A Simple Variational Bayes Detector for Orthogonal Time Frequency Space (OTFS) Modulation

Weijie Yuan, Zhiqiang Wei, *Member, IEEE,* Jinhong Yuan, *Fellow, IEEE,* and Derrick Wing Kwan Ng, *Senior Member, IEEE*

Abstract—The emerging orthogonal time frequency space (OTFS) modulation technique has shown its superiority to the current orthogonal frequency division multiplexing (OFDM) scheme, in terms of its capabilities of exploiting full time-frequency diversity and coping with channel dynamics. The optimal maximum *a posteriori* (MAP) detection is capable of eliminating the negative impacts of the inter-symbol interference in the delay-Doppler (DD) domain at the expense of a prohibitively high complexity. To reduce the receiver complexity for OTFS scheme, this paper proposes a variational Bayes (VB) approach as an approximation of the optimal MAP detection. Compared to the widely used message passing algorithm, we prove that the proposed iterative algorithm is guaranteed to converge to the global optimum of the approximated MAP detector regardless the resulting factor graph is loopy or not. Simulation results validate the fast convergence of the proposed VB receiver and also show its promising performance gain compared to the conventional message passing algorithm.

Index Terms—Orthogonal time frequency space, receiver design, variational Bayes, low complexity

I. INTRODUCTION

Future wireless communication applications require reliable data transmission in high mobility environments, such as vehicular networks and high speed railway [1]. When facing a fast time-varying channel, the widely adopted orthogonal frequency division multiplexing (OFDM) modulation in currently cellular networks suffers from a dramatic performance degradation due to the inter-carrier interference imposed by Doppler shift [2].

Recently, a newly developed modulation scheme, namely, orthogonal time frequency space (OTFS) has attracted significant attention since it is robust in doubly selective channels [3]. The basic model of OTFS modulation/demodulation was devised in [4]. In [5] and [6], the discrete time formulation of OTFS is developed and it was shown that OTFS can be implemented via appending a pre- and post-processing to conventional OFDM schemes. In particular, OTFS modulator multiplexes information symbols in the delay-Doppler (DD) domain. Based on a two-dimensional inverse symplectic finite Fourier transform (ISFFT), each symbol in the DD domain is spread over all lattices in the time-frequency (TF) domain. Consequently, full time and frequency diversity can be realized and exploited, which leads to a further performance gain compared to conventional OFDM schemes in time-varying channels [7]. Moreover, when the number of reflectors/scatters is limited during propagation, employing DD domain data multiplexing provides a sparse channel representation, enabling a better channel estimation performance and a low-complexity detector.

In practice, the time and frequency diversity can be fully exploited with the optimal maximum *a posteriori* (MAP) de-

The authors are with the School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia.

tection. Yet, the complexity of the MAP detection increases exponentially with the block size of each OTFS frame. As a compromise approach, based on the factor graph and message passing framework [8], the authors in [9] developed a low-complexity receiver applying Gaussian approximation to capture the impact of the inter-symbol interference. Nevertheless, message passing (MP) receivers will converge to a locally optimal point in a loopy factor graph, which is very likely to happen for OTFS in a multipath scenario [10]. Therefore, it is highly desirable to develop a low-complexity receiver with fast convergence for OTFS systems.

Motivated by the aforementioned discussions, in this paper, a low-complexity detector for OTFS systems is designed. We commence from the optimal MAP detection and derive an approximation of the *a posteriori* distribution exploiting Kullback-Leibler (KL) divergence to reduce the detection complexity. Then, a variational Bayes (VB) approach is used to maximize the evidence lower bound (ELBO) iteratively, which yields a marginal distribution for each symbol and thus results in a low-complexity point-wise MAP detection. Compared to the conventional MP receivers, the advantages of the proposed variational Bayes approach are two-fold. Firstly, by choosing the distribution family for the VB method appropriately, the ELBO maximization problem is strictly convex such that the globally optimal solution can be obtained efficiently, leading to a convergence guaranteed receiver. Secondly, the VB approach only requires a substantially lower complexity than that of the message passing algorithms. Furthermore, we demonstrate the rapid convergence and the better detection performance compared to MP receivers through simulations, particularly in a practical multipath propagation environment.

