2\mathrm{Var}[b] + \mathrm{E}[b]^2\mathrm{Var}[a] + \mathrm{Var}[a]\mathrm{Var}[b] \\
&= \mathrm{Var}[a]\mathrm{Var}[b] = \frac{\sigma_{w'}^4}{4} .
\end{aligned}
\tag{73}
$$

General variance properties have been used and $\mathrm{Var}[a] = \mathrm{Var}[b] = \sigma_{w'}^2/2$, $\mathrm{E}[a] = \mathrm{E}[b] = 0$ and $\mathrm{Cov}[ab] = 0$ are given by (12)–(15). By eventually doubling the end result's square root, the variance of w' is given as in (16).

B. DERIVATION OF MFI(ν) AND MFI(τ)
According to [18] the modified Fisher information (MFI) in this case is computed as follows

$$
[\boldsymbol{I}_\mathrm{M}(\boldsymbol{\lambda})]_{ij} = \mathrm{E}_{\boldsymbol{a}}\left[\frac{\alpha^2 T_\mathrm{S}}{N_0} \Re\left\{ \frac{\partial r^*(t, \boldsymbol{a}, \boldsymbol{\lambda})}{\partial \lambda_i} \cdot \frac{\partial r(t, \boldsymbol{a}, \boldsymbol{\lambda})}{\partial \lambda_j} \right\} \right] .
\tag{74}
$$

To obtain the MFI of the CFO and TO, the partial derivatives are chosen to $\lambda_i = \lambda_j = \nu$ and $\lambda_i = \lambda_j = \tau$, respectively.

Eventually the normalized MFIs are computed to

$$
T^{-2}\mathrm{MFI}(\nu) = \frac{2}{3}\pi^2 N_\mathrm{CS} \frac{E_\mathrm{S}}{N_0} \cdot \left(N_\mathrm{CS}^2 + 1\right) ,
\tag{75}
$$

$$
T^2\mathrm{MFI}(\tau) = \frac{2}{3}\pi^2 N_\mathrm{CS} \frac{E_\mathrm{S}}{N_0} \cdot \left(4h^2 G_2(0)T\right)\left(M^2 - 1\right)
\tag{76}
$$

and by inserting exemplary parameters of the 1/3-B2REC waveform with $G_2(0)T = 0.125$, (45) becomes obvious.

C. TO ESTIMATION LIMIT IN VPB
The formula of calculating the timing offset on a symbol basis in [15] is given by

$$
\hat{\varepsilon} = \frac{\arg\left\{\lambda_1(\hat{\nu})\lambda_2^*(\hat{\nu})\right\}}{2(M-1)\pi h}
\tag{77}
$$

with $\varepsilon = \tau/T$ relating to this work's TO definition as stated. Naturally the angle of a complex number is unambiguous only as long as it is in the range $[-\pi, +\pi)$. This gives a maximum unambiguous estimation range of

$$
\hat{\varepsilon} \in \left[-\frac{1}{2(M-1)h}, +\frac{1}{2(M-1)h} \right) ,
\tag{78}
$$

which is stated in (62).

REFERENCES
[1] C. Bockelmann et al., "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[2] E. Casini, M. Street, P. Vigneron, and R. Barfoot, "SDR-ready standardized waveforms for tactical VHF and UHF communications for NATO," Tech. Rep., NATO C3 AGENCY THE HAGUE (NETHERLANDS), 2010.

[3] R. Mehlan, Y.-E. Chen, and H. Meyr, "A fully digital feedforward MSK demodulator with joint frequency offset and symbol timing estimation for burst mode mobile radio," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 434–443, Nov. 1993.

[4] M. Morelli and U. Mengali, "Feedforward carrier frequency estimation with MSK-type signals," *IEEE Commun. Lett.*, vol. 2, no. 8, pp. 235–237, Aug. 1998.

[5] M. Morelli and G. M. Vitetta, "Joint phase and timing synchronization algorithms for MSK-type signals," in *Proc. IEEE Commun. Theory Mini- Conf.*, 1999, pp. 146–150.

[6] A. D'andrea, A. Ginesi, and U. Mengali, "Frequency detectors for CPM signals," *IEEE Trans. Commun.*, vol. 43, no. 2–4, pp. 1828–1837, Feb.–Apr. 1995.

