REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-03-

0405

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
01-09-2003	Final	Feb. 1, 2001 - Dec. 31, 2002

4. TITLE AND SUBTITLE
Short-data-record Adaptive Receivers for Rapidly Changing Communications Environments

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-01-1-0176

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Batalama, Stella N.
Pados, Dimitris A.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research Foundation
University at Buffalo
State University of New York
211 UB Commons 520
Lee Entrance
Amherst, NY 14228

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM Room 713
4015 Wilson Blvd
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/NM

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
N/A
Approved for public release, distribution unlimited

20031028 202

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT

We defined and pursued a novel line of research that lies in a multidisciplinary intersection of Estimation Theory, Communications Theory, and Mean-Square optimum linear filtering. Consider an arbitrary input signal vector space and a given information bearing signal vector to be protected or recovered in the presence of multiuser or other forms of heavy interference. Based strictly on statistical conditional optimization principles, we developed an iterative algorithm that starts from the conventional matched-filter correlator and generates a sequence of linear filters ("auxiliary-vector" filters) that converges to the exact MS-optimum solution. At each iteration step, the filter is given as a direct function of the input autocorrelation matrix, the signal vector waveform to be protected, and the filter at the previous iteration. When the autocorrelation matrix is sample-average estimated from a short data record, this procedure offers the means for effective control over the filter estimator bias versus (co-)variance trade-off. For a fixed data record size the filter estimators in this sequence have rapidly decreasing bias and gracefully increasing variance. They outperform other known estimators such as Sample-Matrix-Inversion (SMI), Diagonal-Loading (DL) SMI, RLS, LMS, reduced-rank eigenvector decomposition, and "multistage" nested Wiener filter. While all of the above estimators converge to the optimum MMSE/MVDR solution for infinitely long data records, for any given finite data set there is at least one AV filter estimator in the sequence that outperforms all SMI, DL-SMI, RLS, LMS, reduced-rank eigenvector and multistage nested Wiener filter estimators. The theoretical and practical implications of these results are far reaching. Biased estimators and algorithms that offer full control over the bias/variance balance are rarely reported in the literature, if any in a communications applicable context. We pursued thorough theoretical analysis of the filter estimator sequence and applications for rapid adaptive interference suppression to: spread-spectrum/CDMA communications systems, spread-spectrum/CDMA systems with adaptive antenna arrays, interference resistant synchronization, and jam-resistant GPS.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Stella N. Batalama
U	U	U	UU	17	19b. TELEPHONE NUMBER (include area code) (716) 645-3115 ext. 2164

Title:
Short-Data-Record Adaptive Receivers for Rapidly Changing Communications Environments

AFOSR Grant F49620-01-1-0176

Final Report: Sept. 1, 2003

TABLE OF CONTENTS

1 OBJECTIVES

We defined and we pursued a novel line of research that lies in a multidisciplinary intersection of Estimation Theory, Communications Theory, and Mean-Square optimum linear filtering. Consider an arbitrary input signal vector space and a given information bearing signal vector to be protected or recovered in the presence of multiuser or other forms of heavy interference. Based strictly on statistical conditional optimization principles, we developed an iterative algorithm that starts from the conventional matched-filter correlator and generates a sequence of linear filters ("auxiliary-vector" filters) that converges to the exact MS-optimum solution. At each iteration step, the filter is given as a direct function of the input autocorrelation matrix, the signal vector waveform to be protected, and the filter at the previous iteration. When the autocorrelation matrix is sample-average estimated from a short data record, this procedure offers the means for effective control over the filter estimator bias versus (co-)variance trade-off. For a fixed data record size the filter estimators in this sequence have rapidly decreasing bias and slowly increasing variance. They outperform other known estimators such as Sample-Matrix-Inversion (SMI), Diagonal-Loading (DL) SMI, RLS, LMS, reduced-rank eigenvector decomposition and "multistage" nested Wiener filter. While all of the above estimators converge to the optimum MMSE/MVDR solution for infinitely long data record, for *any* given finite data set there is at least one AV filter estimator in the sequence that outperforms the SMI, DL-SMI, RLS, LMS, reduced-rank eigenvector decomposition as well as the multistage nested Wiener filter.

The theoretical and practical implications of these results are far reaching. Biased estimators and algorithms that offer full control over the bias/variance balance are rarely reported in the literature, if any in a communications applicable context. We pursued thorough theoretical analysis of the filter estimator sequence and application for rapid adaptive interference suppression to: spread-spectrum/CDMA communications systems, spread-spectrum/CDMA systems with adaptive antenna arrays, interference resistant synchronization, and jam-resistant GPS.

This work was taken one step further and the short-data-record adaptive optimization problem was studied in the context of non-linear processing with applications to multi-layer perceptron receiver designs for spread-spectrum communications. The importance of the theoretical developments lies in the fact that they are independent of the specific receiver optimizations criterion

and thus many existing receiver optimization algorithms can benefit by these developments directly.

Finally, we looked at the problem of binary code design. Although this investigation is at an initial stage, our theoretical developments so far advanced the prior-state-of-the-art in the context of fundamental binary code design for code division signal multiplexing. We developed novel, doubly optimal designs of minimum total-squared-correlation (TSC) and minimum maximum-squared-correlation (MSC) binary codes and we pursued theoretical analysis of the codes and application for interference avoidance at the transmitter-end of a communication link.

2 SUMMARY OF RESEARCH EFFORT

The milestones reached during the duration of this research project (2 years) are as follows.

(i) We addressed successfully the problem of determining an automated theoretically derived rule for the selection of the best number of auxiliary vectors for a given data record of input observations.

(ii) We developed a recursive algorithm for the on-line estimation of the AV filter and a modified RLS-type algorithm for the estimation of the MMSE/ MVDR filter that are both based on the interference-plus-noise covariance matrix estimate and we established formally their convergence properties.

(iii) We investigated theoretically the coarse synchronization performance of blind adaptive linear self-synchronized receivers for asynchronous direct-sequence spread-spectrum communications under finite data record adaptation. We derived analytical expressions that approximate the probability of coarse synchronization error of matched-filter-type (MF) and minimum-variance-distortionless-response-type (MVDR) receivers based on transformation noise modeling techniques.

(iv) We studied theoretically the data-record-size requirements of MMSE/ MVDR-type adaptive algorithms to meet a given performance objective in joint space-time signal detection and direction-of-arrival (DoA) estimation problems for direct-sequence spread-spectrum systems. We derived closed form expressions that provide the data record size that is necessary to achieve a given performance confidence level in a neighborhood of the optimal performance point, as well as expressions that identify the performance level that can be reached for a given data record size.

(v) As an application of great importance to AFRL, we addressed the problem of navigation data demodulation by an adaptive GPS receiver that utilizes a bank of single-satellite linear-tap-delay filters and employs antenna-array reception.
(vi) We investigated the relative output SINR performance of the full decorrelator and the partial decorrelator spread-spectrum receiver. For each receiver, we considered two implementations that are equivalent under perfectly known input statistics. The first implementation utilizes the signature matrix while the second implementation is based on the eigen-decomposition of the ideal input covariance matrix. While the full decorrelator aims at decorrelating the complete spread-spectrum (multiple access) interference, the partial decorrelator aims at decorrelating only a part of it by excluding one or more user-signatures or eigenvectors from the corresponding implementation method. We derived necessary and sufficient conditions on the signal energy and signature cross-correlation levels under which the partial decorrelator outperforms the full decorrelator in the output SINR sense. Numerical results demonstrated the validity of the above conditions and simulation studies illustrated the relative SINR and BER performance of the full and partial decorrelators under sample-average-estimated input statistics.
(vii) The short-data-record adaptive filtering problem was taken one step further and studied in the context of non-liner signal processing. We examined multi-layer perceptron DS-CDMA receiver designs and we developed a fast converging training methodology that is independent of the specific optimization criterion used to train the adaptive receiver.
(viii) In direct-sequence spread-spectrum systems, the pre-detection signal-to-interference -plus-noise ratio (SINR) at the output of the single-user minimum-mean-square-error (MMSE) or auxiliary-vector (AV) filter is a function of the specific user spreading code (signature). We considered the adaptive optimization of the user signature assignment such that the output SINR of the MMSE/AV filter is maximized under a transmitter power constraint. In the context of binary signatures, the complexity of the signature optimization procedure is exponential in the processing gain. We derived a low-cost suboptimum adaptive binary signature assignment algorithm based on conditional optimization principles. We used this algorithm to design an efficient system-wide multiuser adaptive signature set assignment scheme. The performance of the proposed scheme was evaluated under asynchronous multipath fading DS-CDMA channel models and was compared to the performance of systems with arbitrarily chosen signature sets.

(ix) Transmitter optimization was further investigated in the context of optimal spreading code design. We designed doubly optimal binary code sets, that is, sets of binary spreading sequences that can be used for code division multiplexing purposes that exhibit minimum total-squared-correlation (TSC) and minimum maximum-squared-correlation (MSC) at the same time. *(x)* We showed that we can efficiently and effectively approach the error rate performance of the optimum multiuser detector as follows. We utilized a multiuser decorrelating or MMSE/AV filter as a pre-processor and we established that the output magnitudes, when properly scaled, provide a reliability measure for each user bit decision. An ordered reliability-based error search sequence of length linear in the number of users returns the most likely user bit vector among all visited options. Numerical and simulation studies for moderately loaded systems that permit exact implementation of the optimum detector indicated that the error rate performance of the optimum and the proposed detector are nearly indistinguishable over the whole pre-detection signal-to-noise ratio (SNR) range of practical interest. Similar studies for higher user loads (that prohibit comparisons with the optimum detector) demonstrated error rate performance gains of orders of magnitude in comparison with straight decorrelating or MMSE multiuser detection.

3 ACCOMPLISHMENTS/NEW FINDINGS

Selecting the most successful (in some appropriate sense) AV filter estimator in the generated sequence for a given data record was a problem that had not been addressed so far. We dealt exactly with this problem and we proposed two data driven selection criteria [1],[2],[3]. The first criterion is specific to digital communications (binary hypothesis testing) and maximizes the estimated J-divergence of the AV-filter output conditional distributions given the transmitted information bit. The second criterion applies to all filter estimation problems and minimizes the cross-validated sample average variance of the AV-filter output.

We observed that the presence of the desired signal during estimation of the minimum-mean-square-error (MMSE)/minimum-variance-distortionless-response (MVDR) and auxiliary-vector (AV) filters under limited data support leads to significant signal-to-interference-plus-noise ratio (SINR) performance degradation. We quantified this observation in the context of

DS/CDMA communications by deriving close approximations for the mean-square filter estimation error, the probability density function of the output SINR and the probability density function of the symbol-error-rate (SER) of the sample-matrix-inversion (SMI) receiver evaluated using both desired-signal-"present" and desired-signal-"absent" input covariance matrix. To avoid such performance degradation we proposed a DS/CDMA receiver that utilizes a simple pilot-assisted algorithm that estimates and then subtracts the desired signal component from the received signal prior to filter estimation. Then, to accommodate decision directed operation we developed two recursive algorithms for the on-line estimation of the AV and MMSE/MVDR filter and we studied their convergence properties [4].

Equipped with these important developments, we considered blind adaptive linear receivers for the demodulation of DS/SS signals in asynchronous transmissions. The proposed structures are self-synchronized in the sense that adaptive synchronization and demodulation are viewed and treated as an integrated receiver operation. Two computationally efficient *combined* synchronization/demodulation schemes were proposed, developed and analyzed [5]. The first scheme is based on the principles of minimum-variance-distortionless-response (MVDR) processing, while the second scheme follows the principles of auxiliary-vector filtering and exhibits enhanced performance in short data record scenarios. In both cases the resulting receiver is a linear structure of order exactly equal to the system processing gain. Simulation studies demonstrated the coarse synchronization, as well as the bit-error-rate performance of the proposed strategies. In addition, we investigated theoretically the coarse synchronization performance of blind adaptive linear self-synchronized receivers for asynchronous direct-sequence code-division-multiple-access communications under finite data record adaptation. We derived [6] analytical expressions that approximate the probability of coarse synchronization error of matched-filter-type (MF) and minimum-variance-distortionless-response-type (MVDR) receivers based on transformation noise modeling techniques. The expressions are explicit functions of the data record size N and the filter order p and reveal the effect of short-data-record Sample-Matrix-Inversion (SMI) implementations on the coarse synchronization performance. Besides their theoretical value, the derived expressions provide simple, highly-accurate alternatives to computationally demanding performance evaluation through simulations. Numerical and simulation studies examined the accuracy of the theoretical developments and showed that the derived expressions approximate closely the actual coarse synchronization

performance.

We also investigated the data-record-size requirements of minimum-variance-distortionless-response-type adaptive algorithms to meet a given performance objective in joint space-time signal detection and direction-of-arrival (DoA) estimation problems for direct-sequence spread-spectrum systems. We derived closed form expressions that provide the data record size that is necessary to achieve a given performance confidence level in a neighborhood of the optimal performance point, as well as expressions that identify the performance level that can be reached for a given data record size [7]. This was done by utilizing close approximations of the involved probability density functions. The practical significance of the derived expressions lies in the fact that the expressions are functions of the number of antenna elements and the system spreading gain only, while they are independent of the ideal input covariance matrix which is not known in most realistic applications.

An important application for AFRL is the problem of navigation data demodulation by an adaptive GPS receiver that utilizes a bank of single-satellite linear-tap-delay filters and employs antenna-array reception. The presence of an antenna array allows the receiver to operate in the spatial domain in addition to the temporal (code) domain. We investigated disjoint-domain as well as joint-domain space-time GPS signal processing techniques and we considered design criteria of conventional matched-filter (MF) type, minimum-variance-distortionless-response (MVDR) type and auxiliary-vector (AV) type [8]. The proposed structures utilize filters that operate at a fraction of the navigation data bit period (1 msec) and are followed by hard-decision detectors. Hard decisions taken over a navigation data bit period are then combined according to a simple combining rule for further bit-error-rate (BER) performance improvements. Analytic, numerical and simulation comparisons illustrated the relative merits of the investigated design alternatives.

In the presence of additive white Gaussian noise (AWGN), complete elimination of the spread-spectrum interference introduces enhanced noise variance at the decorrelator's output.[1] It was observed, however, that at times a partial decorrelator (PDEC) can strike an improved balance between interference suppression and noise enhancement when compared to the full decorrelator (DEC) while maintaining a decorrelating structure. That is,

[1]Noise variance enhancement at the decorrelator's output refers to the filtered noise variance experienced at the output of a decorrelator that is distortionless in the direction of a given user-signature of interest as it compares to the noise variance at the output of the corresponding MF.

instead of decorrelating the complete multiple-access-interference (MAI) in the system, the PDEC chooses purposefully to decorrelate only a part of the MAI so that it exhibits a higher output signal-to-interference-plus-noise-ratio (SINR) than the full DEC. We understand that when unknown interference is present in the system, decorrelating receivers may operate unintentionally under partial decorrelating conditions. We identified necessary and sufficient conditions on the signal energy and signature cross-correlation levels under which the PDEC outperforms the full DEC in the output SINR sense, and we showed why decorrelation of interferers that satisfy such conditions should be purposefully avoided if at all possible [9], [10].

In the context of short-data-record adaptive non-linear signal processing of direct-sequence spread-spectrum (DS-SS) signals, we proved formally that the optimum (nonlinear) DS-CDMA single-user decision boundary exhibits the following properties [11]: *(a)* It is symmetric with respect to the origin and *(b)* as it is traversed away from the origin it converges to a hyperplane parallel to the matched filter (MF) decision boundary. Then, we translated these properties to a set of constraints that can be used by any optimization algorithm for the selection (training) of the parameters of a general multilayer perceptron neural network receiver. Using these constraints, the number of parameters to be optimized is reduced by nearly 50% for large-size networks which effectively doubles the speed of the training procedure. Furthermore, we utilized the above properties to develop a new initialization scheme that provides additional improvements on the convergence rate and can be used by any recursive optimization algorithm. As a representative case study, we considered the familiar backpropagation (BP) algorithm and developed a new version of BP that incorporates both the proposed constraints and the proposed initialization. Convergence rate enhancement of about two orders of magnitude was achieved. We also developed a fast converging adaptive training algorithm that minimizes the bit error rate (BER) at the output of the receiver [12], [13], [14]. It is important to note that, in addition to utilizing the constraints that were derived from the properties of the optimum single-user decision boundary the minimum BER adaptive algorithm *(a)* incorporates the BER, i.e. the ultimate performance evaluation measure, directly into the learning process, and *(b)* embeds importance sampling principles directly into the receiver optimization process.

We considered the problem of *binary* signature set design for DS-CDMA users that transmit over multipath fading channels. First, we considered the problem of selecting optimally the signature of a single user in the presence

of spread-spectrum interference (MAI), inter-symbol-interference (ISI) due to the multipath channel, and background noise. The signature is chosen to maximize the output SINR of the ("maximum-SINR") RAKE/MMSE filter. Since optimum binary signature selection has complexity exponential in the processing gain of the system, we proposed [15] a suboptimum computationally efficient algorithm based on the Cholesky decomposition method. The proposed algorithm optimizes the coordinates of the user signature conditionally and is characterized by linear complexity. Finally, we generalized this approach to cover the system-wide multiuser signature set selection problem, where each user's signature is adjusted iteratively according to the above algorithm. The performance of this scheme was evaluated through simulations where we initialized the algorithm at an arbitrary signature set and we observed its convergence to a new signature set for which the output SINR of each user's RAKE/MMSE filter is significantly increased.

To further improve the performance of a communication link we also considered the problem of interference avoidance at the transmitter end by designing optimal CDMA spreading codes. In particular, we designed doubly optimal binary spreading code sets of sequences that exhibit minimum TSC and minimum MSC at the same time [16], [17]. We focused on the performance of sets with signatures of odd length and we derived explicit closed-form formulas for the signature cross-correlation matrix of such sets, its eigenvalues, and its inverse. Based on these formulas, we were able to obtain analytical expressions for the BER of decorrelating filters, the (maximum) output SINR of MMSE filters, and the total asymptotic efficiency of synchronous code division multiplexing systems. The calculated across-the-board superiority of the "$4m+1$" signature sets is intriguing. Combined with the indicated similar to Gold sets performance under asynchronous transmissions, this strong proven superiority for (near-)synchronous links of the "$4m+1$" sets questions seriously the status quo and the popularity of Gold sets (or any other "$4m-1$" doubly optimal candidate set to that respect). Regretfully, neither the "$4m+1$" codes nor the Gold codes have what we might call satisfactory performance for heavy-loaded asynchronous transmissions.

Finally, we developed a new multiuser detection algorithm in the context of asynchronous AWGN multiple-access channels [18],[19]. A decorrelating or MMSE/AV multiuser filter is used as a pre-processor that provides initial decisions and reliability measurements based on which an ordered error pattern sequence of tunable length is formed. The error pattern sequence is followed to its end and the most likely bit vector among all visited op-

tions is returned. When the length of the pattern sequence is of the order of the number of bits to be detected, the additional imposed computational cost compared to straight decorrelating or MMSE/AV detection is rather insignificant. Still, we saw that this proposed multiuser detection algorithm maintains near-ML bit-error-rate performance over the whole studied SNR range of interest. There is strong resemblance between this scheme and "efficient" decoding algorithms for binary linear block codes [20].

4 PERSONNEL SUPPORTED

Faculty:
S. N. Batalama
D. A. Pados

Graduate Students:
John Matyjas
I. N. Psaromiligkos
Ping Xiong
Jeffrey Farrell
Haoli Qian
George Karystinos
Zhenyu Liu

Postdoctoral Fellow:
Marita Gkizeli

5 PUBLICATIONS

[1] G. N. Karystinos, H. Qian, M. J. Medley, and S. N. Batalama, "Short-data-record adaptive filtering: The auxiliary-vector algorithm," *Digital Signal Processing,* Special Issue on Defense Applications of Signal Processing, vol. 12, pp. 193-222, July 2002.

[2] G. N. Karystinos, H. Qian, M. J. Medley, and S. N. Batalama, "Short-data-record adaptive filtering: The auxiliary-vector algorithm," in *Proceedings DASP 2001/2002*, Workshop on Defense Applications of Signal Processing, Adelaide, Australia, June 2002.

[3] H. Qian and S. N. Batalama, "Data-record-based criteria for the selection of an auxiliary-vector estimator of the MMSE/MVDR filter," *IEEE Transactions on Communications,* to appear Oct. 2003.

[4] 2. I. N. Psaromiligkos and S. N. Batalama, "Recursive AV and MVDR filter estimation for maximum SINR adaptive space-time processing," *IEEE Transactions on Communications,* to appear.

[5] I. N. Psaromiligkos, S. N. Batalama, and M. J. Medley, "Rapid synchronization and combined demodulation for DS/CDMA communications. Part I: Algorithmic developments," *IEEE Transactions on Communications,* vol. 51, pp. 983-994, June 2003.

[6] I. N. Psaromiligkos and S. N. Batalama, "Rapid synchronization and combined demodulation for DS/CDMA communications. Part II: Finite data-record-size performance analysis," *IEEE Transactions on Communications,* vol. 51, pp. 1162-1172, July 2003.

[7] I. N. Psaromiligkos and S. N. Batalama, "Data record size requirements for adaptive space-time DS/CDMA signal detection," *IEEE Transactions on Communications,* accepted for publication, May 2003.

[8] P. Xiong, M. J. Medley, and S. N. Batalama, "Spatial and temporal processing for global navigation satellite systems: The GPS receiver paradigm," *IEEE Transactions on Aerospace and Electronic Systems,* accepted for pub-

lication, Aug. 2003.

[9] P. Xiong, I. N. Psaromiligkos, and S. N. Batalama, "On the relative output SINR of full and partial decorrelators," in *Proceedings IEEE MILCOM '01 - Conference on Military Communications,* McLean, VA, Oct. 2001.

[10] P. Xiong, I. N. Psaromiligkos, and S. N. Batalama, "On the relative output SINR of full and partial decorrelators," *IEEE Transactions on Communications*, to appear Oct. 2003.

[11] J. D. Matyjas, G. N. Karystinos, and S. N. Batalama, "On the training of DS-CDMA neural-network receivers," *IEEE Transactions on Communications*, submitted Dec. 2002.

[12] J. D. Matyjas, I. N. Psaromiligkos, S. N. Batalama, and M. J. Medley, "Fast converging minimum probability of error neural network receivers for DS-CDMA communications," *IEEE Transactions on Neural Networks,* to appear.

[13] J. D. Matyjas, I. N. Psaromiligkos, S. N. Batalama, and M. J. Medley, "Fast converging minimum probability of error neural network receivers for DS-CDMA communications," in *Proceedings SPIE 17th Annual International Symposium, Digital Wireless Communication Conference,* Orlando, FL, April 2003.

[14] I. N. Psaromiligkos and S. N. Batalama, "Short-data-record bit-error-rate estimation of DS/CDMA receivers," in *Proceedings 2001 Conference on Information Sciences and Systems,* Johns Hopkins University, Baltimore, MD, Mar. 2001, pp. 27-31.

[15] G. N. Karystinos and D. A. Pados, "Adaptive assignment of binary user spreading codes in DS-CDMA systems," in *Proceedings of SPIE 15th Annual International Symposium, Digital Wireless Communications Conference,* Orlando, FL, April 2001, vol. 4395, pp. 137-144.

[16] G. N. Karystinos and D. A. Pados, "Theoretic evaluation of the code division multiplexing abilities of odd-length doubly-optimal signature sets," *IEEE Transactions on Communications*, submitted Oct. 2002.

[17] G. N. Karystinos and D. A. Pados, "Performance analysis of doubly optimal CDMA spreading codes with odd length," in *Proceedings SPIE 17th Annual International Symposium, Digital Wireless Communication Conference*, Orlando, FL, Apr. 2003, vol. 5100.

[18] Z. Liu and D. A. Pados, "Reliability-based near maximum-likelihood multiuser detection," in *Proceedings IEEE MILCOM '01 - Conference on Military Communications,* Fairfax, VA, Oct. 2001, pp. 576-580.

[19] Z. Liu and D. A. Pados, "Near ML multiuser detection with linear filters and reliability-based processing," *IEEE Transactions on Communications,* to appear Sept. 2003.

[20] Y. Wu and D. A. Pados, "An adaptive two-stage algorithm for ML and sub-ML decoding of linear block codes," *IEEE Transactions on Information Theory,* vol. 49, pp. 261-269, Jan. 2003.

6 INTERACTIONS/TRANSITIONS

I. Participation/presentations at meetings, conferences

1. S. N. Batalama, "Short-data-record adaptive filtering: The auxiliary-vector algorithm," participation/presentation at 2001/2002 Workshop on Defense Applications of Signal Processing, Adelaide, Australia, June 2002.

2. P. Xiong and S. N. Batalama, "On the relative output SINR of the conventional and the partial decorrelator," participation/presentation at IEEE MILCOM '01 - Conference on Military Communications, Vienna, VA, Oct. 28-31, 2001.

3. G. N. Karystinos and D. A. Pados, "Adaptive assignment of binary user spreading codes in DS-CDMA systems," participation/presentation at Digital Wireless Communications Conference, SPIE 15th Annual International Symposium, Orlando, FL, April 17-18, 2001.

4. Z. Liu and D. A. Pados, "Reliability-based near maximum-likelihood multiuser detection," participation/presentation at IEEE MILCOM '01 - Conference on Military Communications, Vienna, VA, Oct. 28-31, 2001.

5. G. N. Karystinos and D. A. Pados, "Performance analysis of doubly optimal CDMA spreading codes with odd length," participation/presentation at Digital Wireless Communications Conference, SPIE 17th Annual International Symposium, Orlando, FL, Apr. 2003.

6. J. D. Matyjas, I.N. Psaromiligkos, S. N. Batalama, and M. J. Medley, "Fast converging minimum probability of error neural network receivers for DS-CDMA communications," participation/presentation at Digital Wireless Communications Conference, SPIE's 17th Annual International Symposium Orlando, FL, April 2003.

7. I. N. Psaromiligkos and S. N. Batalama, "Short-data-record bit-error-rate estimation of DS/CDMA receivers," participation/presentation at 2001 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, Mar. 2001

II. Consultative and advisory functions to laboratories and agencies

S. N. Batalama, Visiting Summer Faculty, U.S. Air Force Research Laboratory, IFGC, Rome, NY, May-July 2001, May-July 2002, May-July 2003.

S. N. Batalama and D. A. Pados, "Short-data-record adaptive filtering", One-day Workshop, U.S. Air Force Research Laboratory, IFGC, Rome, NY, May 2002.

III. Transitions

1. Performer: S. N. Batalama

Customer: Dr. Michael J. Medley, [IFGC, AFRL, Rome, NY, telephone: (315) 330-4830, e-mail: medleym@rl.af.mil]. Results: Spread-spectrum receivers - Methods for adaptive protection from non-Gaussian disturbances and unknown correlated spread-spectrum interference with limited input observations. Applications: Robust adaptive SS receivers, robust adaptive SS receivers with antenna arrays, adaptive jam-resistant GPS.

2. Performer: S. N. Batalama

Customer: Dr. Michael J. Medley, [IFGC, AFRL, Rome, NY, telephone: (315) 330-4830, e-mail: medleym@rl.af.mil]. Results: Adaptive GPS array receivers - Methods for jam resistant navigation data demodulation employing space-time interference suppression techniques. Applications: GPS navigation and GPS guided delivery.

2. Performer: S. N. Batalama and D. A. Pados

Customer: Dr. Michael J. Medley, [IFGC, AFRL, Rome, NY, telephone: (315) 330-4830, e-mail: medleym@rl.af.mil]. Results: Adaptive binary signature design for spread-spectrum signals. Applications: Avoidance of interferers.

REPORT OF INVENTIONS AND SUBCONTRACTS

(Pursuant to "Patent Rights" Contract Clause) (See Instructions on back)

Form Approved
OMB No. 9000-0095
Expires Oct 31, 2004

PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THIS ADDRESS. RETURN COMPLETED FORM TO THE CONTRACTING OFFICER.

1.a. NAME OF CONTRACTOR/SUBCONTRACTOR	c. CONTRACT NUMBER	2.a. NAME OF GOVERNMENT PRIME CONTRACTOR	c. CONTRACT NUMBER	3. TYPE OF REPORT *(X one)*
The Research Foundation of SUNY	F49620-01-1-0176	US Air Force Office of Scientific Research	F49620-01-1-0176	a. INTERIM [ ] b. FINAL [X]
b. ADDRESS *(Include ZIP Code)*	**d. AWARD DATE** *(YYYYMMDD)*	**b. ADDRESS** *(Include ZIP Code)*	**d. AWARD DATE** *(YYYYMMDD)*	**4. REPORTING PERIOD** *(YYYYMMDD)*
520 Lee Entrance, Suite 211, Amherst, NY 14228	2001/02/15	801 N. Randolph Street, Room 732 Arlington, VA 22203-1977	2001/02/15	a. FROM 2001/02/15 b. TO 2002/12/31

SECTION I - SUBJECT INVENTIONS

5. "SUBJECT INVENTIONS" REQUIRED TO BE REPORTED BY CONTRACTOR/SUBCONTRACTOR *(If "None," so state)*

NAME(S) OF INVENTOR(S) *(Last, First, Middle Initial)* a.	TITLE OF INVENTION(S) b.	DISCLOSURE NUMBER, PATENT APPLICATION SERIAL NUMBER OR PATENT NUMBER c.	ELECTION TO FILE PATENT APPLICATIONS *(X)* d.				CONFIRMATORY INSTRUMENT OR ASSIGNMENT FORWARDED TO CONTRACTING OFFICER *(X)* e.	
			(1) UNITED STATES		(2) FOREIGN			
			(a) YES	(b) NO	(a) YES	(b) NO	(a) YES	(b) NO
None								

f. EMPLOYER OF INVENTOR(S) NOT EMPLOYED BY CONTRACTOR/SUBCONTRACTOR		g. ELECTED FOREIGN COUNTRIES IN WHICH A PATENT APPLICATION WILL BE FILED	
(1) (a) NAME OF INVENTOR *(Last, First, Middle Initial)*	(2) (a) NAME OF INVENTOR *(Last, First, Middle Initial)*	(1) TITLE OF INVENTION	(2) FOREIGN COUNTRIES OF PATENT APPLICATION
(b) NAME OF EMPLOYER	(b) NAME OF EMPLOYER		
(c) ADDRESS OF EMPLOYER *(Include ZIP Code)*	(c) ADDRESS OF EMPLOYER *(Include ZIP Code)*		

SECTION II - SUBCONTRACTS *(Containing a "Patent Rights" clause)*

6. SUBCONTRACTS AWARDED BY CONTRACTOR/SUBCONTRACTOR *(If "None," so state)*

NAME OF SUBCONTRACTOR(S) a.	ADDRESS *(Include ZIP Code)* b.	SUBCONTRACT NUMBER(S) c.	FAR "PATENT RIGHTS" d.		DESCRIPTION OF WORK TO BE PERFORMED UNDER SUBCONTRACT(S) e.	SUBCONTRACT DATES *(YYYYMMDD)* f.	
			(1) CLAUSE NUMBER	(2) DATE *(YYYYMM)*		(1) AWARD	(2) ESTIMATED COMPLETION
None							

SECTION III - CERTIFICATION

7. CERTIFICATION OF REPORT BY CONTRACTOR/SUBCONTRACTOR *(Not required if: (X as appropriate))* [ ] SMALL BUSINESS or [ ] NONPROFIT ORGANIZATION

I certify that the reporting party has procedures for prompt identification and timely disclosure of "Subject Inventions," that such procedures have been followed and that all "Subject Inventions" have been reported.

a. NAME OF AUTHORIZED CONTRACTOR/SUBCONTRACTOR OFFICIAL *(Last, First, Middle Initial)*	b. TITLE	c. SIGNATURE	d. DATE SIGNED
Batalama, Stella N.	Associate Professor	S. Batalama	09/01/03

DD FORM 882, DEC 2001 PREVIOUS EDITION IS OBSOLETE.

Digital Signal Processing 12, 193–222 (2002)
doi:10.1006/dspr.2002.0450

Short Data Record Adaptive Filtering: The Auxiliary-Vector Algorithm [1]

George N. Karystinos,* Haoli Qian,* Michael J. Medley,† and Stella N. Batalama*

*Department of Electrical Engineering, State University of New York at Buffalo, 332 Bonner Hall, Buffalo, New York 14260; and †Air Force Research Laboratory, IFGC, 525 Brook Road, Rome, New York 13441
E-mail: cary@eng.buffalo.edu, haoqian@eng.buffalo.edu, michael.medley@rl.af.mil, batalama@eng.buffalo.edu

---

Karystinos, G. N., Qian, H., Medley, M. J., and Batalama, S. N., Short Data Record Adaptive Filtering: The Auxiliary-Vector Algorithm, *Digital Signal Processing* **12** (2002) 193–222.

Based on statistical conditional optimization criteria, we developed an iterative algorithm that starts from the matched filter (or constraint vector) and generates a sequence of filters that converges to the minimum variance distortionless response (MVDR) solution for any positive definite input autocorrelation matrix. Computationally, the algorithm is a simple recursive procedure that avoids explicit matrix inversion, decomposition, or diagonalization operations. When the input autocorrelation matrix is replaced by a conventional sample-average (positive definite) estimate, the algorithm effectively generates a sequence of MVDR filter estimators: The bias converges rapidly to zero and the covariance trace rises slowly and asymptotically to the covariance trace of the familiar sample matrix inversion (SMI) estimator. For short data records, the early, nonasymptotic, elements of the generated sequence of estimators offer favorable bias–covariance balance and are seen to outperform in mean-square estimation error constraint-LMS, RLS-type, and orthogonal multistage decomposition estimates (also called nested Wiener filters) as well as plain and diagonally loaded SMI estimates. The problem of selecting the most successful (in some appropriate sense) filter estimator in the sequence for a given data record is addressed and two data-driven selection criteria are proposed. The first criterion minimizes the cross-validated sample average variance of the filter estimator output. The second criterion maximizes the estimated J-divergence of the filter estimator output conditional distributions. Illustrative interference suppression examples drawn from the communications literature are followed throughout this presentation.

[1] This work was supported by the National Science Foundation under Grant ECS-0073660 and the U.S. Air Force Office of Scientific Research under Grant F49620-01-1-0176.

1051-2004/02 $35.00

Key Words: adaptive filters; biased estimators; code division multiple access; cross-validation; interference suppression; iterative methods; J-divergence; least mean square methods; auxiliary-vector filters; MMSE filters; MVDR filters; filter estimation; small sample support; finite sample support; short data record estimators; Wiener filters; antenna arrays; smart antenna

1. INTRODUCTION

Minimum variance distortionless response (MVDR) filtering refers to the problem of identifying a linear filter that minimizes the variance at its output, while at the same time the filter maintains a distortionless response toward a specific input vector direction of interest. If $\mathbf{r}$ is a random, zero mean without loss of generality, complex input vector of dimension L, $\mathbf{r} \in \mathbb{C}^L$, that is processed by an L-tap filter $\mathbf{w} \in \mathbb{C}^L$, then the filter output variance is $\mathbf{w}^H\mathbf{R}\mathbf{w}$, where $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$ is the input autocorrelation matrix ($E\{\cdot\}$ denotes the statistical expectation operation and $\mathbf{x}^H$ denotes the Hermitian, that is, the transpose conjugate of $\mathbf{x}$). The MVDR filter minimizes $\mathbf{w}^H\mathbf{R}\mathbf{w}$ and simultaneously satisfies $\mathbf{w}^H\mathbf{v} = 1$, or more generally $\mathbf{w}^H\mathbf{v} = \rho \in \mathbb{C}$, where $\mathbf{v} \in \mathbb{C}^L$ is the signal vector direction to be protected. In this setup, MVDR filtering is a standard linear constraint optimization problem and a conventional Lagrange multipliers procedure leads to the well-known solution [1, 2]

$$\mathbf{w}_{\text{MVDR}} = \rho^* \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}, \tag{1}$$

where ρ^* denotes the conjugate of the desired response $\mathbf{w}^H\mathbf{v} = \rho$. MVDR filtering has been used extensively in unsupervised signal processing applications where a desired scalar filter output $d \in \mathbb{C}$ cannot be identified or cannot be assumed available for each input $\mathbf{r} \in \mathbb{C}^L$ (for example, in radar and array processing problems where the constraint vector $\mathbf{v}$ is usually referred to as the *target* or *look* direction of interest). We may also observe the close relationship between the MVDR filter and the minimum mean square error (MMSE) or Wiener filter. Indeed, if the constraint vector $\mathbf{v}$ is chosen to be the statistical cross-correlation vector between the desired output d and the input vector $\mathbf{r}$, that is if $\mathbf{v} = E\{\mathbf{r}d^*\}$, then the MVDR and MMSE filters become scaled versions of each other, $c\mathbf{R}^{-1}\mathbf{v}$, $c \in \mathbb{C}$. For this reason, in the rest of the paper we will use the term *MMSE–MVDR filter* to refer to either filter.

