ABSTRACT

In this paper, a novel decision-feedback differential detection (DF-DD) scheme based on linear prediction is proposed for 16-level differentially encoded amplitude-phase-shift keying (16DAPS K). Unlike previously proposed DF-DD schemes, this scheme does not only provide a performance gain over conventional differential detection (DD) under additive white Gaussian noise (AWGN) conditions but also under general Ricean fading conditions. A further important advantage of the novel scheme is that it does not degrade under frequency offset. The linear predictor coefficients may be updated using the recursive least-squares (RLS) algorithm, which can start blind, i.e., without a priori knowledge about the channel statistics and without training sequence. This makes the scheme attractive for application in mobile communications since the statistics of the nonstationary mobile channel can be tracked.

1. INTRODUCTION

Recently, 16-level differentially encoded amplitude-phase-shift keying (16D APS K) has gained high attention in research as well as in practice (see e.g., [1, 2, 3, 4, 5] and references therein). A possible field of application for 16D APS K is future personal communication systems which will require bandwidth efficient modulation formats because of an increasing demand for high data rate services. Especially for mobile communications, differential detection (DD) of 16D APS K is very attractive because of its simplicity of implementation and its robustness against phase variations [1, 2]. The main drawback of conventional DD when applied to D APS K signals is the significant loss in power efficiency under additive white Gaussian noise (AWGN) conditions in comparison to coherent detection (CD). Therefore, improved noncoherent detection schemes have been proposed in literature. Divsalar et al. [6], Suzuki et al. [7], and Machida et al. [8] reported three different multiple-symbol detection (MSD) schemes. However, these MSD schemes are not well suited for implementation because of their complexity. A strategy well suited for implementation is decision-feedback differential detection (DF-DD). The first DF-DD scheme for 16D APS K was proposed by Adachi et al. [2]. Recently, it has been shown by Wei et al. [3] that an improved version of this scheme approaches the error performance of coherent detection (CD) if the number of feedback symbols tends to infinity. Another DF-DD scheme based on multiple-symbol detection (MSD) was proposed in [4, 5]. It provides a performance gain over the scheme of [3] for the AWGN channel if only a finite number of feedback symbols is applied. Although 16D APS K is especially attractive for application in fading channels [1], the above mentioned DF-DD schemes [2, 3, 4, 5] are designed for the AWGN channel and under fast fading conditions they perform even worse than conventional DD. The DF-DD scheme proposed in this paper is based on linear prediction and offers a performance gain over conventional DD in general flat Ricean fading channels. The predictor coefficients may be updated using a recursive least-squares (RLS) algorithm. The adaptive algorithm can start blind, i.e., no training sequence and no a priori knowledge about the channel statistics are required.

2. TRANSMISSION MODEL

Fig. 1 shows a block diagram of the transmission model under consideration. For the following, all signals are represented by their complex-valued baseband equivalents and ideal symbol timing is assumed. 16D APS K may be viewed as a combination of 2DASK (differential amplitude-phase keying) and 8DPSK (differential phase-shift keying), which are independent of each other [2]. The information bearing amplitude ratios \( \Delta R[k] \in \mathcal{A}_{\Delta} = \{ a_0, 1 / a_0 \} \) and phase difference symbols \( a[k] \in \mathcal{A}_a = \{ e^{j\pi\nu / 4} \mid \nu \in \{ 0, 1, \ldots, 7 \} \} \) are differentially encoded at the transmitter; the transmitted amplitude symbols \( R[k] \in \mathcal{A}_R = \{ R_H, R_L \} \) and phase symbols \( b[k] \in \mathcal{A}_b \) are given by

\[
R[k] = \Delta R[k] R[k-1] \tag{1}
\]

