Efficient Near-ML Detection for MIMO Channels: The Sphere-Projection Algorithm

Dominik Seethaler, Harold Artés, and Franz Hlawatsch

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
phone: +43 1 58801 38958, fax: +43 1 58801 38999, email: dominik.seethaler@tuwien.ac.at
web: http://www.nt.tuwien.ac.at/dspgroup/time.html

Abstract— It is well known that suboptimal detection schemes for MIMO spatial multiplexing systems (equalization-based as well as nulling-and-cancelling detectors) cannot exploit all of the available diversity. In this paper, we show that this inferior performance is primarily caused by poorly conditioned channel realizations. We then present the novel *sphere-projection algorithm* (SPA) that is robust to poorly conditioned channels. The SPA is a computationally efficient add-on to standard suboptimal detectors. Simulation results show that the SPA is able to achieve near-ML performance and significantly increased diversity gains. The SPA's computational complexity is comparable to that of nulling-and-cancelling detectors and only a fraction of that of the Fincke-Phost sphere-decoding algorithm for ML detection.

I. INTRODUCTION

It is well known that the diversity offered by multiple-input/multiple-output (MIMO) fading channels cannot be fully exploited by suboptimal detectors (linear equalization and nulling-and-cancelling schemes [2]). Maximum-likelihood (ML) detection exploits all of the available diversity but tends to be computationally intensive. Thus, there is a strong demand for computationally efficient suboptimal detectors that can exploit a large part of the available diversity.

In this paper, we present the novel *sphere-projection algorithm* (SPA) for data detection in spatial multiplexing systems, and we show that the SPA can achieve near-ML performance at low complexity. The SPA is motivated by the observation that the inferior average performance of standard suboptimal detectors is mainly caused by the occurrence of "bad" (i.e., poorly conditioned) channel realizations. The SPA is a simple add-on to standard suboptimal schemes that significantly increases the robustness of these schemes to bad channels.

Our paper is organized as follows. In the remainder of this section, we describe the system model and briefly review existing detection schemes. In Section II, the effect of bad channels on equalization-based detection is analyzed. The SPA is introduced in Section III. Finally, in Section IV the performance of the SPA is assessed through simulation results.

A. System Model

We consider a MIMO channel with M_T transmit antennas and $M_R \geq M_T$ receive antennas (briefly termed an (M_T, M_R)

channel). We assume a spatial multiplexing system such as V-BLAST [2] where the mth data symbol (or *layer*) d_m is directly transmitted on the mth transmit antenna. For any given time instant, this leads to the well-known baseband model

$$\mathbf{r} = \mathbf{Hd} + \mathbf{w}, \tag{1}$$

with the data vector $\mathbf{d} \triangleq (d_1 \cdots d_{M_T})^T$, the $M_R \times M_T$ channel matrix $\mathbf{H}$, the received vector $\mathbf{r} \triangleq (r_1 \cdots r_{M_R})^T$, and the noise vector $\mathbf{w} \triangleq (w_1 \cdots w_{M_R})^T$. The data symbols d_m are assumed zero-mean and white with variance σ_d^2. The components of the noise vector $\mathbf{w}$ are assumed to be independent and circularly symmetric complex Gaussian with variance σ_w^2. The channel $\mathbf{H}$ is considered constant over a block of N consecutive time instants and perfectly known at the receiver.

B. Review of Detection Schemes

Major suboptimal detection schemes for spatial multiplexing systems include linear equalization followed by quantization and nulling-and-cancelling (or decision-feedback) detection [2]. In linear equalization based schemes, the detected data vector is $\hat{\mathbf{d}} = Q\{\mathbf{y}\}$ with $\mathbf{y} = \mathbf{Gr}$, where $Q\{\cdot\}$ denotes component-wise quantization according to the symbol alphabet. The zero-forcing (ZF) equalizer is given by $\mathbf{G} = \mathbf{H}^\# = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$ (assuming that $\mathbf{H}$ has full rank), while the minimum mean-square error (MMSE) equalizer is given by $\mathbf{G} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma_w^2}{\sigma_d^2}\mathbf{I}\right)^{-1}\mathbf{H}^H$ [3]. In contrast to linear equalization schemes, where all layers are detected jointly, nulling and cancelling (NC) uses a serial decision-feedback approach to detect each layer separately. NC is based on the ZF or MMSE approach [2]; the corresponding detectors will be referred to as NC-ZF and NC-MMSE, respectively.