In this article, first we present an iterative algorithm for the calculation of the MMSE–MVDR vector $\mathbf{w}_{\text{MMSE/MVDR}}$ in (1). The algorithm is a noninvasive procedure where no explicit matrix inversion–eigendecomposition–diagonalization is attempted. The MMSE–MVDR computation algorithm creates a sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, that begins from $\mathbf{w}_0 = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and converges to the MMSE–MVDR filter ($\mathbf{w}_\infty = \mathbf{w}_{\text{MMSE/MVDR}}$). At each step $n = 1, 2, \ldots$, $\mathbf{w}_n$ is given as a simple, direct function of $\mathbf{R}$, $\mathbf{v}$, and $\mathbf{w}_{n-1}$.

The development of the iterative algorithm (which we call the *auxiliary-vector algorithm* for reasons that will become apparent) is founded solely on statistical signal processing principles. The motivation behind its development is *adaptive* signal processing where the input autocorrelation matrix $\mathbf{R}$ is assumed unknown and it is sample-average estimated by a data record of M points, $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$:

$$\widehat{\mathbf{R}}(M) = \frac{1}{M} \sum_{m=1}^{M} \mathbf{r}_m \mathbf{r}_m^H. \tag{2}$$

When $\mathbf{R}$ is substituted by $\widehat{\mathbf{R}}(M)$ in the recursively generated sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, the corresponding filter estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, offer the means for effective control over the filter estimator bias versus (co-)variance trade-off [3]. Starting from the *0*-variance, high-bias (for nonwhite inputs) $\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ estimate, we can go all the way up to the unbiased, yet high-variance for small data record sizes M, $\widehat{\mathbf{w}}_\infty(M)$ estimate and anywhere in between, $\widehat{\mathbf{w}}_n(M)$, $1 \leq n < \infty$. As a result, adaptive filters from this newly developed class can be seen to outperform in expected norm-square estimation error, $E\{\|\widehat{\mathbf{w}}(M) - \mathbf{w}_{\text{MMSE/MVDR}}\|^2\}$, (constraint-) LMS [4], sample matrix inversion (SMI) [5] with or without diagonal loading [6], RLS-type [7, 8], and orthogonal multistage decomposition (also called nested Wiener) [9, 10] adaptive filter implementations. It is worth mentioning that the familiar trial-and-error tuning to problem and data-record-size specifics of the real-valued LMS gain or RLS inverse matrix initialization constant or SMI diagonal loading parameter that plagues field practitioners is now replaced by an integer choice of one of the recursively generated filters.

The problem of selecting the best (in some appropriate sense) filter estimator in the sequence for a given data record is addressed and two data-driven selection criteria are proposed. The first criterion is rather general and is motivated by the asymptotic minimum output variance property of the generated sequence of filter estimators. In particular, for a given data record, we select the filter estimator that has minimum cross-validated average output variance (energy). The second rule is built specifically for binary antipodal (BPSK-type) communication signals and is related to the objective of achieving maximum stochastic distance between the two conditional distributions of the filter estimator output. Under this rule, we choose the filter estimator in the sequence that exhibits maximum estimated J-divergence of the conditional output distributions. We pursue and analyze both supervised and unsupervised (blind) implementations of this criterion. Illustrative case studies drawn from the code division multiple access (CDMA) communications literature are followed throughout this article.

The rest of the article is organized as follows. In Section 2 we present the basic algorithmic development and analysis results. Filter estimation issues are discussed in Section 3. The two data-driven criteria for the selection of a filter estimator from the generated sequence are developed in Section 4. In Section 5 we examine the quality of the proposed criteria through simulations. A few concluding remarks are given in Section 6.

2. ALGORITHMIC DEVELOPMENTS AND CONVERGENCE ANALYSIS

For a given constraint vector $\mathbf{v} \in \mathbb{C}^L$ consider the set of filters $\mathcal{D} = \{\mathbf{w} \in \mathbb{C}^L: \mathbf{w} = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} + \mathbf{u}, \mathbf{u} \in \mathbb{C}^L, \text{ and } \mathbf{v}^H\mathbf{u} = 0\}$. $\mathcal{D}$ is the class of all filters $\mathbf{w}$ in $\mathbb{C}^L$ that have a given response ρ in $\mathbf{v}$; that is, $\mathbf{w}^H\mathbf{v} = \rho$. In this section we develop an iterative algorithm for the computation of the $\mathbf{u}$ component of the MMSE–MVDR filter.

Algorithmic designs that focus on the MMSE–MVDR filter part $\mathbf{u}$ that is orthogonal to the constraint vector, or look, direction $\mathbf{v}$ have been widely pursued in the array processing literature and have been known as generalized sidelobe cancelers (GSC) [11] or partially adaptive beamformers [12]. Recent developments have been influenced by principal component analysis reduced-rank processing principles [13]. In general, the MMSE–MVDR filter part $\mathbf{u}$ ($\mathbf{u}^H\mathbf{v} = 0$) has been approximated by $\mathbf{u}_{L\times 1} \simeq -\mathbf{B}_{L\times(L-1)}\mathbf{T}_{(L-1)\times p}\mathbf{w}^{\text{GSC}}_{p\times 1}$, where $\mathbf{B}$ is the so-called *blocking matrix* that satisfies $\mathbf{B}^H\mathbf{v} = \mathbf{0}_{L-1}$ ($\mathbf{B}$ is a full column-rank matrix that can be derived by Gram–Schmidt orthogonalization of an $L \times L$ orthogonal projection operator such as $\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2$, where $\mathbf{I}$ is the identity matrix), $\mathbf{T}$ is the rank reducing matrix with $1 \le p < L - 1$ columns to be designed, and $\mathbf{w}^{\text{GSC}}$ is the MS-optimum vector of weights of the p columns of $\mathbf{T}$ ($\mathbf{w}^{\text{GSC}} = (\rho^*/\|\mathbf{v}\|^2)[\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{B}\mathbf{T}]^{-1}\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{v}$ [12]). In [14] and [15] the p columns of $\mathbf{T}$ were chosen to be the p maximum eigenvalue eigenvectors of the blocked data autocorrelation matrix $\mathbf{B}^H\mathbf{R}\mathbf{B}$. If, however, the columns of $\mathbf{T}$ *have to be* eigenvectors of $\mathbf{B}^H\mathbf{R}\mathbf{B}$ (there is no documented technical optimality to this approach), then the best way to choose them in the minimum output variance p-rank approximation sense was presented in [16]: Select the p eigenvectors $\mathbf{q}_i$ of $\mathbf{B}^H\mathbf{R}\mathbf{B}$, with corresponding eigenvalues λ_i, that maximize $|\mathbf{v}^H\mathbf{R}\mathbf{B}\mathbf{q}_i|^2/\lambda_i$, $i = 1, \ldots, p$. This design algorithm was called *cross-spectral metric* reduced-rank processing in [17]. A different approach from a different point of view is described in this work. A *conditional* statistical optimization procedure is shown to offer the means for *exact* computation of $\mathbf{u}$ as the convergence point of an infinite series of the form $-\sum_{n=1}^{\infty}\mu_n\mathbf{g}_n$, $\mu_n \in \mathcal{R}^+$, $\mathbf{g}_n \in \mathcal{C}^L$, and $\mathbf{g}_n^H\mathbf{v} = 0$, $\forall n = 1, 2, \ldots$.

We begin the algorithmic developments from the conventional matched filter (MF) with desired response $\mathbf{w}^H\mathbf{v} = \rho$

$$\mathbf{w}_0 = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}, \tag{3}$$

which is MMSE–MVDR optimum for white $\mathbb{C}^L$ vector inputs (when $\mathbf{R} = \sigma^2\mathbf{I}$, $\sigma > 0$). We recall that, w.l.o.g. and for notational simplicity, we assume throughout this presentation that the input vectors $\mathbf{r} \in \mathbb{C}^L$ are zero mean. Next, we incorporate in $\mathbf{w}_0$ an "auxiliary" vector component that is orthogonal to $\mathbf{v}$ and we form (Fig. 1)

$$\mathbf{w}_1 = \mathbf{w}_0 - \mu_1\mathbf{g}_1 = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} - \mu_1\mathbf{g}_1, \tag{4}$$

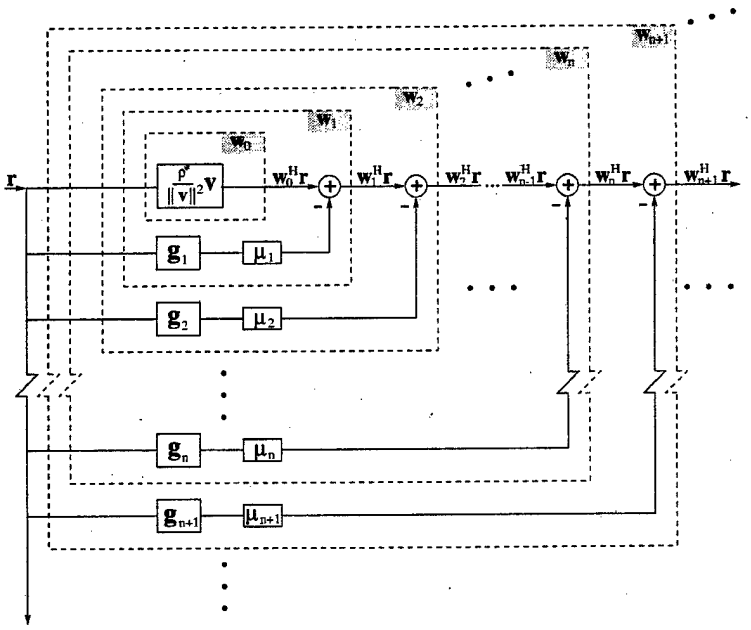


FIG. 1. Block diagram representation of the iteratively generated sequence of filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

where $\mathbf{g}_1 \in \mathbb{C}^L - \{\mathbf{0}\}$, $\mu_1 \in \mathbb{C}$, and $\mathbf{g}_1^H\mathbf{v} = 0$. We assume for a moment that the orthogonal auxiliary vector $\mathbf{g}_1$ is arbitrary but nonzero and fixed and we concentrate on the selection of the scalar μ_1. The value of μ_1 that minimizes the variance of the output of the filter $\mathbf{w}_1$ can be found by direct differentiation of $E\{|\mathbf{w}_1^H\mathbf{r}|^2\}$ or simply as the value that minimizes the MS error between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mu_1^*\mathbf{g}_1^H\mathbf{r}$. This is essentially a scalar version of the GSC weight determination problem and we present the solution in the form of a proposition [18]:

PROPOSITION 1. *The scalar μ_1 that minimizes the variance at the output of $\mathbf{w}_1$ or equivalently minimizes the MS error between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mu_1^*\mathbf{g}_1^H\mathbf{r}$ is*

$$\mu_1 = \frac{\mathbf{g}_1^H\mathbf{R}\mathbf{w}_0}{\mathbf{g}_1^H\mathbf{R}\mathbf{g}_1}, \tag{5}$$

where $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$ is the input autocorrelation matrix.

Since $\mathbf{g}_1$ is set to be orthogonal to $\mathbf{v}$, (5) shows that if the vector $\mathbf{R}\mathbf{w}_0$ happens to be *on* $\mathbf{v}$ (that is, $\mathbf{R}\mathbf{w}_0 = (\mathbf{v}^H\mathbf{R}\mathbf{w}_0)/(\|\mathbf{v}\|^2)\mathbf{v}$ or equivalently $(\mathbf{I} - (\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_0 = \mathbf{0})$, then $\mu_1 = 0$. Indeed, if $\mathbf{R}\mathbf{w}_0 = (\mathbf{v}^H\mathbf{R}\mathbf{w}_0/\|\mathbf{v}\|^2)\mathbf{v}$ then $\mathbf{w}_0$ is *already* the MMSE–MVDR filter. To avoid this trivial case and continue with our developments, we suppose that $\mathbf{R}\mathbf{w}_0 \neq (\mathbf{v}^H\mathbf{R}\mathbf{w}_0/\|\mathbf{v}\|^2)\mathbf{v}$. By inspection, we also observe that for the MS-optimum value of μ_1 the product $\mu_1\mathbf{g}_1 = (\mathbf{g}_1^H\mathbf{R}\mathbf{w}_0/\mathbf{g}_1^H\mathbf{R}\mathbf{g}_1)\mathbf{g}_1$ is independent of the norm of $\mathbf{g}_1$. Hence, so is $\mathbf{w}_1$. At this point, we decide to choose the auxiliary vector $\mathbf{g}_1$ as the normalized vector that maximizes the magnitude of the cross-correlation between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$

and $\mathbf{g}_1^H\mathbf{r}$, under the constraint that $\mathbf{g}_1^H\mathbf{v}=0$ and $\mathbf{g}_1^H\mathbf{g}_1=1$:

$$\mathbf{g}_1 = \arg\max_{\mathbf{g}} \left| E\left\{\mathbf{w}_0^H\mathbf{r}(\mathbf{g}^H\mathbf{r})^*\right\}\right| = \arg\max_{\mathbf{g}} \left|\mathbf{w}_0^H\mathbf{R}\mathbf{g}\right|$$

$$\text{subject to } \mathbf{g}^H\mathbf{v}=0 \text{ and } \mathbf{g}^H\mathbf{g}=1. \tag{6}$$

For the sake of mathematical accuracy, we note that both the criterion function $|\mathbf{w}_0^H\mathbf{R}\mathbf{g}|$ to be maximized and the orthogonality constraints are phase invariant. In other words, if $\mathbf{g}_1$ satisfies (6) so does $\mathbf{g}_1 e^{j\phi}$ for any phase ϕ. Without loss of generality, to avoid any ambiguity in our presentation and to have a uniquely defined auxiliary vector, we seek the one and only auxiliary vector that satisfies (6) and places the cross-correlation value on the positive real line ($\mathbf{w}_0^H\mathbf{R}\mathbf{g} > 0$). This constraint optimization problem was first posed and solved in [19] where the filter $\mathbf{w}_1$ in (4) was used for multiple access interference suppression in multipath CDMA communication channels. Intuitively, the maximum magnitude cross-correlation criterion as defined in (6) strives to identify the auxiliary vector orthonormal to $\mathbf{v}$ that can capture the most interference present in $\mathbf{w}_0^H\mathbf{r}$. The solution, derived through conventional Lagrange multipliers optimization, is given below.

PROPOSITION 2. *Suppose that* $(\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_0 \neq \mathbf{0}$ $(\mathbf{w}_0 \neq \mathbf{w}_{\text{MMSE/MVDR}})$. *Then, the auxiliary vector*

$$\mathbf{g}_1 = \frac{\mathbf{R}\mathbf{w}_0 - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_0}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_0 - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_0}{\|\mathbf{v}\|^2}\mathbf{v}\right\|} \tag{7}$$

maximizes the magnitude of the cross-correlation between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ *and* $\mathbf{g}_1^H\mathbf{r}$, $|\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1|$, *subject to the constraints* $\mathbf{g}_1^H\mathbf{v}=0$ *and* $\mathbf{g}_1^H\mathbf{g}_1=1$. *In addition,* $\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1$ *is real positive* $(\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1 > 0)$.

So far we have defined $\mathbf{w}_0$ in (3) and $\mathbf{w}_1$ in (4) with $\mathbf{g}_1$ and μ_1 given by (7) and (5), respectively. The iterative algorithm for the generation of an infinite sequence of filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$ is already taking shape. Formally, we just need to specify the inductive step. Assuming that the filter $\mathbf{w}_n = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} - \sum_{i=1}^{n}\mu_i\mathbf{g}_i$ has been identified for some $n \geq 1$ and $\mathbf{w}_n \neq \mathbf{w}_{\text{MMSE/MVDR}}$, we argue as in Propositions 1 and 2 and we define

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{n+1}\mathbf{g}_{n+1}, \tag{8}$$

where

$$\mathbf{g}_{n+1} = \frac{\mathbf{R}\mathbf{w}_n - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_n - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{v}\right\|} \tag{9}$$

is the orthonormal auxiliary vector (with respect to $\mathbf{v}$) that, given $\mathbf{w}_n$, maximizes conditionally the cross-correlation magnitude $|E\{\mathbf{w}_n^H\mathbf{r}(\mathbf{g}_{n+1}^H\mathbf{r})^*\}| = |\mathbf{w}_n^H\mathbf{R}\mathbf{g}_{n+1}|$ and

$$\mu_{n+1} = \frac{\mathbf{g}_{n+1}^H\mathbf{R}\mathbf{w}_n}{\mathbf{g}_{n+1}^H\mathbf{R}\mathbf{g}_{n+1}} \tag{10}$$

is the scalar that minimizes the MS error between $\mathbf{w}_n^H\mathbf{r}$ and $\mu_{n+1}^*\mathbf{g}_{n+1}^H\mathbf{r}$ (minimizes $E\{|\mathbf{w}_{n+1}^H\mathbf{r}|^2\}$).

It is important to note that, while the generated auxiliary vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots$ are all constrained to be orthogonal to $\mathbf{v}$, orthogonality among the auxiliary vectors is *not* imposed [20, 21]. This is in sharp contrast to previous work that involved filtering with up to $L-1$ orthogonal to each other and to $\mathbf{v}$ vectors [22–24], where L is the data input vector dimension. We observe, however, that *successive* auxiliary vectors generated by the above recursive conditional optimization procedures (8)–(10) *do* come up orthogonal: $\mathbf{g}_n^H\mathbf{g}_{n+1}=0$, $\forall n = 1, 2, 3, \ldots$ (while $\mathbf{g}_n^H\mathbf{g}_m \neq 0$, $\forall n, m$, $|n-m| \neq 1$). For completeness purposes, we present this observation below in the form of a lemma.

LEMMA 1. *Successive auxiliary vectors generated through* (8)–(10) *are orthogonal:* $\mathbf{g}_n^H\mathbf{g}_{n+1}=0$, $n=1,2,3,\ldots$. *However,* $\mathbf{g}_n^H\mathbf{g}_m \neq 0$, $\forall n, m$, $|n-m| \neq 1$.

The algorithm is summarized in Fig. 2. The conceptual simplicity of the conditional statistical optimization process led to a computationally simple recursion. In Fig. 2 we chose to drop the unnecessary, as previously explained, normalization of the auxiliary vectors and we also factorized their numerator to make the orthogonal projection operator apparent. Formal convergence of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$ to the MMSE–MVDR filter $\rho^*\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$ is established by the following theorem. The proof can be found in [21].

THEOREM 1. *Let* $\mathbf{R}$ *be a Hermitian positive definite matrix. Consider the iterative algorithm of Fig.* 2.

Auxiliary–Vector Algorithm

Input:

Autocovariance matrix $\mathbf{R}$, constraint vector $\mathbf{v}$, desired response $\mathbf{w}^H\mathbf{v} = \rho$.

Initialization:

$\mathbf{w}_0 := \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}$.

Iterative computation:

For n=1, 2, ... do

begin

$\mathbf{g}_n := \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\,\mathbf{w}_{n-1}$

if $\mathbf{g}_n = \mathbf{0}$ then EXIT

$\mu_n := \frac{\mathbf{g}_n^H\mathbf{R}\,\mathbf{w}_{n-1}}{\mathbf{g}_n^H\mathbf{R}\,\mathbf{g}_n}$

$\mathbf{w}_n := \mathbf{w}_{n-1} - \mu_n\mathbf{g}_n$

end

Output:

Filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

FIG. 2. The algorithm for the iterative generation of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

(i) *The generated sequence of auxiliary-vector weights* $\{\mu_n\}$, $n = 1, 2, \ldots$, *is real-valued, positive, and bounded,*

$$0 < \frac{1}{\lambda_{\max}} \leq \mu_n \leq \frac{1}{\lambda_{\min}}, \qquad n = 1, 2, \ldots, \tag{11}$$

where $\lambda_{\max}$ *and* $\lambda_{\min}$ *are the maximum and minimum, correspondingly, eigenvalues of* $\mathbf{R}$.

(ii) *The sequence of auxiliary vectors* $\{\mathbf{g}_n\}$, $n = 1, 2, \ldots$, *converges to the* $\mathbf{0}$ *vector:*

$$\lim_{n\to\infty} \mathbf{g}_n = \mathbf{0}. \tag{12}$$

(iii) *The sequence of auxiliary-vector filters* $\{\mathbf{w}_n\}$, $n = 1, 2, \ldots$, *converges to the MMSE–MVDR filter:*

$$\lim_{n\to\infty} \mathbf{w}_n = \rho^* \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}. \tag{13}$$

We conclude this section with an illustration. We draw a signal model example from the direct-sequence code division multiple access (DS-CDMA) communications literature and we assume a synchronous system where the input signal vector $\mathbf{r} \in \mathcal{R}^L$ is given by

$$\mathbf{r} = \sum_{k=1}^{K} \sqrt{E_k} b_k \mathbf{s}_k + \mathbf{n}. \tag{14}$$

In this setup, K denotes the total number of signals (*users*) present and each signal is defined through an L-dimensional, normalized, binary-antipodal vector waveform (or *user signature*) $\mathbf{s}_k$, $k = 1, 2, \ldots, K$. The signature vector dimension L is usually referred to as the system *spreading gain*. With respect to the kth user signal, E_k is the received signal energy and $b_k \in \{-1, +1\}$ is the information bit modeled as a random variable with equally probable values and assumed to be statistically independent from all other user bits b_j, $j \neq k$. Additive white Gaussian noise contributions are accounted for by $\mathbf{n}$ with autocorrelation matrix $E\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \sigma^2\mathbf{I}_{L\times L}$ ($\mathbf{x}^{\mathrm{T}}$ denotes the transpose of $\mathbf{x}$). With this notation and normalized user signatures, the signal-to-noise ratio of the kth user signal is defined by $\mathrm{SNR}_k \triangleq 10\log_{10} E_k/\sigma^2$ dB, $k = 1, 2, \ldots, K$.

MMSE–MVDR filtering for DS-CDMA type problems has attracted significant interest [25–29]. If we wish to recover the information bits of, say, user *1*, then all other signals constitute multiple-access interference and the MMSE–MVDR filter is built with constraint vector $\mathbf{v} = \mathbf{s}_1$, desired response $\mathbf{w}^{\mathrm{T}}\mathbf{s}_1 = 1$, and autocorrelation matrix $\mathbf{R} = \sum_{k=1}^{K} E_k\mathbf{s}_k\mathbf{s}_k^{\mathrm{T}} + \sigma^2\mathbf{I}$. We choose $L = 32$, $K = 13$, and we draw an arbitrary set of signatures $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{13}$. For purposes of completeness in presentation, the exact signature assignment is given in the Appendix. We fix the SNR of the user of interest at $\mathrm{SNR}_1 = 12$ dB while the *interferers* $k = 2, \ldots, 13$ are at $\mathrm{SNR}_{2\text{–}5} = 10$ dB, $\mathrm{SNR}_{6\text{–}9} = 12$ dB and $\mathrm{SNR}_{10\text{–}13} = 14$ dB. Figure 3 shows how the sequence of filters $\mathbf{w}_0$, $\mathbf{w}_1$, ... generated by the

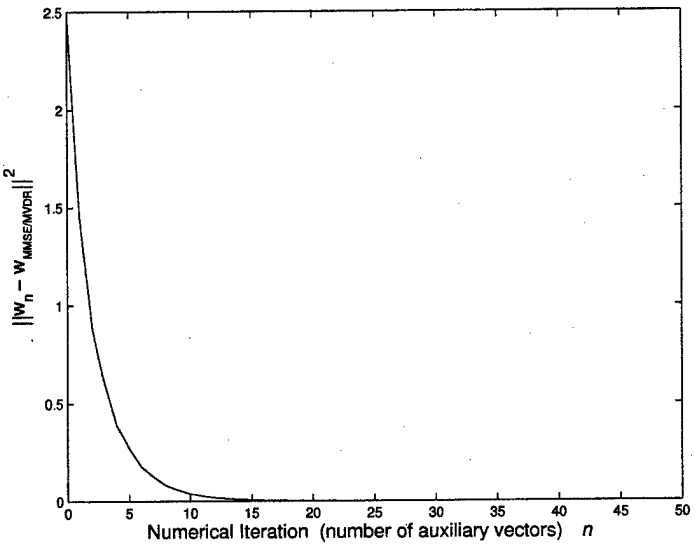


FIG. 3. Convergence of the sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, to the MMSE–MVDR solution for the signal model example in (14).

algorithm in Fig. 2 converges to the MMSE–MVDR solution. The convergence is captured in terms of the norm-square metric $\|\mathbf{w}_n - \mathbf{w}_{\text{MMSE/MVDR}}\|^2$ as a function of the iteration step (index of the AV filter in the sequence or number of auxiliary vectors used) n.

3. FILTER ESTIMATION

Consider a constraint vector $\mathbf{v}$ and a Hermitian positive definite autocorrelation matrix $\mathbf{R}$ of an input vector $\mathbf{r} \in \mathbb{C}^L$. Assume that $\mathbf{R}$ is in fact unknown and it is sample-average estimated from a data record of M points: $\widehat{\mathbf{R}}(M) = (1/M)\sum_{m=1}^{M} \mathbf{r}_m \mathbf{r}_m^H$. For Gaussian inputs, $\widehat{\mathbf{R}}(M)$ is a maximum-likelihood (ML), consistent, unbiased estimator of $\mathbf{R}$ [3, 30]. For a large class of multivariate elliptically contoured input distributions that includes the Gaussian, if $M \geq L$ then $\widehat{\mathbf{R}}(M)$ is positive definite (hence invertible) with probability 1 (w.p. 1) [31–33]. Then, Theorem 1 in Section 2 shows that

$$\widehat{\mathbf{w}}_n(M) \longrightarrow_{n\to\infty} \widehat{\mathbf{w}}_\infty(M) = \rho^* \frac{[\widehat{\mathbf{R}}(M)]^{-1}\mathbf{v}}{\mathbf{v}^H[\widehat{\mathbf{R}}(M)]^{-1}\mathbf{v}}, \tag{15}$$

where $\widehat{\mathbf{w}}_\infty(M)$ is the widely used MMSE–MVDR filter estimator known as the SMI filter [5].

The output sequence begins from $\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$, which is a 0-variance, fixed-valued estimator that may be severely biased ($\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} \neq \mathbf{w}_{\text{MMSE/MVDR}}$) unless $\mathbf{R} = \sigma^2\mathbf{I}$, for some $\sigma > 0$. In the latter trivial case, $\widehat{\mathbf{w}}_0(M)$ is already the perfect MMSE–MVDR filter. Otherwise, the next filter estimator in the sequence, $\widehat{\mathbf{w}}_1(M)$, has a significantly reduced bias due to the optimization

procedure employed, at the expense of nonzero estimator (co-) variance. As we move up in the sequence of filter estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, the bias decreases rapidly to zero[2] while the variance increases slowly to the SMI ($\widehat{\mathbf{w}}_\infty(M)$) levels (cf. (15)). To quantify these remarks, we plot in Fig. 4 the norm-square bias $\|E\{\widehat{\mathbf{w}}_n(M)\} - \mathbf{w}_{\text{MMSE/MVDR}}\|^2$ and the trace of the covariance matrix $E\{[\widehat{\mathbf{w}}_n(M) - E\{\widehat{\mathbf{w}}_n(M)\}][\widehat{\mathbf{w}}_n(M) - E\{\widehat{\mathbf{w}}_n(M)\}]^H\}$ as a function of the iteration step n for the signal model example of Fig. 3 and data record size $M = 256$. Bias and cov-trace values are calculated from 100000 independent filter estimator realizations for each iteration point n. Formal, theoretical statistical analysis of the generated estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, is beyond the scope of this presentation. We do note, however, that for multivariate elliptically contoured input distributions, an analytic expression for the covariance matrix of the SMI estimator $\widehat{\mathbf{w}}_\infty(M)$ can be found in [33]: $E\{[\widehat{\mathbf{w}}_\infty(M) - E\{\widehat{\mathbf{w}}_\infty(M)\}][\widehat{\mathbf{w}}_\infty(M) - E\{\widehat{\mathbf{w}}_\infty(M)\}]^H\} = [|\rho|^2 E(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})(M - L + 1)](\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})$. Since under these input distribution conditions $\widehat{\mathbf{w}}_\infty(M)$ is unbiased, the trace of the covariance matrix is the MS filter estimation error. It is important to observe that the covariance matrix and, therefore, the MS filter estimation error depend on the data record size M and the filter length L, as well as the specifics of the signal processing problem at hand ($\mathbf{R}$ and $\mathbf{v}$). It is also important to note that for the CDMA signal model example in (14) the input is Gaussian-mixture distributed. Therefore, the analytic result in [33] is not directly applicable and can only be thought of as an approximation (a rather close approximation as we concluded in our studies). In any case, from the results in Fig. 4 for $M = 256$, we see that the

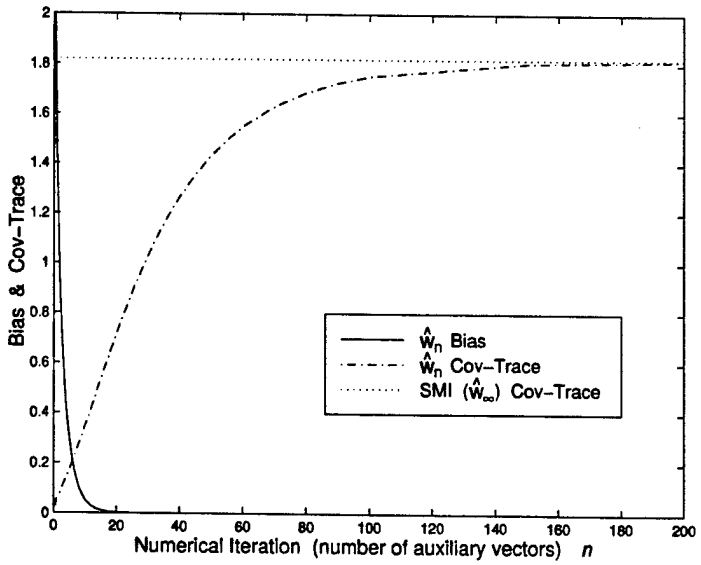


FIG. 4. Norm-square bias and covariance trace for the sequence of estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, \ldots$. The signal model is as in Fig. 3 and $M = 256$.

[2] The SMI estimator is unbiased for multivariate elliptically contoured input distributions [33, 34]: $E\{\widehat{\mathbf{w}}_\infty(M)\} = \mathbf{w}_{\text{MMSE/MVDR}} = \rho^*\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$.

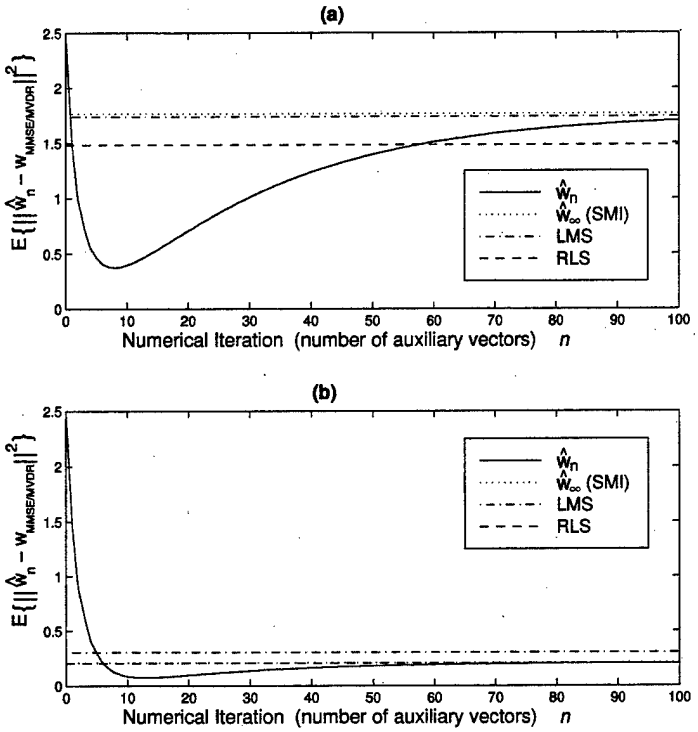


FIG. 5. MS estimation error for the sequence of estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, \ldots$. Data record size (a) $M = 256$, (b) $M = 2048$.

estimators $\widehat{\mathbf{w}}_1(M), \widehat{\mathbf{w}}_2(M), \ldots$, up to about $\widehat{\mathbf{w}}_{20}(M)$ are particularly appealing. In contrast, the estimators $\widehat{\mathbf{w}}_n(M)$ for $n > 20$ do not justify their increased cov-trace cost since they have almost nothing to offer in terms of further bias reduction.

The mean-square estimation error expression $E\{\|\widehat{\mathbf{w}}_n(M) - \mathbf{w}_{\text{MMSE/MVDR}}\|^2\}$ captures the bias–variance balance of the individual members of the estimator sequence $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$. In Fig. 5 we plot the MS estimation error as a function of the iteration step (index of AV filter in the sequence or number of auxiliary vectors) n for the case study in Fig. 4, for $M = 256$ (Fig. 5a) and $M = 2048$ (Fig. 5b). As a reference, we also include the MS-error of the constraint-LMS estimator [4, 35]

$$\widehat{\mathbf{w}}_{\text{LMS}}(m) = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\left[\widehat{\mathbf{w}}_{\text{LMS}}(m-1) - \mu \mathbf{r}_m \mathbf{r}_m^H \widehat{\mathbf{w}}_{\text{LMS}}(m-1)\right] + \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}, \quad m = 1, \ldots, M, \tag{16}$$

with $\widehat{\mathbf{w}}_{\text{LMS}}(0) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and some $\mu > 0$, and the RLS estimator [7, 8] with matrix-inversion-lemma-based $\mathbf{R}^{-1}$ estimation:

$$\widehat{\mathbf{R}}^{-1}(m) = \widehat{\mathbf{R}}^{-1}(m-1) - \frac{\widehat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m \mathbf{r}_m^H \widehat{\mathbf{R}}^{-1}(m-1)}{1 + \mathbf{r}_m^H \widehat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m}, \qquad m = 1, \ldots, M, \tag{17}$$

with $\widehat{\mathbf{R}}^{-1}(0) = (1/\epsilon_0)\mathbf{I}$ for some $\epsilon_0 > 0$. Theoretically, it is known that the LMS gain parameter $\mu > 0$ [36] has to be less than $1/(2 \cdot \lambda_{\max}^{\text{blocked}})$, where

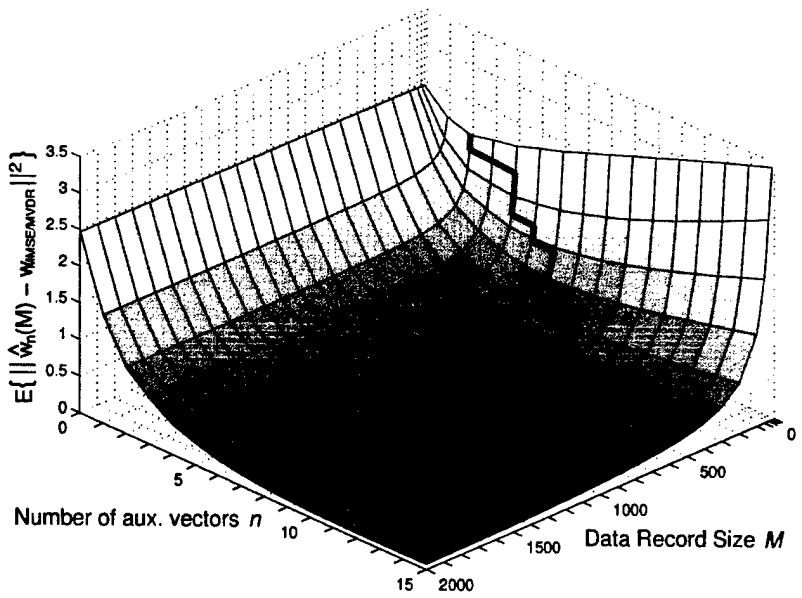


FIG. 6. MS estimation error versus number of auxiliary vectors n and sample support M.