\[
b[k] = a[k] \in \mathcal{A}_a
\]
and
\[ b[k] = a[k]b[k-1], \]
respectively. Here, \( \Delta R[k] = 1 \) corresponds to message bit ‘0’, whereas \( \Delta R[k] = a_0 \) and \( \Delta R[k] = 1/a_0 \) correspond to message bit ‘1’. Note, that \( \Delta R[k] \in \{ 1, a_0 \} \) and \( \Delta R[k] \in \{ 1, 1/a_0 \} \) if \( R[k-1] = R_L \) and \( R[k-1] = R_H \), respectively. For the phase difference symbols \( a[k] \) Gray mapping is applied. The transmitted symbols \( s[k] \) can be written as
\[ s[k] = R[k]b[k]. \]
Both, transmit filter \( H_t(f) \) and receiver filter \( H_r(f) \), have square-root Nyqst characteristic. Hence, no intersymbol interference occurs as long as the continuous-time fading process \( f[f(t) \) with single-sided bandwidth \( B_f \) and the factor \( e^{j2\pi \Delta f T} \), where \( \Delta f \) is the frequency offset between modulator and demodulator, do not change significantly during one symbol interval \( T \). In this paper, we assume that these conditions are fulfilled for \( B_f T \leq 0.04 \) and \( \Delta f T \leq 0.04 \), respectively. Taking into account the above mentioned assumptions, the received signal sample \( r[k] \) may be expressed as
\[ r[k] = e^{j\theta}e^{j2\pi \Delta f T k} f[k]s[k] + n[k], \]
where the discrete-time Rician fading process \( f[k] \) and the noise process \( n[k] \) are correlated and uncorrelated complex Gaussian random processes, respectively. Furthermore, \( f[k] \) and \( n[k] \) are mutually uncorrelated. \( \theta \) denotes an unknown, constant phase shift. Due to an appropriate normalization, \( f[k] \) has mean power \( |f[k]|^2 = 1 \) and \( n[k] \) has variance \( \sigma_n^2 = E \{ |n[k]|^2 \} = \frac{N_0}{E_s} \). Here, \( E \{ \cdot \} \) denotes expectation and \( E_s \) is the mean received energy per symbol, whereas \( N_0 \) is the single-sided power spectral density of the underlying passband noise process. At the receiver, the estimated amplitude ratio \( \Delta \hat{R}[k] \in \Delta A_R \) and the estimated phase difference symbol \( \hat{a}[k] \in A_a \) are determined by prediction-based DF-DD.

3. PREDICTION-BASED DF-DD SCHEME

In this section, prediction-based DF-DD is introduced. Using Eqs. (1)–(4), the received 16DAPSK signal sample \( r[k] \) may be rewritten as
\[ r[k] = e^{j\theta}e^{j2\pi \Delta f T k} f[k]s[k-1] + n[k]. \]
For optimum coherent detection (CD) of \( a[k] \) and \( R[k] \), the reference symbol
\[ r_{ref}[k-1] \triangleq e^{j\theta}e^{j2\pi \Delta f T k} f[k]s[k-1] \]
has to be known. For conventional DD [1], \( r[k-1] \) is used as an estimate for \( r_{ref}[k-1] \) and because of the noisy character of \( r[k-1] \), a performance degradation compared to CD is inevitable. Here, we propose to estimate \( r_{ref}[k-1] \) not only from \( r[k-1] \) but from the last \( N-1 \) observed symbols \( r[k-\nu], 1 \leq \nu \leq N-1 \), i.e., the decision is based on \( N \) received signal samples. The coefficients \( p_{\nu}^2[k], 1 \leq \nu \leq N-1 \), of the proposed linear estimator are obtained by minimizing the mean-squared error (MSE) between \( r_{ref}[k-1] \) and an estimated reference symbol
\[ r_{e}[k-1] \triangleq \sum_{\nu=1}^{N-1} p_{\nu}^2[k] r[k-\nu]. \]

It can be shown that the resulting estimator coefficients for estimation of \( r_{ref}[k-1] \) are also the solution to the problem of estimating \( r[k] \) from \( \Delta R[k]a[k] r_{ref}[k-1] \) with minimum MSE [9].