[7] A. D'andrea, A. Ginesi, and U. Mengali, "Digital carrier frequency estimation for multilevel CPM signals," in *Proc. IEEE Int. Conf. Commun.*, 1995, vol. 2, pp. 1041–1045.

[8] A. N. D'Andrea, U. Mengali, and M. Morelli, "Symbol timing estimation with CPM modulation," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1362–1372, Oct. 1996.

[9] P. Bianchi, P. Loubaton, and F. Sirven, "On the blind estimation of the parameters of continuous phase modulated signals," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 5, pp. 944–962, May 2005.

[10] A. D'Amico, A. N. D'Andrea, and U. Mengali, "Feedforward synchronization schemes for MSK-type signals," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 597–607, 1999.

[11] A. Lang and B. Lankl, "Carrier frequency offset estimation in burst-type CPM via the EM algorithm," in *Proc. IEEE 91st Veh. Technol. Conf.*, 2020, pp. 1–6.

[12] A. Lang and B. Lankl, "Blind joint timing and carrier phase estimation in burst-type CPM," in *Proc. IEEE 94th Veh. Technol. Conf.*, 2021, pp. 1–6.

[13] J. Huber and W. Liu, "Data-aided synchronization of coherent CPM-Receivers," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 178–189, Jan. 1992.

[14] N. Hajiabdolrahim, S. R. Aghdam, and T. Eriksson, "An extended Kalman filter framework for joint phase noise, CFO and sampling time error estimation," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2020, pp. 1–6.

[15] E. Hosseini and E. Perrins, "Timing, carrier, and frame synchronization of burst-mode CPM," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5125–5138, Dec. 2013.

[16] P. Moqvist and T. M. Aulin, "Serially concatenated continuous phase modulation with iterative decoding," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1901–1915, Nov. 2001.

[17] J. Fischer, *Physical Layer Link Modeling for Mobile Ad-Hoc Networks (MANET)*. Stuttgart, Germany: Fraunhofer Verlag, 2016.

[18] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramer-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 52–60, Jan. 1998.

[19] J. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2008.

[20] J. Huber, "Trelliscodierung," in *Trelliscodierung*. Berlin, Germany: Springer, 1992, pp. 65–98.

[21] K.-D. Kammeyer, *Nachrichtenübertragung*. Berlin, Germany: Springer-Verlag, 2013.

[22] A. Lang and B. Lankl, "A comprehensive study of CPM trellis initialization methods," in *Proc. IEEE Mil. Commun. Conf.*, 2022, pp. 1–6.

[23] A. Lang and B. Lankl, "Very short channel codes for burst-type transmissions," in *Proc. IEEE Mil. Commun. Conf.*, 2018, pp. 1–6.

[24] V. Bioglio, F. Gabry, and I. Land, "Low-complexity puncturing and shortening of polar codes," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops*, 2017, pp. 1–6.

[25] U. U. Fayyaz and J. R. Barry, "Low-complexity soft-output decoding of polar codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 958–966, May 2014.

[26] M. Moeneclaey, "On the true and the modified Cramer-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1536–1544, Nov. 1998.

[27] A. Pollok and R. McKilliam, "Modified Cramér-Rao bounds for continuous-phase modulated signals," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1681–1690, May 2014.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[29] S. Stein, "Algorithms for ambiguity function processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 588–599, Jun. 1981.

[30] L. Svilainis, "Review on time delay estimate subsample interpolation in frequency domain," *IEEE Trans. Ultrasonics, Ferroelect., Freq. Control*, vol. 66, no. 11, pp. 1691–1698, Nov. 2019.

[31] G. Goertzel et al., "An algorithm for the evaluation of finite trigonometric series," *Amer. Math. Monthly*, vol. 65, no. 1, pp. 34–35, 1958.

[32] E. Hosseini and E. Perrins, "The Cramer-Rao bound for training sequence design for burst-mode CPM," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2396–2407, Jun. 2013.

[33] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," *TMO Prog. Rep.*, vol. 42, no. 133, pp. 1–23, 1998.

[34] L. G. Baltar, F. Schaich, M. Renfors, and J. A. Nossek, "Computational complexity analysis of advanced physical layers based on multicarrier modulation," in *Proc. IEEE Future Netw. Mobile Summit*, 2011, pp. 1–8.