$\lambda_{\max}^{\text{blocked}}$ is the maximum eigenvalue of the blocked-data autocorrelation matrix $(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)$. While this is a theoretical upper bound, practitioners are well aware that empirical, data-dependent optimization or tuning of the LMS gain $\mu > 0$ or the RLS initialization parameter $\epsilon_0 > 0$ [37] is necessary to achieve acceptable performance (in our study we set $\mu = 1/(200 \cdot \lambda_{\max}^{\text{blocked}})$ and $\epsilon_0 = 20$, respectively). This data specific tuning frequently results in misleading, overoptimistic conclusions about the short data record performance of the LMS and RLS algorithms. In contrast, when the filter estimators $\widehat{\mathbf{w}}_n$ generated by the algorithm of Fig. 2 are considered instead, tuning of the real-valued parameters μ and ϵ_0 is virtually replaced by an integer choice among the first several members of the $\{\widehat{\mathbf{w}}_n\}$ sequence. Adaptive, data-dependent criteria for the selection of the most appropriate AV filter in the sequence for a given data record are developed in the next section. In Fig. 5a, for $M = 256$ all estimators $\widehat{\mathbf{w}}_n$ from $n = 2$ up to about $n = 55$ outperform in MS-error their RLS, LMS, and SMI ($\widehat{\mathbf{w}}_\infty$) counterparts. $\widehat{\mathbf{w}}_8$ ($n = 8$ auxiliary vectors) has the least MS-error of all (best bias–variance trade-off). When the data record size is increased to $M = 2048$ (Fig. 5b), we can afford more iterations (more auxiliary vectors) and $\widehat{\mathbf{w}}_{13}$ offers the best bias–variance trade-off (lowest MS-error). All filter estimators $\widehat{\mathbf{w}}_n$ for $n > 8$ outperform the LMS, RLS, and SMI ($\widehat{\mathbf{w}}_\infty$) estimators. For such large data record sets ($M = 2048$), the RLS and the SMI ($\widehat{\mathbf{w}}_\infty$) MS-errors are almost identical. Figure 6 offers a 3-dimensional plot of the MS estimation error as a function of the number of auxiliary vectors n and the sample support M. The dark line that traces the bottom of the MS estimation error surface identifies the best number of auxiliary vectors for any given data record size M.

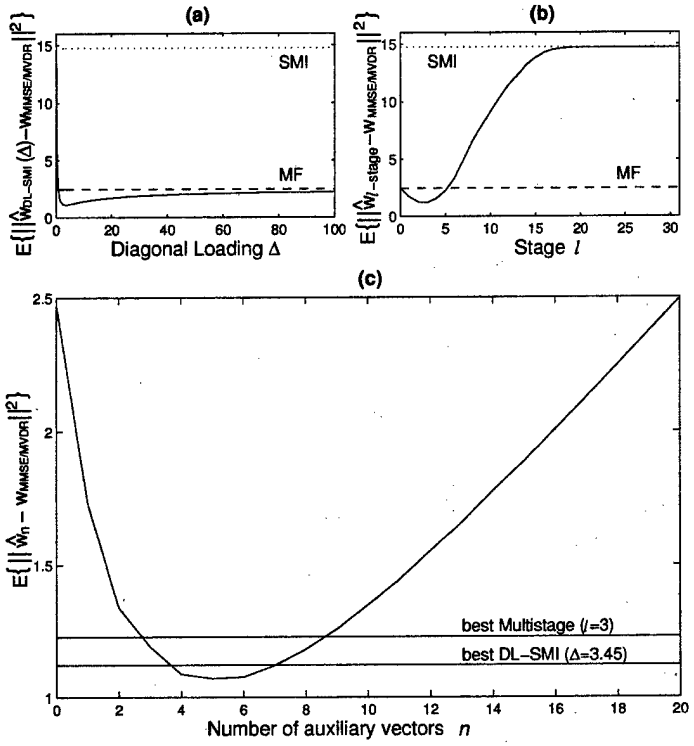


FIG. 7. MS estimation error studies for (a) diagonally loaded SMI, (b) multistage, and (c) auxiliary-vector estimators ($M = 60$).

An alternative bias–variance trading mechanism through real-valued tuning is the diagonally loaded (DL) SMI estimator [6]

$$\widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta) = \rho^* \frac{[\widehat{\mathbf{R}}(M) + \Delta \mathbf{I}]^{-1}\mathbf{v}}{\mathbf{v}^H[\widehat{\mathbf{R}}(M) + \Delta \mathbf{I}]^{-1}\mathbf{v}}, \tag{18}$$

where $\Delta \geq 0$ is the diagonal loading parameter. We observe that $\widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta = 0)$ is the regular SMI estimator, while $\lim_{\Delta \to \infty} \widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ which is the properly scaled matched filter. In Fig. 7a we plot the MS estimation error of the DL-SMI estimator as a function of the diagonal loading parameter Δ ($M = 60$). We identify the *best possible* diagonal loading value $\Delta \simeq 3.45$ (at significant computational cost) and in Fig. 7c we compare the best DL-SMI estimator against the AV estimator sequence for which *no* diagonal loading is performed. Interestingly, the AV estimators $\widehat{\mathbf{w}}_n$ from $n = 4$ to 7 outperform in MS-error the best possible DL-SMI estimator ($\Delta \simeq 3.45$).

Finally, a *finite set* of L filter estimators with varying bias–covariance balance can be obtained through the use of the orthogonal multistage filter decomposition procedure in [9, 10] (the resulting filters have been also referred to as nested Wiener filters). It can be shown theoretically that the l-stage filter, $\mathbf{w}_{l\text{-stage}}$, $0 \leq l \leq L-1$, is equivalent to the following structure. First, change the auxiliary-vector generation recursion in (9) or Fig. 2 to impose orthogonality

not only with respect to the constraint vector $\mathbf{v}$ but also with respect to *all previously defined* auxiliary vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}$, $n \leq L-1$:

$$\mathbf{y}_n = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2} - \sum_{i=1}^{n-1} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\|\mathbf{y}_i\|^2} \right) \mathbf{R}\mathbf{w}_{n-1}. \tag{19}$$

Next, terminate the recursion at $n = l$, $0 \leq l \leq L-1$, and organize the l orthogonal to each other and to $\mathbf{v}$ vectors $\mathbf{y}_1, \ldots, \mathbf{y}_l$ in the form of a blocking matrix $\mathbf{B}_{L\times l} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l]$. Then,

$$\mathbf{w}_{l\text{-stage}} = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} - \mathbf{B}_{L\times l}\vec{\alpha}_{l\times 1}, \tag{20}$$

where

$$\vec{\alpha} = \frac{\rho^*}{\|\mathbf{v}\|^2}[\mathbf{B}^H\mathbf{R}\mathbf{B}]^{-1}\mathbf{B}^H\mathbf{R}\mathbf{v} \tag{21}$$

is the MS vector-optimum (unconditionally optimum) set of weights of the vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l$.[3] In the context of MMSE–MVDR filter estimation from a data record of size M, $\widehat{\mathbf{w}}_{0\text{-stage}}(M)$ is the matched filter and $\widehat{\mathbf{w}}_{(L-1)\text{-stage}}(M)$ is the SMI estimator. In Fig. 7b we plot the MS estimation error of $\widehat{\mathbf{w}}_{l\text{-stage}}(M)$ as a function of l, $0 \leq l \leq L-1 = 31$ ($M = 60$). We identify the *best* multistage estimator ($l = 3$ stages) and in Fig. 7c we compare against the AV estimator sequence. We see that all AV estimators $\widehat{\mathbf{w}}_n$ from $n = 3$ to 8 outperform in MS-error the best multistage estimator ($l = 3$ stages). Finally, as a last study, in Fig. 8 we plot the MS-error of the $\Delta = 3.45$ DL-SMI estimator together with the MS-error of the *best* multistage and AV estimators over the data support range $M = L/2 = 16$ to $M = 3L = 96$.

4. HOW TO CHOOSE THE NUMBER OF AUXILIARY VECTORS

In this section we present two data-driven rules for the selection of the number of auxiliary vectors n [39]. The first rule selects the AV filter estimator with n auxiliary vectors that has minimum cross-validated average filter output energy. The second selection rule is specific to BPSK communications receivers that employ a sign detector at the output of the linear auxiliary-vector filter. Details are given below.

[3] Therefore, the multistage filter in [9, 10] is identical to the filter $\mathbf{w}_B$ as it appears in [22–24]. The multistage decomposition algorithm is a computationally efficient procedure for the calculation of this filter tailored to the particular structure of $\mathbf{B}^H\mathbf{R}\mathbf{B}$ (tridiagonal matrix). The same computational savings can be achieved by the general forward calculation algorithm of Liu and Van Veen [38] that returns all intermediate stage filters along the way, up to the stage of interest l (total computational complexity of order $O((M+l)L^2)$). The AV algorithm in Fig. 2 has computational complexity $O((M+n)L^2)$ where n is the desired number of auxiliary vectors. Again, all intermediate AV filters are returned. Estimators of practical interest have $l \ll M$ or $n \ll M$. Therefore, the complexity of all such algorithms is dominated by $O(ML^2)$ which is required for the computation of $\widehat{\mathbf{R}}(M)$.

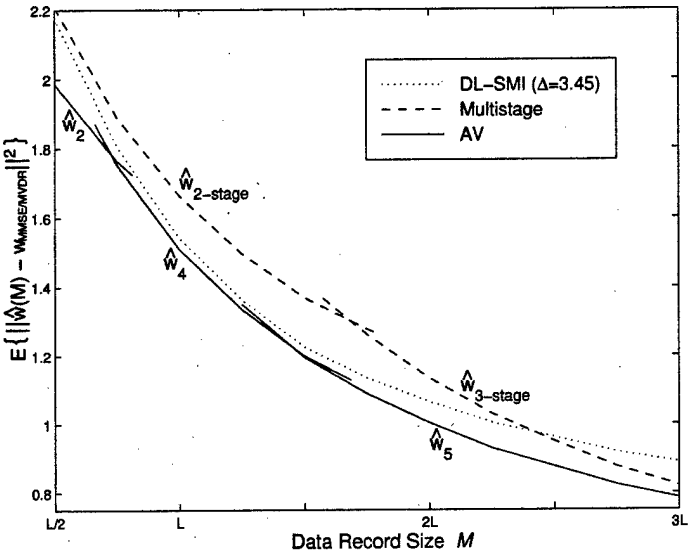


FIG. 8. MS estimation error for the *best* multistage and AV estimators over the data support range $M = L/2 = 16$ to $M = 3L = 96$. The MS estimation error of the $\Delta = 3.45$ DL-SMI estimator is also included as a reference.

4.1. Cross-Validated Minimum Output Variance Rule (CV-MOV)

Cross-validation is a well-known statistical method [40]. Here, we use cross-validation to select the filter parameter of interest (number of auxiliary vectors n) that minimizes the output variance which is estimated based on input observations that have not been used in the process of building the filter estimator itself.

A particular case of cross-validation that we use in this work is the leave-one-out method. The following criterion defines the CV-MOV AV filter estimator selection process.

CRITERION 1. *For a given data record of size M, the cross-validated minimum output variance AV filter estimator selection rule chooses the AV filter estimator $\widehat{\mathbf{w}}_{n_1}(M)$ that minimizes the cross-validated sample average output variance, i.e.,*

$$n_1 = \arg\min_n \left\{ \sum_{m=1}^{M} \widehat{\mathbf{w}}_n^H(M\backslash m)\mathbf{r}_m\mathbf{r}_m^H\widehat{\mathbf{w}}_n(M\backslash m) \right\}, \tag{22}$$

where $(M\backslash m)$ identifies the AV filter estimator that is evaluated from the available data record after removing the mth sample.

It may be important to emphasize the need for invoking the cross-validation technique for the evaluation of the sample-average output variance. If a sample-average evaluation using *all* data were attempted, then the selection rule would take the form $\min_n\{\widehat{\mathbf{w}}_n^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_n(M)\}$, where $\widehat{\mathbf{R}}(M) = (1/M)\sum_{m=1}^{M}\mathbf{r}_m\mathbf{r}_m^H$ and $\widehat{\mathbf{w}}_n(M)$ is given by the algorithm of Fig. 2 with $\widehat{\mathbf{R}}(M)$ in place of $\mathbf{R}$. Such minimization, however, would result in $n = \infty$ since we know that $\widehat{\mathbf{w}}_n^H(M)\widehat{\mathbf{R}}(M)$

$\widehat{\mathbf{w}}_n(M) \underset{n\to\infty}{\longrightarrow} \widehat{\mathbf{w}}_\infty^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_\infty(M)$ and the SMI estimator $\widehat{\mathbf{w}}_\infty(M)$ achieves minimum sample average output variance $\widehat{\mathbf{w}}_\infty^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_\infty(M)$ (but not minimum true output variance $\widehat{\mathbf{w}}_\infty^H(M)\mathbf{R}\widehat{\mathbf{w}}_\infty(M)$, of course).

4.2. *Output J-divergence Rule*

For illustration purposes we reconsider the BPSK CDMA signal model example in (14). The output J-divergence rule selects the AV filter estimator from the sequence of AV estimators that maximizes the J-divergence of the Gaussian approximated conditional filter-output distributions (conditioned on the transmitted information bit $b_1 = +1$ or $b_1 = -1$). The appropriateness of such a criterion as well as implementation details are presented below.

For a given AV filter estimator $\widehat{\mathbf{w}}_n(M)$, we denote by p_m the real part [4] of the filter output with input the mth data vector $\mathbf{r}_m$, $m = 1, 2, \ldots, M$,

$$p_m \triangleq \mathrm{Re}\big[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m\big] = \sqrt{E_1}\,\mathrm{Re}\big[b_1(m)\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_1\big] + \sum_{k=2}^{K}\sqrt{E_k}\,\mathrm{Re}\big[b_k(m)\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k\big] + \mathrm{Re}\big[\widehat{\mathbf{w}}_n^H(M)\mathbf{n}_m\big], \tag{23}$$

where the information bits $b_k(m)$, $k = 1, 2, \ldots, K$, $m = 1, 2, \ldots, M$, are assumed to be independent identically distributed (i.i.d.) with equally probable outcomes and $\mathbf{n}_m$ is a 0-mean complex white Gaussian random vector with autocovariance matrix $\sigma^2\mathbf{I}$. Then, the scalars p_m, $m = 1, 2, \ldots, M$, are i.i.d. with common distribution $f_P(x)$ given by

$$f_P(x) = \frac{1}{2^K\sqrt{2\pi}\sigma}\sum_{i=1}^{2^K}\exp\left\{\frac{-\{x - \sum_{k=1}^{K}\sqrt{E_k}\,\mathrm{Re}[b_k^{(i)}\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k]\}^2}{2\sigma^2\|\widehat{\mathbf{w}}_n(M)\|^2}\right\}, \tag{24}$$

where $b_k^{(i)}$, $i = 1, 2, \ldots, 2^K$, is the bit of user k in the ith bit-combination.

Conditioned on the transmitted information bit of the user of interest, user *1*, the pdf of the filter output is a mixture of 2^{K-1} Gaussian distributions. However, for "effective" interference suppressive filters we can safely approximate the conditional output distribution by a Gaussian distribution as argued in [41] for MMSE–MVDR linear filtering. Under this approximation, the filter output conditional distributions given that $+1$ or -1 is transmitted are $f_{1,n} \sim \mathcal{N}[\mu(n), \sigma_{I+N}^2(n)]$ and $f_{0,n} \sim \mathcal{N}[-\mu(n), \sigma_{I+N}^2(n)]$, respectively, where $\mu(n) \triangleq \sqrt{E_1}\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_1]$ and $\sigma_{I+N}^2(n) \triangleq \sum_{k=2}^{K} E_k\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k]^2 + \sigma^2\|\widehat{\mathbf{w}}_n(M)\|^2$ is the conditional variance due to multiple access interference and additive white Gaussian noise (AWGN) (the index "I+N" denotes comprehensively the disturbance contribution). The effect of the above approximation on the performance of the output J-divergence selection rule will be examined in Section 5.

The J-divergence distance $J(f_{1,n}, f_{0,n})$ between the distributions $f_{1,n}(\cdot)$ and $f_{0,n}(\cdot)$ is defined as the sum of the Kullback–Leibler (K-L) distances between

[4] While the signal model in (14) is real-valued, we choose to carry out this presentation in the more general context of complex input vectors and filters.

$f_{1,n}$ and $f_{0,n}$

$$J(f_{1,n}, f_{0,n}) \triangleq D(f_{1,n}, f_{0,n}) + D(f_{0,n}, f_{1,n}), \tag{25}$$

where the K–L distance of $f_{1,n}$ from $f_{0,n}$ is defined by $D(f_{1,n}, f_{0,n}) \triangleq \int_{-\infty}^{\infty} f_{1,n}(x) \log(f_{1,n}(x)/f_{0,n}(x))\,dx$ [42]. Since $J(f_{1,n}, f_{0,n})$ is a function of the AV filter-estimator parameter n (number of auxiliary vectors), in the rest of this paper we will use the notation $J(n)$ to represent the J-divergence distance between $f_{1,n}(x)$ and $f_{0,n}(x)$. For the Gaussian approximated pdf's $f_{1,n}$ and $f_{0,n}$, we have $D(f_{1,n}, f_{0,n}) = D(f_{0,n}, f_{1,n}) = [2\mu(n)]^2/2\sigma^2_{I+N}(n)$ and the J-divergence simplifies to

$$J(n) = \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}. \tag{26}$$

Expression (26) justifies our choice of the output J-divergence as one of the underlying rules for the selection of the best AV filter estimator. We recall that under the same Gaussian approximation of the conditional filter-output pdf's the filter output signal-to-interference-plus-noise-ratio (SINR) can be expressed as $J(n)/4$ and, consequently, the bit-error-rate (BER) as $Q(\sqrt{J(n)}/2)$, where $Q(x) \triangleq \int_x^{\infty} 1/\sqrt{2\pi}\exp(-u^2/2)\mathrm{d}u$. To this extent, maximization of the output J-divergence in (26) implies minimization of the BER. Therefore, we propose to select the estimator from the generated sequence of AV filter estimators that exhibits maximum *estimated* J-divergence.

(1) *Supervised output J-divergence rule.* Exploiting the symmetry of $f_{1,n}(\cdot)$ and $f_{0,n}(\cdot)$, we can show in a straightforward manner that

$$J(n) = \frac{4E^2\{b_1 \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}}{\operatorname{Var}\{b_1 \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}} \tag{27}$$

$$= \frac{4[\sum_{i=\pm 1} E\{i \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|b_1 = i\} \Pr(b_1 = i)]^2}{\sum_{i=\pm 1} \operatorname{Var}\{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|b_1 = i\} \Pr(b_1 = i)}, \tag{28}$$

where $\operatorname{Var}\{\cdot\}$ and $\Pr\{\cdot\}$ denote variance and probability, respectively. Assuming availability of a pilot information bit sequence $\{b_1(m)\}_{m=1}^{M}$, we propose to estimate $J(n)$ by estimating statistical expectations and probabilities via sample averaging and frequencies of occurrence, respectively. We note that although (27) and (28) are ideally equivalent (when all statistical quantities are known), this is not the case in general when estimated measures are considered. So, let $\{p_1^+, p_2^+, p_3^+, \ldots, p_{M_1}^+\} \triangleq \{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] : b_1(m) = +1,\ m = 1, 2, 3 \ldots, M\}$ and $\{p_1^-, p_2^-, p_3^-, \ldots, p_{M_2}^-\} \triangleq \{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] : b_1(m) = -1,\ m = 1, 2, 3, \ldots, M\}$ be the sets of all filter outputs under $b_1(m) = +1$ and $b_1(m) = -1$, respectively ($M_1, M_2 \neq 0$ and $M_1 + M_2 = M$). First, it can be shown in a straightforward manner that the estimator of the numerator of (28) (that is, the weighted average of the sample average mean of the set $\{p_m^+\}_{m=1}^{M_1}$ and the sample average mean of the set $\{p_m^-\}_{m=1}^{M_2}$ weighted by the frequency of each set) is equivalent to the estimator of the numerator of (27) $4\hat{\mu}^2(n)$ where $\hat{\mu}(n) \triangleq \sum_{m=1}^{M_1+M_2} b_1(m)p_m/(M_1 + M_2)$ (in fact, $\hat{\mu}(n)$ is the minimum variance unbiased estimator of $\mu(n)$ [43, p. 178]).

Estimators of the denominator of (27) and (28) are examined in the following proposition. The proof is included in the Appendix.

PROPOSITION 3. *Consider the estimator of* $\sigma^2_{I+N}(n)$, *which is the weighted average of the sample average variance of the set* $\{p_m^+\}_{m=1}^{M_1}$, *and the sample average variance of the set* $\{p_m^-\}_{m=1}^{M_2}$, *weighted by the frequency of occurrence of each set*

$$\hat{\sigma}_1^2(n) = \frac{M_1}{M_1+M_2}\hat{\sigma}^{2+}(n) + \frac{M_2}{M_1+M_2}\hat{\sigma}^{2-}(n), \tag{29}$$

where

$$\hat{\sigma}^{2+}(n) \triangleq \frac{1}{M_1}\sum_{m=1}^{M_1}(p_m^+ - \hat{\mu}^+(n))^2, \qquad \hat{\mu}^+(n) \triangleq \frac{1}{M_1}\sum_{m=1}^{M_1} p_m^+,$$

$$\hat{\sigma}^{2-}(n) \triangleq \frac{1}{M_2}\sum_{m=1}^{M_2}(p_m^+ - \hat{\mu}^-(n))^2, \qquad \hat{\mu}^-(n) \triangleq \frac{1}{M_2}\sum_{m=1}^{M_2} p_m^-,$$

and $\{p_m^+\}_{m=1}^{M_1}$, $\{p_m^-\}_{m=1}^{M_2}$, M_1, *and* M_2 *are as defined previously. Consider also the direct sample average estimator of* $\sigma^2_{I+N}(n)$,

$$\hat{\sigma}_2^2(n) = \frac{1}{M_1+M_2}\sum_{m=1}^{M_1+M_2}[b_1(m)p_m - \hat{\mu}(n)]^2, \tag{30}$$

where p_m *is given by* (23). *The estimators* $\hat{\sigma}_1^2(n)$ *and* $\hat{\sigma}_2^2(n)$ *exhibit the following properties:* (i) *They are both biased and* (ii) $\hat{\sigma}_2^2(n)$ *exhibits smaller MSE from the true value than* $\hat{\sigma}_1^2(n)$.

Utilizing Proposition 3, estimators for the filter-output J-divergence become readily available. The following theorem identifies their relative merits. The proof is included in the Appendix.

THEOREM 2. *Define the two supervised estimators of the output J-divergence* $\hat{J}_{S,1}(n) \triangleq 4\hat{\mu}^2(n)/\hat{\sigma}_1^2(n)$ *and* $\hat{J}_{S,2}(n) \triangleq 4\hat{\mu}^2(n)/\hat{\sigma}_2^2(n)$, *where the subscript "S" identifies a supervised implementation. Then, for a given information bit pilot sequence of size* $M = M_1 + M_2$, *where* M_1 *and* M_2 *are the cardinalities of the sets* $\{b_1(m) = +1\}$ *and* $\{b_1(m) = -1\}$, *respectively, both estimators are biased while the MSE of* $\hat{J}_{S,2}(n)$ *is less than the MSE of* $\hat{J}_{S,1}(n)$; *i.e.,*

$$E\left\{\left[\hat{J}_{S,2}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\} < E\left\{\left[\hat{J}_{S,1}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\}, \tag{31}$$

where $4\mu^2(n)/\sigma^2_{I+N}(n)$ *is the true value of* $J(n)$ *as given by* (26).

Using the preferred estimator $\hat{J}_{S,2}(n)$, the supervised implementation of the output J-divergence AV filter estimator selection rule takes the final form given by the following criterion.

CRITERION 2. *For a given information bit pilot sequence of size M, the supervised J-divergence AV filter estimator selection rule chooses the estimator $\widehat{\mathbf{w}}_{n_2}(M)$ with n_2 auxiliary vectors where*

$$n_2 = \arg\max_n \{\hat{J}_{S,2}(n)\}$$
$$= \arg\max_n \left\{ \frac{4[\frac{1}{M}\sum_{m=1}^{M} b_1(m)\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]]^2}{\frac{1}{M}\sum_{m=1}^{M}[b_1(m)\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] - \hat{\mu}(n)]^2} \right\}. \tag{32}$$

(2) *Unsupervised (blind) output J-divergence rule.* The blind implementation of the rule is obtained by substituting the information bit b_1 in (27) by the detected bit $\hat{b}_1 = \mathrm{sgn}[\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]]$ (output of the sign detector that follows the linear filter). In particular, using $\hat{b}_1$ in place of b_1 in (27) we obtain the following J-divergence expression

$$J_B(n) = \frac{4E^2\{\hat{b}_1\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}}{\mathrm{Var}\{\hat{b}_1\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}} = \frac{4E^2\{|\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|\}}{\mathrm{Var}\{|\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|\}}, \tag{33}$$

where the subscript "B" identifies the blind version of the J-divergence function. The following proposition provides the conditions under which $J_B(n)$ is nearly equal to $J(n)$. The proof is included in the Appendix.

PROPOSITION 4. *If $\mu(n)/(\sigma_{I+N}(n)) \gg 1$, i.e., the filter output SINR is significantly higher than 0 dB, then $J_B(n) \approx J(n)$.*

To estimate $J_B(n)$ from a data record of finite size, we substitute the statistical expectations in (33) by sample averages. The following criterion summarizes the corresponding AV filter estimator selection rule.

CRITERION 3. *For a given data record of size M, the unsupervised (blind) J-divergence AV filter estimator selection rule chooses the estimator $\widehat{\mathbf{w}}_{n_3}(M)$ with n_3 auxiliary vectors where*

$$n_3 = \arg\max_n \{\hat{J}_B(n)\}$$
$$= \arg\max_n \left\{ \frac{4[\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|]^2}{\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|^2 - [\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|]^2} \right\}. \tag{34}$$

5. SIMULATION STUDIES

We examine the performance of the proposed short data record AV filter estimator selection rules for a DS-CDMA system with K users, spreading gain L, and multipath fading reception by a narrowband antenna array with N elements. All elements experience identical fading. Let J denote the number of chip interval spaced paths per baseband user signal. After conventional carrier demodulation, chip-matched filtering, and sampling at the chip rate over a multipath extended symbol interval of $L+J-1$ chips, the $L+J-1$ data samples

from the ith antenna element, $i = 1, 2, \ldots, N$, are organized in the form of a vector $\mathbf{r}_m^{(i)}$ given by

$$\mathbf{r}_m^{(i)} = \sum_{k=1}^{K}\sum_{t=1}^{J} c_{k,t}\sqrt{E_k}(b_k(m)\mathbf{s}_{k,t} + b_k^-(m)\mathbf{s}_{k,t}^- + b_k^+(m)\mathbf{s}_{k,t}^+)a_{k,t}[i] + \mathbf{n}_m^{(i)},$$
$$m = 1, \ldots, M,\ i = 1, \ldots, N. \tag{35}$$

In (35), with respect to the kth user signal, E_k is the transmitted energy, $b_k(m)$, $b_k^-(m)$, and $b_k^+(m)$ are the present, the previous, and the following transmitted bits, respectively, and $c_{k,t}$ is the coefficient of the tth path of the kth user signal. The channel coefficients are modeled as independent zero-mean complex Gaussian random variables that are assumed to remain constant over the filter adaptation data record of size M. $\mathbf{s}_{k,t}$ represents the $(J-1)$-zero-padded and $(t-1)$-right-shifted version of the signature of the kth user $\mathbf{s}_k$; $\mathbf{s}_{k,t}^-$ is the 0-filled, L-left-shifted version of $\mathbf{s}_{k,t}$; and $\mathbf{s}_{k,t}^+$ is the 0-filled, L-right-shifted version of $\mathbf{s}_{k,t}$. Finally, $\mathbf{n}_m^{(i)}$ represents additive complex white Gaussian noise and $a_{k,t}[i]$ denotes the ith coordinate of the array response vector $\mathbf{a}_{k,t}$ that corresponds to the tth path of the kth user signal

$$a_{k,t}[i] = \exp\left\{ j2\pi(i-1)\frac{\sin\theta_{k,t}q}{\lambda} \right\}, \qquad i = 1, \ldots, N, \tag{36}$$

where $\theta_{k,t}$ is the angle of arrival, λ is the carrier wavelength, and q is the inter-element spacing (in our studies we set $q \triangleq \lambda/2$).

We vectorize the $(L + J - 1) \times N$ space–time received data matrix $[\mathbf{r}_m^{(1)}, \mathbf{r}_m^{(2)}, \ldots, \mathbf{r}_m^{(N)}]$ to form the joint space–time data vector $\mathbf{r}_m$, which is a $(L + J - 1)N$-long column vector:

$$\mathbf{r}_m = \mathrm{Vec}\{[\mathbf{r}_m^{(1)}, \mathbf{r}_m^{(2)}, \ldots, \mathbf{r}_m^{(N)}]_{(L+J-1)\times N}\}. \tag{37}$$

The joint space–time RAKE filter for user 1 is $\mathbf{v}_1 \triangleq E_{b_1}\{\mathbf{r}_m b_1(m)\} = \mathrm{Vec}\{[\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \ldots, \mathbf{v}_{1,N}]\}$ ($E_{b_1}\{\cdot\}$ denotes statistical expectation with respect to $b_1(m)$), and $\mathbf{v}_{1,i} \triangleq \sum_{t=1}^{J} c_{1,t}\mathbf{s}_{1,t}a_{1,t}[i]$, $i = 1, 2, \ldots, N$. The MMSE–MVDR filter is built with constraint vector $\mathbf{v} = \mathbf{v}_1$, desired response $\mathbf{w}^H\mathbf{v}_1 = 1$, and autocorrelation matrix $\mathbf{R} = E\{\mathbf{r}_m\mathbf{r}_m^H\}$.

We choose $K = 20$, $N = 5$, $J = 3$ paths with independent zero-mean complex Gaussian fading coefficients of variance one (i.e., $E\{|c_{k,t}|^2\} = 1$) and Gold signatures with processing gain $L = 31$. The total SNR's (over the three paths) of the 19 interferers are set at $\mathrm{SNR}_{2\text{–}6} = 6$ dB, $\mathrm{SNR}_{7\text{–}8} = 7$ dB, $\mathrm{SNR}_{9\text{–}13} = 8$ dB, $\mathrm{SNR}_{14\text{–}15} = 9$ dB, and $\mathrm{SNR}_{16\text{–}20} = 10$ dB. The space–time product (filter length) equals $(L + J - 1)N = (31 + 2)5 = 165$. All experimental results that follow are averages over 100 different channel realizations and 10 independent data record generations per channel.

We first examine the performance of the AV filter estimator selection rules under the assumption that no info-bit pilot sequence is available. The data record size is set equal to $M = 230$ while the total SNR of the user of interest

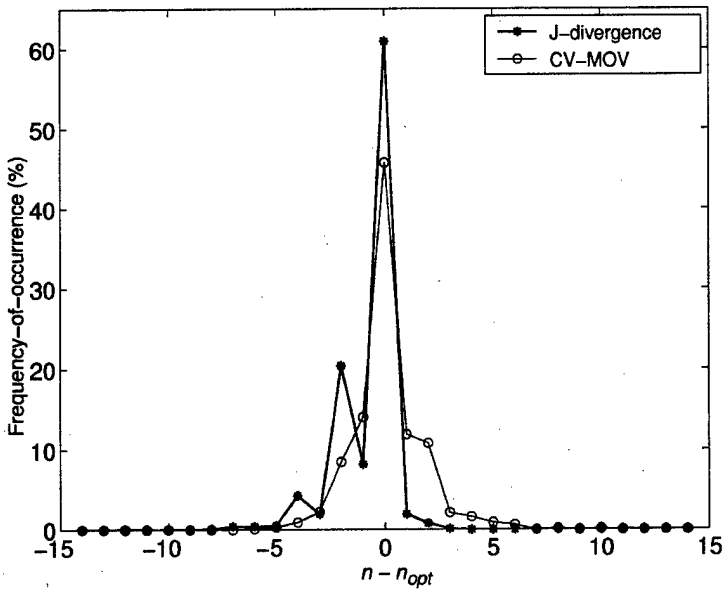


FIG. 9. Histogram of the two differences $n_1 - n_{\text{opt}}$ and $n_3 - n_{\text{opt}}$ where n_1 is the CV-MOV choice and n_3 is the blind J-divergence choice ($M = 230$, $\text{SNR}_1 = 8$ dB).

is set at $\text{SNR}_1 = 8$ dB. In Fig. 9, we plot the empirical pdf of the differences $(n_1 - n_{\text{opt}})$ and $(n_3 - n_{\text{opt}})$ where n_1 and n_3 denote selections according to Criteria 1 and 3, respectively, while n_{opt} denotes the "genie" maximum SINR optimum choice of the number of auxiliary vectors. We observe that both criteria provide a reliable estimate of the genie-assisted optimum number of auxiliary vectors.

The overall short data record adaptive filter performance is examined in Figs. 10 and 11. In Fig. 10, we plot the BER[5] of the AV filter estimators $\widehat{\mathbf{w}}_{n_1}(M)$ and $\widehat{\mathbf{w}}_{n_3}(M)$ as a function of the SNR of the user of interest for data records of size $M = 230$. The BER curve of the genie-assisted maximum SINR optimum filter choice $\widehat{\mathbf{w}}_{n_{\text{opt}}}(M)$ as well as the corresponding curves of the ideal MMSE–MVDR filter $\mathbf{w}_{\text{MMSE/MVDR}}$, the SMI filter estimator $\widehat{\mathbf{w}}_{\infty}(M)$, the S-T RAKE matched filter (MF) $\widehat{\mathbf{w}}_0(M) = \mathbf{v}_1$, and the multistage filter [9, 10] with the preferred number of stages[6] $l = 7$ are also included for comparison purposes. We observe that both $\widehat{\mathbf{w}}_{n_1}(M)$ and $\widehat{\mathbf{w}}_{n_3}(M)$ are very close to the genie optimum AV filter estimator choice and outperform significantly the SMI filter estimator, the multistage filter estimator, and the matched filter. We also observe that for moderate to high SNRs of the user of interest, the J-divergence selection rule is slightly superior to the CV-MOV selection rule. The opposite is true in the low SNR range. This is explained by the fact that the J-divergence approximation

[5] The BER of each filter under consideration is approximated by $Q\left(\sqrt{\text{SINR}_{\text{out}}}\right)$ [41], since the computational complexity of the BER expression for this antenna array CDMA system prohibits exact analytic evaluation.

[6] In [44], it is argued that $l = 7$ ($D = 8$ in the notation of [44]) stages are "nearly optimal over a wide range of loads and SNRs."

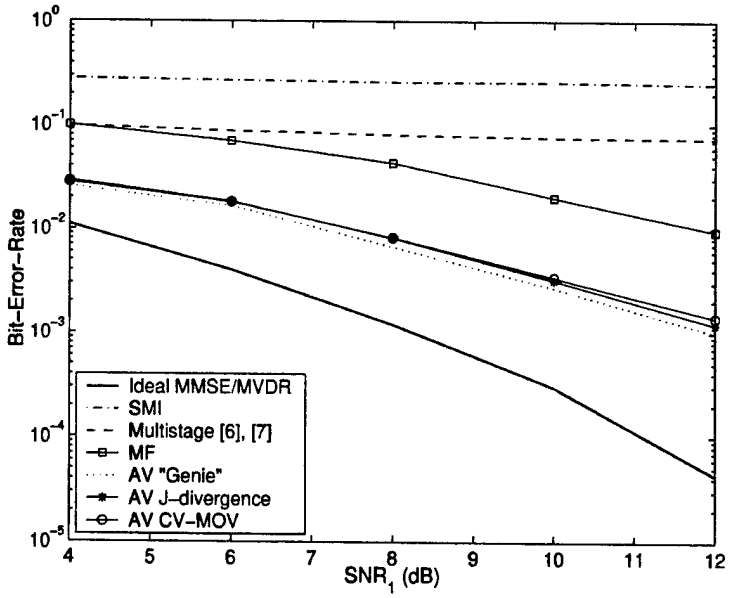


FIG. 10. BER versus SNR for the user signal of interest ($M = 230$).