Finally, ML detection [4], [5] yields minimum error probability for equally likely data vectors. For our system model (1) and our assumptions, the ML detector is given by

$$\hat{\mathbf{d}}_{\mathrm{ML}} = \arg\min_{\mathbf{d} \in \mathcal{D}} \left\{ \|\mathbf{r} - \mathbf{Hd}\|^2 \right\}, \tag{2}$$

with $\mathcal{D}$ the set of data vectors. The computational complexity of ML detection grows exponentially with M_T. Using the Fincke-Phost sphere-decoding algorithm for ML detection [4], an *average* complexity of roughly $\mathcal{O}(M_T^3)$ is obtained [6].
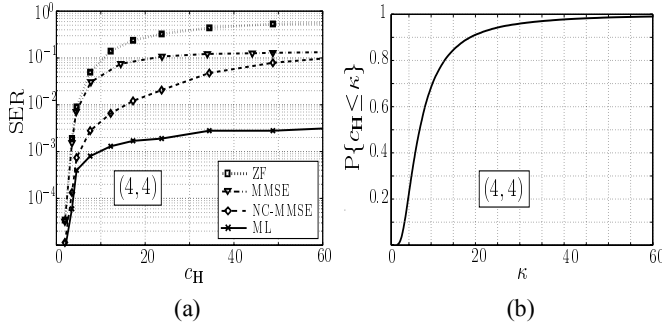
Fig. 1. Detector performance and channel condition number: (a) SER performance of various detection schemes versus the condition number $c_{\mathbf{H}}$ of the respective channel realization $\mathbf{H}$ for a $(4, 4)$ MIMO channel, 4-QAM modulation, and a fixed SNR of $15\,\mathrm{dB}$, (b) cumulative distribution function of $c_{\mathbf{H}}$. $\mathbf{H}$ is modeled iid Gaussian.
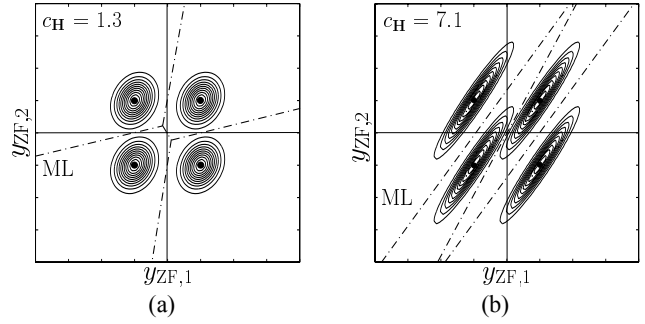

Fig. 2. Pdf contour lines and ZF and ML decision regions in the ZF-equalized domain for a $(2, 2)$ channel and BPSK modulation: (a) "Good" channel realization with $c_{\mathbf{H}} = 1.3$, (b) "bad" channel realization with $c_{\mathbf{H}} = 7.1$.

II. THE "BAD CHANNEL" EFFECT

As a motivation for the SPA, we first study the effect of bad channel realizations on equalization-based detection. We will use the singular value decomposition (SVD) $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where the diagonal matrix $\boldsymbol{\Sigma}$ contains the singular values σ_m of $\mathbf{H}$ (assumed indexed in nonincreasing order) and the columns of $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors of $\mathbf{H}$, respectively [7]. The *condition number* $c_{\mathbf{H}} = \sigma_1/\sigma_{M_{\mathrm{T}}}$ is the ratio of the largest and smallest singular values; it is large for a "bad" channel.

A. Detector Performance and Bad Channels

Experiments suggest that the performance of suboptimal detection for a given channel realization $\mathbf{H}$ strongly depends on the condition number $c_{\mathbf{H}}$. In Fig. 1(a), we show the symbol error rate (SER) of various detection schemes versus $c_{\mathbf{H}}$. For this simulation, we used a $(4, 4)$ channel with iid Gaussian channel matrix entries and a fixed SNR of $15\,\mathrm{dB}$. For $c_{\mathbf{H}}$ larger than about 20, we observe a significant performance gap between suboptimal detection and ML detection. Fig. 1(b) shows that channel realizations with $c_{\mathbf{H}}$ larger than about 20 occur frequently enough to cause a significant degradation of the average performance of suboptimal detection.

Because ML detection usually is too complex (in this context, we note that the Fincke-Phost sphere-decoding algorithm for ML detection has substantially increased complexity in the case of bad channels [4]), there is a strong demand for computationally efficient suboptimal detectors that are able to achieve near-ML performance. The SPA, to be presented in Section III, is designed to satisfy this demand.