The error variance \( \sigma_{MSE}^2(s[k]) \) with \( s[k] \triangleq [s[k], s[k-1], \ldots, s[k-N+1]]^T \) [\( (\cdot)^T \) denotes transposition] for estimation of \( r[k] \) is given by
\[ \sigma_{MSE}^2(s[k]) = \mathcal{E} \left\{ \left[ r[k] - \Delta R[k] a[k] r_{ref}[k-1] \right]^2 \right\} \]
\[ \triangleq \mathcal{E} \left\{ \sum_{\nu=1}^{N-1} p_{\nu}^2[k] \left[ e^{j\theta}e^{j2\pi \Delta f T(k-\nu)} f[k-\nu] s[k-\nu] + n[k-\nu] \right]^2 \right\}, \]
where \( s[k] \) is assumed to be known.

In order to get a simple adaptation algorithm for the estimator coefficients we introduce the substitution
\[ p_{\nu}^2[k] = \frac{p_{\nu} s[k-1]}{s[k-\nu]} = \frac{p_{\nu} \prod_{\mu=1}^{\nu-1} \Delta R[k-\mu] a[k-\mu]}{1 \leq \nu \leq N-1}, \]
with fixed coefficients \( p_{\nu} \), in Eq. (8). In addition, we average \( \sigma_{MSE}^2(s[k]) \) also over \( s[k] \). This is necessary since otherwise the resulting estimator coefficients would depend on \( s[k-\nu], 0 \leq \nu \leq N-1 \), and therefore the RLS algorithm could not be used for adaptation. Now the MSE variance \( \sigma_{MSE}^2 \) is given by
\[ \sigma_{MSE}^2 = \mathcal{E} \left\{ \frac{[s[k]]^2}{E_s} \cdot \left[ c[k] - \sum_{\nu=1}^{N-1} p_{\nu} c[k-\nu] \right]^2 \right\}, \]
where averaging is also performed over \( s[k] \) and the definition
\[ c[k] \triangleq e^{j2\pi \Delta f T k} f[k] + e^{-j\theta} n[k] \]
is used. Eq. (10) shows that minimizing \( \sigma_{MSE}^2 \) is identical with determining the coefficients \( p_{\nu} \), \( 1 \leq \nu \leq N-1 \), of a linear Wiener predictor [10] for process \( c[k] \) since the factor \( [s[k]]^2 \) in Eq. (10) has no influence on the estimator (predictor) coefficients [9]. With this predictor, the initially desired estimate for the reference symbol \( r_{ref}[k-1] \) can be constructed. Due to
solution is not strictly optimum, but it is the optimum time-invariant solution. Moreover, it is well suited for practical implementation and yields high performance as will be shown by computer simulations. The coefficients \( p_\nu \) can be obtained from the Yule–Walker equations. However, since 16D APSK is particularly interesting for application in mobile communications [1], the channel may be nonstationary and the channel statistics are not known a priori; thus, it is advantageous to use an adaptive algorithm for determination of the predictor coefficients (cf. Section 4).

In practice, the transmitted symbols \( \Delta R[k - \nu], a[k - \nu], 1 \leq \nu \leq N - 2 \), required in Eq. (9) are not known at the receiver and have to be replaced by previously detected symbols \( \hat{\Delta R}[k - \nu], \hat{a}[k - \nu], 1 \leq \nu \leq N - 2 \),

\[
\hat{r}[k - 1] = \sum_{\nu=1}^{N-1} p_\nu \prod_{\mu=1}^{\nu-1} \Delta \hat{R}[k - \mu]\hat{a}[k - \mu]r[k - \nu] \quad (12)
\]

has to be used as an estimate for the reference symbol \( r_{\text{ref}}[k - 1] \) instead of \( r_r[k - 1] \); cf. Eqs. (7), (9).