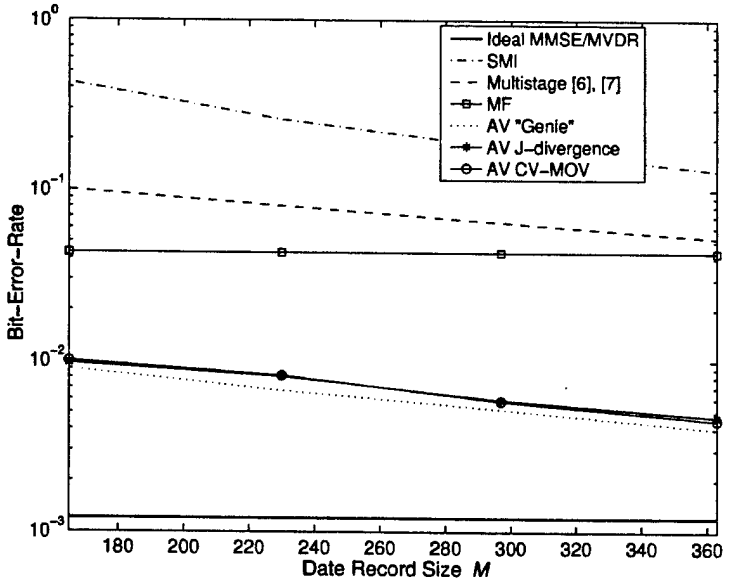


FIG. 11. BER versus data record size ($\text{SNR}_1 = 8$ dB).

$J(n) \approx J_B(n)$ used in Proposition 4 is less accurate for low filter output SINR values. On the other hand, for high filter output SINRs the discrimination capability of the CV-MOV rule is not as sharp.

Finally, Fig. 11 repeats the study of Fig. 10 as a function of the data record size. The SNR of the user of interest is fixed at 8 dB.

6. CONCLUDING REMARKS

In this article we relied strictly on statistical *conditional* optimization principles to derive an iterative algorithm that starts from the white-noise matched filter and converges to the MMSE–MVDR filter solution for any given positive definite input autocorrelation matrix. The conceptual simplicity of the employed conditional optimization criteria led to a computationally simple iteration step. We analyzed basic algorithmic properties and we established formal convergence to the MMSE–MVDR filter.

When the input autocorrelation matrix is substituted by a sample-average (positive definite) estimate, the algorithm generates a sequence of filter estimators that converges to the familiar sample matrix inversion unbiased estimator. The bias of the generated estimator sequence decreases rapidly to zero while the estimator covariance trace increases slowly from zero (for the initial, fixed-valued, matched-filter estimator) to the asymptotic covariance trace of SMI. Sequences of practical estimators that offer such exceptional control over favorable bias–covariance balance points are always a prime objective in the estimation theory literature. Indeed, for finite data record sets, members of the generated sequence of estimators were seen to outperform in MS estimation error LMS and RLS types, SMI and diagonally loaded SMI, and orthogonal multistage decomposition filter estimators. In addition, the troublesome, data-dependent tuning of the real-valued LMS learning gain parameter, the RLS initialization parameter, or the SMI diagonal loading parameter is replaced by an integer choice among the first several members of the estimator sequence. Two data-driven criteria were proposed for the identification of the best AV filter estimator in the sequence. The first criterion calls for the minimization of the cross-validated filter-estimator output variance. The second criterion calls for the maximization of the J-divergence of the filter-estimator output-conditional distributions. Simulation studies examined and compared the operational characteristics of the proposed selection methods. With respect to the relative merits of the minimum cross-validated output variance and the maximum output J-divergence selection rules, we observed that for moderate to high output SINRs the latter method appears superior to the former (for high SINRs the cross-validated minimum output variance rule is not as sharp in discrimination ability). In contrast, in low output SINR the J-divergence method is somewhat lacking in performance (technically, the approximation in Proposition 4 is less accurate for near 0 dB or lower output SINR values). As a final general comment, the use of a sufficiently long antenna array in combination with an "effective" interference suppressive filter can result in high output SINR which favors the J-divergence selection rule, even when the *transmitted* energy of the user of interest is much lower than that of the interferers.

The auxiliary-vector algorithm in Fig. 2 together with Criteria 1, 2, and 3 form a complete toolbox for state-of-the-art estimation of MMSE–MVDR filters. The developments are of particular interest in high-dimensional adaptive signal processing applications that rely on data records of limited size.

APPENDIX

Signature Assignment for the DS-CDMA Example of Sections 3 and 4

The matrix $\mathbf{S}_{32\times 13} = [\mathbf{s}_1\ \mathbf{s}_2\ \cdots\ \mathbf{s}_{13}]$ with columns the signature vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{13}$ is given below.

$$\mathbf{S} = \frac{1}{\sqrt{32}} \begin{bmatrix}
-1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & +1 \\
-1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 \\
-1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 \\
+1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 \\
-1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 \\
-1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & +1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & +1 & -1 \\
-1 & -1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 \\
-1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 \\
-1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 \\
+1 & +1 & +1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 \\
-1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 \\
+1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 \\
-1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 \\
-1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 \\
+1 & -1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 \\
-1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\
-1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 \\
-1 & +1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1
\end{bmatrix}. \quad \text{(A.1)}$$

Proof of Proposition 3. The quantities $(1/M_1)\sum_{m=1}^{M_1}(p_m^+ - \hat{\mu}^+(n))^2$ and $(1/M_2)\sum_{m=1}^{M_2}(p_m^- - \hat{\mu}^-(n))^2$ are the ML estimators for the variance of the filter output conditioned on $b_1 = +1$ and $b_1 = -1$, respectively [43, p. 179]. Both estimators are biased. (In fact, their unbiased counterparts that have multiplying factors $1/(M_1 - 1)$ and $1/(M_2 - 1)$ instead of $1/M_1$ and $1/M_2$, respectively, exhibit higher MSE.) The MSEs of the estimators of interest $\hat{\sigma}_1^2(n)$ and $\hat{\sigma}_2^2(n)$ are as follows:

$$\begin{aligned}
\mathrm{MSE}_{\hat{\sigma}_1^2(n)} &= E\left\{\left(\hat{\sigma}_1^2(n) - \sigma_{I+N}^2(n)\right)^2\right\} \\
&= E\left\{\left[\left(\frac{\sum_{m=1}^{M_1}(p_m^+ - \hat{\mu}^+(n))^2}{M_1 + M_2} - \frac{M_1\sigma_{I+N}^2(n)}{M_1 + M_2}\right)\right.\right. \\
&\qquad \left.\left. + \left(\frac{\sum_{m=1}^{M_2}(p_m^- - \hat{\mu}^-(n))^2}{M_1 + M_2} - \frac{M_2\sigma_{I+N}^2(n)}{M_1 + M_2}\right)\right]^2\right\}
\end{aligned}$$

$$= \frac{(2M_1-1)\sigma_{I+N}^4(n)}{(M_1+M_2)^2} + \frac{(2M_2-1)\sigma_{I+N}^4(n)}{(M_1+M_2)^2} + \frac{2\sigma_{I+N}^4(n)}{(M_1+M_2)^2}$$

$$= \frac{2\sigma_{I+N}^4(n)}{M_1+M_2} \tag{A.2}$$

and

$$\mathrm{MSE}_{\hat{\sigma}_2^2(n)} = E\left\{\left(\hat{\sigma}_2^2(n) - \sigma_{I+N}^2(n)\right)^2\right\} = \frac{[2(M_1+M_2)-1]\sigma_{I+N}^4(n)}{(M_1+M_2)^2}. \tag{A.3}$$

Thus,

$$\mathrm{MSE}_{\hat{\sigma}_1^2(n)} - \mathrm{MSE}_{\hat{\sigma}_2^2(n)} = \frac{\sigma_{I+N}^4(n)}{(M_1+M_2)^2} > 0. \quad \blacksquare \tag{A.4}$$

Proof of Theorem 2. $\hat{\mu}^+(n)$ and $\hat{\sigma}^{2+}(n)$ are independent random variables with distributions $\hat{\mu}^+(n) \sim \mathcal{N}(\mu(n), \sigma_{I+N}^2(n)/M_1)$ and $(M_1/\sigma_{I+N}^2(n))\hat{\sigma}^{2+}(n) \sim \chi_{M_1-1}^2$ [45] ($\chi_{M_1-1}^2$ denotes the chi-square distribution with (M_1-1) degrees of freedom). Similarly, $\hat{\mu}^-(n)$ and $\hat{\sigma}^{2-}(n)$ are independent random variables with distributions $\hat{\mu}^-(n) \sim \mathcal{N}(-\mu(n), \sigma_{I+N}^2(n)/M_2)$ and $(M_2/\sigma_{I+N}^2(n))\hat{\sigma}^{2-}(n) \sim \chi_{M_2-1}^2$. Furthermore, $(\hat{\mu}^+(n), \hat{\sigma}^{2+}(n))$ and $(\hat{\mu}^-(n), \hat{\sigma}^{2-}(n))$ are mutually independent because $(\hat{\mu}^+(n), \hat{\sigma}^{2+}(n))$ and $(\hat{\mu}^-(n), \hat{\sigma}^{2-}(n))$ are evaluated from two independent training sets.

Since $\hat{\mu}(n) = (M_1/M)\hat{\mu}^+(n) - (M_2/M)\hat{\mu}^-(n)$ and $\hat{\sigma}_1^2(n) = (M_1/M)\hat{\sigma}^{2+}(n) + (M_2/M)\hat{\sigma}^{2-}(n)$, $\hat{\mu}(n)$ and $\hat{\sigma}_1^2(n)$ are independent random variables with distributions $\hat{\mu}(n) \sim \mathcal{N}(\mu(n), \sigma_{I+N}^2(n)/M)$ and $(M/\sigma_{I+N}^2(n))\hat{\sigma}_1^2(n) \sim \chi_{M-2}^2$ where $M = M_1 + M_2$ is the size of the given information bit pilot sequence. Therefore,

$$E\{\hat{J}_{S,1}(n)\} = 4E\left\{\hat{\mu}^2(n)\right\}E\left\{\frac{1}{\hat{\sigma}_1^2(n)}\right\} = 4\left[\frac{\sigma_{I+N}^2(n)}{M} + \mu^2(n)\right]\frac{M}{\sigma_{I+N}^2(n)(M-4)}$$

$$= \frac{4\mu^2(n)}{\sigma_{I+N}^2(n)}\frac{M}{M-4} + \frac{4}{M-4}, \tag{A.5}$$

$$E\left\{\hat{J}_{S,1}^2(n)\right\} = 16E\left\{\hat{\mu}^4(n)\right\}E\left\{\frac{1}{\hat{\sigma}_1^4(n)}\right\} = 16\left[\mu^4(n) + \frac{6\mu^2(n)\sigma_{I+N}^2(n)}{M} + \frac{3\sigma_{I+N}^4(n)}{M^2}\right]$$

$$\times \frac{M^2}{\sigma_{I+N}^4(n)(M-4)(M-6)}. \tag{A.6}$$

The inverse moments $E\{1/\hat{\sigma}_1^2(n)\}$ and $E\{1/\hat{\sigma}_1^4(n)\}$ exist for $M > 4$ and $M > 6$, respectively, and are given by [46]

$$E\left\{\frac{1}{\hat{\sigma}_1^2(n)}\right\} = \frac{M}{\sigma_{I+N}^2(n)(M-4)}, \tag{A.7}$$

$$E\left\{\frac{1}{\hat{\sigma}_1^4(n)}\right\} = \frac{M^2}{\sigma_{I+N}^4(n)(M-4)(M-6)}. \tag{A.8}$$

Thus, the MSE of $\hat{J}_{S,1}(n)$ is

$$\mathrm{MSE}_1 \triangleq E\left\{\left[\hat{J}_{S,1}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\} = \frac{16}{(M-4)(M-6)}a - \frac{16}{M-4}b + \frac{16\mu^4(n)}{\sigma^4_{I+N}(n)}, \tag{A.9}$$

where

$$a \triangleq \left[\mu^4(n) + \frac{6\mu^2(n)\sigma^2_{I+N}(n)}{M} + \frac{3\sigma^4_{I+N}(n)}{M^2}\right]\frac{M^2}{\sigma^4_{I+N}(n)} \quad \text{and}$$
$$b \triangleq \left[\frac{\sigma^2_{I+N}(n)}{M} + \mu^2(n)\right]\frac{2M\mu^2(n)}{\sigma^4_{I+N}(n)}. \tag{A.10}$$

On the other hand, $\hat{\mu}(n)$ and $\hat{\sigma}^2_2(n)$ are independent random variables with distributions $\hat{\mu}(n) \sim \mathcal{N}(\mu(n), \sigma^2_{I+N}(n)/M)$ and $(M/\sigma^2_{I+N}(n))\hat{\sigma}^2_2(n) \sim \chi^2_{M-1}$, respectively. Therefore,

$$E\{\hat{J}_{S,2}(n)\} = 4\left[\frac{\sigma^2_{I+N}(n)}{M} + \mu^2(n)\right]\frac{M}{\sigma^2_{I+N}(n)(M-3)}$$
$$= \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\frac{M}{M-3} + \frac{4}{M-3}, \tag{A.11}$$

$$E\{\hat{J}^2_{S,2}(n)\} = 16\left[\mu^4(n) + \frac{6\mu^2(n)\sigma^2_{I+N}(n)}{M} + \frac{3\sigma^4_{I+N}(n)}{M^2}\right]$$
$$\times \frac{M^2}{\sigma^4_{I+N}(n)(M-3)(M-5)}. \tag{A.12}$$

Thus, the MSE of $\hat{J}_{S,2}(n)$ is

$$\mathrm{MSE}_2 \triangleq E\left\{\left[\hat{J}_{S,2}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\}$$
$$= \frac{16}{(M-3)(M-5)}a - \frac{16}{M-3}b + \frac{16\mu^4(n)}{\sigma^4_{I+N}(n)}, \tag{A.13}$$

where the first and second inverse moments $E\{1/\hat{\sigma}^2_2(n)\}$ and $E\{1/\hat{\sigma}^4_2(n)\}$ exist for $M > 3$ and $M > 5$, respectively, and a and b are given by (A.10).

From (A.9) and (A.13), assuming $M > 6$ which is the condition for all inverse moments to exist, we obtain

$$\mathrm{MSE}_1 - \mathrm{MSE}_2 = \frac{16a}{(M-4)(M-6)} - \frac{16a}{(M-3)(M-5)} + \frac{b}{M-3} - \frac{b}{M-4}$$
$$= \frac{16}{(M-3)(M-4)}\left[\frac{2M-9}{(M-5)(M-6)}a - b\right]$$
$$> \frac{16}{(M-3)(M-4)}\left[\frac{2(M-5)}{(M-5)(M-6)}a - b\right]$$
$$= \frac{32[6M\mu^4(n) + 5M\mu^2(n)\sigma^2_{I+N}(n) + 6\mu^2(n)\sigma^2_{I+N}(n) + 3\sigma^4_{I+N}(n)]}{(M-3)(M-4)(M-6)\sigma^4_{I+N}(n)}$$
$$> 0. \tag{A.14}$$

Proof of Proposition 4. Define $Y \triangleq |\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|$. The pdf of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_{I+N}(n)}\left[\exp\frac{-(y-\mu(n))^2}{2\sigma_{I+N}^2(n)} + \exp\frac{-(y+\mu(n))^2}{2\sigma_{I+N}^2(n)}\right]U(y), \tag{A.15}$$

where $U(y)$ is the unit step function. The mean of Y is

$$E\{Y\} = \int y f_Y(y)\,dy = \mu(n) + \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(n)}{2\sigma_{I+N}^2}\right) - 2\mu(n)Q\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right). \tag{A.16}$$

The series expansion of $Q(x)$ is [47]

$$Q(x) = \frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}\right)\left\{1 - \frac{1}{x^2} + \frac{1\cdot 3}{x^4} + \cdots + \frac{(-1)^n 1\cdot 3\cdots(2n-1)}{x^{2n}}\right\} + R_n, \tag{A.17}$$

where $R_n = (-1)^{n+1}1\cdot 3\cdots(2n+1)\int_x^\infty \frac{1}{\sqrt{2\pi}t^{2n+2}}\exp(-t^2/2)\,dt$ is the remainder which is always less (in absolute value) than the first neglected term. If $\mu(n)/\sigma_{I+N}(n) \gg 1$, (A.16) can be written as

$$\begin{aligned} E\{Y\} &= \mu(n)\left\{1 + \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}\mu(n)}\exp\left(-\frac{\mu^2(n)}{2\sigma_{I+N}^2(n)}\right) - \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}\mu(n)}\exp\left(-\frac{\mu^2(n)}{2\sigma_{I+N}^2(n)}\right)\left[1 - O\left[\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right)^{-2}\right]\right]\right\} \\ &= \mu(n)\left\{1 + \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(n)}{2\sigma_{I+N}^2(n)}\right)O\left[\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right)^{-3}\right]\right\} \\ &\approx \mu(n). \end{aligned} \tag{A.18}$$

The variance of Y is

$$\mathrm{Var}\{Y\} = E\{Y^2\} - E^2\{Y\} = \mu^2(n) + \sigma_{I+N}^2(n) - E^2\{Y\} \tag{A.19}$$

and using (A.18) we may approximate

$$\mathrm{Var}\{Y\} \approx \sigma_{I+N}^2(n). \tag{A.20}$$

Finally, from (A.18) and (A.20) we obtain

$$J_B(n) = \frac{4E^2\{Y\}}{\mathrm{Var}\{Y\}} \approx \frac{4\mu^2(n)}{\sigma_{I+N}^2(n)} = J(n). \tag{A.21}$$

REFERENCES

1. Capon, J., High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57** (1969), 1408–1418.
2. Owsley, N. L., A recent trend in adaptive spatial processing for sensor arrays: Constraint adaptation. In *Signal Processing* (Griffiths, J. W. R., *et al.*, Ed.). pp. 591–604, Academic Press, New York, 1973.
3. Nahi, N. E., *Estimation Theory and Applications.* Krieger, Huntington, NY, 1976.
4. Widrow, B., Mantey, P. E., Griffiths, L. J., and Goode, B. B., Adaptive antenna systems. *Proc. IEEE* **55** (1967), 2143–2158.
5. Reed, I. S., Mallet, J. D., and Brennan, L. E., Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerospace Electron. Systems* **10** (1974), 853–863.
6. Carlson, B. D., Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Trans. Aerospace Electron. Systems* **24** (1988), 397–401.
7. Plackett, R. L., Some theorems in least squares. *Biometrika* **37** (1950), 149.
8. Wiggins, R. A. and Robinson, E. A., Recursive solution to the multichannel filtering problem. *J. Geophys. Res.* **70** (1965), 1885–1891.
9. Goldstein, J. S., Reed, I. S., Zulch, P. A., and Melvin, W. L., A multistage STAP CFAR detection technique. In *Proc. IEEE Radar Conf.* pp. 111–116, Dallas, TX, May 1998.
10. Goldstein, J. S., Reed, I. S., and Scharf, L. L., A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Inform. Theory* **44** (1998), 2943–2959.
11. Griffiths, L. J. and Jim, C. W., An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagation* **30** (1982), 27–34.
12. Van Veen, B. D. and Roberts, R. A., Partially adaptive beamformer design via output power minimization. *IEEE Trans. Acoust. Speech Signal Process.* **35** (1987), 1524–1532.
13. Scharf, L. L. and Tufts, D. W., Reduced rank methods for modeling stationary signals. *IEEE Trans. Acoust. Speech Signal Process.* **35** (1987), 350–354.
14. Van Veen, B. D., Eigenstructure based partially adaptive array design. *IEEE Trans. Antennas Propagation* **36** (1988), 357–362.
15. Haimovich, A. M. and Bar-Ness, Y., An eigenanalysis interference canceler. *IEEE Trans. Signal Process.* **39** (1991), 76–84.
16. Byerly, K. A. and Roberts, R. A., Output power based partially adaptive array design. In *Proc. Asilomar Conf. Signals, Syst., Computers.* pp. 576–580, Pacific Grove, CA, 1989.
17. Goldstein, J. S. and Reed, I. S., Reduced-rank adaptive filtering. *IEEE Trans. Signal Process.* **45** (1997), 492–496.
18. Pados, D. A. and Batalama, S. N., Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers. *IEEE Trans. Commun.* **45** (1997), 1586–1594.
19. Kansal, A., Batalama, S. N., and Pados, D. A., Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels. *IEEE J. Select. Areas Commun.* **16** (1998), 1765–1773.
20. Batalama, S. N., Medley, M. J., and Pados, D. A., Robust adaptive recovery of spread-spectrum signals with short data records. *IEEE Trans. Commun.* **48** (2000), 1725–1731.
21. Pados, D. A. and Karystinos, G. N., An iterative algorithm for the computation of the MVDR filter. *IEEE Trans. Signal Process.* **49** (2001), 290–300.
22. Pados, D. A. and Batalama, S. N., Joint space–time auxiliary-vector filtering for DS/CDMA systems with antenna arrays. In *Proc. Conf. on Inform. Sci. and Systems.* Vol. II, pp. 1007–1013, Princeton University, Princeton, NJ, March 1998.
23. Pados, D. A., Michels, J. H., Tsao, T., and Wicks, M. C., Joint domain space–time adaptive processing with small training data sets. In *Proc. IEEE Radar Conf.* pp. 99–104, Dallas, TX, May 1998.
24. Pados, D. A. and Batalama, S. N., Joint space–time auxiliary-vector filtering for DS/CDMA systems with antenna arrays. *IEEE Trans. Commun.* **47** (1999), 1406–1415.
25. Rapajic, P. B. and Vucetic, B. S., Adaptive receiver structures for asynchronous CDMA systems. *IEEE J. Select. Areas Commun.* **12** (1994), 685–697.
26. Madhow, U. and Honig, M. L., MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. Commun.* **42** (1994), 3178–3188.
27. Miller, S. L., An adaptive DS-CDMA receiver for multiuser interference rejection. *IEEE Trans. Commun.* **43** (1995), 1746–1755.

28. Honig, M. L., Madhow, U., and Verdu, S., Blind adaptive multiuser detection. *IEEE Trans. Inform. Theory* **41** (1995), 944–960.
29. Pateros, C. N. and Saulnier, G. J., An adaptive correlator receiver for direct-sequence spread-spectrum communication. *IEEE Trans. Commun.* **44** (1996), 1543–1552.
30. Richmond, C. D., Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance. *IEEE Trans. Signal Process.* **44** (1996), 305–315.
31. Dykstra, R. L., Establishing the positive definiteness of the sample covariance matrix. *Ann. Math. Statist.* **41** (1970), 2153–2154.
32. Steinhardt, A. O., Adaptive multisensor detection and estimation. In *Adaptive Radar Detection and Estimation* (Haykin, S. and Steinhardt, A. O., Eds.). Chap. 3, Wiley, New York, 1992.
33. Richmond, C. D., PDF's, confidence regions, and relevant statistics for a class of sample covariance-based array processors. *IEEE Trans. Signal Process* **44** (1996), 1779–1793.
34. Steinhardt, A. O., The PDF of adaptive beamforming weights. *IEEE Trans. Signal Process.* **39** (1991), 1232–1235.
35. Godara, L. C. and Cantoni, A., Analysis of constrained LMS algorithm with application to adaptive beamforming using perturbation sequences. *IEEE Trans. Antennas Propagation* **34** (1986), 368–379.
36. Solo, V., The limiting behavior of LMS. *IEEE Trans. Acoust., Speech, Signal Process.* **37** (1989), 1909–1922.
37. Cioffi, J. M. and Kailath, T., Fast recursive-least-squares transversal filters for adaptive filtering. *IEEE Trans. Acoust., Speech, Signal Process.* **32** (1984), 304–337.
38. Liu, T.-C. and Van Veen, B., A modular structure for implementation of linearly constrained minimum variance beamformers. *IEEE Trans. Signal Process.* **39** (1991), 2343–2346.
39. Qian, H. and Batalama, S. N., Data-record-based criteria for the selection of an auxiliary-vector estimator of the MVDR filter. In *Proc. Asilomar Conf. Signals, Systems., Computers.* pp. 802–807, Pacific Grove, CA, Oct. 2000.
40. Rao, C. R., *Handbook of Statistics 9.* Elsevier, New York, 1993.
41. Poor, H. V. and Verdu, S., Probability of error in MMSE multiuser detection. *IEEE Trans. Inform. Theory* **43** (1997), 858–871.
42. Cover, T. M. and Thomas, J. A., *Elements of Information Theory.* Wiley, New York, 1991.
43. Poor, H. V., *An Introduction to Signal Detection and Estimation.* Springer-Verlag, New York, 1994.
44. Honig, M. L. and Xiao, W., Performance of reduced-rank linear interference suppression. *IEEE Trans. Inform. Theory* **47** (2001), 1928–1946.
45. DeGroot, M. H., *Probability and Statistics.* Addison-Wesley, Reading, MA, 1984.
46. Jones, M. C., Expressions for inverse moments of positive quadratic forms in normal variables. *Austral. J. Statist.* **28** (1986), 242–250.
47. Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions.* Dover, New York, 1965.

GEORGE N. KARYSTINOS was born in Athens, Greece, on April 12, 1974. He received the Diploma degree in computer engineering and science from the University of Patras, Patras, Greece, in 1997. Since 1998, he has been a research assistant with the Communications and Signals Group in the Department of Electrical Engineering at the State University of New York at Buffalo where he is currently a Ph.D. candidate. His research interests are in the areas of wireless multiple access communications and statistical signal processing. Mr. Karystinos is a student member of the IEEE Communications Society, Signal Processing Society, and Information Theory Society and a member of Eta Kappa Nu.

HAOLI QIAN was born in Sichuan, P.R. China, on June 2, 1977. He received the B.Sc. in electronic engineering from Tsinghua University, Beijing, P.R. China, in 1998 and the M.Sc. in electrical engineering from the State University of New York at Buffalo in 2000. Since 2000, he has been a research assistant with the Communications and Signals Group, Department of Electrical Engineering, State University of New York at Buffalo, where he is currently working toward the Ph.D. degree. His research interests are in the areas of wireless communications and

statistical signal processing. Mr. Qian is a student member of the IEEE Communications Society and Information Theory Society.

MICHAEL J. MEDLEY received his B.S., M.S., and Ph.D. in electrical engineering from Rensselaer Polytechnic Institute in 1990, 1991, and 1995, respectively. Since 1991, he has been a research engineer for the United States Air Force Research Laboratory in Rome, NY where he has been actively involved in HF channel modeling, interference suppression in spread spectrum communications, spread spectrum waveform design and multicarrier modulation. He is currently a member of the Smart Networking Radio research group developing advanced modulation and signal processing technology for next-generation military radios. His research interests include transform domain signal processing and adaptive filtering, multicarrier modulation and multiple access communication techniques. Dr. Medley is a member of the IEEE Communications Society, Eta Kappa Nu, and Tau Beta Pi.

STELLA N. BATALAMA received the Diploma degree in computer engineering and science from the University of Patras, Greece in 1989 and the Ph.D. in electrical engineering from the University of Virginia, Charlottesville, VA, in 1994. From 1989 to 1990 she was with the Computer Technology Institute, Patras, Greece. From 1990 to 1994 she was a research assistant in the Communication Systems Laboratory, Department of Electrical Engineering, University of Virginia. In 1995 she joined the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY, where she is presently an associate professor. During the summers of 1997–2002 she was Visiting Faculty in the U.S. Air Force Research Laboratory, Rome, NY. Her research interests include small sample support adaptive filtering and receiver design, adaptive multiuser detection, robust spread-spectrum communications, supervised and unsupervised optimization, and distributed detection. Dr. Batalama is currently an associate editor for the *IEEE Communications Letters* and the *IEEE Transactions on Communications*.

Data-record-based Criteria for the Selection of an Auxiliary-Vector Estimator of the MMSE/MVDR Filter

Haoli Qian, *Student Member, IEEE,* and Stella N. Batalama, *Member, IEEE*

***Abstract*— When the auxiliary-vector (AV) filter generation algorithm utilizes sample average estimated input data statistics, it provides a sequence of estimates of the ideal minimum-mean-square-error (MMSE) or minimum-variance-distortionless-response (MVDR) filter for the given signal processing/receiver design application. Evidently, early non-asymptotic elements of the sequence offer favorable bias/variance balance characteristics and outperform in mean-square filter estimation error the unbiased sample-matrix-inversion (SMI) estimator as well as the (constraint)-LMS, RLS, "multistage nested Wiener filter," and diagonally-loaded (DL)-SMI filter estimators. Selecting the most successful (in some appropriate sense) AV-filter estimator in the sequence for a given data record is a critical problem that has not been addressed so far. In this paper we deal exactly with this problem and we propose two data driven selection criteria. The first criterion minimizes the cross-validated sample average variance of the AV-filter output and can be applied to general filter estimation problems; the second criterion maximizes the estimated J-divergence of the AV-filter-output conditional distributions and is tailored to binary phase-shift-keying (PSK)-type detection problems.**

***Index Terms*— Adaptive filters, antenna arrays, code division multiaccess, cross-validation, interference suppression, J-divergence, low sample support, short-data-record estimation.**

I. Introduction and Background

In this paper, we deal with the broad and challenging problem of designing adaptive receivers from a limited data record of size commensurate to the channel coherence time. The coherence time of the channel, which is reciprocal to the channel rate-of-change, will determine the size of the data record (block) available for receiver adaptation and redesign (the shorter the coherence time of the channel, the smaller the available data record). For presentation purposes we consider a simple block fading channel model that assumes fixed fading over the symbols in the block and independent fading across blocks. After receiver adaptation based on a specific data record (block) is completed, we utilize the obtained receiver/filter estimate to detect the information bits contained in that same block. Our study focuses on the auxiliary-vector (AV) algorithm [1]-[3] that generates a sequence of linear filters/receivers that we call AV-filters/receivers. Members of this sequence -when evaluated (estimated) from a data record of limited size- were seen to exhibit superior short-data-record adaptive filtering performance. Our objective in this paper is to develop data driven criteria for the selection of the most successful AV-filter estimator in the sequence for a given record of input observations of size commensurate to the rate-of-change of the receiver operational environment.

This work was supported by the National Science Foundation under Grant ECS-0073660 and Grant CCR-0129903 and by the Air Force Office of Scientific Research under Grant F49620-01-1-0176. This paper was presented in part at the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2000.

The authors are with the Department of Electrical Engineering, 332 Bonner Hall, State University of New York at Buffalo, Buffalo, NY 14260 (e-mail: haoqian@eng.buffalo.edu; batalama@eng.buffalo.edu).

The auxiliary-vector (AV) filter generation algorithm given in [1]–[3] is a statistical optimization procedure that produces a sequence of linear filters (AV-filters) that converges to the MMSE/MVDR solution under an ideal setup (perfectly known input data autocovariance matrix). The developments in [1]–[3] can be viewed as an extension of the work in [4]–[8] to non-orthogonal multiple AV MMSE/MVDR filter synthesis. The AV filter generation algorithm exhibits the following key attributes: *(i)* Utilizes *non-orthogonal* auxiliary vectors [1]–[3]; *(ii)* the auxiliary vectors are selected according to the *maximum cross-correlation* (MCC) criterion [5]; *(iii)* the filter sequence is created based on statistical *conditional* optimization [6]–[8]; and *(iv)* each (non-asymptotic) filter in the sequence has a form that does not involve any explicit or implicit matrix inversion operation (in contrast for example to reduced-rank filtering methods such as in [9],[10]).

The AV algorithm is summarized in Fig. 1. In terms of notation, $\mathbf{w} \in \mathbb{C}^L$ represents a complex L-tap linear filter and $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$ is the input data autocorrelation matrix ($E\{\cdot\}$ denotes statistical expectation, H denotes the Hermitian operation, and $\mathbf{r} \in \mathbb{C}^L$ is the random, zero mean without loss of generality, complex input vector of dimension L); $\mathbf{v}$ is the input signal vector direction to be protected during filtering (for general MMSE filters $\mathbf{v}$ is set equal to the statistical cross-correlation between the desired scalar filter output $b \in \mathbb{C}$ and the input vector $\mathbf{r}$, $\mathbf{v} = E\{\mathbf{r}b^*\}$). For presentation purposes, in Fig. 1 we assume that the filtering operation is supposed to be "distortionless" in the $\mathbf{v}$ vector direction, i.e. $\mathbf{w}^H\mathbf{v} = 1$ (the developments can be generalized in a straightforward manner to cover filter constraints of the form $\mathbf{w}^H\mathbf{v} = \rho$ for any complex scalar ρ [2]). The vectors $\mathbf{g}_d$, $d = 1, 2, \ldots$, are referred to as "auxiliary vectors." The auxiliary vector $\mathbf{g}_d$ is required to be orthogonal *only* to the constraint vector $\mathbf{v}$ itself and is not necessarily orthogonal to the earlier auxiliary vectors $\mathbf{g}_i, i = 1, 2, \ldots, d-1$; the auxiliary vector $\mathbf{g}_d$ is chosen to maximize the magnitude of the statistical cross-correlation between the previous filter output and the projection of the input data onto the auxiliary vector

itself, i.e. $\mathbf{g}_d = \arg\max_{\mathbf{g}} |\mathbf{w}_{d-1}^H \mathbf{R}\mathbf{g}|$ subject to $\mathbf{g}_d^H \mathbf{v} = 0$ and $\mathbf{g}_d^H \mathbf{g}_d = 1$. The corresponding scalar AV weight μ_d is chosen to minimize the new filter output variance, i.e. $\mu_d = \arg\min_{\mu} \mathbf{w}_d^H \mathbf{R}\mathbf{w}_d$.[1] We emphasize that by relaxing the orthogonality condition among auxiliary vectors, the length of the generated sequence of filters is not limited by the vector space dimension as in the case where orthogonality between auxiliary vectors is maintained [6]–[8]. The AV algorithm in Fig. 1 generates a sequence of filters $\mathbf{w}_d = \mathbf{w}_0 - \sum_{i=1}^{d} \mu_i \mathbf{g}_i$, $d = 1, 2, \ldots$. Formal theoretical analysis of the sequence of auxiliary-vector filters was pursued in [2].