B. The Bad Channel Effect in the ZF Domain

After ZF equalization (i.e., in the "ZF domain"), we obtain

$$\mathbf{y}_{\mathrm{ZF}} = \mathbf{H}^{\#}\mathbf{r} = \mathbf{d} + \tilde{\mathbf{w}}.$$

This is the *undistorted* data $\mathbf{d}$ corrupted by the noise $\tilde{\mathbf{w}} = \mathbf{H}^{\#}\mathbf{w}$ that is *correlated* with covariance matrix

$$\mathbf{R}_{\tilde{\mathbf{w}}} = \sigma_w^2(\mathbf{H}^H\mathbf{H})^{-1} = \sigma_w^2\,\mathbf{V}\boldsymbol{\Sigma}^{-2}\mathbf{V}^H.$$

Hence, the contour surfaces of the probability density function (pdf) of $\tilde{\mathbf{w}}$ are hyperellipsoids whose mth principal axis is given by the mth eigenvector $\mathbf{v}_m$ of $\mathbf{R}_{\tilde{\mathbf{w}}}$ (the mth column of $\mathbf{V}$), with its length proportional to $\sigma_{\tilde{w},m} = \sigma_w/\sigma_m$ [8]. Thus, ZF equalization results in a *distortion* of the noise pdf relative to the spherical pdf geometry of $\mathbf{w}$. For illustration, Fig. 2 shows the pdf of the received signal after ZF equalization for a "good" and a "bad" realization of a real-valued $(2, 2)$ channel. The modulation format is BPSK. Also shown are the ZF decision regions (the four quadrants) and the ML decision regions (indicated by dash-dotted lines). For the good channel, the ZF and ML decision regions are similar. For the bad channel, however, they are very different; in particular, two of the boundary lines separating the ML decision regions feature an offset perpendicular to the dominant principal axis $\mathbf{v}_{M_{\mathrm{T}}}$ (which is the direction of the dominant noise component).

Also the MMSE and NC schemes are unable to properly adapt to the noise distortion caused by bad channels, even though their decision regions are somewhat better matched to the distorted pdf of $\tilde{\mathbf{w}}$ than the ZF decision regions [1].

Experiments indicate that for a bad channel, the largest ZF-domain noise component (whose variance is the largest eigenvalue $\sigma_{\tilde{w},M_{\mathrm{T}}}^2$ of $\mathbf{R}_{\tilde{\mathbf{w}}}$ and whose direction is given by the principal axis $\mathbf{v}_{M_{\mathrm{T}}}$) tends to dominate all the other noise components. Hence, this dominant noise component causes the main part of the bad channel effects that are responsible for the poor performance of linear and NC detection.

III. THE SPHERE-PROJECTION ALGORITHM

For simplicity, we hereafter assume a symbol alphabet with constant modulus σ_d. Here, all data vectors $\mathbf{d}$ are located on an M_{T}-dimensional *data hypersphere* $\mathcal{H}$ of radius $R = \sigma_d\sqrt{M_{\mathrm{T}}}$ about the origin. This assumption allows arbitrary PSK constellations. (Our algorithms can be extended to other symbol alphabets by using several hyperspheres.)

We also assume that the symbol alphabet is "line-structured" in that the boundaries of the symbol quantization regions in the 2-D real symbol domain are straight lines. This holds for arbitrary ASK, QAM, and PSK alphabets but not, e.g., for hexagonal constellations. Let P denote the number of boundary lines. For instance, 4-QAM has $P = 2$ orthogonal

boundary lines (see Fig. 3).

We will now describe the SPA. The SPA is an add-on to an arbitrary suboptimal detector, in the following sense. Recall from (2) that the ML detector would minimize $\|\mathbf{r}-\mathbf{Hd}\|^2$ over the entire data vector set $\mathcal{D}$. Let $\hat{\mathbf{d}}_{\text{sub}} \in \mathcal{D}$ denote the result of the given suboptimal detector. This result can be expected to be reasonably good for a good channel. In order to improve the performance for bad channels, we additionally consider a suitably chosen set $\mathcal{D}_+ \subset \mathcal{D}$ of valid data vectors $\mathbf{d}$ that are potentially better than $\hat{\mathbf{d}}_{\text{sub}}$ in the sense of smaller $\|\mathbf{r}-\mathbf{Hd}\|^2$. We then minimize $\|\mathbf{r}-\mathbf{Hd}\|^2$ over the search set $\mathcal{D}_{\text{SP}}$ that consists of $\hat{\mathbf{d}}_{\text{sub}}$ and all data vectors in $\mathcal{D}_+$:

$$\hat{\mathbf{d}}_{\text{SP}} \triangleq \arg\min_{\mathbf{d}\in\mathcal{D}_{\text{SP}}} \left\{\|\mathbf{r}-\mathbf{Hd}\|^2\right\}, \quad \text{with } \mathcal{D}_{\text{SP}} \triangleq \{\hat{\mathbf{d}}_{\text{sub}}\} \cup \mathcal{D}_+ .$$

It remains to discuss the construction of $\mathcal{D}_+$.