The decisions on \( \hat{a}[k] \) and \( \hat{\Delta R}[k] \) are made by using \( r[k] \) and \( \hat{r}[k - 1] \):

- **Phase Decision**

For the phase decision, the complex plane is divided into 8 sectors corresponding to the 8 possible values of \( \hat{a}[k] \). Then, the sector, into which the phase decision variable

\[
g_p[k] = r[k]r^*_r[k - 1] \quad (13)
\]

falls, determines the estimate \( \hat{a}[k] \).

- **Amplitude Decision**

For the amplitude decision, the amplitude decision variable

\[
g_a[k] = \left| \frac{r[k]}{\hat{r}_r[k - 1]} \right| \quad (14)
\]

is introduced. Now, the estimated amplitude \( \hat{R}[k - 1] = \hat{\Delta R}[k - 1] \hat{R}[k - 2] \) of the previously transmitted symbol \( s[k - 1] \) has to be taken into account and two cases have to be distinguished:

- \( \hat{R}[k - 1] = R_L \)

Here, \( \hat{\Delta R}[k] \) is obtained from

\[
\Delta \hat{R}[k] = \begin{cases} a_0, & g_a[k] \geq \gamma_L, \\ 1, & g_a[k] < \gamma_L, \end{cases} \quad (15)
\]

where \( \gamma_L \) is the decision threshold if \( \hat{R}[k - 1] = R_L \).

- \( \hat{R}[k - 1] = R_H \)

In this case, \( \hat{\Delta R}[k] \) is given by

\[
\Delta \hat{R}[k] = \begin{cases} 1, & g_a[k] \geq \gamma_H, \\ 1/a_0, & g_a[k] < \gamma_H, \end{cases} \quad (16)
\]

with threshold \( \gamma_H \).

Although the arithmetic means \( \gamma_L = 0.5(1 + a_0) \) and \( \gamma_H = 0.5(1 + 1/a_0) \) are not the optimum decision thresholds, they provide a good performance [9] and will be used for the simulations in Section 5.

4. ADAPTATION OF THE PREDICTOR COEFFICIENTS USING THE RLS ALGORITHM

An adaptive algorithm may be employed for determination of the predictor coefficients. Since the eigenvalue spread of the autocorrelation matrix of \( c[\cdot] \) can become quite large [9], the convergence speed of the simple least-mean-square (LMS) algorithm is slow [10]. Here the RLS algorithm is proposed, i.e., instead of the error variance \( \sigma^2_{\text{ME}} \) according to Eq. (10), the related cost function

\[
J[k] \triangleq \sum_{\nu=1}^{N} w^{k-\nu} \left| r[\nu] - p^T_k \hat{r}_\nu \right|^2, \quad (17)
\]

with

\[
\hat{r}_\nu \triangleq \left[ r[\nu, 1] \ r[\nu, 2] \ldots r[\nu, N - 1] \right]^T, \quad (18)
\]

\[
r[\nu, \eta] \triangleq r[\nu - \eta] \prod_{\mu=0}^{\eta-1} \Delta \hat{R} [\nu - \mu] \hat{a}[\nu - \mu], \quad (19)
\]

\[
p_k \triangleq [p_1[k] \ p_2[k] \ldots p_{N-1}[k]]^T, \quad (20)
\]

is minimized. \( J[k] \) follows from Eqs. (8) and (9).\(^2\) \( w, 0 < w < 1, \) is the forgetting factor of the RLS algorithm.

The RLS algorithm corresponding to \( J[k] \) consists of the following equations [10]

\[
p_k = \frac{P_{k-1} \hat{r}_k}{w + \hat{r}^H_k P_{k-1} \hat{r}_k}, \quad (21)
\]

\[
\zeta[k] = r[k] - p^T_k \hat{r}_k, \quad (22)
\]

\[
P_k = w^{-1} P_{k-1} - w^{-1} q_k \zeta^H_k P_{k-1}, \quad (23)
\]

\[
p_k = p_{k-1} + q_k \zeta[k]^T, \quad (24)
\]

where \( [\cdot]^H \) denotes Hermitian transposition. We propose the initialization

\[
P_0 = \delta^{-1} I, \quad (25)
\]

\[
p_0 = [1 \ 0 \ 0 \ldots \ 0]^T, \quad (26)
\]

where \( \delta \) is a small positive constant [10]. The initialization of \( p_k \) according to Eq. (26) ensures that the algorithm can start blind, without a priori knowledge of the channel statistics and without training sequence, since the structure of the proposed DF-DD scheme with coefficient vector \( p_0 \) is almost identical with the scheme of Wei et al. [3] resulting for \( N = 2 \).\(^3\)