For a given constraint (or cross-correlation) vector $\mathbf{v}$ and perfectly known input autocorrelation matrix $\mathbf{R}$, the sequence of ideal AV filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$ converges to the ideal MVDR (MMSE) optimum filter

$$\mathbf{w}_{MMSE/MVDR} = \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H \mathbf{R}^{-1}\mathbf{v}}. \quad (1)$$

We recall that the filter in (1) minimizes $\mathbf{w}^H \mathbf{R}\mathbf{w}$ and simultaneously satisfies $\mathbf{w}^H \mathbf{v} = 1$. Not withstanding the satisfying convergence properties of the ideal AV filter sequence, the main motivation for the development of the AV filtering framework is improved adaptive receiver performance when receiver adaptation is based on a short record of N input data. When the input autocorrelation matrix $\mathbf{R}$ is estimated by averaging the outer products of N received vectors, $\widehat{\mathbf{R}}(N) \triangleq \frac{1}{N}\sum_{n=1}^{N} \mathbf{r}_n \mathbf{r}_n^H$, the MMSE/MVDR filter estimator obtained by using the sample-average estimate $\widehat{\mathbf{R}}(N)$ in place of $\mathbf{R}$ in (1) is known as the sample-matrix-inversion (SMI) filter, $\widehat{\mathbf{w}}_{SMI}(N) \triangleq \frac{\widehat{\mathbf{R}}(N)^{-1}\mathbf{v}}{\mathbf{v}^H \widehat{\mathbf{R}}^{-1}(N)\mathbf{v}}$. The filter $\widehat{\mathbf{w}}_{SMI}(N)$ exhibits all the disadvantages associated with the inversion of a matrix that is estimated from a finite-size data record (for example, the inverse exists w.p. 1 only if the size of the data record N is greater than or equal to L) and data record sizes many times the filter length L are necessary for $\widehat{\mathbf{w}}_{SMI}(N)$ to approach reasonably well the performance characteristics of the ideal $\mathbf{w}_{MMSE/MVDR}$ filter. Unfortunately, in many communications systems with typical input data dimension values and data transmission rates, the data record (block) size (that fits within a channel coherence period) is relatively small for $\widehat{\mathbf{w}}_{SMI}(N)$ to perform well. In contrast to $\widehat{\mathbf{w}}_{SMI}(N)$ and other potential reduced-rank filter estimators such as the "multistage nested Wiener filters" in [9], [10], the AV-filter estimators obtained by utilizing the sample-average estimate $\widehat{\mathbf{R}}(N)$ in place of $\mathbf{R}$ in Fig. 1 are not the result of any form of *direct or indirect* matrix inversion operation. As illustrated in [2], for a fixed finite data record of size N, the AV-filter sequence $\{\widehat{\mathbf{w}}_d(N)\}_d$ based on the N-point autocorrelation matrix estimate $\widehat{\mathbf{R}}(N)$ can be viewed not only as a sequence of estimators of the corresponding ideal AV filters but also as a sequence of estimators of the ideal MMSE/MVDR filter that has varying bias versus variance characteristics and converges to $\widehat{\mathbf{w}}_{SMI}(N)$ (if the latter exists). For small N, the early non-asymptotic elements of the sequence of AV estimators offer favorable bias/variance balance and outperform in mean-square filter estimation error $\widehat{\mathbf{w}}_{SMI}(N)$, (constraint)-LMS, RLS, multistage nested Wiener filter [9], [10], and diagonally-loaded (DL)-SMI estimators. In the context of digital wireless communication receivers, the latter translates to superior bit-error-rate performance under short data record receiver adaptation.

We conclude this introductory section and we motivate the developments that follow with an illustration of the significance of the AV parameter d (that is the exact number of auxiliary vectors used in building the AV-filter estimator) and the dependence of the proper choice of d on the true signal model, the constraint vector direction, the data record size, as well as the specific data record realization. We draw a signal model example from the direct-sequence code-division-multiple-access (DS/CDMA) literature and we consider a system with $K = 20$ users with Gold signatures of length $T = 31$ and multipath fading reception by a narrowband linear antenna array with $M = 5$ elements (all elements are assumed to experience identical fading). We assume that each signal experiences $J = 3$ chip-interval-spaced paths. The path fading coefficients are modeled as independent identically distributed (i.i.d.) zero-mean complex Gaussian random variables. The directions of arrival for all user paths are i.i.d. uniformly distributed in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. We pick a "user of interest," say user *0*, and we fix the total SNR's (over the three paths) of the 19 interferers at $\text{SNR}_{1-5} = 6\text{dB}$, $\text{SNR}_{6-7} = 7\text{dB}$, $\text{SNR}_{8-12} = 8\text{dB}$, $\text{SNR}_{13-14} = 9\text{dB}$, $\text{SNR}_{15-19} = 10\text{dB}$. The space-time product for this communication system equals $(T+J-1)M = 165$. After conventional carrier demodulation and chip-matched filtering and sampling at the chip rate over a multipath extended symbol interval of $T + J - 1$ chips, the $(T + J - 1) \times M$ data samples ($T + J - 1$ data samples from each antenna element) are organized in the form of an $L = 165$-long joint space-time received data vector $\mathbf{r}$. The input vector direction to be protected during MMSE/MVDR space-time filtering is the effective space-time signature of user *0* or in other words the cross-correlation between the input data vector $\mathbf{r}$ and the desired user information bit b_0, $\mathbf{v}_0 \triangleq E_{b_0}\{\mathbf{r}b_0^*\}$.

With this setup, we conduct the following experiment. We generate an input data record of size N and we estimate the elements of the AV-filter sequence. Then, for each AV-filter estimate we evaluate the output SINR using the ideal (true) input statistics and we record the *"genie" assisted* optimum choice of the number d that corresponds to the AV-filter estimate in the sequence with the maximum SINR value. We emphasize that the optimum number d is a random variable rather than a deterministic function of N; for a given data record realization of size N, the "genie" assisted choice of d optimizes (in the maximum output SINR sense) the filter estimate that is evaluated based on this specific data record realization. We recall that in practice the data record size N is controlled by system specifics and the coherence time of the channel (for example with CDMA chip rates at 1.25MHz, processing gain 32, carrier frequency 900MHz, and vehicle speed 45mph, the channel coherence time is approximately 7ms [11] and the fading channel fluctuates decisively at least every 275

[1] Since $\mu_d \mathbf{g}_d$ is independent of the norm of $\mathbf{g}_d$, then so is $\mathbf{w}_d$. Based on this observation, in Fig. 1 we dropped the unnecessary normalization of the auxiliary vectors.

data symbols [7]). We repeat our experiment 1,000 times (100 different channel realizations and 10 independent data record generations per channel) and we show the empirical probability density function of the random variable d in Fig. 2. In Fig. 2(a), the data record size is set equal to $N = 230$ while the SNR of the user of interest is set at $SNR_0 = 8dB$. Figs. 2(b) and 2(c) show an example of how this empirical density changes when we increase the sample support ($N = 363$) or decrease the user SNR ($SNR_0 = 4dB$), respectively.

While it is clear that the above "genie" assisted procedure for the selection of d cannot be used in practice (due to the unknown ideal input statistics), these studies illustrate the importance of identifying a "good" d in some appropriate "goodness" sense. Accomplishing such a task is equivalent to developing a data-driven, fully adaptive AV filter receiver that adapts even its own structure (i.e. number of auxiliary vectors used) for any given input data record.

Section II deals exactly with this problem, that is the selection of the best (in some appropriate sense) AV filter from the sequence of AV filter estimates based on a given finite data record. Two data driven selection criteria are proposed and analyzed. The first selection rule is motivated by the asymptotic minimum output variance property of the AV filters and can be applied to general filter estimation problems. The second rule is related to the objective of achieving maximum stochastic distance between the AV filter output conditional distributions and is tailored to applications that can be formulated as antipodal hypothesis testing problems (for example, detection of PSK-type signals processed by arbitrary channels and corrupted by AWGN). In particular, for a given data record, the first criterion selects the AV-filter estimator (or, equivalently, the number of auxiliary vectors used in forming the AV estimator) that minimizes the *cross-validated* average filter output variance (energy). The second criterion selects the number of auxiliary vectors that maximizes the estimated *J-divergence* of the filter output conditional distributions. Both supervised and unsupervised (blind) implementations of the J-divergence criterion are pursued and analyzed. Simulation studies and comparisons are presented in Section III and some conclusions are drawn in Section IV.

II. AV-Filter Selection Rules

A. Cross-Validated Minimum-Output-Variance Rule (CV-MOV)

The CV-MOV rule is motivated by the fact that minimization of the output variance of filters that are constrained to be distortionless in the vector direction of a signal of interest is equivalent to maximization of the output SINR. Cross-validation is a well-known statistical method [12]. It is used here to select the filter parameter of interest (number of AV's d) that minimizes the output variance which is estimated based on observations (training data) that have not been used in the process of building the filter itself. A particular case of this general method used in this paper is the "leave-one-out" method. The following criterion outlines the CV-MOV AV-filter selection process.

Criterion 1: For a given data record of size N, the cross-validated minimum-output-variance AV-filter selection rule chooses the AV-filter estimator $\widehat{\mathbf{w}}_{\hat{d}_1}$ that minimizes the cross-validated sample average output variance, i.e.

$$\hat{d}_1 = \arg\min_d \left\{ \sum_{i=1}^{N} \widehat{\mathbf{w}}_{d(N\backslash i)}^H \mathbf{r}_i \mathbf{r}_i^H \widehat{\mathbf{w}}_{d(N\backslash i)} \right\} \qquad (2)$$

where the subscript $(N\backslash i)$ identifies the AV-filter estimator that is evaluated from the available data record after removing the i-th sample. ■

It is straightforward to show that the cross-validated output variance estimate of the AV-filter estimator that utilizes d auxiliary vectors converges to the output variance of the corresponding ideal AV-filter $\mathbf{w}_d$ as the data record size increases, i.e. $\frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{w}}_{d(N\backslash i)}^H \mathbf{r}_i \mathbf{r}_i^H \widehat{\mathbf{w}}_{d(N\backslash i)} \overset{N\to\infty}{\longrightarrow} \mathbf{w}_d^H \mathbf{R} \mathbf{w}_d$. In addition, it may be important to emphasize the need for invoking the cross-validation technique for the evaluation of the sample-average output variance. If sample-average evaluation using *all data* were attempted, then the selection rule would take the form $\arg\min_d \widehat{\mathbf{w}}_d^H \widehat{\mathbf{R}} \widehat{\mathbf{w}}_d$, where $\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{r}_n \mathbf{r}_n^H$ and $\widehat{\mathbf{w}}_d$ is obtained by substituting $\widehat{\mathbf{R}}$ in place of $\mathbf{R}$ in Fig. 1. Such minimization, however, would result in $d = \infty$ since we know that $\widehat{\mathbf{w}}_d^H \widehat{\mathbf{R}} \widehat{\mathbf{w}}_d \overset{d\to\infty}{\longrightarrow} \widehat{\mathbf{w}}_{SMI}^H \widehat{\mathbf{R}} \widehat{\mathbf{w}}_{SMI}$ and $\widehat{\mathbf{w}}_{SMI}$ is the filter that has minimum sample-output-variance $\widehat{\mathbf{w}}_{SMI}^H \widehat{\mathbf{R}} \widehat{\mathbf{w}}_{SMI}$ (but, of course, the true output variance of $\widehat{\mathbf{w}}_{SMI}$, $\widehat{\mathbf{w}}_{SMI}^H \mathbf{R} \widehat{\mathbf{w}}_{SMI}$, is not minimum).

B. Output J-divergence Rule

While the CV-MOV criterion presented in the previous section can be applied to general filter estimation problems, in this section we present and analyze an alternative criterion that is tailored to applications that can be formulated as binary hypothesis testing problems on AV-filtered data.

We recall that for any scalar binary hypothesis testing problem characterized by the detector input conditional distributions f_0 and f_1 under hypothesis H_0 and H_1, respectively, the J-divergence distance between f_0 and f_1 is defined as the sum of the Kullback-Leibler (K-L) distances between f_0 and f_1 [13]

$$J(f_0, f_1) \triangleq D(f_1, f_0) + D(f_0, f_1) \qquad (3)$$

where the K-L distance of f_1 from f_0 is defined as $D(f_1, f_0) \triangleq \int_{-\infty}^{\infty} f_1(x) \log \frac{f_1(x)}{f_0(x)} dx$.

Our choice of the output J-divergence as one of the underlying rules for the selection of the AV filter is motivated by the fact that the probability of error of the optimum Bayesian detector for any binary hypothesis testing problem is lower bounded by a monotonically decreasing function of the J-divergence between the conditional distributions of the detector input [13]. For scalar input, the lower bound is given by

$$P_e \geq \pi_0 \pi_1 exp\{-J(f_0, f_1)/2\} \qquad (4)$$

where π_0, π_1 are the a priori probabilities of H_0 and H_1, respectively. Thus, the larger the J-divergence the smaller the lower bound on the probability of error. When the conditional distributions under H_0 and H_1 are Gaussian with the same variance, the probability of error is an exact monotonically decreasing function of the J-divergence. In particular, let

the conditional distributions of the detector input be $f_0 \sim \mathcal{N}\left[\mu_0, \sigma^2\right]$ and $f_1 \sim \mathcal{N}\left[\mu_1, \sigma^2\right]$ where μ_0 and μ_1 are the conditional means under hypothesis H_0 and H_1, respectively, and σ^2 is the conditional variance under either hypothesis. Then,

$$J(f_0, f_1) = (\mu_1 - \mu_0)^2/\sigma^2 \tag{5}$$

and the probability of error of the optimum detector is equal to

$$\begin{aligned} P_e &= \pi_1 Q\left(\frac{\sqrt{J(f_0,f_1)}}{2} + \frac{\ln(\pi_1/\pi_0)}{\sqrt{J(f_0,f_1)}}\right) \\ &\quad + \pi_0 Q\left(\frac{\sqrt{J(f_0,f_1)}}{2} - \frac{\ln(\pi_1/\pi_0)}{\sqrt{J(f_0,f_1)}}\right) \end{aligned} \tag{6}$$

where $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$. We can show that the right hand side of (6) is a monotonically decreasing function of $J(f_0, f_1)$ for any π_0 and π_1. Thus, the larger the J-divergence the easier the detection problem or, equivalently, maximization of the J-divergence implies minimization of the probability of error. Due to the above properties and its relationship with the probability of error of the optimum detector, J-divergence has been extensively used in the detection literature as a hypothesis discriminant function.

In the AV filtering context, we denote the conditional distributions of the AV scalar filter output under H_0 and H_1 by $f_{0,d}(\cdot)$ and $f_{1,d}(\cdot)$ where the index d indicates the dependence of the distributions on the AV filter parameter d. Then, the J-divergence between $f_{0,d}(\cdot)$ and $f_{1,d}(\cdot)$ is also a function of d and in the rest of the paper will be denoted as $J(d)$. The following criterion outlines the J-divergence AV-filter selection process.

Criterion 2: For a given data record of size N, the J-divergence AV-filter selection rule chooses the AV-filter estimator $\widehat{\mathbf{w}}_{\widehat{d}_J}$ that maximizes the estimated J-divergence $\widehat{J}(d)$ between the AV-filter output conditional distributions, i.e.

$$\widehat{d}_J = \arg\max_d \left\{\widehat{J}(d)\right\}. \tag{7}$$

■

To reach more specific solutions, we consider scalar binary hypothesis testing problems that result from AV-filtered versions of binary antipodal signals. Along these lines, let the received signal vector be of the general form

$$\mathbf{r} = b\sqrt{E}\mathbf{v} + \mathbf{n} \tag{8}$$

where $\mathbf{v} \in \mathbb{C}^L$ is a known deterministic signal vector, $E > 0$ represents the unknown energy scalar, $\mathbf{n} \in \mathbb{C}^L$ is a comprehensive zero-mean disturbance vector (i.e., it may incorporate inter-symbol-interference (ISI), multiple-access-interference (MAI) and additive noise effects), and b is $+1$ or -1 with equal probability. Then, the decision on H_0 $(b = -1)$ or H_1 $(b = +1)$ is based on the real part of the AV filter output

$$p \triangleq Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right] \tag{9}$$

where $\widehat{\mathbf{w}}_d$ is the AV-filter estimator that utilizes d auxiliary vectors. To the extent that the conditional distributions of the filter output p under H_0 and H_1 can be approximated by Gaussian distributions, we have $f_{0,d} \sim \mathcal{N}\left[-\mu(d), \sigma^2(d)\right]$ and $f_{1,d} \sim \mathcal{N}\left[\mu(d), \sigma^2(d)\right]$, where $-\mu(d)$ and $\mu(d)$ are the conditional means under hypothesis H_0 and H_1, respectively, and $\sigma^2(d)$ is the conditional variance under either hypothesis. This Gaussian approximation of the actual conditional distributions of the filter output can be considered safe for "effective" interference suppressive filters, as argued, for example, in [14] for the MMSE/MVDR linear filter operating on CDMA signals. Such a case study will be considered later in Section III where the effect of the above approximation on the performance of the output J-divergence selection rule will be examined through comparisons with the optimum "genie" assisted selection rule. In the rest of this section we develop a supervised and a blind implementation of Criterion 2 for the selection of the best AV-filter in binary hypothesis testing problems that result from AV-filtered versions of binary antipodal signals.

1) Supervised Output J-divergence Rule

Exploiting the symmetry of the conditional densities of p, $f_{0,d}(\cdot)$ and $f_{1,d}(\cdot)$, we can show in a straightforward manner that

$$J(d) = \frac{4E^2\left\{bRe\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right\}}{Var\left\{bRe\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right\}} \tag{10}$$

$$= \frac{4\left[\sum\limits_{j=+1,-1} E\left\{bRe\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right] | b = j\right\} Pr(b = j)\right]^2}{\sum\limits_{j=+1,-1} Var\left\{bRe\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right] | b = j\right\} Pr(b = j)} \tag{11}$$

where $Var(\cdot)$ and $Pr(\cdot)$ denote variance and probability, respectively. Assuming availability of a pilot bit sequence $\{b(i)\}_{i=1}^N$, our objective is to estimate $J(d)$. We propose to achieve this by estimating statistical expectations and probabilities via sample averaging and frequencies of occurrence, respectively. We note that although (10) is equivalent to (11) ideally (when all statistical quantities are known), this is not in general the case when estimated measures are considered. So, let $\{p_1^-, \cdots, p_{N_1}^-\} \triangleq \{Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right] : b(i) = -1,\ i = 1, \cdots, N\}$ and $\{p_1^+, \cdots, p_{N_2}^+\} \triangleq \{Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right] : b(i) = +1,\ i = 1, \cdots, N\}$ be the sets of all filter outputs under hypotheses H_0 and H_1, respectively ($N_1, N_2 \neq 0$ and $N_1 + N_2 = N$). First, we can show in a straightforward manner that the estimator of the numerator of (11) (that is, the weighted average of the sample average mean of the set $\{p_i^-\}$ and the sample average mean of the set $\{p_j^+\}$ weighted by the frequency of each set) is *equivalent* to the estimator of the numerator of (10) $4\widehat{\mu}^2(d)$ where $\widehat{\mu}(d) \triangleq \frac{\sum_{i=1}^{N_1+N_2} b(i)p_i}{N_1+N_2}$ (in fact, $\widehat{\mu}(d)$ is the minimum variance unbiased estimator of $\mu(d)$ [15, p. 178]). Estimators of the denominator of (10) and (11) are examined in the following proposition. The proof is included in the Appendix.

Proposition 1: Consider the estimator of $\sigma^2(d)$ which is the weighted average of the sample average variance of the set $\{p_i^+\}$ and the sample average variance of the set $\{p_j^-\}$, weighted by the frequency of each set:

$$\widehat{\sigma}_1^2(d) = \frac{N_1}{N_1 + N_2}\ \widehat{\sigma}^{2-}(d) + \frac{N_2}{N_1 + N_2}\ \widehat{\sigma}^{2+}(d) \tag{12}$$

where

$$\widehat{\sigma}^{2-}(d) \triangleq \frac{1}{N_1}\sum_{i=1}^{N_1}\left[p_i^- - \widehat{\mu}^-(d)\right]^2, \qquad \widehat{\mu}^-(d) \triangleq \frac{1}{N_1}\sum_{i=1}^{N_1} p_i^-,$$

$$\widehat{\sigma}^{2+}(d) \triangleq \frac{1}{N_2}\sum_{j=1}^{N_2}\left[p_j^+ - \widehat{\mu}^+(d)\right]^2, \qquad \widehat{\mu}^+(d) \triangleq \frac{1}{N_2}\sum_{j=1}^{N_2} p_j^+,$$

and $\{p_i^-\}$, $\{p_j^+\}$, N_1 and N_2 are as defined previously. Consider also the direct sample average estimator of $\sigma^2(d)$:

$$\widehat{\sigma}_2^2(d) = \frac{1}{N_1+N_2}\sum_{i=1}^{N_1+N_2}\left[b(i)p_i - \widehat{\mu}(d)\right]^2 \qquad (13)$$

where $p_i \triangleq Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right]$. The estimators $\widehat{\sigma}_1^2(d)$ and $\widehat{\sigma}_2^2(d)$ exhibit the following properties: (i) They are both biased and (ii) $\widehat{\sigma}_2^2(d)$ exhibits smaller MSE from the true value than $\widehat{\sigma}_1^2(d)$. ■

Utilizing Proposition 1, estimators for the filter-output J-divergence become readily available. The following theorem identifies their relative merits. The proof is included in the Appendix.

Theorem 1: Define the two supervised estimators of the output J-divergence $\widehat{J}_{S,1}(d) \triangleq \frac{4\widehat{\mu}^2(d)}{\widehat{\sigma}_1^2(d)}$ and $\widehat{J}_{S,2}(d) \triangleq \frac{4\widehat{\mu}^2(d)}{\widehat{\sigma}_2^2(d)}$, where the subscript "S" identifies a supervised (pilot assisted) implementation. Then, for a given information bit pilot sequence of size $N = N_1 + N_2$, where N_1 and N_2 are the cardinalities of the sets $\{b(i) = -1\}$ and $\{b(i) = +1\}$ respectively, both estimators are biased while the MSE of $\widehat{J}_{S,2}(d)$ is less than the MSE of $\widehat{J}_{S,1}(d)$, i.e.

$$E\left\{\left[\widehat{J}_{S,2}(d) - \frac{4\mu^2(d)}{\sigma^2(d)}\right]^2\right\} < E\left\{\left[\widehat{J}_{S,1}(d) - \frac{4\mu^2(d)}{\sigma^2(d)}\right]^2\right\} \qquad (14)$$

where $\frac{4\mu^2(d)}{\sigma^2(d)}$ is the true value of $J(d)$. ■

Using the preferred estimator $\widehat{J}_{S,2}(d)$, the supervised implementation of the output J-divergence AV selection rule takes the final form given by the following criterion.

Criterion 2a: For a given information bit pilot sequence of size N, the supervised J-divergence AV-filter selection rule chooses the AV-filter estimator $\widehat{\mathbf{w}}_{\hat{d}_2}$ such that

$$\begin{aligned}\widehat{d}_2 &= \arg\max_d \left\{\widehat{J}_{S,2}(d)\right\} \\ &= \arg\max_d \left\{\frac{4\left[\frac{1}{N}\sum_{i=1}^{N} b(i) Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right]\right]^2}{\frac{1}{N}\sum_{i=1}^{N}\left[b(i) Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right] - \widehat{\mu}(d)\right]^2}\right\} \qquad (15)\end{aligned}$$

■

2) Unsupervised (Blind) Output J-divergence Rule

The blind implementation of the rule is obtained by substituting the information bit b in (10) by the detected bit $\widehat{b} = sgn\left(Re\left[\widehat{\mathbf{w}}_{AV}^{(d)H}\mathbf{r}\right]\right)$ (output of the sign detector that follows the linear filter). In particular, using $\widehat{b}$ in place of b in (10) we obtain the following J-divergence expression:

$$J_B(d) = \frac{4E^2\left\{\widehat{b} Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right\}}{Var\left\{\widehat{b} Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right\}} = \frac{4E^2\left\{\left|Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right|\right\}}{Var\left\{\left|Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}\right]\right|\right\}} \qquad (16)$$

where the subscript "B" identifies the blind version of the J-divergence function. The following proposition identifies conditions under which $J_B(d)$ is nearly equal to $J(d)$. The proof is included in the Appendix.

Proposition 2: If $\frac{\mu(d)}{\sigma(d)} >> 1$, i.e. the filter output $SINR$ is significantly higher than $0dB$, then $J_B(d) \approx J(d)$. ■

To estimate $J_B(d)$ from a data record of finite size, we substitute the expectations in (16) by sample averages. The following criterion summarizes the corresponding AV-filter selection rule.

Criterion 2b: For a given data record of size N, the unsupervised (blind) J-divergence AV-filter selection rule chooses the AV-filter estimator $\widehat{\mathbf{w}}_{\hat{d}_3}$ such that

$$\begin{aligned}\widehat{d}_3 &= \arg\max_d \left\{\widehat{J}_B(d)\right\} \qquad (17) \\ &= \arg\max_d \left\{\frac{4\left[\frac{1}{N}\sum_{i=1}^{N}\left|Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right]\right|\right]^2}{\frac{1}{N}\sum_{i=1}^{N}\left|Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right]\right|^2 - \left[\frac{1}{N}\sum_{i=1}^{N}\left|Re\left[\widehat{\mathbf{w}}_d^H \mathbf{r}_i\right]\right|\right]^2}\right\}\end{aligned}$$

■

III. Simulation Studies

We revisit the signal model of Fig. 2. Using the same set-up as in Fig. 2(a) (i.e. $N = 230$, $SNR_0 = 8dB$), we examine the performance of the AV-filter selection rules under the assumption that no pilot sequence $\{b(i)\}_i$ of the user signal of interest, user *0*, is available. In Fig. 3, we plot the empirical probability density function (pdf) of the differences $\left(\widehat{d}_1 - d_{opt}\right)$ and $\left(\widehat{d}_3 - d_{opt}\right)$ where $\widehat{d}_1$ and $\widehat{d}_3$ denote selections according to Criterion 1 and Criterion 2b, respectively, while d_{opt} denotes the ("genie" found) maximum SINR optimum choice of the number of auxiliary vectors d (as in Fig. 2, we conduct $1,000$ experiments with 100 different channel realizations and 10 independent data record generations per channel). The pdf's of the two differences $\left(\widehat{d}_1 - d_{opt}\right)$ and $\left(\widehat{d}_3 - d_{opt}\right)$ shown in Fig. 3 capture the level of failure in estimating the optimum AV parameter d. The heavy concentration of both pdf's around 0 implies that both criteria provide reliable estimates of the optimum number of auxiliary vectors.

In Fig. 4, we plot the BER[2] of the AV-filter estimators $\widehat{\mathbf{w}}_{\hat{d}_1}$ and $\widehat{\mathbf{w}}_{\hat{d}_3}$ as a function of the SNR of the user of interest for data records of size $N = 230$. The BER curve of the "genie"

[2]The BER of each filter under consideration is approximated by $Q\left(\sqrt{SINR_{out}}\right)$ [14] since the computational complexity of the BER expression for this antenna array CDMA system prohibits exact analytical evaluation.

assisted BER optimum filter choice $\widehat{\mathbf{w}}_{d_{opt}}$ as well as the corresponding curves of the ideal MMSE/MVDR, SMI, RAKE matched-filter (MF), and multistage nested Wiener filter [9], [10] with the preferred number of eight stages [16] are also included for comparison purposes. We observe that both $\widehat{\mathbf{w}}_{\hat{d}_1}$ and $\widehat{\mathbf{w}}_{\hat{d}_3}$ are very close to the "genie" optimum AV-filter choice and outperform significantly the SMI, RAKE matched filter,[3] and multistage nested Wiener filter estimators. We also observe that for moderate to high SNR's of the user of interest, the J-divergence selection rule is slightly superior to the CV-MOV selection rule. The opposite is true in the low SNR range. This is explained by the fact that the J-divergence approximation $J(\mathbf{w}) \approx J_B(\mathbf{w})$ used in Proposition 2 is less accurate for low filter output SINR values. On the other hand, for high filter output SINR's the discrimination capability of the CV-MOV rule is not as sharp.

Finally, Fig. 5 repeats the study of Fig. 4 as a function of the data record size N. The SNR of the user signal of interest is fixed at $8dB$. To illustrate the benefits that we gain by actively seeking the most appropriate number of AV's for *each given input data record*, we include the BER curve of AV filter estimators that either use always some fixed number of auxiliary vectors (in this study we consider $d = 15$) or for each data record size N use $d_{mean}(N)$ auxiliary vectors where $d_{mean}(N)$ denotes the integer nearest to the empirical mean of the pdf of the random variable $d_{opt}(N)$. We observe that both $\widehat{\mathbf{w}}_{\hat{d}_1}$ and $\widehat{\mathbf{w}}_{\hat{d}_3}$ outperform $\widehat{\mathbf{w}}_{15}$ and $\widehat{\mathbf{w}}_{d_{mean}}$; of course, d_{mean} cannot be calculated/estimated in practice without long pilot signaling (not needed by either one of the $\hat{d}_1$ and $\hat{d}_3$ proposed selection rules).

IV. Summary and Discussion

In [2], [3] it was shown that *for a given finite data record* the sequence of AV-filter estimators initialized at the matched-filter converges to the SMI estimator of the MMSE/MVDR filter. It was also noted that the early non-asymptotic elements in the sequence provide favorable bias/variance balance and outperform in MS filter estimation error the SMI, (constraint)-LMS, RLS, multistage nested Wiener filter (MNWF), and DL-SMI filter estimators. In this paper we considered the problem of identifying the best AV-filter estimator in the sequence. We proposed two selection criteria, one based on the filter output variance and one based on the J-divergence of the filter output conditional distributions. The first AV-filter selection rule aims at the minimization of the estimated filter output variance using the well established statistical method of cross-validation (in particular "leave-one-out" cross-validation). This selection criterion can be applied to general filter estimation problems. The need to invoke cross-validation was established formally. The second AV-filter selection rule is tailored to applications that can be formulated as binary hypothesis testing problems defined on AV-filtered data. This selection rule aims at the maximization of the estimated J-divergence between the filter-output conditional distributions under each hypothesis. We presented a supervised and a blind implementation of the J-divergence selection criterion for problems where the received data correspond to binary antipodal signals in noise.

Simulation studies examined and compared the operational characteristics of the proposed methods. With respect to the relative merits of the output J-divergence and the cross-validation minimum output variance selection rule, we observed that for moderate to high output SINR's the former method appears somewhat superior to the latter (for high SINR's the cross-validated minimum output variance rule is not as sharp in discrimination ability). In contrast, in low output SINR the J-divergence method is slightly lagging in performance (technically, the approximation in Proposition 2 is less accurate for near $0dB$ or lower filter output SINR values).

[3]We recall that the ideal MMSE/MVDR filter is in fact the AV filter $\mathbf{w}_\infty$ ($d = \infty$ AV's), the SMI filter is $\widehat{\mathbf{w}}_\infty(N)$, and the RAKE matched-filter is $\mathbf{w}_0$ ($d = 0$ AV's). As a side note, the ideal MMSE/MVDR filter ($d = \infty$) lies in the signal subspace (see for example [17]), while both the SMI filter ($d = \infty$) and the AV-filter estimates ($0 < d < \infty$) do not necessarily belong to either the true or the estimated signal subspace.

Appendix A
Proof of Proposition 1

Proof: The quantities $\frac{1}{N_1}\sum_{i=1}^{N_1}\left(p_i^- - \widehat{\mu}^-(d)\right)^2$ and $\frac{1}{N_2}\sum_{j=1}^{N_2}\left(p_j^+ - \widehat{\mu}^+(d)\right)^2$ are the ML estimators of the filter output variance conditioned on $b = -1$ and $b = +1$, respectively [15, p. 179]. Both estimators are biased. (In fact, their unbiased counterparts that have multiplying factors $\frac{1}{N_1-1}$ and $\frac{1}{N_2-1}$ instead of $\frac{1}{N_1}$ and $\frac{1}{N_2}$, respectively, exhibit higher MSE). The MSE's of the estimators of interest $\widehat{\sigma}_1^2(d)$ and $\widehat{\sigma}_2^2(d)$ are as follows:

$$
\begin{aligned}
MSE_{\widehat{\sigma}_1^2(d)} &= E\left\{\left(\widehat{\sigma}_1^2(d) - \sigma^2(d)\right)^2\right\} \\
&= E\left\{\left[\left(\frac{\sum_{i=1}^{N_1}(p_i^- - \widehat{\mu}^-(d))^2}{N_1+N_2} - \frac{N_1\sigma^2(d)}{N_1+N_2}\right)\right]^2\right\} \\
&\quad + E\left\{\left[\left(\frac{\sum_{j=1}^{N_2}(p_j^+ - \widehat{\mu}^+(d))^2}{N_1+N_2} - \frac{N_2\sigma^2(d)}{N_1+N_2}\right)\right]^2\right\} \\
&\quad + 2\times E\left\{\frac{\sum_{i=1}^{N_1}(p_i^- - \widehat{\mu}^-(d))^2}{N_1+N_2} - \frac{N_1\sigma^2(d)}{N_1+N_2}\right\} \\
&\quad \times E\left\{\frac{\sum_{j=1}^{N_2}(p_j^+ - \widehat{\mu}^+(d))^2}{N_1+N_2} - \frac{N_2\sigma^2(d)}{N_1+N_2}\right\} \\
&= \frac{(2N_1-1)\,\sigma^4(d) + (2N_2-1)\,\sigma^4(d) + 2\sigma^4(d)}{(N_1+N_2)^2} \\
&= \frac{2\sigma^4(d)}{N_1+N_2} \qquad (18)
\end{aligned}
$$

and

$$
\begin{aligned}
MSE_{\widehat{\sigma}_2^2(d)} &= E\left\{\left(\widehat{\sigma}_2^2(d) - \sigma^2(d)\right)^2\right\} \\
&= \frac{[2(N_1+N_2)-1]\,\sigma^4(d)}{(N_1+N_2)^2}. \qquad (19)
\end{aligned}
$$

Thus,

$$
MSE_{\widehat{\sigma}_1^2(d)} - MSE_{\widehat{\sigma}_2^2(d)} = \frac{\sigma^4(d)}{(N_1+N_2)^2} > 0. \qquad (20)
$$

■

APPENDIX B
PROOF OF THEOREM 1

Proof: $\widehat{\mu}^-(d)$ and $\widehat{\sigma}^{2-}(d)$ are independent random variables with distributions $\widehat{\mu}^-(d) \sim \mathcal{N}\left(-\mu(d), \frac{\sigma^2(d)}{N_1}\right)$ and $\frac{N_1}{\sigma^2(d)}\widehat{\sigma}^{2-}(d) \sim \chi^2_{N_1-1}$ [18] ($\chi^2_{N_1-1}$ denotes the chi-square distribution with $N_1 - 1$ degrees of freedom). Similarly, $\widehat{\mu}^+(d)$ and $\widehat{\sigma}^{2+}(d)$ are independent random variables with distributions $\widehat{\mu}^+(d) \sim \mathcal{N}\left(\mu(d), \frac{\sigma^2(d)}{N_2}\right)$ and $\frac{N_2}{\sigma^2(d)}\widehat{\sigma}^{2+}(d) \sim \chi^2_{N_2-1}$. Furthermore, $\left(\widehat{\mu}^-(d), \widehat{\sigma}^{2-}(d)\right)$ and $\left(\widehat{\mu}^+(d), \widehat{\sigma}^{2+}(d)\right)$ are mutually independent because $\left(\widehat{\mu}^-(d), \widehat{\sigma}^{2-}(d)\right)$ and $\left(\widehat{\mu}^+(d), \widehat{\sigma}^{2+}(d)\right)$ are evaluated from two independent training sets. Since $\widehat{\mu}(d) = \frac{N_1}{N}\widehat{\mu}^-(d) - \frac{N_2}{N}\widehat{\mu}^+(d)$ and $\widehat{\sigma}_1^2(d) = \frac{N_1}{N}\widehat{\sigma}^{2-}(d) + \frac{N_2}{N}\widehat{\sigma}^{2+}(d)$, it is implied that $\widehat{\mu}(d)$ and $\widehat{\sigma}_1^2(d)$ are independent random variables with distributions $\widehat{\mu}(d) \sim \mathcal{N}\left(\mu(d), \frac{\sigma^2(d)}{N}\right)$ and $\frac{N}{\sigma^2(d)}\widehat{\sigma}_1^2(d) \sim \chi^2_{N-2}$ where $N = N_1 + N_2$ is the size of the given pilot bit sequence. Therefore,

$$\begin{aligned} E\{\widehat{J}_{S,1}(d)\} &= 4E\{\widehat{\mu}^2(d)\}E\left\{\frac{1}{\widehat{\sigma}_1^2(d)}\right\} \\ &= 4\left[\frac{\sigma^2(d)}{N} + \mu^2(d)\right]\frac{N}{\sigma^2(d)\,(N-4)} \\ &= \frac{4\mu^2(d)}{\sigma^2(d)}\,\frac{N}{N-4} + \frac{4}{N-4}, \end{aligned} \tag{21}$$

$$\begin{aligned} E\{\widehat{J}_{S,1}^2(d)\} &= 16E\{\widehat{\mu}^4(d)\}E\left\{\frac{1}{\widehat{\sigma}_1^4(d)}\right\} \\ &= \frac{16\left[N^2\mu^4(d) + 6N\mu^2(d)\sigma^2(d) + 3\sigma^4(d)\right]}{\sigma^4(d)\,(N-4)\,(N-6)}. \end{aligned} \tag{22}$$