Notations: We use a boldface lowercase letter and boldface capital letter to denote a vector and a matrix, respectively. The superscript T, $*$, and H denote the transpose, conjugate, and the Hermitian operations, respectively; $\propto$ represents both sides of the equation are multiplicatively connected to a constant; $|\cdot|$ denotes the modulus of a complex number or the cardinality of a set; $[\cdot]_N$ denotes the modulo operation with divisor N; $\|\cdot\|$ denotes the ℓ^2 norm; $\mathcal{R}\{x\}$ denotes the real part of complex number x; $\mathbb{E}_p[x]$ denotes the expectation of x with respect to distribution p; ∂ denotes the partial derivative operator; the big-O notation $\mathcal{O}$ is an asymptotic notation describes the order of complexity. The circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu, \sigma^2)$.

II. SYSTEM MODEL

The OTFS modulation and demodulation are realized by employing two dimensional pre- and post-processing. We focus on a point-to-point scenario. In particular, the transmitted coded bits **c** are mapped to data symbols $d_{k,l} \in \chi$ in the DD

domain with constellation set χ, where $k \in [0, N-1]$ and $l \in [0, M-1]$ denotes the index of Doppler shift and delay, respectively. Integers N and M denote the numbers of available time slots and subcarriers, respectively. The symbols $\{d_{k,l}\}$ are transformed to time-frequency (TF) domain symbols $X_{n,m}$ through ISFFT, given by [4]

$$\mathbf{X} = \mathbf{F}_N^{\mathrm{H}} \mathbf{D} \mathbf{F}_M, \tag{1}$$

where $\mathbf{X}$ and $\mathbf{D}$ are matrices with elements $X_{n,m}$ and $d_{k,l}$ and $\mathbf{F}_N$ denotes the normalized discrete Fourier transform (DFT) matrix of dimension N. Then, the TF domain symbols $\mathbf{X}$ are converted to a continuous transmitted signal $s(t)$, expressed as

$$s(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{n,m} e^{j2\pi m \Delta_f (t-nT)} g_{\mathrm{tx}}(t-nT), \tag{2}$$

where Δ_f is subcarrier spacing and $T = \frac{1}{\Delta_f}$ is the symbol duration. The above equation can be viewed as a two-step process that first transforms the TF domain symbols to time domain symbols and then shapes the pulse with function $g_{\mathrm{tx}}(t)$. For a linear time-variant channel, the received signal can be expressed as

$$r(t) = \int \int h(\tau, \nu) e^{j2\pi \nu(t-\tau)} s(t-\tau) \, d\tau d\nu + n(t), \tag{3}$$

where the channel impulse response in the DD domain with P resolvable paths is given by [4]

$$h(\tau, \nu) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i) \delta(\nu - \nu_i), \tag{4}$$

where τ_i and ν_i denote the delay and Doppler shift associated with the ith path.

At the receiver side, we adopt $g_{\mathrm{rx}}^*(t)$ as the receiving filter and transform the received signal to the TF domain via performing DFT. By sampling at $t = nT$ and $f = m\Delta_f$, the received samples $Y_{n,m}$ is given by

$$Y_{n,m} = \int r(t) g_{\mathrm{rx}}^*(t-nT) e^{-j2\pi m \Delta f(t-nT)} dt. \tag{5}$$