A. Combatting the Bad Channel Effect

Let us assume a bad channel. As discussed in Section II-B, in the ZF domain there is a dominant noise component in the direction of $\mathbf{v}_{M_{\text{T}}}$ (see Fig. 2(b)). The ZF-equalized vector $\mathbf{y}_{\text{ZF}} = \mathbf{H}^{\#}\mathbf{r} = \mathbf{d} + \tilde{\mathbf{w}}$ can be decomposed as $\mathbf{y}_{\text{ZF}} = \mathbf{y}_{\mathbf{v}_{M_{\text{T}}}} + \mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$, where $\mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$ is orthogonal to $\mathbf{v}_{M_{\text{T}}}$. We define the *reference line* $\mathcal{L}$ as the straight line parallel to the dominant noise axis $\mathbf{v}_{M_{\text{T}}}$ whose offset from the origin is $\mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$:

$$\mathcal{L}: \quad \mathbf{y}_{\text{ref}}(k) \triangleq k\,\mathbf{v}_{M_{\text{T}}} + \mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}, \qquad k \in \mathbb{C}. \quad (3)$$

$\mathcal{L}$ passes through $\mathbf{y}_{\text{ZF}} = \mathbf{d} + \tilde{\mathbf{w}}$, which is obtained for $k = \mathbf{v}_{M_{\text{T}}}^{H}\mathbf{y}_{\text{ZF}}$. Now consider the distance $\psi_i \triangleq \|\mathbf{r}-\mathbf{Hd}^{(i)}\|$ for some $\mathbf{d}^{(i)} \in \mathcal{D}_+$. Since we assumed a bad channel, the dominant noise component in the direction of $\mathbf{v}_{M_{\text{T}}}$ (i.e., of $\mathcal{L}$) is much larger than all other noise components. Therefore, it makes sense to exclude this dominant noise direction from our distance calculation. Furthermore, to simplify things, we will do as if all of the small noise components had equal variance. With these approximations, it can be shown [1] that $\mathbf{d}^{(i)}$ has a small distance ψ_i if and only if it is close to $\mathcal{L}$. (In fact, if all nondominant noise components were exactly zero, $\mathcal{L}$ would pass right through the transmitted data vector $\mathbf{d}$.)

This suggests to construct $\mathcal{D}_+$ as a set of data vectors that are close to the reference line $\mathcal{L}$. We will next present an efficient algorithm for constructing the data vectors $\mathbf{d}^{(i)} \in \mathcal{D}_+$ and determining their distances $\psi_i = \|\mathbf{r}-\mathbf{Hd}^{(i)}\|$.

B. Efficient Calculation of $\mathcal{D}_+$ and ψ_i^2: Case 1

We first assume that $\mathcal{L}$ intersects the data hypersphere $\mathcal{H}$ (the opposite case will be considered in Section III-C). We here choose $\mathcal{D}_+$ to consist of data vectors $\mathbf{d} \in \mathcal{H}$ close to the intersection set $\mathcal{L} \cap \mathcal{H}$. This intersection set is defined by the equation $\|\mathbf{y}_{\text{ref}}(k)\|^2 = R^2$ or, equivalently,

$$|k|^2 + \left\|\mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}\right\|^2 = R^2, \quad (4)$$

where R is the radius of $\mathcal{H}$. In the complex k plane, this represents an *intersection circle* $\mathcal{I}$ with radius $\tilde{R} \triangleq \sqrt{R^2 - \|\mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}\|^2}$. That is, (4) is satisfied by $k = \tilde{R}\,e^{j\phi}$ with
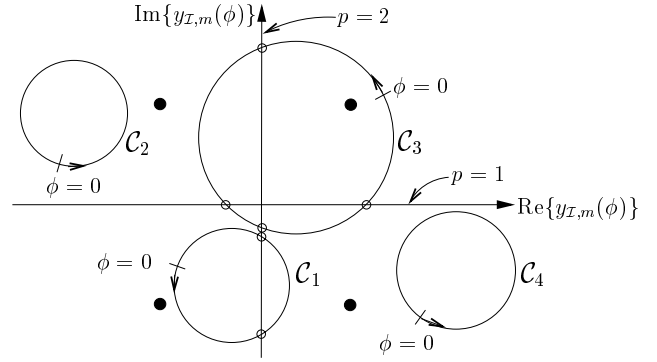

Fig. 3. Component circles $\mathcal{C}_m$ for a $(4, 4)$ channel and 4-QAM modulation ($P = 2$). (All 4 component circles are drawn in a common complex plane.) When moving along the intersection circle $\mathcal{I}$ (by varying ϕ), the mth component $y_{\mathcal{I},m}(\phi)$ of $\mathbf{y}_{\mathcal{I}}(\phi)$ moves along the component circle $\mathcal{C}_m$.

arbitrary angle ϕ. The $\mathbf{y}$ vectors corresponding to $\mathcal{I}$ are thus given by $\mathbf{y}_{\mathcal{I}}(\phi) \triangleq \mathbf{y}_{\text{ref}}(\tilde{R}e^{j\phi}) = \tilde{R}e^{j\phi}\mathbf{v}_{M_{\text{T}}} + \mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$.