The inverse moments $E\left\{\frac{1}{\widehat{\sigma}_1^2(d)}\right\}$ and $E\left\{\frac{1}{\widehat{\sigma}_1^4(d)}\right\}$ exist for $N > 4$ and $N > 6$, respectively, and are given by [19]

$$E\left\{\frac{1}{\widehat{\sigma}_1^2(d)}\right\} = \frac{N}{\sigma^2(d)\,(N-4)}, \tag{23}$$

$$E\left\{\frac{1}{\widehat{\sigma}_1^4(d)}\right\} = \frac{N^2}{\sigma^4(d)\,(N-4)\,(N-6)}. \tag{24}$$

Thus, the MSE of $\widehat{J}_{S,1}(d)$ is

$$\begin{aligned} MSE_1 &\triangleq E\left\{\left[\widehat{J}_{S,1}(d) - \frac{4\mu^2(d)}{\sigma^2(d)}\right]^2\right\} \\ &= \frac{16a}{(N-4)\,(N-6)} - \frac{16b}{N-4} + \frac{16\mu^4(d)}{\sigma^4(d)} \end{aligned} \tag{25}$$

where

$$a \triangleq \left[\mu^4(d) + \frac{6\mu^2(d)\sigma^2(d)}{N} + \frac{3\sigma^4(d)}{N^2}\right]\frac{N^2}{\sigma^4(d)} \text{ and}$$

$$b \triangleq \left[\frac{\sigma^2(d)}{N} + \mu^2(d)\right]\frac{2N\mu^2(d)}{\sigma^4(d)}.$$

On the other hand, $\widehat{\mu}(d)$ and $\widehat{\sigma}_2^2(d)$ are independent random variables with distributions $\widehat{\mu}(d) \sim \mathcal{N}\left(\mu(d), \frac{\sigma^2(d)}{N}\right)$ and $\frac{N}{\sigma^2(d)}\widehat{\sigma}_2^2(d) \sim \chi^2_{N-1}$, respectively. Therefore,

$$\begin{aligned} E\{\widehat{J}_{S,2}(d)\} &= 4\left[\frac{\sigma^2(d)}{N} + \mu^2(d)\right]\frac{N}{\sigma^2(d)\,(N-3)} \\ &= \frac{4\mu^2(d)}{\sigma^2(d)}\,\frac{N}{N-3} + \frac{4}{N-3}, \end{aligned} \tag{26}$$

$$E\{\widehat{J}_{S,2}^2(d)\} = \frac{16\left[N^2\mu^4(d) + 6N\mu^2(d)\sigma^2(d) + 3\sigma^4(d)\right]}{\sigma^4(d)\,(N-3)\,(N-5)}. \tag{27}$$

Thus, the MSE of $\widehat{J}_{S,2}(d)$ is

$$\begin{aligned} MSE_2 &\triangleq E\left\{\left[\widehat{J}_{S,2}(d) - \frac{4\mu^2(d)}{\sigma^2(d)}\right]^2\right\} \\ &= \frac{16a}{(N-3)\,(N-5)} - \frac{16b}{N-3} + \frac{16\mu^4(d)}{\sigma^4(d)} \end{aligned} \tag{28}$$

where the first and second inverse moments $E\left\{\frac{1}{\widehat{\sigma}_2^2(d)}\right\}$ and $E\left\{\frac{1}{\widehat{\sigma}_2^4(d)}\right\}$ exist for $N > 3$ and $N > 5$, respectively, and a and b are defined above. From (25) and (28), assuming $N > 6$ which is the condition for all inverse moments to exist, we obtain

$$\begin{aligned} &MSE_1 - MSE_2 \\ &= \frac{16a}{(N-4)(N-6)} - \frac{16a}{(N-3)(N-5)} + \frac{b}{N-3} - \frac{b}{N-4} \\ &= \frac{16}{(N-3)(N-4)}\left[\frac{2N-9}{(N-5)(N-6)}a - b\right] \\ &> \frac{16}{(N-3)(N-4)}\left[\frac{2(N-5)}{(N-5)(N-6)}a - b\right] \\ &= \frac{32\left[6N\mu^4(d) + 5N\mu^2(d)\sigma^2(d) + 6\mu^2(d)\sigma^2(d) + 3\sigma^4(d)\right]}{(N-3)(N-4)(N-6)\sigma^4(d)} > 0. \end{aligned} \tag{29}$$

■

APPENDIX C
PROOF OF PROPOSITION 2

Proof: Define $Y \triangleq |Re(\widehat{\mathbf{w}}_d^H\mathbf{r})|$. The pdf of Y is

$$f_Y(y) = \frac{U(y)\exp\left(\frac{-(y-\mu(d))^2}{2\sigma^2(d)}\right)}{\sqrt{2\pi}\sigma(d)} + \frac{U(y)\exp\left(\frac{-(y+\mu(d))^2}{2\sigma^2(d)}\right)}{\sqrt{2\pi}\sigma(d)} \tag{30}$$

where $U(y)$ is the unit step function. The mean of Y is

$$\begin{aligned} E(Y) &= \int y f_Y(y)dy \\ &= \mu(d) + \frac{2\sigma(d)}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(d)}{2\sigma^2}\right) - 2\mu(d)Q\left(\frac{\mu(d)}{\sigma(d)}\right). \end{aligned} \tag{31}$$

The series expansion of $Q(x)$ is [20]

$$\begin{aligned} Q(x) = &\frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}x}\left\{1 - \frac{1}{x^2} + \cdots + \frac{(-1)^n 1\cdot 3\cdots(2n-1)}{x^{2n}}\right\} \\ &+ R_n \end{aligned} \tag{32}$$

where $R_n = (-1)^{n+1}1 \cdot 3\cdots(2n+1)\int_x^\infty \frac{1}{\sqrt{2\pi}t^{2n+2}}\exp\left(-\frac{t^2}{2}\right)dt$ is the remainder which is

always less (in absolute value) than the first neglected term. If $\frac{\mu(d)}{\sigma(d)} >> 1$, (31) can be written as

$$\begin{aligned} E(Y) &= \mu(d)+\frac{2\mu(d)}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(d)}{2\sigma^2(d)}\right)\frac{\sigma(d)}{\mu(d)} \\ &\quad -\frac{2\mu(d)}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(d)}{2\sigma^2(d)}\right)\left[1-O\left[\left(\frac{\mu(d)}{\sigma(d)}\right)^{-2}\right]\right]\frac{\sigma(d)}{\mu(d)} \\ &= \mu(d)\left\{1+\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(d)}{2\sigma^2(d)}\right)O\left[\left(\frac{\mu(d)}{\sigma(d)}\right)^{-3}\right]\right\} \\ &\approx \mu(d). \end{aligned} \tag{33}$$

The variance of Y is

$$Var(Y) = E(Y^2) - E^2(Y) = \mu^2(d) + \sigma^2(d) - E^2(Y) \tag{34}$$

and using (33) we may approximate

$$Var(Y) \approx \sigma^2(d). \tag{35}$$

Finally, from (33) and (35) we obtain

$$J_B(d) = \frac{4E^2(Y)}{Var(Y)} \approx \frac{4\mu^2(d)}{\sigma^2(d)} = J(d). \tag{36}$$

■

REFERENCES

[1] S. N. Batalama, M. J. Medley, and D. A. Pados, "Robust adaptive recovery of spread-spectrum signals with short data records," *IEEE Trans. Commun.*, vol. 48, pp. 1725-1731, Oct. 2000.

[2] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Processing*, vol. 49, pp. 290-300, Feb. 2001.

[3] D. A. Pados and G. N. Karystinos, "Short-data-record estimators of the MVDR/MMSE filters," *in Proc. Intern. Conf. on Acoust., Speech, and Signal Proc. (ICASSP 2000)*, Istanbul, Turkey, June 2000, vol. 1, pp. 384-387.

[4] D. A. Pados and S. N. Batalama, "Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers," *IEEE Trans. Commun.*, vol. 45, pp. 1586-1594, Dec. 1997.

[5] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1765-1773, Dec. 1998.

[6] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," in *Proc. 1998 Conf. Info. Sc. and Syst.*, Princeton Univ., Princeton, NJ, Mar. 1998, vol. II, pp. 1007-1013.

[7] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406-1415, Sept. 1999.

[8] D. A. Pados, T. Tsao, J. Michels, and M. Wicks, "Joint domain space-time adaptive processing with small training data sets," in *Proc. IEEE Radar Conf.*, Dallas, TX, May 1998, pp. 99-104.

[9] J. S. Goldstein, I. S. Reed, P. A. Zulch, and W. L. Melvin, "A multistage STAP CFAR detection technique," in *Proc. IEEE Radar Conf.*, Dallas, TX, May 1998, pp. 111-116.

[10] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2943-2959, Nov. 1998.

[11] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall Inc., 2nd edition, 2002.

[12] C. R. Rao, *Handbook of Statistics 9*. New York, NY: Elsevier, 1993.

[13] D. Kazakos and P. Papantoni-Kazakos, *Detection and Estimation*. New York, NY: Computer Science Press, 1990.

[14] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858-871, May 1997.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY: Springer-Verlag, 1994.

[16] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1928-1946, July 2001.

[17] D. Gesbert, J. Sorelius, P. Stoica, and A. Paulraj, "Blind multiuser MMSE detector for CDMA signals in ISI channels," *IEEE Commun. Letters*, vol. 3, pp. 233-235, Aug. 1999.

[18] M. H. DeGroot, *Probability and Statistics*. Reading, MA.: Addison-Wesley, 1984.

[19] M. C. Jones, "Expressions for inverse moments of positive quadratic forms in normal variables," *Austral. J. Statist.*, vol. 28, pp. 242-250, Feb. 1986.

[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York, NY: Dover Publications, 1965.

Input:

Autocovariance matrix $\mathbf{R}$, constraint vector $\mathbf{v}$, desired response $\mathbf{w}^H\mathbf{v} = 1$.

Initialization:

$$\mathbf{w}_0 := \frac{\mathbf{v}}{\|\mathbf{v}\|^2}.$$

Iterative computation:

For d=1, 2, ... do

begin

$$\mathbf{g}_d := \mathbf{R}\,\mathbf{w}_{d-1} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\mathbf{R}\,\mathbf{w}_{d-1}$$

if $\mathbf{g}_d = \mathbf{0}$ then EXIT

$$\mu_d := \frac{\mathbf{g}_d^H \mathbf{R}\,\mathbf{w}_{d-1}}{\mathbf{g}_d^H \mathbf{R}\,\mathbf{g}_d}$$

$$\mathbf{w}_d := \mathbf{w}_{d-1} - \mu_d \mathbf{g}_d$$

end

Output:

Filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

Fig. 1. The "AV algorithm" for the iterative generation of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

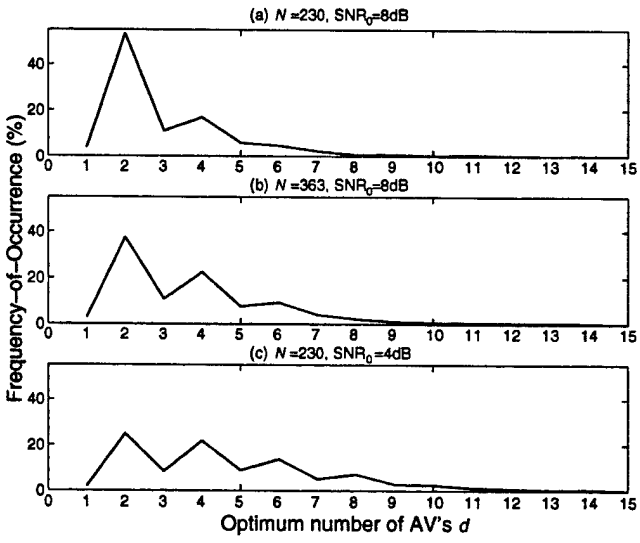


Fig. 2. Histogram of the number of AV's d utilized by the best (max SINR) AV-filter: (a) $N = 230$, $SNR_0 = 8dB$, (b) $N = 363$, $SNR_0 = 8dB$, and (c) $N = 230$, $SNR_0 = 4dB$.

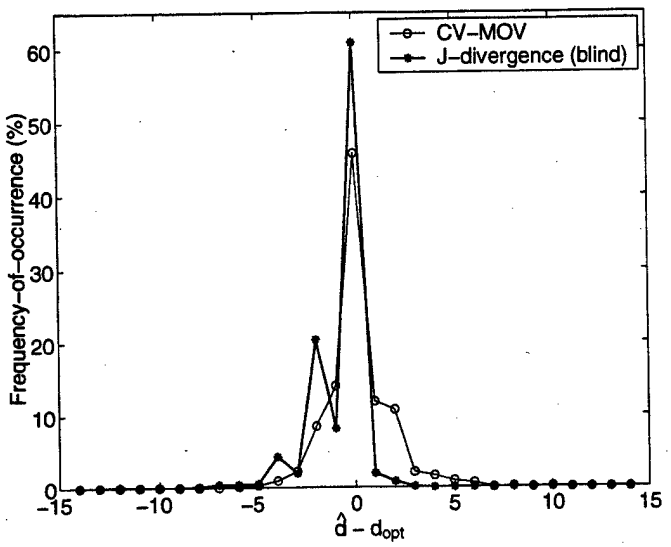


Fig. 3. Histogram of the two differences $\widehat{d}_1 - d_{opt}$ and $\widehat{d}_3 - d_{opt}$ where $\widehat{d}_1$ is the CV-MOV choice, $\widehat{d}_3$ is the blind J-divergence choice, and d_{opt} is the maximum-SINR optimal choice of the number of auxiliary vectors ($N = 230$, $SNR_0 = 8dB$).

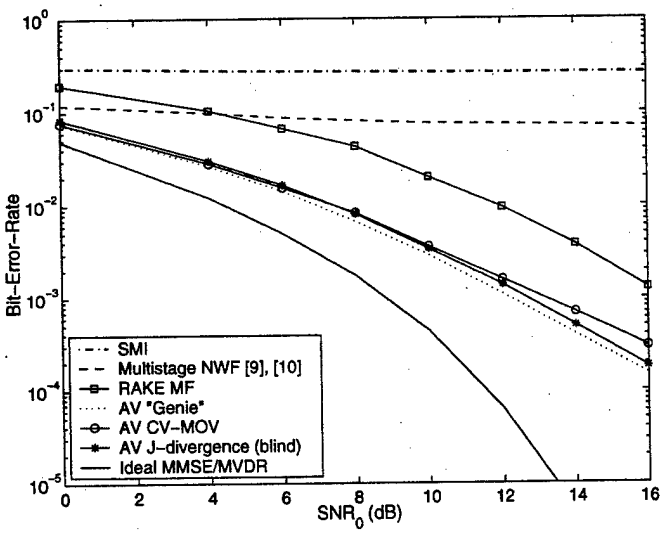


Fig. 4. BER versus SNR for the user signal of interest (data record size $N = 230$).

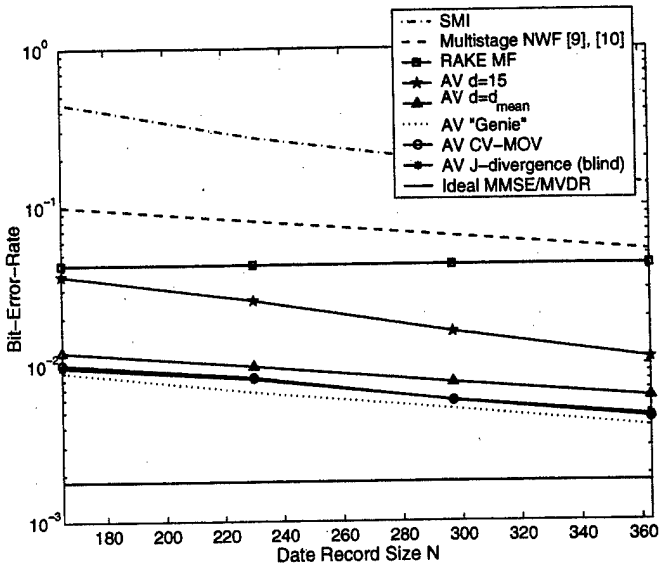


Fig.5. BER versus data record size ($SNR_0 = 8dB$).

Rapid Combined Synchronization/Demodulation Structures for DS-CDMA Systems—Part II: Finite Data-Record Performance Analysis

Ioannis N. Psaromiligkos, *Member, IEEE,* and Stella N. Batalama, *Member, IEEE*

***Abstract*—We investigate the coarse synchronization performance of blind adaptive linear *self-synchronized* receivers for asynchronous direct-sequence code-division multiple-access communications under finite data record adaptation. Based on transformation noise modeling techniques, three alternative methods are developed, leading to analytical expressions that approximate the probability of coarse synchronization error of matched-filter-type and minimum-variance distortionless-response-type receivers. The expressions are explicit functions of the data record size and the filter order and reveal the effect of short data-record sample matrix-inversion implementations on the coarse synchronization performance. Besides their theoretical value, the derived expressions provide simple, highly-accurate alternatives to computationally demanding performance evaluation through simulations. The effect of the data record size on the probability of coarse synchronization error is further quantified through the use of a receiver synchronization resolution metric. Numerical and simulation studies examine the accuracy of the theoretical developments and show that the derived expressions approximate closely the actual coarse synchronization performance.**

***Index Terms*—Adaptive filters, approximation methods, code-division multiple access (CDMA), finite data analysis, minimum mean-square error (MMSE) filtering, spread-spectrum communication, synchronization.**

I. Introduction

THE effectiveness of a receiver designed for rapidly changing wireless direct-sequence code-division multiple-access (DS-CDMA) communication environments depends on the following design attributes: 1) low computational complexity; 2) multiple-access interference (MAI) near–far resistance; and 3) system adaptivity with superior performance under limited data support. Adaptive short data-record designs appear as the natural next step [1]–[4] to a matured discipline that has extensively addressed the first two design objectives in ideal setups (perfectly known or asymptotically estimated statistical properties) [5], [6]. System adaptivity based on short data records is necessary for the development of practical receivers that exhibit superior bit-error rate (BER) performance when they operate in rapidly changing communication environments that limit substantially the available data support.

In Part I [7], we considered self-synchronized receivers (integrated synchronizers/demodulators) and we presented three schemes along the classical receiver evolution path. We started with a matched-filter-type (MF) structure, we continued with a minimum-variance distortionless-response-type (MVDR) scheme and, finally, we developed an auxiliary-vector-type (AV) alternative, all of order either twice or equal to the system spreading gain L. Simulation studies illustrated the performance of the proposed structures in terms of BER versus data support for a given signal-to-noise ratio (SNR) of the user of interest or in terms of BER versus SNR for a given data record size.

In this paper, we develop three methods for the analysis of the finite data-record behavior of self-synchronized linear receivers and, in particular, the effect of the filter order and the data-record size on the coarse-synchronization error rate (CSER). We derive analytical expressions that approximate closely the probability of coarse synchronization error of MF-type and MVDR-type schemes. Probability of coarse-synchronization error analysis for AV-type schemes based on a finite data-record size is prohibitively complex, and thus, not attempted at this time.

The paper is organized as follows. In Section II, we present the system model and we summarize briefly the combined synchronization/demodulation algorithms developed in [7]. In Section III, we develop analytical expressions that approximate the coarse-synchronization error probability (P_{cse}), while in Section IV, we derive a sequence of increasingly tight lower bounds on P_{cse}. In Section V, we discuss the effect of the filter order and the data-record size on the CSER. The accuracy of the analytical approximations of Sections III and IV is examined through simulations in Section VI. Some final conclusions are drawn in Section VII.

II. System Model and Algorithmic Description

We consider an asynchronous DS-CDMA system populated by K active users transmitting over a common additive white

Paper approved by M. Brandt-Pearce, the Editor for Modulation and Signal Design of the IEEE Communications Society. Manuscript received December 15, 1999; revised August 15, 2001 and October 15, 2002. This work was supported in part by the Air Force Office of Scientific Research under Grant F49620-01-1-0176 and in part by the National Science Foundation under Grant ECS-0073660 and under Grant CCR-0219903. This paper was presented in part at the 2000 International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, June 2000 and in part at the International Conference on Communications, New Orleans, LA, June 2000.

I. N. Psaromiligkos was with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA. He is now with the Department of Electrical and Computer Engineering, McGill University, Montréal, QC H3H 2N4, Canada (e-mail: yannis@ece.mcgill.ca).

S. N. Batalama is with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: batalama@eng.buffalo.edu).

Digital Object Identifier 10.1109/TCOMM.2003.814199

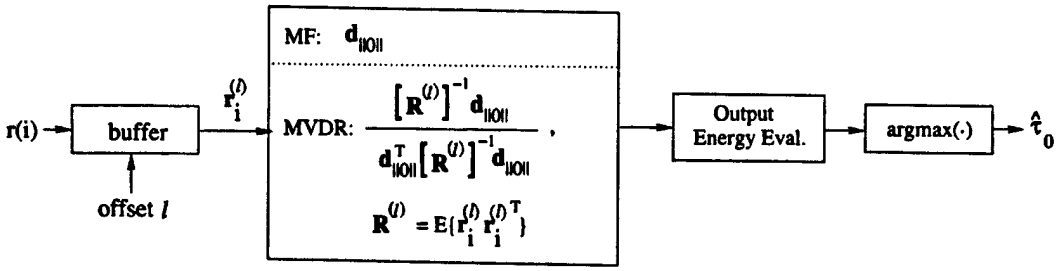


Fig. 1. Coarse synchronization based on filters of order L.

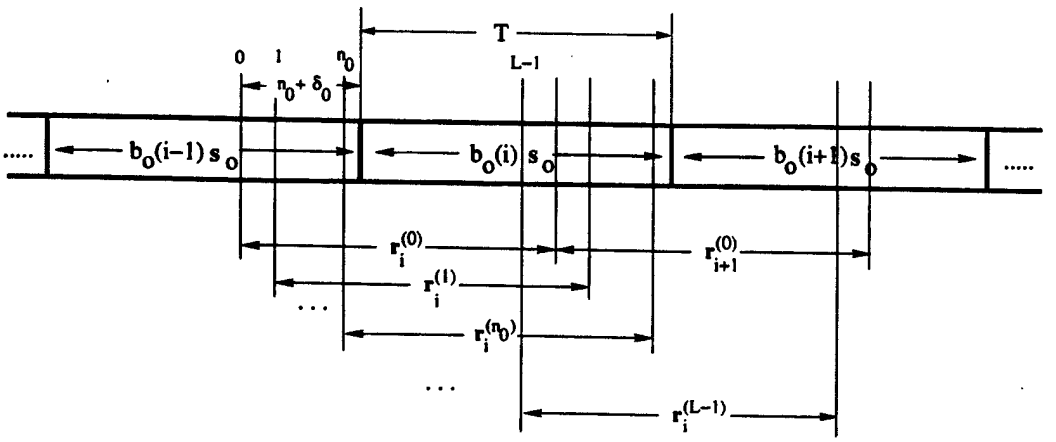


Fig. 2. Construction of the sequences $\mathbf{r}^{(l)}, l = 0\ldots, L-1$.

Gaussian noise (AWGN) channel. The received signal $r(t)$ is the superposition of the K transmissions corrupted by channel noise, i.e.,

$$r(t) = \sum_{k=0}^{K-1} \sum_{i=-\infty}^{\infty} \sqrt{E_k} b_k(i) S_k(t - iT - \tau_k) + n(t). \quad (1)$$

In the above expression, with respect to the kth user, τ_k is the signal delay relative to the receiver's time reference point, $b_k(i) \in \{-1, +1\}$ is the ith information bit, E_k is the transmitted energy, T is the information bit period, and $n(t)$ is AWGN. The normalized signature waveform $S_k(t)$ assigned to the kth user has the form

$$S_k(t) = \sum_{l=0}^{L-1} s_k(l) P_{Tc}(t - lT_c), \qquad k = 0, \ldots, K-1 \quad (2)$$

where T_c is the chip duration, $P_{Tc}(\cdot)$ is a rectangular pulse with support $[0, T_c]$, $L = T/T_c$ is the system spreading gain, and $s_k(l)$ is the lth element of the normalized bipolar signature or spreading vector $\mathbf{s}_k \triangleq [s_k(0), s_k(1), \ldots, s_k(L-1)]^T$ that uniquely identifies the kth user. Without loss of generality, we assume that $\tau_k \in [0, T)$, $k = 0, \ldots, K-1$. Thus, we may write the delay τ_k as a sum of an integer multiple of chips plus a fraction of a chip: $\tau_k = (n_k + \delta_k)T_c$ where $n_k \in \{0, \ldots, L-1\}$ and $\delta_k \in [0, 1)$. After chip-matched filtering and chip-rate sampling, the continuous-time signal $r(t)$ in (1) is converted to a discrete-time sequence $\{r(n)\}_{n=-\infty}^{\infty}$, where $r(n) = \int_{nT_c}^{(n+1)T_c} r(t)dt$.

In this paper, we analyze the coarse synchronization performance of L and $2L$-order algorithms of MF and sample matrix-inversion (SMI) MVDR type. A brief description of the coarse synchronization step of these algorithms follows (details can be found in [7]). Fig. 1 shows the block diagram representation of an L-order coarse synchronizer/demodulator. The buffer follows a chip-rate sampler and groups the received samples into L sequences of vectors, $\{\mathbf{r}_i^{(l)}\}_i \in \Re^L$, $l = 0, \ldots, L-1$, as depicted in Fig. 2 where

$$\mathbf{r}_i^{(l)} = [r(iL + l), r(iL + l + 1), \ldots, r((i+1)L + l - 1)]^T, \quad l = 0, \ldots, L-1. \quad (3)$$

In the case of $2L$-order algorithms, the received samples $\{r(n)\}_n$ are grouped into a sequence of overlapping vectors, $\bar{\mathbf{r}}_i \in \Re^{2L}$, obtained by a sliding window such that successive vectors share L common data samples

$$\bar{\mathbf{r}}_i = [r(iL), r(iL + 1), \ldots, r((i+2)L - 1)]^T. \quad (4)$$

To facilitate our presentation, overlined variables always refer to $2L$-order processing to distinguish themselves from the corresponding variables used by L-order algorithms. Table I provides the definition of both the ideal and estimated decision statistic (filter output energy) of each filter under consideration, namely SMI-MVDR and MF of order L or $2L$. Table II provides additional definitions needed in the expressions of Table I.

For the L-order case of Table I, $\mathbf{R}^{(l)} \triangleq E\left\{\mathbf{r}_i^{(l)}\mathbf{r}_i^{(l)T}\right\}$ is the input covariance matrix and $\hat{\mathbf{R}}^{(l)} \triangleq 1/N \sum_{i=0}^{N-1} \mathbf{r}_i^{(l)}\mathbf{r}_i^{(l)T}$ is its sample average estimate (N is the total number of samples or data-record size). Similarly, for the $2L$-order case, $\bar{\mathbf{R}} \triangleq E\{\bar{\mathbf{r}}_i\bar{\mathbf{r}}_i^T\}$ and $\hat{\bar{\mathbf{R}}} \triangleq 1/N \sum_{i=0}^{N-1} \bar{\mathbf{r}}_i\bar{\mathbf{r}}_i^T$. In Table II, $\mathbf{s}_0^{(+1)} \triangleq [0, s_0(0), \ldots, s_0(L-2)]^T$ denotes the one-chip right-shifted zero-filled version of $\mathbf{s}_0$, while $\bar{\mathbf{s}}_0^{(1)}$,

TABLE I
DECISION STATISTIC

L-order	
SMI-MVDR	
$U_j = \left(\mathbf{d}_{\|0\|}^{(j)T} \mathbf{R}^{(\lfloor j/2 \rfloor)^{-1}} \mathbf{d}_{\|0\|}^{(j)}\right)^{-1}$	$\hat{U}_j = \left(\mathbf{d}_{\|0\|}^{(j)T} \hat{\mathbf{R}}^{(\lfloor j/2 \rfloor)^{-1}} \mathbf{d}_{\|0\|}^{(j)}\right)^{-1}$
MF	
$U_j = \mathbf{d}_{\|0\|}^{(j)T} \mathbf{R}^{(\lfloor j/2 \rfloor)} \mathbf{d}_{\|0\|}^{(j)}$	$\hat{U}_j = \mathbf{d}_{\|0\|}^{(j)T} \hat{\mathbf{R}}^{(\lfloor j/2 \rfloor)} \mathbf{d}_{\|0\|}^{(j)}$
$2L$-order	
SMI-MVDR	
$U_j = \left(\bar{\mathbf{d}}_{\|0\|}^{(j)T} \bar{\mathbf{R}}^{-1} \bar{\mathbf{d}}_{\|0\|}^{(j)}\right)^{-1}$	$\hat{U}_j = \left(\bar{\mathbf{d}}_{\|0\|}^{(j)T} \hat{\bar{\mathbf{R}}}^{-1} \bar{\mathbf{d}}_{\|0\|}^{(j)}\right)^{-1}$
MF	
$U_j = \bar{\mathbf{d}}_{\|0\|}^{(j)T} \bar{\mathbf{R}} \bar{\mathbf{d}}_{\|0\|}^{(j)}$	$\hat{U}_j = \bar{\mathbf{d}}_{\|0\|}^{(j)T} \hat{\bar{\mathbf{R}}} \bar{\mathbf{d}}_{\|0\|}^{(j)}$

TABLE II
EFFECTIVE SIGNATURE

L-order	
$\mathbf{d}_{\|0\|}^{(j)} \triangleq \begin{cases} \mathbf{s}_{\|0\|}(0), & \text{if } j \text{ is even} \\ \mathbf{s}_{\|0\|}(0.5) & \text{if } j \text{ is odd,} \end{cases}$	$\mathbf{s}_{\|0\|}(\delta) \triangleq \frac{(1-\delta)\mathbf{s}_0 + \delta \mathbf{s}_0^{(+1)}}{\|(1-\delta)\mathbf{s}_0 + \delta \mathbf{s}_0^{(+1)}\|}$
$2L$-order	
$\bar{\mathbf{d}}_{\|0\|}^{(j)} \triangleq \begin{cases} \bar{\mathbf{s}}_{\|0\|}^{(\lfloor j/2 \rfloor)}(0), & \text{if } j \text{ is even} \\ \bar{\mathbf{s}}_{\|0\|}^{(\lfloor j/2 \rfloor)}(0.5) & \text{if } j \text{ is odd,} \end{cases}$	$\bar{\mathbf{s}}_{\|0\|}^{(l)}(\delta) \triangleq \frac{(1-\delta)\bar{\mathbf{s}}_0^{(l)} + \delta \bar{\mathbf{s}}_0^{(l+1)}}{\|(1-\delta)\bar{\mathbf{s}}_0^{(l)} + \delta \bar{\mathbf{s}}_0^{(l+1)}\|}$

$l = 0, \ldots, L-1$, is the l-shifted version of the zero-padded $2L$-long vector $[\mathbf{s}_0^T, 0, \ldots, 0]^T$.

An SMI-MVDR-type or MF-type coarse estimate of the delay τ_0 can be determined as follows:

$$\hat{\tau}_0 = \begin{cases} \left(\left\lfloor \frac{j_{\max}}{2} \right\rfloor\right) T_c, & \text{if } j_{\max} \text{ is even} \\ \left(\left\lfloor \frac{j_{\max}}{2} \right\rfloor + 0.5\right) T_c, & \text{if } j_{\max} \text{ is odd} \end{cases} \tag{5}$$

where

$$j_{\max} = \arg\max_j \left\{ \hat{U}_j | j = 0, \ldots, 2L-1 \right\} \tag{6}$$

and $\hat{U}_j$ is obtained from Table I for each filter, respectively.

III. FINITE DATA-RECORD PERFORMANCE ANALYSIS

The probability of coarse synchronization error is defined as the probability that the coarse synchronization algorithm fails to provide an estimate within the "pull-in" range. In this paper, we adopt the commonly used assumption that the pull-in range is equal to $(1/2)T_c$ around the correct delay τ_0. However, the performance analysis presented in this section can be carried out with straightforward modifications for any given pull-in range value. The probability of coarse synchronization error is given by

$$P_{\text{cse}} \triangleq 1 - \Pr\left[|\hat{\tau}_0 - \tau_0| < \frac{1}{2} T_c \right] \tag{7}$$

where the coarse estimate of $\hat{\tau}_0$ is given by (5) and (6). We recall that $j_{\max}$ in (5) and (6) is the index of the filter with the highest output energy. If we define $\mathcal{H}(\tau_0)$ as the set of the filter indexes that yield a timing estimate $\hat{\tau}_0$ within $(1/2)T_c$ about τ_0, then the probability of coarse synchronization error will be equal to

$$P_{\text{cse}} = 1 - \Pr\left[j_{\max} \in \mathcal{H}(\tau_0) \right]. \tag{8}$$

Different choices of the pull-in range can be accommodated by appropriately modifying the definition of $\mathcal{H}(\tau_0)$. In the rest of this section, we derive close approximations of the coarse synchronization performance of the four algorithms under consideration.

A. MF-Type Receivers ($2L$ Order)

We recall that in this case, $U_j = \bar{\mathbf{d}}_{\|0\|}^{(j)T} \bar{\mathbf{R}} \bar{\mathbf{d}}_{\|0\|}^{(j)}$ and $\hat{U}_j = \bar{\mathbf{d}}_{\|0\|}^{(j)T} \hat{\bar{\mathbf{R}}} \bar{\mathbf{d}}_{\|0\|}^{(j)}$, $j = 0, 1, \ldots, 2L-1$. Then, the CSER $P_{\text{cse,MF}}^{(2L)}$ of the $2L$-order MF algorithm can be expressed as the probability that the index of the largest decision variable $\hat{U}_j$, $j = 0, 1, \ldots, 2L-1$, is not contained in $\mathcal{H}(\tau_0)$; that is

$$P_{\text{cse,MF}}^{(2L)} = 1 - \Pr\left[\arg\max_j \left\{ \hat{U}_j | j = 0, \ldots, 2L-1 \right\} \in \mathcal{H}(\tau_0) \right] \tag{9}$$

$$= 1 - \Pr\left[\max\left\{ \hat{U}_k | k \in \mathcal{H}(\tau_0) \right\} > \max\left\{ \hat{U}_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0) \right\} \right]. \tag{10}$$

The latter form of $P_{\text{cse,MF}}^{(2L)}$ is particularly easy to handle, as we will see. We observe that the evaluation of the joint probability density function (pdf) of the random quantities $\hat{U}_j$, $j = 0, \ldots, 2L-1$, requires knowledge of the pdf of the random matrix $\hat{\bar{\mathbf{R}}}$. We recall that $\hat{\bar{\mathbf{R}}}$ is the sample average estimate of the autocorrelation matrix $\bar{\mathbf{R}}$ of the received vectors $\bar{\mathbf{r}}$, which are distributed according to a mixture of up to 2^{3K} Gaussian distributions, and they are, by construction, statistically dependent. Unfortunately, the use of exact closed-form expressions for the pdf of the sample covariance matrix that is formed by vector samples drawn from a Gaussian mixture distribution [8] leads to mathematically intractable expressions for the probability of synchronization error. An additional complicating factor for the analysis of the finite data-support behavior of the MF-type (and SMI-MVDR-type) receivers is the underlying dependence of the vector samples. In that respect, we make the following simplifying assumptions. We assume that the received vectors are uncorrelated and identically Gaussian $\mathcal{N}(\mathbf{0}, \bar{\mathbf{R}})$ distributed [thus independent and identically distributed (i.i.d.)] [9], [10]. The latter Gaussianity postulation can be considered as a wishful approximation of the true distribution, while the former independence postulation can be justified if we consider "guard" bands of size L between the received vectors, so that successive input vectors do not share common information bits. The effect of the above assumptions will be examined in the simulations section through comparisons with the exact P_{cse}.