According to [4], invoking ideal shaping and receiving pulses $g_{\mathrm{tx}}(t)$ and $g_{\mathrm{tx}}(t)$ can simplify the TF domain input-output relationship as

$$Y_{n,m} = H_{n,m} X_{n,m} + W_{n,m}, \tag{6}$$

where $H_{n,m}$ is the equivalent channel in the TF domain and $W_{n,m} \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise with a noise power of σ^2. Finally, we arrive at the DD domain received sample $y_{k,l}$ by employing the symplectic finite Fourier transform (SFFT), formulating as

$$y_{k,l} = \sum_{k'=0}^{N-1} \sum_{l'=0}^{M-1} d_{k',l'} h_{k,l}[k', l'] + w_{k,l}, \tag{7}$$

where

$$h_{k,l}[k', l'] = \sum_{i=1}^{P} h_i w[k-k'-k_{\nu_i}, l-l'-l_{\tau_i}] e^{-j2\pi \frac{k_{\nu_i} l_{\tau_i}}{NM}} \tag{8}$$

and $w[k,l] = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} e^{-j2\pi nT \frac{k}{NT}} e^{j2\pi m \Delta f \frac{l}{M \Delta f}},$ (9)

where k_{ν_i} and l_{τ_i} denote the Doppler frequency shift index and the delay index associated with the ith path, respectively, i.e., $\nu_i = \frac{k_{\nu_i}}{NT}$ and $\tau_i = \frac{l_{\tau_i}}{M \Delta f}$.

III. RECEIVER DESIGN

A. Variational Bayes Approach

To facilitate the presentation of the proposed variational Bayes approach, we stack the transmitted symbols and equivalent channels into vectors yielding the following compact form of (7):

$$y_{k,l} = \mathbf{h}_{k,l}^{\mathrm{T}} \mathbf{d} + w_{k,l}, \tag{10}$$

where $\mathbf{d}$ and $\mathbf{h}_{k,l}$ denotes the data symbol vector and equivalent channel vector, respectively, whose the $(k'M + l')$th entry is $d_{k',l'}$ and $h_{k,l}[k', l']$. As shown in Sec. IV. B of [11], using both ideal pulse and practical rectangular pulse provides the same input-output relationship form as in (10) by slightly changing the equivalent channels. Therefore, it is worth to note that the proposed detection in this work is applicable to both cases. Furthermore, by stacking $y_{k,l}$ into a vector $\mathbf{y}$, the optimal MAP detection can be formulated as:

$$\hat{\mathbf{d}} = \arg \max_{\mathbf{d} \in \chi} p(\mathbf{d}|\mathbf{y}). \tag{11}$$

Solving the optimization problem in (11) requires a computational complexity order of $|\chi|^{NM}$, which increases exponentially with the size of $\mathbf{d}$. As a compromise approach, we focus on the variational Bayes method to handle (11) in the sequel. Specifically, the idea behind variational Bayes aims for finding a distribution $q(\mathbf{d})$ from a tractable distribution family $\mathcal{Q}$ as an optimized approximation of the *a posteriori* distribution $p(\mathbf{d}|\mathbf{y})$. The approximation $q(\mathbf{d})$ can be obtained by minimizing the Kullback-Leibler divergence [12] $\mathcal{D}(q||p)$, i.e.,

$$
\begin{aligned}
q^*(\mathbf{d}) &= \arg \min_{q \in \mathcal{Q}} \mathcal{D}(q||p) \\
&= \arg \max_{q \in \mathcal{Q}} \underbrace{\mathbb{E}_q[-\ln q(\mathbf{d}) + \ln p(\mathbf{d}|\mathbf{y})]}_{\mathcal{L}},
\end{aligned} \tag{12}
$$

where the expectation is taken over $\mathbf{d}$ according to the probability density function $q(\mathbf{d})$. The functional $\mathcal{L}$ on the right-hand side of (12) is referred to as the evidence lower bound (ELBO), which characterizes the difference of the distribution of latent variables and the distribution of respective observed variables [13]. Obviously, the family $\mathcal{Q}$ manages the complexity of the optimization problem in (12). In the sequel, we consider a family $\mathcal{Q}$ that all variables in $q(\mathbf{d})$ are mutually independent, satisfying $q(\mathbf{d}) = \prod_{k,l} q_{k,l}(d_{k,l})$, also known as mean filed approximation. Consequently, each latent variable $d_{k,l}$ is characterized by its own variational factor $q_{k,l}(d_{k,l})$.