Evidently, the data vector nearest to $\mathbf{y}_{\mathcal{I}}(\phi)$ is obtained by component-wise quantization of $\mathbf{y}_{\mathcal{I}}(\phi)$. Thus, we construct $\mathcal{D}_+$ to consist of all data vectors obtained by quantizing $\mathbf{y}_{\mathcal{I}}(\phi)$ for all possible angles ϕ. Equivalently, $\mathcal{D}_+$ *consists of all data vectors whose ZF decision regions (i.e., the decision regions of the ZF detector) are pierced by the intersection set* $\mathcal{L} \cap \mathcal{H}$.

Let $v_{M_{\text{T}},m} = a_m e^{j\phi_m}$ denote the mth component of $\mathbf{v}_{M_{\text{T}}}$, and let b_m denote the mth component of $\mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$. Then the mth component of $\mathbf{y}_{\mathcal{I}}(\phi) = \tilde{R}e^{j\phi}\mathbf{v}_{M_{\text{T}}} + \mathbf{y}^{\perp}_{\mathbf{v}_{M_{\text{T}}}}$ can be written as

$$y_{\mathcal{I},m}(\phi) = \tilde{R}a_m e^{j(\phi+\phi_m)} + b_m, \qquad m = 1,\ldots,M_{\text{T}}.$$

For $\phi \in [0, 2\pi)$, this represents a circle $\mathcal{C}_m$ with radius $\tilde{R}a_m$ and center point b_m in the symbol plane. When we move along $\mathcal{I}$ (by varying ϕ), all components $y_{\mathcal{I},m}(\phi)$ of $\mathbf{y}_{\mathcal{I}}(\phi)$ move along their associated circles $\mathcal{C}_m$ (see Fig. 3). If one of the M_{T} components of $\mathbf{y}_{\mathcal{I}}(\phi)$ crosses a boundary line of the symbol quantization regions, $\mathbf{y}_{\mathcal{I}}(\phi)$ moves into a neighboring ZF decision region. Hereafter, ϕ_m^p will denote the angle ϕ for which $\mathcal{C}_m$ intersects the pth boundary line.

It follows that the number of intersections of the M_{T} circles $\mathcal{C}_m$ with the P boundary lines of the line-structured symbol constellation equals the number of ZF decision regions that are pierced by the intersection circle $\mathcal{I}$, which equals the number of data vectors in $\mathcal{D}_+$. (Here, we have assumed that at least one circle $\mathcal{C}_m$ intersects one boundary line. If no such intersection exists although $\mathcal{L}$ intersects $\mathcal{H}$, we refer to the discussion of Case 2 in Section III-C.) Thus, an upper bound on the number of data vectors in $\mathcal{D}_+$ is $|\mathcal{D}_+|_{\text{max}} = 2M_{\text{T}}P$ (each one of the M_{T} component circles $\mathcal{C}_m$ has 0 or 2 intersections with each one of the P boundary lines). Fig. 3 shows the circles $\mathcal{C}_m$ for a $(4, 4)$ channel and 4-QAM modulation. In this example, $M_{\text{T}} = 4$, $P = 2$, $|\mathcal{D}_+| = 6$, and $|\mathcal{D}_+|_{\text{max}} = 16$.

For efficient calculation of the data vectors $\mathbf{d}^{(i)} \in \mathcal{D}_+$, we first determine all "intersection angles" ϕ_m^p and sort them such that they increase monotonically. A first data vector is chosen as $\mathbf{d}^{(1)} = Q\{\mathbf{y}_{\mathcal{I}}(\phi)\}$ for $\phi = 0$. The remaining data vectors $\mathbf{d}^{(2)}, \mathbf{d}^{(3)}, \cdots \in \mathcal{D}_+$ are then constructed by means

of the following recursive algorithm. Assume that we just obtained $\mathbf{d}^{(i)} \in \mathcal{D}_+$ and wish to determine $\mathbf{d}^{(i+1)} \in \mathcal{D}_+$ that corresponds to the adjacent ZF decision region reached by $\mathbf{y}_{\mathcal{I}}(\phi)$ when ϕ is increased beyond the corresponding intersection angle ϕ_m^p. Moving into this neighboring ZF decision region corresponds to updating $\mathbf{d}^{(i)}$ to $\mathbf{d}^{(i+1)} = \mathbf{d}^{(i)} + \Delta\mathbf{d}^{(i)}$. Here, the vector $\Delta\mathbf{d}^{(i)}$ is zero except for the mth component, denoted $\Delta d_m^{(i)}$, that depends on $(\mathbf{d}^{(i)})_m$ and on the pth boundary. For example, for 4-QAM we have (cf. Fig. 3)