It is known [11] that if $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1} \in \Re^p$ are i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ vectors, then the distribution of the matrix $\mathbf{X} \triangleq \sum_{i=0}^{N-1} \mathbf{x}_i \mathbf{x}_i^T$ is

$$f(\mathbf{X}) = \frac{|\mathbf{X}|^{(N-p-1)/2} \exp\{-\frac{1}{2}\text{tr}\boldsymbol{\Sigma}^{-1}\mathbf{X}\}}{2^{pN/2} \pi^{p(p-1)/4} |\boldsymbol{\Sigma}|^{N/2} \prod_{i=1}^{p} \Gamma\left[\frac{(N+1-i)}{2}\right]}, \quad p > N \tag{11}$$

where $\Gamma(\cdot)$ is the Gamma function. The distribution in (11) is called the Wishart distribution with N degrees of freedom (DOFs) and is denoted by $W_p(\Sigma; N)$. A random matrix distributed according to (11) is called a Wishart matrix. In our case, $\hat{\bar{\mathbf{R}}}$ is approximately Wishart with distribution $W_{2L}(\bar{\mathbf{R}}/N; N)$. If we define the matrix

$$\bar{\mathbf{D}} \triangleq \left(\bar{\mathbf{d}}_{\|0\|}^{(0)}, \bar{\mathbf{d}}_{\|0\|}^{(1)}, \ldots, \bar{\mathbf{d}}_{\|0\|}^{(2L-1)}\right) \tag{12}$$

then the random variables $\hat{U}_j$, $j = 0, \ldots, 2L-1$, form the main diagonal of $\bar{\mathbf{D}}^T \hat{\bar{\mathbf{R}}} \bar{\mathbf{D}}$. For a full-rank matrix, $\bar{\mathbf{D}}$, $\bar{\mathbf{D}}^T \hat{\bar{\mathbf{R}}} \bar{\mathbf{D}}$ is Wishart $W_{2L}((1/N)\bar{\mathbf{D}}^T \bar{\mathbf{R}} \bar{\mathbf{D}}; N)$ [12], and the main diagonal elements $\hat{U}_j$, $j = 0, \ldots, 2L-1$, are jointly distributed according to a multivariate Gamma distribution [13]. However, $\bar{\mathbf{D}}$ is not full rank in general, and even if it were, it would have been of little help since closed-form expressions of the multivariate Gamma pdf do not exist except for some special cases [14]. Thus, appropriate approximations of the joint pdf of the diagonal elements of $\bar{\mathbf{D}}^T \hat{\bar{\mathbf{R}}} \bar{\mathbf{D}}$ with tractable multivariate integrals are needed.

Motivated by the transformation noise-modeling techniques used in [15], we approximate the joint pdf of the MF decision variables $\hat{U}_j$, $j = 0, \ldots, 2L-1$, by the joint pdf of a nonlinear transformation of Gaussian random variables Y_j, $j = 0, \ldots, 2L-1$. The nonlinear transformation and the covariance of Y_j, $j = 0, \ldots, 2L-1$, are chosen such that the outputs of the nonlinear transformations have identical covariances with the corresponding $\hat{U}_j$, $j = 0, \ldots, 2L-1$, and marginal pdfs that are close approximations to the pdfs of $\hat{U}_j$, $j = 0, \ldots, 2L-1$. The following lemma identifies this nonlinear transformation and the statistics of the random variables Y_j, $j = 0, \ldots, 2L-1$. The proof is included in the Appendix.

Lemma 1: The pdf of $\hat{U}_j$ can be approximated by the pdf of $U_j Y_j^3$, $j = 0, \ldots, 2L-1$, where the random variables Y_j, $j = 0, \ldots, 2L-1$, are uncorrelated Gaussian with mean $1 - 2/9N$ and variance $2/9N$. ■

The probability of coarse synchronization error can now be evaluated as follows. Let the random variables Z_j, $j = 0, \ldots, 2L-1$ be defined as

$$Z_j = \frac{9N}{9N-2} U_j^{1/3} Y_j, \qquad j = 0, \ldots, 2L-1. \tag{13}$$

It is straightforward to show that

$$\begin{aligned}
&\Pr\left[\max\left\{\hat{U}_k | k \in \mathcal{H}(\tau_0)\right\}\right. \\
&\qquad \left. > \max\left\{\hat{U}_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\right\}\right] \\
&\simeq \Pr\left[\max\left\{U_k Y_k^3 | k \in \mathcal{H}(\tau_0)\right\}\right. \\
&\qquad \left. > \max\left\{U_j Y_j^3 | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\right\}\right] \\
&= \Pr\left[\max\left\{Z_k | k \in \mathcal{H}(\tau_0)\right\}\right. \\
&\qquad \left. > \max\left\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\right\}\right] \\
&= \Pr\left[\max\left\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\right\}\right. \\
&\qquad \left. - \max\left\{Z_k | k \in \mathcal{H}(\tau_0)\right\} < 0\right].
\end{aligned} \tag{14}$$

The random variables Z_j, $j = 0, \ldots, 2L-1$, are uncorrelated Gaussian (thus, independent) with mean $\mu_j = U_j^{1/3}$ and variance $\sigma_j^2 = (18N/(9N-2)^2) U_j^{2/3}$. The pdf of Z_j, $j = 0, \ldots, 2L-1$ is

$$f_j(z) = \frac{1}{\sigma_j} g\left(\frac{z-\mu_j}{\sigma_j}\right) \tag{15}$$

where $g(z) \triangleq 1/\sqrt{2\pi} \exp(-z^2/2)$ is the pdf of a Gaussian random variable with mean 0 and variance 1, while the cumulative distribution function (cdf) of Z_j is

$$\mathcal{F}_j(z) = \mathcal{G}\left(\frac{z-\mu_j}{\sigma_j}\right) \tag{16}$$

where $\mathcal{G}(z)$ is the cdf of a Gaussian random variable with mean 0 and variance 1. The cdf of $\max\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\}$ is given by [17]

$$\mathcal{F}_{\max, \backslash \mathcal{H}(\tau_0)}(z) \triangleq \prod_{j=0, j \notin \mathcal{H}(\tau_0)}^{2L-1} \mathcal{F}_j(z) \tag{17}$$

where the index "$\backslash \mathcal{H}(\tau_0)$" indicates that the maximum is taken over a set that does not include the elements of $\mathcal{H}(\tau_0)$. Therefore, the corresponding pdf is given by

$$f_{\max, \backslash \mathcal{H}(\tau_0)}(z) \triangleq \sum_{l=0, l \notin \mathcal{H}(\tau_0)}^{2L-1} \left(f_l(z) \prod_{j=0, j \notin \mathcal{H}(\tau_0), j \neq l}^{2L-1} \mathcal{F}_j(z) \right). \tag{18}$$

Similarly, the cdf and pdf of $\max\{Z_k | k \in \mathcal{H}(\tau_0)\}$ are given by

$$\mathcal{F}_{\max, \mathcal{H}(\tau_0)}(z) \triangleq \prod_{j \in \mathcal{H}(\tau_0)} \mathcal{F}_j(z) \tag{19}$$

and

$$f_{\max, \mathcal{H}(\tau_0)}(z) \triangleq \sum_{l \in \mathcal{H}(\tau_0)} \left(f_l(z) \prod_{j \in \mathcal{H}(\tau_0), j \neq l} \mathcal{F}_j(z) \right) \tag{20}$$

respectively. Since Z_j, $j = 0, \ldots, 2L-1$, are independent, the two random variables $\max\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\}$ and $\max\{Z_k | k \in \mathcal{H}(\tau_0)\}$ are also independent, therefore, the pdf of their difference $\max\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\} - \max\{Z_k | k \in \mathcal{H}(\tau_0)\}$ is given by

$$f_{\text{diff}}(z) \triangleq \int_{-\infty}^{+\infty} f_{\max, \backslash \mathcal{H}(\tau_0)}(z-y) f_{\max, \mathcal{H}(\tau_0)}(-y) dy. \tag{21}$$

We conclude that the probability of coarse synchronization error for the $2L$-long MF-type coarse synchronization algorithm is

$$P_{\text{cse,MF}}^{(2L)} \simeq 1 - \int_{-\infty}^{0} f_{\text{diff}}(z) dz \tag{22}$$

$$= 1 - \int_{-\infty}^{0} \int_{-\infty}^{+\infty} f_{\max, \backslash \mathcal{H}(\tau_0)}(z-y) \times f_{\max, \mathcal{H}(\tau_0)}(-y) dy dz \tag{23}$$

$$=1-\int_{-\infty}^{0}\int_{-\infty}^{+\infty}\left[\sum_{l=0,l\notin\mathcal{H}(\tau_0)}^{2L-1}\left(f_l(z-y)\times\prod_{j=0,j\neq l,j\notin\mathcal{H}(\tau_0)}^{2L-1}\mathcal{F}_j(z-y)\right)\right]\times\left[\sum_{l\in\mathcal{H}(\tau_0)}\left(f_l(-y)\prod_{j\in\mathcal{H}(\tau_0),j\neq l}\mathcal{F}_j(-y)\right)\right]dydz. \quad (24)$$

Expression (24) approximates the probability of coarse synchronization error as a double integral.

B. MF-Type Receivers (L Order)

In the case of the MF-type coarse synchronization algorithm that utilizes filters of order L, the probability of coarse synchronization error $P_{\text{cse,MF}}^{(L)}$ can be evaluated by (24), where $f_j(\cdot)$ and $\mathcal{F}_j(\cdot)$ denote the pdf and cdf, respectively, of Z_j, $j=0,1,\ldots,2L-1$ (Z_j is defined as in (13), with the difference that U_j is now given by $U_j=\mathbf{d}_{\|0\|}^{(j)^T}\mathbf{R}^{(\lfloor j/2\rfloor)}\mathbf{d}_{\|0\|}^{(j)}$, $j=0,1,\ldots,2L-1$). The proof follows similar reasoning as in Section III-A, with the difference that $\hat{U}_j=\mathbf{d}_{\|0\|}^{(j)^T}\hat{\mathbf{R}}^{(\lfloor j/2\rfloor)}\mathbf{d}_{\|0\|}^{(j)}$, $j=0,1,\ldots,2L-1$ and $\hat{\mathbf{R}}^{(\lfloor j/2\rfloor)}$ is Wishart $W_L((1/N)\mathbf{R}^{(\lfloor j/2\rfloor)};N)$ with N DOFs.

C. SMI-MVDR-Type Receivers (2L Order)

The MVDR-type SMI-implemented coarse synchronization algorithm that utilizes filters of order $2L$ can be analyzed in a similar manner. In this case, U_j is given by $U_j=\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(j)}\right)^{-1}$, $j=0,1,\ldots,2L-1$. If we define the random variables $Z_j=(9(N-2L)/(9(N-2L)-2))U_j^{1/3}Y_j$, $j=0,1,\ldots,2L-1$, where the random variables Y_j, $j=0,1,\ldots,2L-1$, are uncorrelated Gaussian with mean $1-2/9(N-2L)$ and variance $2/9(N-2L)$, then Z_j, $j=0,1,\ldots,2L-1$ are uncorrelated Gaussian random variables with mean $\mu_j=U_j^{1/3}$ and variance $\sigma_j^2=U_j^{2/3}(18(N-2L)/[9(N-2L)-2]^2)$. The probability of coarse synchronization error $P_{\text{cse,MVDR}}^{(2L)}$ is given by (24), where $\mathcal{F}_j(\cdot)$ and $f_j(\cdot)$ denote the cdf and pdf of the random variable Z_j, $j=0,1,\ldots,2L-1$. The proof again follows similar reasoning as in Section III-A, with the difference that $\hat{U}_j=\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\hat{\bar{\mathbf{R}}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(j)}\right)^{-1}$, $j=0,1,\ldots,2L-1$, and $\hat{\bar{\mathbf{R}}}$ is Wishart $W_{2L}((1/N)\bar{\mathbf{R}};N)$ with N DOFs. The latter implies that $\hat{U}_j$ can be expressed as a multiple of a χ^2 random variable with $N-2L$ DOFs.

D. SMI-MVDR-Type Receivers (L Order)

For the MVDR-type SMI-implemented coarse synchronization algorithm that utilizes filters of order L, we have $U_j=\left(\mathbf{d}_{\|0\|}^{(j)^T}\mathbf{R}^{(\lfloor j/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(j)}\right)^{-1}$, $j=0,1,\ldots,2L-1$. Then, the probabilty of coarse synchronization error $P_{\text{cse,MVDR}}^{(L)}$ is still given by (24), where $\mathcal{F}_j(\cdot)$ and $f_j(\cdot)$ denote the cdf and pdf of the random variable $Z_j=(9(N-L)/(9(N-L)-2))U_j^{1/3}Y_j$, $j=0,\ldots,2L-1$. The random variables Y_j, $j=0,1,\ldots,2L-1$, are uncorrelated Gaussian with mean $1-(2/9(N-L))$ and variance $(2/9(N-L))$ therefore, the random variables Z_j, $j=0,1,\ldots,2L-1$ are uncorrelated Gaussian random variables with mean $\mu_j=U_j^{1/3}$ and variance $\sigma_j^2=U_j^{2/3}(18(N-L)/[9(N-L)-2]^2)$. The proof follows similar reasoning as in Section III-A, with the only difference that $\hat{U}_j$ is now given by $\hat{U}_j=\left(\mathbf{d}_{\|0\|}^{(j)^T}\hat{\mathbf{R}}^{(\lfloor j/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(j)}\right)^{-1}$ and $\hat{\mathbf{R}}^{(\lfloor j/2\rfloor)^{-1}}$ is Wishart $W_L((1/N)\mathbf{R}^{(\lfloor j/2\rfloor)^{-1}};N)$ with N DOFs. The latter implies that $\hat{U}_j$ can be expressed as a multiple of a χ^2 random variable with $N-L$ DOFs.

IV. A Recursive Method for the Approximation of the Probability of Coarse Synchronization Error

In this section, we develop a recursive procedure for the evaluation of the probability of coarse synchronization error, P_{cse}. The method evaluates P_{cse} as a sequence of increasingly tight lower bounds on the approximate expression (24) and is computationally simpler than the latter expression that evaluates P_{cse} as a convolution of pdfs. We recall that (24) with appropriate definitions of $f_j(\cdot)$ and $\mathcal{F}_j(\cdot)$ can be used for the evaluation of $P_{\text{cse,MF}}^{(2L)}$, $P_{\text{cse,MF}}^{(L)}$, $P_{\text{cse,MVDR}}^{(2L)}$, and $P_{\text{cse,MVDR}}^{(L)}$ as discussed in Section III.

We begin with the observation that the cardinality of the set $\mathcal{H}(\tau_0)$ is almost always[1] equal to two. Thus, without loss of generality, we assume that $\mathcal{H}(\tau_0)=\{0,1\}$. In addition, we assume that the random variables Z_j, $j=2,\ldots,2L-1$ are ordered in terms of decreasing variance,[2] i.e., $\text{Var}\{Z_j\}\geq\text{Var}\{Z_{j+1}\}$ for $j=2,\ldots,2L-2$. Let

$$P_{\text{cse}}^{(k)}\triangleq 1-\Pr\left[\left(\arg\max_j\{Z_j|j=0,1,\ldots,k+1\}\right)\in\mathcal{H}(\tau_0)\right] \quad (25)$$

$$=1-\Pr[\max\{Z_0,Z_1\}>\max\{Z_2,Z_3,\ldots,Z_{k+2}\}] \quad (26)$$

where $k=0,\ldots,2L-3$. It is straightforward to check that for any $k=0,\ldots,2L-4$, we have $P_{\text{cse}}^{(k)}<P_{\text{cse}}^{(k+1)}$. Indeed, including the extra element Z_{k+3} to $\{Z_2,\ldots,Z_{k+2}\}$ can only increase the probability that the maximum element of $\{Z_2,\ldots,Z_{k+2},Z_{k+3}\}$ will be larger than $\max\{Z_0,Z_1\}$. Sorting the elements by decreasing variance implies that the k elements with the highest variance (and mean) will produce the tightest lower bound $P_{\text{cse}}^{(k)}$ of all other choices of k elements from $\{Z_2,\ldots,Z_{k+2}\}$. For $k=2L-3$, all variables Z_j, $j=0,\ldots,2L-1$, are included in the evaluation of the kth bound, so

$$P_{\text{cse}}^{(2L-3)}=1-\Pr[\max\{Z_0,Z_1\}>\max\{Z_1,Z_2,\ldots,Z_{2L-1}\}]. \quad (27)$$

[1]The cardinality is one if and only if $\tau_0=nT_c$ for some $n=0,\ldots,L-1$. However, the probability of these events is zero.

[2]Sorting the variables Z_j in terms of decreasing variance is equivalent to sorting them in terms of decreasing mean.

Thus [cf. (14)]

$$P_{\text{cse}}^{(0)} < P_{\text{cse}}^{(1)} < \cdots < P_{\text{cse}}^{(2L-3)} = 1 - \Pr\left[\max\{Z_j | j = 0, \ldots, 2L-1, j \notin \mathcal{H}(\tau_0)\} - \max\{Z_k | k \in \mathcal{H}(\tau_0)\} < 0\right]. \quad (28)$$

The probability bounds $P_{\text{cse}}^{(k)}$, $k = 0, \ldots, 2L-3$, can be evaluated as follows. The pdf of $\max\{Z_0, Z_1\}$ is

$$f_{\max\{Z_0,Z_1\}}(z) = f_0(z)F_1(z) + f_1(z)F_0(z) \quad (29)$$

while the pdf of $\max\{Z_2, \ldots, Z_{k+2}\}$ is

$$f_{\max\{Z_2,\ldots,Z_{k+2}\}}(z) \triangleq \sum_{l=2}^{k+2}\left(f_l(z)\prod_{j=2,\, j\neq l}^{k+2}\mathcal{F}_j(z)\right). \quad (30)$$

Then, the probability $P_{\text{cse}}^{(k)}$ is equal to

$$P_{\text{cse}}^{(k)} = 1 - \int_{-\infty}^{0}\int_{-\infty}^{+\infty} f_{\max\{Z_2,\ldots,Z_{k+2}\}}(z-y) \times f_{\max\{Z_0,Z_1\}}(-y)dydz. \quad (31)$$

Alternatively, a simpler recursive procedure for the evaluation of the probability bounds $P_{\text{cse}}^{(k)}$, $k = 0, \ldots, 2L-3$, is proposed below. The probability $P_{\text{cse}}^{(k)}$, $k = 0, \ldots, 2L-3$, is equal to

$$P_{\text{cse}}^{(k)} = 1 - P\left[Z_{\max}^{(k)} - Z_{\max}^{\{0,1\}} < 0\right] \quad (32)$$

where

$$Z_{\max}^{(k)} \triangleq \max\{Z_2, \ldots, Z_{k+2}\} \quad (33)$$

and

$$Z_{\max}^{\{0,1\}} \triangleq \max\{Z_0, Z_1\}. \quad (34)$$

Assuming that $Z_{\max}^{(k)}$ and $Z_{\max}^{\{0,1\}}$ are normally distributed, we may calculate $P_{\text{cse}}^{(k)}$ as follows:

$$P_{\text{cse}}^{(k)} \simeq \mathcal{G}\left(\frac{E\left\{Z_{\max}^{(k)}\right\} - E\left\{Z_{\max}^{\{0,1\}}\right\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k)}\right\} + \text{Var}\left\{Z_{\max}^{\{0,1\}}\right\}}}\right). \quad (35)$$

The mean and variance of $Z_{\max}^{(k)}$ can be evaluated recursively as described by the following lemma.

Lemma 2: For $k = 1, \ldots, 2L-3$

$$E\left\{Z_{\max}^{(k)}\right\} \triangleq E\left\{Z_{\max}^{(k-1)}\right\} \times \mathcal{G}\left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) + E\{Z_{k+2}\} \times \mathcal{G}\left(-\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) + \sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}} \times g\left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) \quad (36)$$

$$\text{Var}\left\{Z_{\max}^{(1)}\right\} \triangleq \text{Var}\left\{Z_{\max}^{(k-1)}\right\} \times \mathcal{G}\left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) + \text{Var}\{Z_{k+2}\} \times \mathcal{G}\left(-\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) + \left[\left(E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}\right)\right.$$
$$\mathcal{G} \times \left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) + \sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}} \times g\left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right)\Bigg] \times \Bigg[\left(E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}\right)$$
$$\mathcal{G} \times \left(-\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right) - \sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}} \times g\left(\frac{E\left\{Z_{\max}^{(k-1)}\right\} - E\{Z_{k+2}\}}{\sqrt{\text{Var}\left\{Z_{\max}^{(k-1)}\right\} + \text{Var}\{Z_{k+2}\}}}\right)\Bigg] \quad (37)$$

with $E\left\{Z_{\max}^{(0)}\right\} = E\{Z_2\}$ and $\text{Var}\left\{Z_{\max}^{(0)}\right\} = \text{Var}\{Z_2\}$.

Proof: Since

$$Z_{\max}^{(k)} = \max\left\{Z_2, \ldots, Z_{k+2}\right\} = \max\{Z_{\max}^{(k-1)}, Z_{k+2}\} \quad (38)$$

(36) and (37) provide, respectively, the mean and variance of the maximum of two uncorrelated Gaussian random variables [18]. ■

The mean and variance of $Z_{\max}^{\{0,1\}}$ can be evaluated in exactly the same way. As a final comment, we recall that (24) was derived based on the assumption that the pdf of the received vectors can be closely approximated by the pdf of i.i.d. Gaussian random vectors. The recursive method for the evaluation of P_{cse} developed in this section is based on the additional assumption that the maximum of two independent Gaussian random variables is approximately Gaussian. The effect of both assumptions is examined in the simulations section through comparisons with the exact P_{cse}.

TABLE III
DEFINITION OF $C(N)$

L-order	
SMI-MVDR	MF
$C(N)=\frac{18(N-L)}{[9(N-L)-2]^2}$	$C(N)=\frac{18N}{(9N-2)^2}$
$2L$-order	
SMI-MVDR	MF
$C(N)=\frac{18(N-2L)}{[9(N-2L)-2]^2}$	$C(N)=\frac{18N}{(9N-2)^2}$

V. DISCUSSION

Summarizing the developments so far, the probability of coarse synchronization error was approximated as follows:

$$P_{\text{cse}} \simeq 1-\Pr[\arg\max_j\{Z_j|\, j=0,\ldots,2L-1\}\in\mathcal{H}(\tau_0)] \tag{39}$$

where $\mathcal{H}(\tau_0)$ is the set that contains the indexes of the filters that yield a timing estimate within the pull-in range. The random variables $Z_0,\ldots,Z_{2L-1}$ are uncorrelated Gaussian random variables with means and variances given by

$$E\{Z_j\}=U_j^{1/3},\quad \text{Var}\{Z_j\}=C(N)U_j^{2/3} \tag{40}$$

where the quantities U_j, $j=0,\ldots,2L-1$, and the functions $C(N)$ are presented in Tables I and III, respectively, for each filter under consideration (SMI-MVDR or MF of order L or $2L$).

Thus, the probability of coarse synchronization error was calculated as the probability that the index of the largest Z_j, $j=0,\ldots,2L-1$, is not contained in $\mathcal{H}(\tau_0)$. This approach led to the P_{cse} expressions for the order-L and order-$2L$ MF-type or SMI-MVDR-type receivers presented in Section III, as well as to the evaluation of P_{cse} through the sequence of lower bounds presented in Section IV. An intuitively simpler but computationally more complex method for the calculation of P_{cse} is presented in the Appendix. Expression (40) and Table III imply that the mean and variance of Z_j, $j=0,\ldots,2L-1$, and, consequently, the performance of the SMI-MVDR-type algorithms depends explicitly on the filter order. Moreover, the data record size N and the filter order appear only in the terms $(N-2L)$ and $(N-L)$ of $C(N)$ for the filters of order $2L$ and L, respectively. Although the filter order does not affect the asymptotic (as $N\to\infty$) performance of the SMI-MVDR-type algorithms, it does cause a performance drop in the small data-record-size operating region, since the filter-order term is subtracted from the data-record-size term. In other words, the data-record size is effectively reduced by a number of data vector samples equal to the filter order. The longer the employed filter is, the greater the effective data-record-size reduction is. For the SMI-MVDR-type coarse synchronization algorithms, this translates to poorer performance under short data-support conditions as the filter order increases.

An alternative way to quantify the effect of the data-record size on the probability of coarse synchronization error is to examine the short data-record synchronization resolution of the SMI-MVDR-type algorithms that we define as follows. For the pair of decision variables (Z_l,Z_m), $l=0,\ldots,2L-1$, $m=0,\ldots,2L-1$, with $E\{Z_l\}\neq E\{Z_m\}$, the resolution metric is defined by

$$\rho_{l,m}\triangleq\begin{cases}\Pr[Z_l>Z_m], & \text{if } E\{Z_l\}>E\{Z_m\}\\ \Pr[Z_l<Z_m], & \text{otherwise.}\end{cases} \tag{41}$$

That is, the resolution $\rho_{l,m}$ is the probability that the coarse synchronization algorithm will decide in favor of the asymptotically largest decision variable. In other words, the synchronization resolution is a measure of how close the performance of the finite data-record SMI-based algorithm is to the asymptotic performance. It is straightforward to show that the resolution of the SMI-MVDR algorithms is given by

$$\rho_{l,m}=\mathcal{G}\left(C(N)^{-1/2}\frac{\left|U_l^{1/3}-U_m^{1/3}\right|}{\sqrt{U_l^{2/3}+U_m^{2/3}}}\right) \tag{42}$$

where $\mathcal{G}(x)$ is the cdf of a normal random variable with mean zero and variance one. The following proposition compares the resolution performance of the L-order and $2L$-order algorithms for a given pair of hypotheses. The proof is straightforward and thus omitted.

Proposition 1: For a given pair of hypotheses (l,m), $l=0,\ldots,2L-1$, $m=0,\ldots,2L-1$, we have

$$\rho_{l,m}^{(L)}>\rho_{l,m}^{(2L)} \tag{43}$$

if and only if

$$\frac{9(N-2L)-2}{9(N-L)-2}\frac{\sqrt{N-L}}{\sqrt{N-2L}}\le S_{l,m} \tag{44}$$

where $S_{l,m}$ is given by (45) as shown at the bottom of the page. ■

Let $S_{\min}\triangleq\min\{S_{l,m}|l,m=0,\ldots,2L-1\}$. Then, the resolution performance of the L-order algorithm will be better

$$S_{l,m}\triangleq\frac{\left|\left(\mathbf{d}_{\|0\|}^{(l)T}\mathbf{R}^{(\lfloor l/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(l)}\right)^{-1/3}-\left(\mathbf{d}_{\|0\|}^{(m)T}\mathbf{R}^{(\lfloor m/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(m)}\right)^{-1/3}\right|}{\left|\left(\bar{\mathbf{d}}_{\|0\|}^{(l)T}\bar{\mathbf{R}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(l)}\right)^{-1/3}-\left(\bar{\mathbf{d}}_{\|0\|}^{(m)T}\bar{\mathbf{R}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(m)}\right)^{-1/3}\right|}\times\frac{\sqrt{\left(\bar{\mathbf{d}}_{\|0\|}^{(l)T}\bar{\mathbf{R}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(l)}\right)^{-1/3}+\left(\bar{\mathbf{d}}_{\|0\|}^{(m)T}\bar{\mathbf{R}}^{-1}\bar{\mathbf{d}}_{\|0\|}^{(m)}\right)^{-2/3}}}{\sqrt{\left(\mathbf{d}_{\|0\|}^{(l)T}\mathbf{R}^{(\lfloor l/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(l)}\right)^{-2/3}-\left(\mathbf{d}_{\|0\|}^{(m)T}\mathbf{R}^{(\lfloor m/2\rfloor)^{-1}}\mathbf{d}_{\|0\|}^{(m)}\right)^{-2/3}}} \tag{45}$$

than the performance of the $2L$-order algorithm (in the sense that the L-order algorithm performance will be closer to the corresponding asymptotic performance) for *any* pair of hypotheses, provided that

$$\frac{9(N-2L)-2}{9(N-L)-2}\frac{\sqrt{N-L}}{\sqrt{N-2L}} \leq S_{\min}. \tag{46}$$

Equation (46) identifies a condition on the system parameters and the data-record size under which SMI-MVDR-type L-order algorithms outperform the $2L$-order algorithms.

On the other hand, in the case of MF-type algorithms, the mean and variance of Z_j do not depend explicitly on the filter order. Theoretically, this conclusion is a direct consequence of the fact that for any Wishart $W_p(\Sigma; N)$ matrix $\mathbf{X}$ and any nonzero vector $\mathbf{v} \in \Re^p$, the quantity $\mathbf{v}^T\mathbf{X}\mathbf{v}$ is a χ^2 random variable that maintains N DOFs (while $(\mathbf{v}^T\mathbf{X}^{-1}\mathbf{v})^{-1}$ is χ^2 with $N-p$ DOFs). Therefore, algorithms that do not require the inversion of a Wishart matrix do not suffer from effective data-record-size reduction. Consequently, it is expected that the MF-type and AV-type coarse synchronization algorithms are fairly insensitive to the choice of filter length.

As a final note, we recall that an alternative to direct inversion of $\hat{\mathbf{R}}^{(l)}$ is the use of a matrix-inversion lemma-based recursive least-squares (RLS) algorithm [16]. We can show that the RLS algorithm computes the inverse of a diagonally loaded version of $\hat{\mathbf{R}}^{(l)}$, $(\hat{\mathbf{R}}^{(l)} + (\epsilon/N)\mathbf{I})^{-1}$, where $(1/\epsilon)\mathbf{I}$ is the initialization matrix of RLS. Therefore, although RLS provides an asymptotically $(N \to \infty)$ unbiased estimate of $(\mathbf{R}^{(l)})^{-1}$, for a finite data-record size N it provides a biased estimate of $(\mathbf{R}^{(l)})^{-1}$ with a bias that depends on ϵ/N. Since in practice, ϵ is small, the effect of the diagonal loading quickly becomes negligible and, consequently, the performance is close to the SMI performance analyzed in this paper.

VI. Simulation Results and Comparisons

We consider a 10-user asynchronous DS-CDMA system that utilizes Gold sequences of length $L = 31$. We compare the CSER evaluated using the expressions derived in Sections III and IV with the exact CSER of MF and SMI-MVDR-type algorithms. The exact CSER is evaluated by averaging over 10 000 independent runs. The delays of all users are chosen randomly and kept constant for the duration of the experiment.

In Figs. 3 and 4, the SNR[3] of the user of interest is 10 dB and the data support size ranges from 35 to 80 samples. The SNRs of the interferers are fixed at 6, 8, 10, 12, 15, 16, 18, 20, and 22 dB. In Fig. 3, we plot as a function of k the sequence of bounds $P_{\text{cse}}^{(k)}$ developed in Section IV for the order L and order $2L$ SMI-MVDR-type algorithms [given by (35)–(37)]. In Fig. 4, we compare the exact coarse synchronization performance of order-L and order-$2L$ algorithms to the derived approximations of Sections III and IV, that is, the direct approximation through (24) and the recursive approximation through (35)–(37). Regardless of the filter order, we see that both approximations offer close coarse synchronization performance estimates. We observe that the SMI-MVDR L-order algorithm exhibits improved

[3]The SNR of user k, $k = 0, \ldots, K-1$, is defined as E_k/σ^2, where σ^2 is the variance of the chip-matched filtered and sampled AWGN.

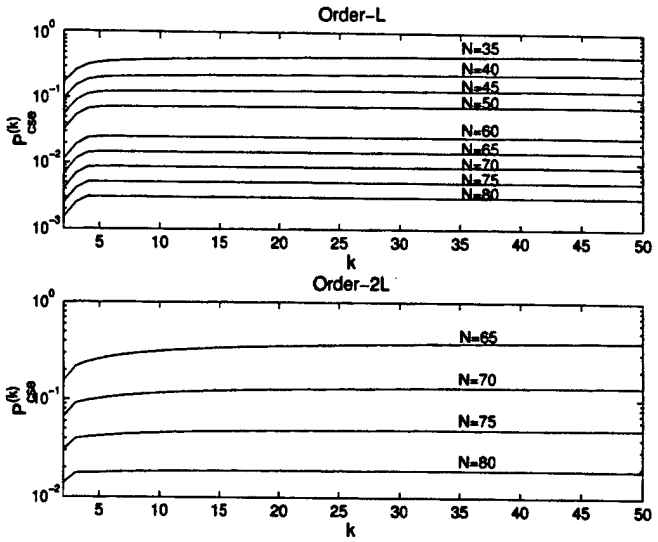


Fig. 3. Recursive approximation of the CSER of SMI-MVDR-type receivers of order L and order $2L$ given by (35)–(37) ($\text{SNR}_0 = 10$ dB).

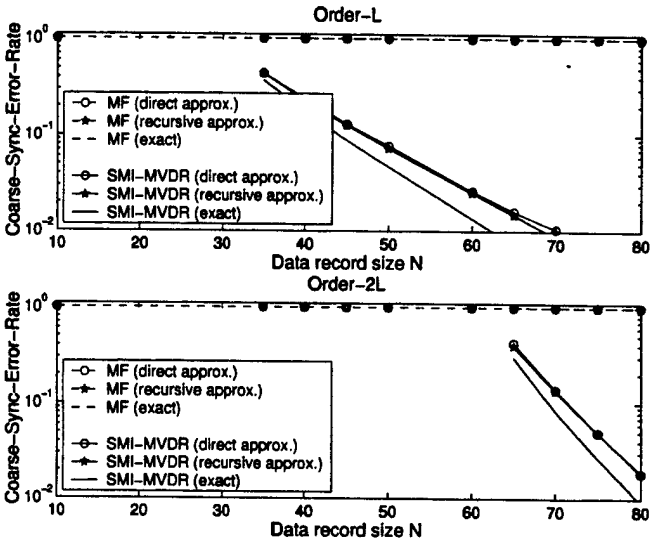


Fig. 4. CSER of linear receivers of order L and $2L$ as a function of the data-record size N ($\text{SNR}_0 = 10$ dB).

performance when compared with the SMI-MVDR $2L$-order scheme for data-record sizes above $L = 31$. In fact, in the small data support region, the L-order algorithm can achieve the same CSER as the $2L$-order algorithm, about L samples faster.

In Figs. 5 and 6, we repeat the studies of Figs. 3 and 4 for different values of the SNR of the user of interest. The data support for the estimation of the order-L and order-$2L$ SMI-MVDR filters is $N = 55$ and 80 samples, respectively. The results parallel the findings in Figs. 3 and 4.

The insensitivity of the noninversion-based coarse synchronization algorithms to the filter order is illustrated by the simulation study of Fig. 7, where the probability of coarse synchronization error of the L-order algorithms is plotted as a function of the system processing gain L (the figure also includes the AV-type coarse synchronization scheme presented in [7]). The user SNR values are identical to those in Fig. 3 and the data support for filter estimation is $N = 125$ samples. The signature vectors of the users are constructed by concatenating Gold

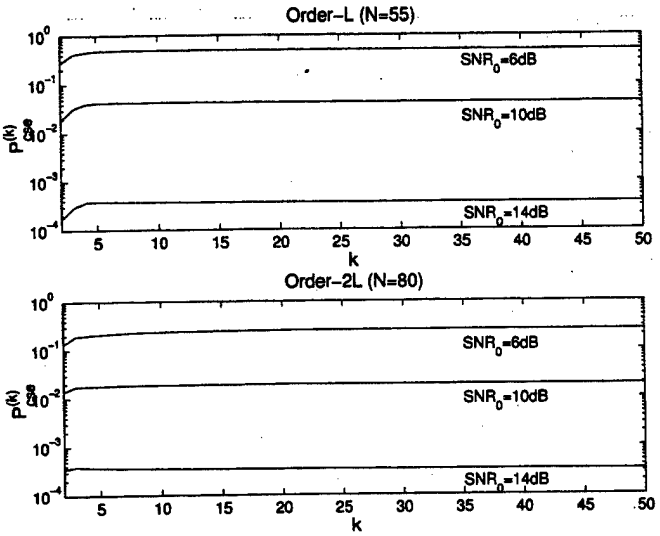


Fig. 5. Recursive approximation of the CSER of SMI-MVDR-type receivers of order L and order $2L$ given by (35)–(37) ($N = 55$ and 80, respectively).