With the mean field approximation, $q(\mathbf{d})$ can be determined iteratively by maximizing the ELBO. Since the noise samples $w_{k,l}$ and data symbols $d_{k,l}$ in the DD domain, $\forall k, l$, are independent, $p(\mathbf{d}|\mathbf{y})$ can be rewritten as

$$p(\mathbf{d}|\mathbf{y}) = \prod_{k,l} p(d_{k,l}) p(y_{k,l}|\mathbf{d}). \tag{13}$$

Since SFFT is an orthogonal transformation that does not change the statistics of noise samples, $w_{k,l}$ is still Gaussian distributed with zero mean and variance σ^2, leading to a Gaussian likelihood function $p(y_{k,l}|\mathbf{d})$. After some straightforward manipulations, we can rewrite $p(\mathbf{d}|\mathbf{y})$ in a pairwise form, i.e.,

$$p(\mathbf{d}|\mathbf{y}) \propto \prod_{k,l} \zeta_{k,l}(d_{k,l}) \prod_{k,l} \psi_{k,l}(d_{k,l}, d_{k',l'}), \tag{14}$$

where

$$\zeta_{k,l}(d_{k,l}) = p(d_{k,l}) \exp\left(-\frac{\rho_{k,l}|d_{k,l}|^2 + \epsilon_{k,l}d_{k,l}}{\sigma^2}\right) \quad (15)$$

$$\text{and } \psi_{k,l}(d_{k,l}, d_{k',l'}) = \exp\left(-\frac{\varrho_{k,l,k',l'}d_{k,l}d_{k',l'}}{\sigma^2}\right), \quad (16)$$

with $\rho_{k,l} = \sum_{k',l'} |h_{k',l'}[k,l]|^2$, $\epsilon_{k,l} = 2\sum_{k',l'} \mathcal{R}\{h_{k',l'}[k,l] \cdot y_{k',l'}\}$, and $\varrho_{k,l,k',l'} = 2\mathcal{R}\{h_{k,l}[k,l]h_{k,l}^*[k',l']\}$. Having $p(\mathbf{d}|\mathbf{y})$ in hand, we substitute (14) and $q(\mathbf{d})$ into $\mathcal{L}$ which yields

$$\mathcal{L} = \mathbb{E}_q\left[\sum_{k,l} \ln \psi_{k,l}(d_{k,l}, d_{k',l'}) - \sum_{k,l} \ln \frac{q_{k,l}(d_{k,l})}{\zeta_{k,l}(d_{k,l})}\right]$$

$$= \mathbb{E}_q\left[-\frac{\sum_{k,l}\varrho_{k,l,k',l'}d_{k,l}d_{k',l'}}{\sigma^2} - \sum_{k,l} \ln \frac{q_{k,l}(d_{k,l})}{\zeta_{k,l}(d_{k,l})}\right]. \quad (17)$$

Next, we aim for finding a stationary point of the variational $\mathcal{L}$, which is given by the solution of $\partial\mathcal{L}/\partial q(\mathbf{d}) = 0$. According to Euler-Lagrange equation [14], this requires the partial derivatives of $\mathcal{L}$ with respect to all local functions $q_{k,l}(d_{k,l})$, $\forall k,l$, being zero. To this end, we propose a simple iterative algorithm to update each local function. Let us take the latent variable $d_{k,l}$ in the iter-th iteration as an example to state the proposed algorithm. Given the obtained local functions in the $(\text{iter}-1)$-th iteration $q_{k',l'}^{\text{iter}-1}(d_{k',l'})$, $\forall\{k',l'\} \neq \{k,l\}$, setting the partial derivative $\partial\mathcal{L}/\partial q_{k,l}(d_{k,l})$ to zero leads to

$$\mathbb{E}_{q\backslash k,l}\left[-\frac{1}{\sigma^2}\sum_{k',l'}\varrho_{k,l,k',l'}d_{k,l}d_{k',l'}\right]$$