$$\Delta d_m^{(i)} = \begin{cases} -2j\,\mathrm{Im}\{(\mathbf{d}^{(i)})_m\}, & \text{if } p=1 \\ -2\,\mathrm{Re}\{(\mathbf{d}^{(i)})_m\}, & \text{if } p=2, \end{cases}$$

that is, the update simply amounts to flipping the imaginary part (for $p=1$) or the real part (for $p=2$) of $(\mathbf{d}^{(i)})_m$.

This recursive data vector calculation leads to a recursive calculation of the distances $\psi_i = \|\mathbf{r} - \mathbf{H}\mathbf{d}^{(i)}\|$ to be minimized. The distances are updated as $\psi_{i+1}^2 = \psi_i^2 + \Delta_i$, with

$$\Delta_i = |\Delta d_m^{(i)}|^2 \|\mathbf{h}_m\|^2 - 2\,\mathrm{Re}\{\Delta d_m^{(i)}(\mathbf{r} - \mathbf{H}\mathbf{d}^{(i)})^H \mathbf{h}_m\},$$

where $\mathbf{h}_m$ denotes the mth column of the channel matrix $\mathbf{H}$. The term $(\mathbf{r} - \mathbf{H}\mathbf{d}^{(i)})^H$ can be calculated recursively, too, and the squared norms $\|\mathbf{h}_m\|^2$ can be precalculated.

C. Efficient Calculation of $\mathcal{D}_+$ and ψ_i^2: Case 2

If $\mathcal{L}$ and $\mathcal{H}$ do not intersect (i.e., $R^2 \leq \|\mathbf{y}_{\mathbf{v}_{M_\mathrm{T}}}^\perp\|^2$), then $\mathbf{y}_{\mathbf{v}_{M_\mathrm{T}}}^\perp$ can be shown [1] to be the point of $\mathcal{L}$ closest to $\mathcal{H}$. We thus take $\mathbf{d}^{(1)} = Q\{\mathbf{y}_{\mathbf{v}_{M_\mathrm{T}}}^\perp\}$ as the first data vector in $\mathcal{D}_+$. Simulation results indicate that it is advantageous to include also the nearest neighbors of $\mathbf{d}^{(1)}$ in $\mathcal{D}_+$. These data vectors can easily be found by substituting in $\mathbf{d}^{(1)}$ the nearest-neighbor symbols for the individual components of $\mathbf{d}^{(1)}$. For PSK constellations, we obtain 2 nearest neighbors for each data vector component, yielding $|\mathcal{D}_+| = 2M_\mathrm{T} + 1$. The distances $\psi_i^2 = \|\mathbf{r} - \mathbf{H}\mathbf{d}^{(i)}\|^2$ can again be calculated recursively.

The same procedure is used if $\mathcal{L}$ and $\mathcal{H}$ intersect but no component circle $\mathcal{C}_m$ intersects a boundary line.

D. Summary of the SPA

The principal steps of the SPA can now be summarized as follows:

- calculation of $(\mathbf{H}^H\mathbf{H})^{-1}$, $\mathbf{y}_\mathrm{ZF} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{r}$, the dominant eigenvector $\mathbf{v}_{M_\mathrm{T}}$ of $(\mathbf{H}^H\mathbf{H})^{-1}$, and $\hat{\mathbf{d}}_\mathrm{sub}$;
- determination of $\mathcal{D}_+$ and calculation of the corresponding distances ψ_i^2 (see Sections III-B and III-C);
- calculation of the distance for $\hat{\mathbf{d}}_\mathrm{sub}$; determination of the minimum distance.

Here, $\hat{\mathbf{d}}_\mathrm{sub}$ may be the result of ZF detection, MMSE detection, ZF-based NC detection, or MMSE-based NC detection; the resulting SPA variants will be referred to as ZF/SP, MMSE/SP, NC-ZF/SP, and NC-MMSE/SP detection, respectively. For MMSE/SP and NC-MMSE/SP detection, the reference line $\mathcal{L}$ in (3) is replaced by an "MMSE-based" reference line $\mathcal{L}'$ that is still parallel to $\mathbf{v}_{M_\mathrm{T}}$ but whose offset is the component of $\mathbf{y}_\mathrm{MMSE}$ (rather than $\mathbf{y}_\mathrm{ZF}$) perpendicular to $\mathbf{v}_{M_\mathrm{T}}$.