\(^2\)Note, that in Eq. (17) averaging over \( s[\cdot] \) (cf. Section 3) is included automatically, i.e., equivalence to Eq. (10) is given.

\(^3\)The amplitude decision variable resulting for \( p_0 \) is different from that of [3], however, this has no influence on performance.
5. SIMULATION RESULTS

In this section, the performance of adaptive prediction-based DF-DD for 16DAPSK is evaluated for AWGN and (Rayleigh and Ricean) fading conditions. Moreover, the scheme is compared with MSD-based DF-DD [4, 5] and the detector proposed by Wei et al. [3]. Although the amplitude decision rules for conventional DD [1] and the scheme of Wei et al. [3] for \( N = 2 \) are different, both yield almost the same bit error rate (BER) for all cases. For fixed \( N = 2 \), the schemes of Wei et al. yield a small gain (\( \Delta fT > 0 \) dB) for an AWGN channel with \( 10\log_{10}(E_b/N_0) = 15 \) dB and a ring ratio of \( a_0 = 1.8 \). Adaptive DF-DD is unaffected by a constant frequency offset, whereas the other schemes degrade considerably for \( \Delta fT > 0 \). This degradation is the more severe, the larger \( N \). This means for MSD-based DF-DD and the DF-DD scheme proposed by Wei et al., there is a trade-off between performance under pure AWGN conditions and performance in the presence of a frequency offset. For adaptive DF-DD, however, this trade-off does not exist since the adapted predictor coefficients compensate the frequency offset (see also [9]). Hence, an additional frequency regulation as required by the other schemes is not necessary here.

![Figure 2: BER vs. 10\log_{10}(E_b/N_0) for DF-DD proposed by Wei et al., MSD-based DF-DD, adaptive (prediction-based) DF-DD (\( w = 1.0 \)), gene-aided adaptive (prediction-based) DF-DD (simulation (\( w = 1.0 \)) and theory), and CD for an AWGN channel.](image)

Figure 2: BER vs. 10\log_{10}(E_b/N_0) for DF-DD proposed by Wei et al., MSD-based DF-DD, adaptive (prediction-based) DF-DD (\( w = 1.0 \)), gene-aided adaptive (prediction-based) DF-DD (simulation (\( w = 1.0 \)) and theory), and CD for an AWGN channel.

better than the scheme of Wei et al. and adaptive DF-DD. However, as \( N \) and \( E_b/N_0 \) increase, the difference becomes smaller. For \( N = 10 \) the three techniques perform almost equally well and approach CD. Note, that for \( N = 2 \) the performance of adaptive prediction-based DF-DD and that of the scheme of Wei et al. are almost identical and thus, adaptive DF-DD with \( N = 2 \) is not included in Fig. 2. For \( N > 2 \), the scheme of Wei et al. yields a small gain (\( \approx 0.2 \) dB for \( N = 3 \), \( \approx 0.1 \) dB for \( N = 10 \)) over adaptive DF-DD. For gene-aided adaptive DF-DD there is a very good agreement between simulation and theory.