$$+ \ln \zeta_{k,l}(d_{k,l}) - \ln q_{k,l}^{\text{iter}}(d_{k,l}) + C = 0, \quad (18)$$

where $q\backslash\{k,l\} = \prod_{\{k',l'\}\neq\{k,l\}} q_{k',l'}^{\text{iter}-1}(d_{k,l})$ and C denotes a constant. Then, solving (18) for $q_{k,l}(d_{k,l})$ results in the local distribution, expressed as

$$q_{k,l}^{\text{iter}}(d_{k,l}) \propto \zeta_{k,l}(d_{k,l})\exp\left(\mathbb{E}_{q\backslash\{k,l\}}\left[-\frac{1}{\sigma^2}\sum_{k',l'}\varrho_{k,l,k',l'}d_{k,l}d_{k',l'}\right]\right)$$

$$\propto p(d_{k,l}) \exp\left(-\frac{\rho_{k,l}|d_{k,l}|^2 - m_{k,l}d_{k,l}}{\sigma^2}\right), \quad (19)$$

where $m_{k,l} = \epsilon_{k,l} - \sum_{\{k',l'\}\neq\{k,l\}} \varrho_{k,l,k',l'}\mathbb{E}_{q_{k',l'}^{\text{iter}-1}}[d_{k',l'}]$. In a similar way, we repeat the above procedure to approximate *a posteriori* distributions for all the data symbols iteratively, resulting in the approximate marginals $q_{k,l}^*(d_{k,l})$, $\forall k,l$. Then, the sophisticated MAP detection in (11) is simplified to a problem of finding the maximum of marginal distribution $q_{k,l}^*(d_{k,l})$, i.e.,

$$\hat{d}_{k,l} = \arg\max_{d_{k,l}\in\chi} q_{k,l}^*(d_{k,l}). \quad (20)$$

It is easy to see that the complexity of the detection is reduced to the order of $|\chi|NM$, which is significantly lower than that of the conventional MAP detection.

B. Comparison with MP Receiver

It is well known that the MP receiver converges to local optimum if the factor graph has loops, especially in a rich-scattering environments. Actually, even when the number of paths is as small as 2, there may exist a large number of girth-4

loops. However, for the proposed variational Bayes approach, it can be observed that the second order partial derivative of $\mathcal{L}$ with respect to $q_{k,l}(d_{k,l})$, $\forall k,l$, is

$$\frac{\partial^2\mathcal{L}}{\partial q_{k,l}(d_{k,l})^2} = -\frac{1}{q_{k,l}(d_{k,l})} \leq 0, \quad (21)$$

while $\frac{\partial^2\mathcal{L}}{\partial q_{k,l}(d_{k,l})\partial q_{k',l'}(d_{k,l})} = 0$. Hence, the Hessian matrix of the ELBO is diagonal with the $(kM + l)$th diagonal entry being $-\frac{1}{q_{k,l}(d_{k,l})}$. This observation indicates that $\mathcal{L}$ is a concave functional and its local maximum is the global one. Therefore, iteratively updating each local distribution $q_{k,l}(d_{k,l})$ according to (19) is guaranteed to converge to an optimal approximate distribution $q^*(\mathbf{d})$ that maximizes the ELBO.

Moreover, in terms of computational complexity, both the MP receiver and the proposed detector achieve a linear complexity with the number of symbols. Nevertheless, we can observe from (19) that the update of $q_{k,l}(d_{k,l})$ depends on the marginal means of all the other symbols, which implies that the marginal means need to be calculated only once in a single iteration. Therefore, for a time-varying channel consisting of P propagation paths, the total complexity of the proposed algorithm is $\mathcal{O}(|\chi|NMP)$. On the contrary, for the MP receivers, the messages have to be calculated individually for different connected factor vertices, leading to a computational complexity of $\mathcal{O}(|\chi|NMP^2)$.