Thus, MMSE/SP and NC-MMSE/SP detection are obtained simply by replacing in the ZF/SP and NC-ZF/SP method, respectively, $(\mathbf{H}^H\mathbf{H})^{-1}$ with $\left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_w^2}{\sigma_d^2}\mathbf{I}\right)^{-1}$. The rationale for using $\mathcal{L}'$ is that on average $\mathbf{y}_\mathrm{MMSE}$ will be closer to the transmitted data vector $\mathbf{d}$ than $\mathbf{y}_\mathrm{ZF}$.

E. Computational Complexity

We will briefly analyze the computational complexity of the SPA add-on (without the calculation of $\hat{\mathbf{d}}_\mathrm{sub}$) for $M_\mathrm{T} = M_\mathrm{R}$. The preparatory operations that have to be performed only once for an entire data block during which the channel is constant have a complexity of $\mathcal{O}(M_\mathrm{T}^3)$ (due to the calculation of $(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$ or $\left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_w^2}{\sigma_d^2}\mathbf{I}\right)^{-1}\mathbf{H}^H$). The complexity of the operations that have to be performed once for each data vector is $\mathcal{O}(M_\mathrm{T}^2 P)$ (due to the calculation of at most $2M_\mathrm{T}P - 1$ efficient distance updates). Note that calculation of the dominant eigenvector $\mathbf{v}_{M_\mathrm{T}}$ has complexity $\mathcal{O}(M_\mathrm{T}^2)$ if the power method [7] is used.

IV. SIMULATION RESULTS

We will now assess the symbol-error rate (SER) performance and computational complexity of the SPA by means of simulation results. In our simulations, we used 4-QAM modulation and MIMO channels with iid Gaussian entries with unit variance. The dominant eigenvector $\mathbf{v}_{M_\mathrm{T}}$ was computed by means of the power method [7] using 4 iterations.

A. SER Performance

Figs. 4(a) and (b) show the SER versus SNR performance[1] of the various SPA and standard detectors for a $(4,4)$ and $(5,5)$ channel, respectively. The following conclusions can be drawn from these results.

- All SPA detectors perform substantially better than the standard suboptimal detectors. Even the simplest SPA detector (ZF/SP) outperforms the best standard suboptimal detector (NC-MMSE with layer ordering [2][9]).
- All SPA detectors achieve higher diversity orders than the standard suboptimal detectors.
- For the $(4,4)$ channel, the MMSE/SP and NC-MMSE/SP detectors perform as well as the ML detector (the corresponding SER curves in Fig. 4(a) are indistinguishable), and the ZF/SP detector achieves near-ML performance.
- For the $(6,6)$ channel, the NC-MMSE/SP detector achieves near-ML performance, and the performance of the other SPA detectors is intermediate between that of the ML detector and that of the best standard suboptimal detector (NC-MMSE with layer ordering).

B. Computational Complexity

A rough picture of the computational complexity of the various detectors is provided in Table I for three different channels ($(4,4)$, $(6,6)$, and $(8,8)$) and 4-QAM modulation. The kflop estimates displayed in Table I were determined

---

[1]The SNR is defined as $\mathrm{E}\{\|\mathbf{H}\mathbf{d}\|^2\}/\mathrm{E}\{\|\mathbf{w}\|^2\} = M_\mathrm{T}\,\sigma_d^2/\sigma_w^2$.
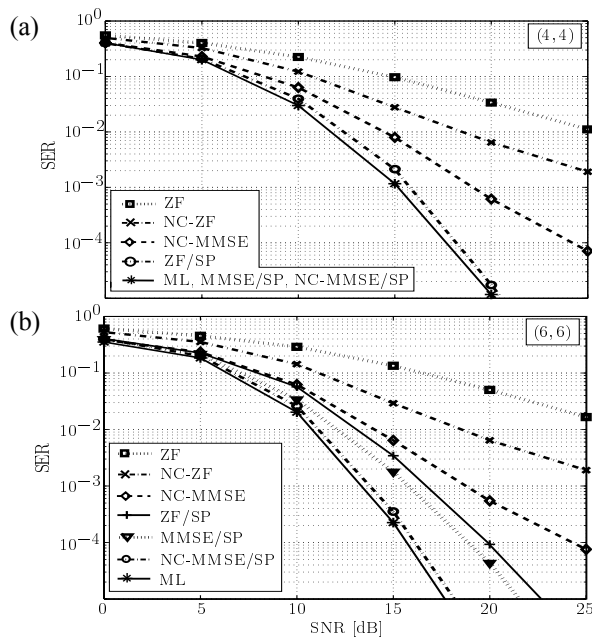
Fig. 4. SER vs. SNR performance of various SPA and standard detectors using 4-QAM modulation: (a) $(4, 4)$ system, (b) $(6, 6)$ system.