It has been shown in Fig. 2, that for zero frequency offset and fixed \( N \) MSD-based DF-DD and the scheme of Wei et al. perform better than adaptive (prediction-based) DF-DD (\( w = 1.0 \)). Now, it is interesting to investigate how the different schemes are affected by frequency offset. Thus, Fig. 3 shows BER vs. normalized frequency offset \( \Delta fT \) for an AWGN channel with \( 10\log_{10}(E_b/N_0) = 15 \) dB and a ring ratio of \( a_0 = 1.8 \). Adaptive DF-DD is unaffected by a constant frequency offset, whereas the other schemes degrade considerably for \( \Delta fT > 0 \). This degradation is the more severe, the larger \( N \). This means for MSD-based DF-DD and the DF-DD scheme proposed by Wei et al., there is a trade-off between performance under pure AWGN conditions and performance in the presence of a frequency offset. For adaptive DF-DD, however, this trade-off does not exist since the adapted predictor coefficients compensate the frequency offset (see also [9]). Hence, an additional frequency regulation as required by the other schemes is not necessary here.

![Figure 3: BER vs. \( \Delta fT \) for the DF-DD scheme proposed by Wei et al., MSD-based DF-DD, and adaptive (prediction-based) DF-DD (\( w = 1.0 \)). 10\log_{10}(E_b/N_0) = 15 \) dB and \( a_0 = 1.8 \) are valid in all cases.](image)

Figure 3: BER vs. \( \Delta fT \) for the DF-DD scheme proposed by Wei et al., MSD-based DF-DD, and adaptive (prediction-based) DF-DD (\( w = 1.0 \)). 10\log_{10}(E_b/N_0) = 15 \) dB and \( a_0 = 1.8 \) are valid in all cases.

In Fig. 4, a Rayleigh fading channel with normalized fading bandwidth \( B_yT = 0.03 \) and a ring ratio of \( a_0 = 2.0 \) are considered. Note, that again the BER for adaptive DF-DD with \( N = 2 \) is not shown in Fig. 4 since the same performance as for the scheme of Wei et al. for \( N = 2 \) is obtained. It can be observed, that for adaptive DF-DD the irreducible error floor decreases significantly with increasing \( N \), whereas it increases for the scheme of Wei et al.

The MSD-based scheme is not considered in Fig. 4 since it performs poor for Rayleigh and Ricean fading. The reason for this behaviour is that MSD-based DF-DD is designed for the noncoherent AWGN channel, i.e., only an unknown phase shift is tolerated. Thus, the resulting amplitude decision rule according to [4,
Eq. (6) is only optimum if the channel gain is constant. For fading this is not the case and the scheme degrades.

Figure 4: BER vs. $10 \log_{10}(E_b/N_0)$ for DF-DD proposed by Wei et al., adaptive (prediction-based) DF-DD ($w = 1.0$), genie-aided adaptive (prediction-based) DF-DD (simulation ($w = 1.0$) and theory), and CD for a Rayleigh fading channel with $B_f T = 0.03$.

Fig. 5 shows the performance of the same schemes for transmission over a Ricean fading channel with a Ricean factor [11] of 7 dB and normalized fading bandwidth $B_f T = 0.03$. For $N = 2$, the scheme of Wei et al. and adaptive DF-DD (not shown here) yield the same performance. For $N = 3$, adaptive DF-DD lowers the error floor whereas the scheme of Wei et al. performs poorer than for $N = 2$.

6. CONCLUSIONS

DF-DD based on linear prediction for 16DPSK signals transmitted over Ricean fading channels has been introduced in this paper. The predictor coefficients may be updated using the RLS algorithm. By this, nonstationary channel statistics can be tracked. The adaptation algorithm can start blind, i.e., no a priori knowledge about the channel statistics and no training sequence are required. This makes the novel scheme particularly interesting for application in mobile communications. Our simulations have confirmed that adaptive prediction-based DF-DD can improve the performance of conventional DD under AWGN, Rayleigh, and Ricean conditions.

REFERENCES


Figure 5: BER vs. $10 \log_{10}(E_b/N_0)$ for DF-DD proposed by Wei et al., adaptive (prediction-based) DF-DD ($w = 1.0$), genie-aided adaptive (prediction-based) DF-DD (simulation ($w = 1.0$) and theory), and CD for a Ricean fading channel with a Ricean factor of 7 dB and $B_f T = 0.03$.