To sum up, the proposed variational Bayes approach is attractive in rich-scattering environments, owing to its low-complexity and guaranteed convergence feature.

IV. NUMERICAL RESULTS

This section illustrates the numerical results via Monte Carlo simulations. All simulation results are averaged from 2×10^4 OTFS frames and for each OTFS frame, we set $M = 512$ and $N = 128$ indicating that there are 128 time slots and 512 subcarriers in the TF domain. The carrier frequency is 4 GHz and the subcarrier spacing is 15 kHz. Quadrature phase shift keying (QPSK) modulation is utilized for symbol mapping. The speed of the mobile user is set to be $v = 120$ km/h, leading to a maximum Doppler frequency shift index $k_{\nu_{\max}} = 4$. We assume that the channel information is perfectly known at the receiver and the maximum delay index is $l_{\tau_{\max}} = 10$. The Doppler index of the ith path is drawn with equal probabilities from the set $[-k_{\nu_{\max}}, k_{\nu_{\max}}]$ and the delay index belongs to $[1, l_{\tau_{\max}}]$ excluding the first path which always satisfies $l_{\tau_1} = 0$. The channel coefficients h_i are generated according to the distribution $h_i \sim \mathcal{CN}(0, q^{l_{\tau_i}})$, where the normalized power delay profile $q^{l_{\tau_i}} = \frac{\exp(-0.1l_{\tau_i})}{\sum_i \exp(-0.1l_{\tau_i})}$. The number of iterations is set to 10.

We compare the bit error rate (BER) performance versus $\mathrm{E_s}/\sigma^2$ corresponding to the proposed variational Bayes (VB) method and the conventional MP for OTFS modulation over delay-Doppler channels in Fig. 1, where $\mathrm{E_s}$ denotes the symbol energy. Two scenarios with $P = 4$ and $P = 9$ paths are considered. It is observed that in the 4-path scenario, the proposed algorithm and the MP method have virtually the same performance. When $P = 9$, the performance of the proposed VB method is significantly better than that of the MP receiver. This is because in the $P = 9$ scenario, the MP receiver converges to a local optimum due to the large amount of
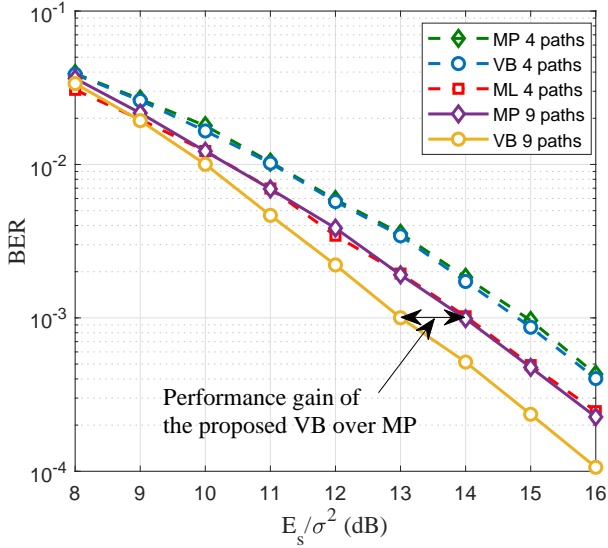
Fig. 1. Impact of number of paths on the BER performance.

Fig. 2. BER performance versus the number of iterations.

girth-4 loops. Moreover, the performance improvement brought by our proposed VB detector is magnified by the increasing number of paths, compared to the MP receiver. In fact, due to its convergence-guaranteed property, the proposed receiver can exploit the TF diversity gain more efficiently offered by OTFS modulation than that of the MP receiver. As a benchmark algorithm, the performance for the MAP detector with 4 paths is illustrated[1]. We can observe that the proposed VB-based receiver approaches the performance of the MAP detector while the complexity is significantly reduced.