TABLE I
MEASURED COMPLEXITY IN KFLOPS FOR (a) OPERATIONS PERFORMED ONCE PER DATA BLOCK AND (b) OPERATIONS PERFORMED ONCE PER DATA VECTOR.

(a)

channel	measured kflops – block operations				
	FPSD	lin.	NC	lin./SP	NC/SP
$(4, 4)$	2.1	2.1	4.2	2.8	4.9
$(6, 6)$	6.4	6.4	16.2	7.8	17.6
$(8, 8)$	14.2	14.2	45.1	16.8	47.6

(b)

channel	measured kflops – vector operations						
	FPSD		lin.	NC	lin./SP		NC/SP	
	av.	max.			av.	max.	av.	max.
$(4, 4)$	32	82	0.18	0.4	1.3	2.4	1.8	2.8
$(6, 6)$	68	300	0.37	1.1	2.5	4.7	3.6	5.7
$(8, 8)$	122	671	0.62	1.9	3.7	7.6	5.6	9.5

The maximum complexity of lin./SP and NC/SP in Table I(b) refers to the case where all circles $\mathcal{C}_m$ intersect all boundary lines (see Subsection III-B).

From Table I, the following conclusions can be drawn.

- The overall complexity of the SPA detectors is somewhat higher than that of the standard suboptimal detectors but much lower than that of FPSD. In particular, the complexity of the NC-MMSE/SP detector is just a fraction of the average complexity of FPSD, even though the NC-MMSE/SP detector achieves near-ML performance.

- For the lin./SP detectors, the complexity of the computations performed once per data block is lower than for the NC detectors but slightly higher than for the linear detectors and FPSD (due to the additional calculation of $\mathbf{v}_{M_T}$).

- For the lin./SP detectors, the average complexity of the computations performed once per data vector is about twice that of the NC detectors but significantly lower (by a factor of about 25...33) than that of FPSD.

V. SUMMARY AND CONCLUSIONS

We developed the *sphere-projection algorithm* (SPA) that is a computationally efficient add-on to standard suboptimal detectors for MIMO spatial multiplexing systems. The SPA was motivated by the observation that the poor performance of standard suboptimal detectors compared to maximum likelihood (ML) detection is primarily caused by the occurrence of "bad" (i.e., poorly conditioned) channel realizations. The SPA detectors are robust to bad channels and capable of achieving near-ML performance and significantly increased diversity gains at a fraction of the computational complexity of Fincke and Phost's sphere-decoding algorithm for ML detection. The SPA detectors also compare favorably to nulling-and-cancelling schemes because they achieve significantly larger diversity gains at a comparable computational complexity.

using MATLAB V5.3. Table I(a) shows the complexity of the operations that have to be performed once for an entire data block during which the channel is constant (e.g., calculation of $\mathbf{H}^{\#}$ and $\mathbf{v}_{M_T}$). Table I(b) shows the complexity of the operations that have to be performed once for each data vector (e.g., determination of $\mathcal{D}_+$ and ψ_i^2). Each table presents only one value for both linear detectors (ZF and MMSE; denoted as "lin.") and one value for both linear detectors combined with the SPA (ZF/SP and MMSE/SP; denoted as "lin./SP"), because there is virtually no difference in complexity.

The ML detector was implemented by means of Fincke and Phost's sphere-decoding (FPSD) algorithm [4]. The computational complexity of FPSD strongly depends on the actual channel realization and SNR. Therefore, in addition to the average complexity, Table I(b) shows the maximum complexity obtained during 10000 simulation runs at an SNR of $10\,$dB.

REFERENCES

[1] H. Artés, D. Seethaler, and F. Hlawatsch, "Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ML detection," *IEEE Trans. Signal Processing, Special Issue on Signal Processing for Multiple-Input Multiple-Output (MIMO) Wireless Communications Systems*, vol. 51, 2003, to appear.

[2] G. Golden, C. Foschini, R. Valenzuela, and P. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communications architecture," *Elect. Lett.*, vol. 35, pp. 14–16, jan. 1999.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs (NJ): Prentice Hall, 1993.

[4] U. Fincke and M. Phost, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. of Comp.*, vol. 44, pp. 463–471, April 1985.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River (NJ): Prentice Hall, 1998.

[6] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *Proc. 35th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove (CA), Nov. 2001, pp. 1051–1055.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.

[8] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs (NJ): Prentice Hall, 1992.

[9] H. Rupp, M. Guillaud, and S. Das, "On MIMO decoding algorithms for UMTS," in *Proc. 35th Asilomar Conf. Signals, Systems, Computers*, vol. 2, Pacific Grove, CA, Nov. 2001, pp. 975–979.