To illustrate the convergence behaviors, we plot the BER performance of the proposed VB method and the MP method versus the number of iterations in Fig. 2, where E_s/σ^2 is set to 15 dB. Again, two scenarios with different number of paths are included. We see that increasing the number of iterations leads to a better performance for both algorithms. Moreover, benefiting from the convexity of the ELBO, the proposed VB algorithm converges much faster than the MP receiver in the 4-path scenario even they converge to almost the same BER. On the other hand, for the 9-path scenario, the proposed VB algorithm can converge faster and achieve a better detection performance compared to that of the MP receiver. Thus, the proposed VB method is very attractive in OTFS systems, especially when the number of paths is large.

V. CONCLUSIONS

This paper proposed a variational Bayes-based receiver for the promising OTFS modulation technology. To reduce the complexity of the conventional MAP receiver, we approximate the joint *a posteriori* distribution by the product of local marginals via maximizing the ELBO. By choosing a distribution family for the VB method appropriately, the ELBO is convexified which guarantees the convergence of the proposed variational Bayes approach. We further analyzed the complexity of the proposed algorithm and showed that it is lower than that of both MAP and MP receiver. Simulation results confirmed the fast convergence
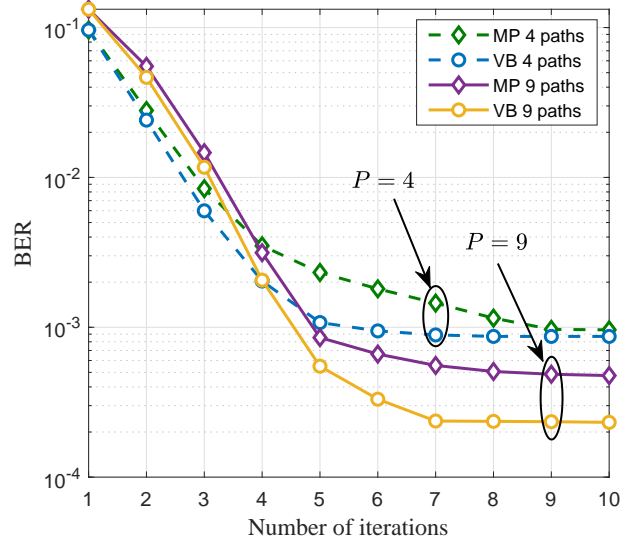
of the proposed algorithm and its superior detection performance compared to the MP algorithm, especially in practical multipath scenarios.

REFERENCES

[1] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge University Press, 2017.

[2] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 957–961, 2018.

[3] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. and Networking Conf.*, 2017, pp. 1–6.

[4] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

[5] A. Farhang, A. RezazadehReyhani, L. E. Doyle, and B. Farhang-Boroujeny, "Low complexity modem structure for OFDM-Based orthogonal time frequency space modulation," *IEEE Commun. Lett.*, vol. 7, no. 3, pp. 344–347, Jun. 2018.

[6] A. RezazadehReyhani, A. Farhang, M. Ji, R. R. Chen, and B. Farhang-Boroujeny, "Analysis of discrete-time MIMO OFDM-Based orthogonal time frequency space modulation," in *Proc. IEEE Intern. Commun. Conf.*, May 2018, pp. 1–6.

[7] G. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, 2019.

[8] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. theory*, vol. 47, no. 2, pp. 498–519, 2001.

[9] P. Raviteja, K. T. Phan, Q. Jin, Y. Hong, and E. Viterbo, "Low-complexity iterative detection for orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, 2018, pp. 1–6.

[10] A. T. Ihler, W. F. John III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, no. May, pp. 905–936, 2005.

[11] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *The A. Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[13] R. Giordano, T. Broderick, and M. I. Jordan, "Covariances, robustness and variational bayes," *J. Mach. Learn. Res.*, vol. 19, no. 1, pp. 1981–2029, 2018.

[14] I. M. Gelfand, R. A. Silverman *et al.*, *Calculus of variations*. Courier Corporation, 2000.

[1]The MAP detection is obtained by exhaustive search. Note that, the curve for the MAP detector with 9 paths is not shown due to the prohibitively high computational complexity.