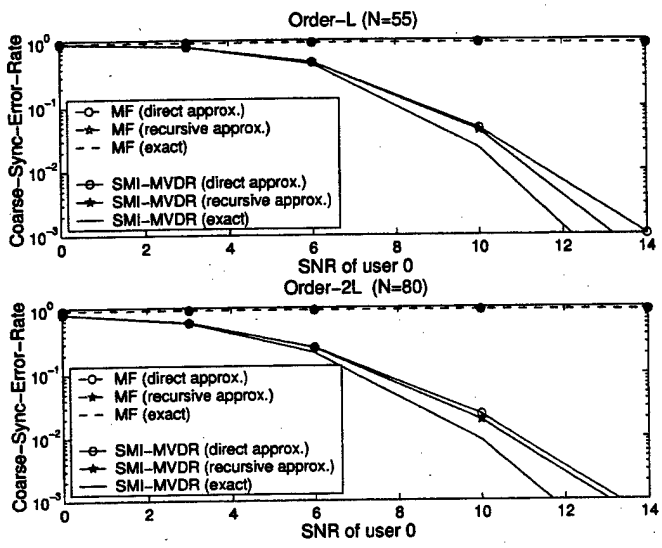


Fig. 6. CSER of linear receivers of order L and $2L$ as a function of the SNR of the user of interest ($N = 55$ and 80, respectively).

codes of length 31. We observe that the SMI-MVDR algorithm is severely affected by the value of the spreading gain (filter length) as L approaches N, while the MF-type algorithm is not affected (but has unacceptable synchronization performance, as also seen in Figs. 4 and 6). The AV-type synchronizer does not suffer from significant performance degradation as the system spreading gain L increases, and offers by far the most reliable synchronization.

VII. Conclusions

In this paper, we investigated the finite data-record coarse synchronization performance of blind adaptive combined synchronizers/demodulators. Using transformation noise modeling techniques, we derived analytical expressions that approximate closely the probability of coarse synchronization error. Three alternative methods were developed. The first method approximates the probability of coarse synchronization error as

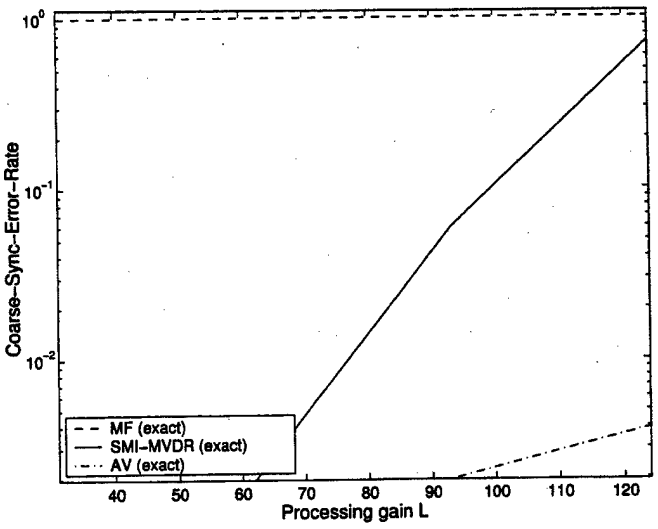


Fig. 7. CSER of linear receivers of order L as a function of the processing gain L.

a double integral. The second approximation method, which is the least computationally complex, provides a sequence of increasingly tight lower bounds on the approximate expressions of the first method. Finally, the third and most intuitive method approximates the probability of coarse synchronization error as a $2L$-order integral (the expression is given in Appendix). The analytical expressions provide simple, highly accurate alternatives to computationally demanding performance evaluation through simulations. We showed that the coarse synchronization performance of the MF-type receiver is a function of the data-record size, while the performance of MVDR-type SMI or RLS-based filter estimators is a function of the difference between the data-record size and the employed filter order. The latter translates to an effective reduction of the data-record size with a more evident negative effect on the performance as the available data record becomes smaller. MF-type and AV-type (which provide active interference suppression) filter estimators do not suffer such data reduction and thus, provide an attractive solution when the environment dictates short data-record synchronization.

Appendix I

A. Proof of Lemma 1

The marginal pdf of $\hat{U}_j$, $j = 0, \ldots, 2L-1$, is

$$h_j(u) = \frac{x^{N/2-1} \exp\left\{\frac{-u}{U_j}\right\}}{\left(\frac{1}{N}U_j\right)^{N/2} \Gamma\left(\frac{N}{2}\right)}. \tag{47}$$

Expression (47) implies that the random variable $\hat{U}_j$, $j = 0, \ldots, 2L-1$, is a multiple of a chi-square (χ^2) random variable. More precisely, $\hat{U}_j$ can be written as

$$\hat{U}_j = \left(\frac{1}{N}U_j\right) X_j, \quad j = 0, \ldots, 2L-1 \tag{48}$$

where X_j is a χ^2 random variable with N DOFs. Since $(X_j/N)^{1/3}$ is approximately Gaussian with mean

$(1-(2/9N))$ and variance $2/9N$ [19], we may approximate the pdf of $\hat{U}_j$ by the pdf of

$$V_j \triangleq N\left(\frac{1}{N}U_j\right)Y_j^3 = U_jY_j^3, \quad j=0,\ldots,2L-1. \quad (49)$$

In the above equation, Y_j is a Gaussian variable with mean $(1-(2/9N))$ and variance $2/9N$. The joint pdf of $Y_0,\ldots,Y_{2L-1}$ will be completely defined if we specify their covariance matrix that we denote by Ψ. The diagonal elements of Ψ are equal to $2/9N$. The nondiagonal elements of Ψ are chosen such that $V_0,\ldots,V_{2L-1}$ have covariances identical to the covariances of the corresponding $\hat{U}_0,\ldots,\hat{U}_{2L-1}$ variables. The latter covariances are [20]

$$\mathrm{Cov}\left[\hat{U}_j,\hat{U}_k\right] = \frac{2}{N}\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(k)}\right)^2. \quad (50)$$

To find the covariances ψ_{jk} of Y_j and Y_k, $j\neq k$, that result in

$$\mathrm{Cov}\left[V_j,V_k\right] = \frac{2}{N}\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(k)}\right)^2 \quad (51)$$

we substitute (49) in (51) and make use of the moment theorem to evaluate $E\left\{Y_j^3Y_k^3\right\}$. Then, we find that the (j,k)th nondiagonal element of Ψ, ψ_{jk}, must satisfy

$$6\psi_{jk}^3+18\left(1-\frac{2}{9N}\right)^2\psi_{jk}^2+9\frac{(4-18N-81N^2)^2}{6561N^4}\psi_{jk} = \frac{2}{N}\frac{\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(k)}\right)^2}{\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(j)}\right)\left(\bar{\mathbf{d}}_{\|0\|}^{(k)^T}\bar{\mathbf{R}}\bar{\mathbf{R}}_{\|0\|}^{(k)}\right)} \quad (52)$$

or, equivalently

$$\psi_{jk}^3+3\left(1-\frac{2}{9N}\right)^2\psi_{jk}^2 +9\frac{(4-18N-81N^2)^2}{6\times 6561N^4}\psi_{jk}-\frac{1}{3N}\alpha=0 \quad (53)$$

where

$$\alpha \triangleq \left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(k)}\right)^2 \Big/ \left(\left(\bar{\mathbf{d}}_{\|0\|}^{(j)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(j)}\right)\left(\bar{\mathbf{d}}_{\|0\|}^{(k)^T}\bar{\mathbf{R}}\bar{\mathbf{d}}_{\|0\|}^{(k)}\right)\right).$$

By the Schwarz inequality $\alpha\in(0,1)$. Equation (53) has three real roots, two of which are always negative, provided that $N>4$. The third root is plotted in Fig. 8 as a function of N and α. We see that the third root is always positive while it drops rapidly to zero as N increases. In fact, for $N>40$, we see that ψ_{jk} is less than 0.0045. Thus, we may assume that $\psi_{jk}=0$, $j\neq k$, $j,k=0,\ldots,2L-1$. In other words, we can safely assume that the Gaussian random variables Y_j, $j=0,\ldots,2L-1$, are uncorrelated and, therefore, independent. ■

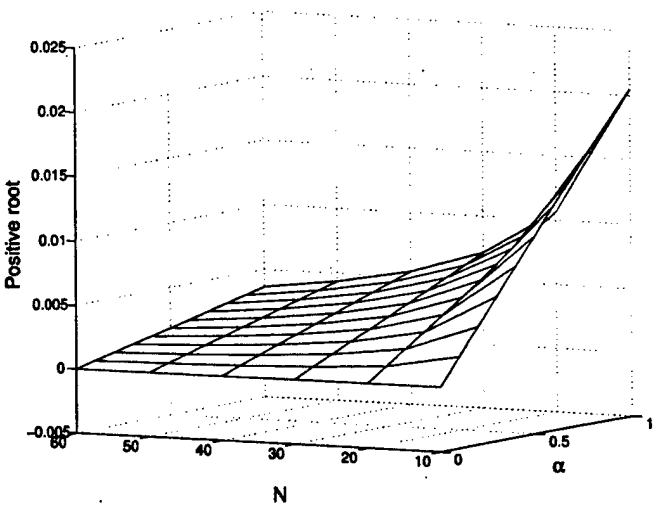


Fig. 8. Positive root of (53) as a function of α and data-record size N.

APPENDIX II

A. An Alternative Approximation of P_{cse}

The probability of coarse synchronization error can be approximated by [cf. (14)]

$$P_{cse}\simeq 1-\Pr[\arg\max\{Z_j|j=0,\ldots,2L-1\}\in\mathcal{H}(\tau_0)] \quad (54)$$

where $\mathcal{H}(\tau_0)$ is the set that contains the indexes of the filters that yield a timing estimate within the pull-in range. The random variables Z_j, $j=0,\ldots,2L-1$, are independent Gaussian with mean $U_j^{1/3}$ and variance $C(N)U_j^{2/3}$, where the function $C(N)$ is given in Table III for each of the receivers under consideration. The pdf of the random vector $\mathbf{Z}\triangleq[Z_0,\ldots,Z_{2L-1}]^T$ is

$$f_{\mathbf{Z}}(\mathbf{z}) \triangleq \frac{1}{(2\pi)^L\prod_{j=0}^{2L-1}\sqrt{C(N)U_j^{2/3}}} \times\exp\left\{-\sum_{j=0}^{2L-1}\frac{\left(z_j-U_j^{1/3}\right)^2}{2C(N)U_j^{2/3}}\right\} \quad (55)$$

where $\mathbf{z}=[z_0,\ldots,z_{2L-1}]^T$. Thus

$$P_{cse}\simeq 1-\int_{\mathcal{D}_0}f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} \quad (56)$$

where

$$\mathcal{D}_0=\bigcup_{k\in\mathcal{H}(\tau_0)}\Big\{(d_0,d_1,\ldots,d_{2L-1})\in\Re^{2L}\text{ with } d_k>d_j, j=0,\ldots,2L-1, j\neq k\Big\}. \quad (57)$$

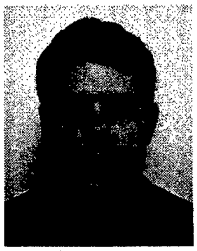
■

REFERENCES

[1] D. A. Pados and S. N. Batalama, "Low-complexity blind detection of DS-CDMA signals: Auxiliary vector receivers," *IEEE Trans. Commun.*, vol. 45, pp. 1586–1594, Dec. 1997.

[2] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1765–1773, Dec. 1998.

[3] D. A. Pados and S. N. Batalama, "Joint space–time auxiliary-vector filtering for DS-CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406–1415, Sept. 1999.
[4] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Processing*, vol. 49, pp. 290–300, Feb. 2001.
[5] S. Verdu, "Adaptive multiuser detection," in *Proc. IEEE Int. Symp. Spread Spectrum Techniques and Applications*, vol. 1, 1994, pp. 43–50.
[6] R. Kohno, P. B. Rapajic, and B. S. Vucetic, "Overview of adaptive techniques for interference minimization in CDMA systems," *Wireless Pers. Commun.*, vol. 1, pp. 3–21, 1994.
[7] I. N. Psaromiligkos and S. N. Batalama, "Rapid combined synchronization/demodulation structures for DS-CDMA systems—Part I: Algorithmic developments," *IEEE Trans. Commun.*, vol. 51, pp. 983–994, June 2003.
[8] W. Y. Tan, "On the distribution of the sample covariance matrix from a mixture of normal densities," *South African Stat. J.*, vol. 12, pp. 47–55, 1978.
[9] S. E. Bensley and B. Aazhang, "Maximum-likelihood synchronization of a single user for code-division multiple-access communication systems," *IEEE Trans. Commun.*, vol. 46, pp. 392–399, Mar. 1998.
[10] Z. Liu, J. Li, and S. L. Miller, "An efficient code-timing estimator for receiver diversity DS-CDMA systems," *IEEE Trans. Commun.*, vol. 46, pp. 826–835, June 1996.
[11] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.
[12] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[13] P. R. Krishnaiah and M. M. Rao, "Remarks on a multivariate gamma distribution," *Amer. Math. Monthly*, vol. 68, pp. 342–346, 1961.
[14] T. Royen, "Expansions for the multivariate chi-square distribution," *J. Multivariate Anal.*, vol. 38, pp. 213–232, 1991.
[15] S. N. Batalama, M. J. Medley, and I. N. Psaromiligkos, "Adaptive robust spread-spectrum receivers," *IEEE Trans. Commun.*, vol. 47, pp. 905–917, June 1999.
[16] L. C. Godara, "Application of antenna arrays to mobile communications—Part II: Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1195–1245, Aug. 1997.
[17] H. A. David, *Order Statistics*. New York: Wiley, 1981.
[18] O. Kella, "On the distribution of the maximum of bivariate normal random variables with general means and variances," *Commun. Statist. A—Theory Methods*, vol. 15, pp. 3265–3276, 1986.
[19] N. C. Severo and M. Zelen, "Normal approximation to the chi-square and noncentral F probability functions," *Biometrika*, vol. 47, pp. 411–416, 1960.
[20] L. R. Haff, "An identity for the Wishart distribution with applications," *J. Multivariate Anal.*, vol. 9, pp. 531–544, 1979.



Ioannis N. Psaromiligkos (S'96–M'03) received the Diploma degree in computer engineering and science from the University of Patras, Patras, Greece in 1995, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo in 1997 and 2001, respectively.

From 1995 to 1997, he was a Teaching Assistant, and from 1997 to 2001, a Research Assistant with the Communications and Signals Group in the Department of Electrical Engineering at the State University of New York at Buffalo. Since 2001, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, McGill University, Montréal, QC, Canada. His research interests are in the areas of adaptive multiuser detection and wireless multiple access communications.

Stella N. Batalama (S'91–M'94) received the Diploma degree in computer engineering and science from the University of Patras, Patras, Greece, in 1989, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1994.

From 1989 to 1990, she was with the Computer Technology Institute, Patras, Greece. From 1990 to 1994, she was a Research Assistant in the Communication Systems Laboratory, Department of Electrical Engineering, University of Virginia. In 1995, she joined the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY, where she is presently an Associate Professor. During the summers of 1997–2003, she was Visiting Faculty in the U.S. Air Force Research Laboratory, Rome, NY. Her research interests include small sample support adaptive filtering and receiver design, adaptive multiuser detection, robust spread-spectrum communications, supervised and unsupervised optimization, and distributed detection.

Dr. Batalama is currently an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS.

Data Record Size Requirements for Adaptive Space-Time DS/CDMA Signal Detection*

Ioannis N. Psaromiligkos
Department of Electrical & Comp. Engineering
McGill University
Montréal, Québec, Canada H3A-2A7
E-Mail: yannis@ece.mcgill.ca

Stella N. Batalama†
Department of Electrical Engineering
State University of New York at Buffalo
Buffalo, NY 14260, USA
E-Mail: batalama@eng.buffalo.edu
Tel: (716) 645-2422 ext. 2164
Fax: (716) 645-3656

Abstract

We investigate the data record size requirements of sample-matrix-inversion-based minimum-variance-distortionless-response and maximum signal-to-interference-plus-noise-ratio adaptive algorithms to meet a given performance objective in joint space-time signal detection problems for direct-sequence code-division-multiple-access systems. We derive closed form expressions that provide the data record size that is necessary to achieve a given performance confidence level in a neighborhood of the optimal performance point as well as expressions that identify the performance level that can be reached for a given data record size. This is done by utilizing close approximations of the involved probability density functions. The practical significance of the derived expressions lies in the fact that the expressions are functions of the number of antenna elements, the number of multipaths, and the system spreading gain only, while they depend neither on the ideal input covariance matrix which is not known in most realistic applications nor on the exact ideal performance value.

Keywords: Antenna arrays, code division multiaccess, direction of arrival estimation, finite data analysis, interference suppression.

*This work was presented in part at the 2000 Conf. on Information Sciences and Systems, Princeton University, Princeton, NJ, March 2000. This work was supported in part by the National Science Foundation under Grants ECS-0073660 and CCR-0219903, and in part by the Air Force Office of Scientific Research under Grant F49620-01-1-0176.

†Corresponding author.

I. Introduction

In this paper we focus on direct-sequence code-division-multiple-access (DS/CDMA) single-user detection. We consider multipath fading additive white Gaussian noise (AWGN) channels and antenna array signal reception, either at the base station or at the mobile end (for example airborne arrays)[1]. We recall that single-user DS/CDMA detection aims at the recovery of the information bits of one user that we call "the user of interest." The only quantity assumed known is the effective (space-time channel processed) signature of the user of interest which is a function of the multipath fading coefficients, delays, and multipath directions of arrival (the interfering signal statistics are assumed unknown).

We consider a receiver that consists of either the minimum-variance-distortionless-response (MVDR) or the maximum-signal-to-interference-plus-noise ratio (MSINR) linear filter followed by a sign detector. The MVDR linear filter is found by minimizing the filter output variance subject to the constraint that the filter remains distortionless in the effective signature vector direction of the user of interest [1], [2]. The filter is a function of the inverse input autocovariance matrix and the effective signature of the user of interest. On the other hand, the MSINR filter is found by maximizing the signal-to-interference-plus-noise ratio at the linear filter output and it is a function of the inverse interference-plus-noise-only autocovariance matrix and the effective signature of the user of interest. The MVDR and MSINR linear filters are scaled versions of each other and, thus, under perfectly known statistics they exhibit identical performance in terms of both filter output SINR and receiver bit-error-rate (BER).

The sample-matrix-inversion (SMI) estimate of the MVDR or the MSINR filter is obtained by substituting, respectively, the inverse of the ideal autocovariance matrix of the input or the interference-plus-noise part only by the inverse of the corresponding sample average estimate. The corresponding filter estimate is called SMI-MVDR or SMI-MSINR

[1]Single-antenna signal detection systems constitute a special case of the general system that we consider here.

filter. These estimates are occasionally referred to in the literature as SMI-"signal-present" or SMI-"signal-absent" filter, respectively. Under asymptotically many data, the SMI-MVDR filter and the SMI-MSINR filter exhibit identical performance in terms of both filter output SINR and receiver BER since they converge in probability, under general conditions, to their ideal counterparts. This is not, however, the case when estimation with finite data records is performed where for small data record sizes the SMI-MSINR outperforms the SMI-MVDR filter in terms of output SINR. Along these lines, we note that the sample-average estimate of the interference-plus-noise autocovariance matrix can be evaluated from data obtained either by probing the channel prior to the beginning of the desired transmission (or during the silent periods of the user of interest) or by estimating and then subtracting the desired user signal from the received signal. In multipath environments the interference component in the former, desired-signal-absent, case does not contain inter-symbol-interference (ISI) and thus is not exactly the same as the interference component in the latter, "estimate and subtract", case (which contains ISI). However, in spread-spectrum systems the channel delay spread is much less than the information symbol period and we can safely ignore the effect of ISI and treat all cases in the same way.

The goal of this paper is twofold. We would like to identify the data record size that is necessary for *(a)* the SMI-MVDR DS/CDMA multiple-access-interference (MAI) suppression filter, and *(b)* the SMI-MSINR DS/CDMA MAI suppression filter to achieve a given performance confidence level in a neighborhood of the optimal performance point. Performance is measured in terms of filter output variance and filter output SINR, respectively. We also investigate the dual problem of quantifying the relative performance of the above estimators with respect to the performance of the corresponding ideal schemes for a given data record size. In pursue of these objectives we define the following relative performance measures for the estimated filters with respect to their ideal counterparts in one-to-one correspondence with (a)-(b) above: *(a)* The ratio of the output variance of the ideal filter over the output variance of the estimated filter, and *(b)* the ratio of the output SINR of the estimated filter

over the output SINR of the ideal filter. Our approach is based on the observation that a near optimum performance level is reached by any estimation scheme when the corresponding ratio is close to unity. Thus, *any given (or desired)* confidence on the performance level can be expressed as the probability that the above measures are within a certain neighborhood of unity. This approach is in contrast to the work in [3] where the measures (a) and (b) were used to determine the data record that the estimated filter needs to achieve *on average* a given percentage of the optimum performance. In addition, in this paper, making use of the functional relationship between the BER and the SINR of a linear receiver, we extend our developments summarized above to evaluate our confidence on the BER performance of DS/CDMA single-user detectors that utilize the SMI-MVDR or SMI-MSINR filter, respectively. The practical significance of the theoretical developments in this paper lies in the fact that by utilizing simple yet close approximations of the involved probability density functions we are able to derive closed form expressions that are functions of the number of antenna elements, the number of multipaths, and the system processing gain and, most importantly, independent of the ideal input statistics.

The paper is organized as follows. In Section II we present the system model. In Section III we outline the MVDR and MSINR joint space-time processors, we analyze the performance of their SMI estimators as a function of the data record size, and we derive their data-record-size requirements to achieve a given performance level. A few conclusions are drawn in Section IV.

II. System Model

We consider a DS/CDMA system with K users transmitting over a multipath fading additive white Gaussian noise (AWGN) channel. The information bits of each user are binary-phase-shift-keying (BPSK) modulated. The kth user baseband transmitted signal is given by

$$u_k(t) = \sum_i b_k(i)\sqrt{E_k}S_k(t - iT), \quad k = 0, \ldots, K-1, \tag{1}$$

where $b_k(i) \in \{-1, 1\}$ is the ith data (information) bit and T is the information bit period. E_k represents transmitted energy per bit and $S_k(t)$ is the signature waveform, assumed to be normalized to unit energy per data bit period, given by

$$S_k(t) = \sum_{l=0}^{L-1} s_k(l)\psi(t - lT_c) \tag{2}$$

where $s_k(l) \in \{-1, 1\}$ is the lth bit of the spreading sequence of the kth user, T_c is the chip period, $\psi(t)$ is the chip waveform (assumed to be a rectangular pulse with support $[0, T_c]$), and $L = T/T_c$ is the system spreading gain. The kth user signal $u_k(t)$ is transmitted over a multipath fading channel with impulse response

$$h_k(t) = \sum_{n=0}^{N_p-1} c_{k,n}\delta(t - \tau_{k,n}) \tag{3}$$

where N_p is the number of paths (assumed the same for all users), $c_{k,n}$, $k = 0, \ldots, K-1$, $n = 0, \ldots, N_p - 1$ are the complex path coefficients, and $\delta(t)$ is the Dirac delta function. The coefficients are assumed to be identical across all antenna elements, that is, antenna diversity effects are not considered. Finally, $\tau_{k,n}$ is the relative transmission delay of the nth path of the kth user signal. For simplicity in presentation, we assume that the receiver is coarsely synchronized with the first path of the user of interest, *user 0*, that the channel delay spread is equal to N_p chips and that no resolvable path delays fall within the same chip period, i.e.

$$\tau_{k,n} = \nu_{k,n} + \delta_{k,n}, \;\; k = 0, \ldots, K-1, \;\; n = 0, \ldots, N_p - 1, \tag{4}$$

where $\nu_{k,n}$ is an integer between 0 and $L-1$ (with $\nu_{0,n} = n$) and $\delta_{k,n}$ lies in $[0, 1)$.

The received signal is collected by a uniform linear antenna array consisting of M elements spaced half-the-wavelength apart. The baseband received signal at the mth antenna element is given by

$$r_m(t) \;\; = \;\; \sum_{k=0}^{K-1}\sum_{n=0}^{N_p-1} c_{k,n}u_k(t - \tau_{k,n})e^{-jm\pi \sin\theta_{k,n}} + n_m(t), \;\; m = 0, \ldots, M-1, \tag{5}$$

where $\theta_{k,n}$ identifies the angle of arrival of the nth path of the kth user signal and $n_m(t)$ represents additive sensor noise modeled as temporally and spatially complex white Gaussian. The received signals $r_0(t), \ldots, r_{M-1}(t)$ can be grouped to form the vector

$$\mathbf{r}(t) \triangleq [r_0(t), r_1(t), \ldots, r_{M-1}(t)]^T = \sum_{k=0}^{K-1}\sum_{n=0}^{N_p-1} c_{k,n}u_k(t - \tau_{k,n})\mathbf{a}(\theta_{k,n}) + \mathbf{n}(t) \tag{6}$$

where

$$\mathbf{a}(\theta_{k,n}) \triangleq [1, e^{-j\pi \sin\theta_{k,n}}, \ldots, e^{-j(M-1)\pi \sin\theta_{k,n}}]^T, \quad k = 0, \ldots, K-1, \quad n = 0, \ldots, N_p - 1, \quad (7)$$

is the steering vector associated with the nth path of the kth user and

$$\mathbf{n}(t) \triangleq [n_0(t), \ldots, n_{M-1}(t)]^T. \quad (8)$$

Chip-matched filtering and sampling at the chip rate, $1/T_c$, of $\mathbf{r}(t)$ over the multipath extended time interval ($L + N_p$ chip periods) prepares the data for one-shot detection of the ith information bit of interest $b_0(i)$. By stacking the vector samples $\mathbf{r}(0), \ldots, \mathbf{r}((L+N_p-1)T_c)$ one below the other we obtain the space-time received data vector

$$\mathbf{r}_{M(L+N_p)\times 1} \triangleq [\mathbf{r}(0)^T \ \mathbf{r}(T_c)^T \ \ldots \mathbf{r}^T((L+N_p-1)T_c)]^T. \quad (9)$$

From now on, $\mathbf{r}$ denotes the joint S-T data in the $\mathcal{C}^{M(L+N_p)}$ complex vector domain.

The cornerstone of joint space-time filtering is the space-time signature which for *user 0* is defined as

$$\mathbf{v}_0 \triangleq \sum_{n=0}^{N_p-1} c_{0,n} \left[(1-\delta_{0,n})\mathbf{s}_0^{(n)} + \delta_{0,n}\mathbf{s}_0^{(n+1)}\right] \otimes \mathbf{a}(\theta_{0,n}) \quad (10)$$

where $\mathbf{s}_0^{(n)} \triangleq [\underbrace{0, \ldots, 0}_{n}, s_0(0), \ldots, s_0(L-1), \underbrace{0, \ldots, 0}_{N_p-n}]^T$ and $\otimes$ denotes the Kronecker product.

In the next section we outline the MVDR and MSINR DS/CDMA joint space-time processors and we analyze the performance of the SMI-MVDR and SMI-MSINR estimators as a function of the data record size. Also, we derive their data-record-size requirements to achieve a given performance level.

III. Data-Record-Size Requirements of Space-Time MVDR and MSINR Receivers

A linear joint S-T DS/CDMA interference suppressing receiver with tap weight vector $\mathbf{w} \in \mathcal{C}^{M(L+N_p)}$ detects the transmitted bit of the user of interest as follows:

$$\hat{b}_0 = \text{sgn}(Re\{\mathbf{w}^H\mathbf{r}\}) \quad (11)$$

where $\hat{b}_0$ denotes the detected information bit of the user of interest, $sgn(\cdot)$ identifies the sign operation, and $Re\{\cdot\}$ extracts the real part of a complex number. In this work we

consider the MVDR and MSINR linear receivers. The tap-weight vector of the MVDR filter is designed to minimize the filter output variance while maintaining unit response in the known[2] vector direction $\mathbf{v}_0$ and equals

$$\mathbf{w}_{MVDR} = \frac{\mathbf{R}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H\mathbf{R}^{-1}\mathbf{v}_0} \tag{12}$$

where $\mathbf{R} \triangleq E\{\mathbf{r}\mathbf{r}^H\}$ is the covariance matrix of the received vector $\mathbf{r}$ ($E\{\cdot\}$ denotes the expectation operation). The MVDR filter output variance is

$$\mathcal{V}(\mathbf{w}_{MVDR}) \triangleq E\{|\mathbf{w}_{MVDR}^H\mathbf{r}|^2\} = \mathbf{w}_{MVDR}^H\mathbf{R}\mathbf{w}_{MVDR}. \tag{13}$$

The tap weight vector of the MSINR filter is found by maximizing the SINR of the linear filter output and is given by

$$\mathbf{w}_{MSINR} = \frac{\mathbf{R}_{I+n}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H\mathbf{R}_{I+n}^{-1}\mathbf{v}_0} \tag{14}$$

where $\mathbf{R}_{I+n} \triangleq \mathbf{R} - E_0\mathbf{v}_0\mathbf{v}_0^H$ is the "interference-plus-noise" covariance matrix. The SINR at the output of the MSINR filter is given by

$$\mathcal{S}(\mathbf{w}_{MSINR}) \triangleq \frac{E_0|\mathbf{w}_{MSINR}^H\mathbf{v}_0|^2}{\mathbf{w}_{MSINR}^H\mathbf{R}_{I+n}\mathbf{w}_{MSINR}}. \tag{15}$$

It can be shown that in ideal situations (i.e., when $\mathbf{R}$ and $\mathbf{R}_{I+n}$ are perfectly known) $\mathbf{w}_{MVDR}$ and $\mathbf{w}_{MSINR}$ exhibit identical performance in terms of both output SINR and receiver BER. However, knowledge of the ideal covariance matrices $\mathbf{R}$ or $\mathbf{R}_{I+n}$ cannot be assumed in practice and, thus, the matrices have to be estimated. We choose to estimate $\mathbf{R}$ and $\mathbf{R}_{I+n}$ by sample averaging[3]. The sample average estimate of $\mathbf{R}$ based on a data record of size N is given by

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{j=0}^{N-1}\mathbf{r}_j\mathbf{r}_j^H \tag{16}$$

while the sample average estimate of $\mathbf{R}_{I+n}$ based on a data record of size N is given by

$$\widehat{\mathbf{R}}_{I+n} = \frac{1}{N}\sum_{j=0}^{N-1}\mathbf{r}_{j,I+n}\mathbf{r}_{j,I+n}^H \tag{17}$$

[2]We note that the vector direction $\mathbf{v}_0$ is the only quantity assumed known (path coefficients, path delays and directions of arrival are modeled as deterministic unknowns).

[3]Other estimators of $\mathbf{R}$ and $\mathbf{R}_{I+n}$ may also be used instead of the sample average estimators. For example, at the base station we can estimate ("reconstruct") the matrices $\mathbf{R}$ and $\mathbf{R}_{I+n}$ using estimates of the energies, channel coefficients, delays, and angles of arrival of all users. However, the performance of the corresponding MVDR and MSINR filter estimator will strongly depend on the quality of all intermediate estimators used and, thus, it will be difficult to analyze under finite data record scenarios.

where the index $(I+n)$ in $\mathbf{r}_{j,I+n}$, $j=0,\ldots,N-1$, identifies desired-signal-free received vectors. To guarantee that the estimates $\widehat{\mathbf{R}}$ or $\widehat{\mathbf{R}}_{I+n}$ are of full rank with probability 1 (w.p. 1) so that the estimates $\left[\widehat{\mathbf{R}}\right]^{-1}$ and $\left[\widehat{\mathbf{R}}_{I+n}\right]^{-1}$ exist, the data record size N must satisfy[4] $N \geq M(L+N_p)$. We also note that the received vectors $\mathbf{r}_j$ used for the computation of $\widehat{\mathbf{R}}$ are distributed according to a mixture of up to 2^{3K} complex Gaussian distributions[5]. Unfortunately, exact closed form expressions for the probability density function (pdf) of the sample covariance matrix that is formed by vector samples drawn from a Gaussian mixture distribution are very complicated even for the real case [4]. An additional complicating factor is the underlying dependence of the vector samples $\mathbf{r}_j$ due to multipath propagation. To that extend, we make the following simplifying assumption. We assume that the received vectors are independent and identically distributed according to a multivariate complex Gaussian distribution $\mathcal{N}(\mathbf{0},\mathbf{R})$. Then, the estimator $\widehat{\mathbf{R}}$ is complex Wishart distributed with N degrees of freedom, $\mathcal{W}_{M(L+N_p)}(\mathbf{R}/N;N)$ [5]. Similarly, we assume that $\widehat{\mathbf{R}}_{I+n}$ is distributed according to a Wishart distribution with N degrees of freedom of the form $\mathcal{W}_{M(L+N_p)}(\mathbf{R}_{I+n}/N;N)$. Numerical studies presented later in this section will demonstrate the effect of the above assumptions on the developments of this paper.

The SMI estimator of the MVDR filter is obtained by substituting (16) into (12), i.e.

$$\widehat{\mathbf{w}}_{SMI-MVDR} = \frac{\widehat{\mathbf{R}}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H\widehat{\mathbf{R}}^{-1}\mathbf{v}_0}. \tag{18}$$

The variance at the SMI-MVDR filter output is

$$\mathcal{V}(\widehat{\mathbf{w}}_{SMI-MVDR}) = \frac{\mathbf{v}_0^H\widehat{\mathbf{R}}^{-1}\mathbf{R}\widehat{\mathbf{R}}^{-1}\mathbf{v}_0}{\left(\mathbf{v}_0^H\widehat{\mathbf{R}}^{-1}\mathbf{v}_0\right)^2}. \tag{19}$$

We understand that the variance $\mathcal{V}(\widehat{\mathbf{w}}_{SMI-MVDR})$ is lower bounded by $\mathcal{V}(\mathbf{w}_{MVDR}) = \left(\mathbf{v}_0^H\mathbf{R}^{-1}\mathbf{v}_0\right)^{-1}$ which is the output variance of the ideal MVDR receiver in (12). This implies that the *vari-*

[4]For a large class of multivariate elliptically contoured input distributions that includes the Gaussian, if $N \geq M(L+N_p)$, then $\widehat{\mathbf{R}}$ (or $\widehat{\mathbf{R}}_{I+n}$) is positive definite (hence invertible) w.p. 1 [6]-[8].

[5]The total number of Gaussian distributions is equal to 2^x, where x is the total number of information bits contributed by all users to each received vector. In our case; user k, $k=0,\ldots,K-1$, contributes to the received vector $\mathbf{r}_j$ two bits ($b_k(j-1)$ and $b_k(j)$) if $\min_n\{\tau_{k,n}\} > N_pT_c$ and three bits ($b_k(j-1)$, $b_k(j)$, and $b_k(j+1)$), otherwise. Thus, the maximum value of x is $3K$.

ance ratio $\rho_{\mathcal{V}}$ defined as

$$\rho_{\mathcal{V}} \triangleq \mathcal{V}(\mathbf{w}_{MVDR})/\mathcal{V}(\widehat{\mathbf{w}}_{SMI-MVDR}) \tag{20}$$

is upper bounded by unity ($\rho_{\mathcal{V}} \leq 1$). Near-optimum performance of the SMI-MVDR filter estimator is attained at a value of $\rho_{\mathcal{V}}$ close to unity.

Similarly, the SMI estimator of the MSINR filter is obtained by substituting (17) into (14) and is given by

$$\widehat{\mathbf{w}}_{SMI-MSINR} = \frac{\widehat{\mathbf{R}}_{I+n}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H\widehat{\mathbf{R}}_{I+n}^{-1}\mathbf{v}_0}. \tag{21}$$

The SINR at the SMI-MSINR filter output is

$$\mathcal{S}(\widehat{\mathbf{w}}_{SMI-MSINR}) \triangleq \frac{E_0(\mathbf{v}_0^H\widehat{\mathbf{R}}_{I+n}^{-1}\mathbf{v}_0)^2}{\mathbf{v}_0^H\widehat{\mathbf{R}}_{I+n}^{-1}\mathbf{R}_{I+n}\widehat{\mathbf{R}}_{I+n}^{-1}\mathbf{v}_0}. \tag{22}$$

The SINR $\mathcal{S}(\widehat{\mathbf{w}}_{SMI-MSINR})$ is upper-bounded by $\mathcal{S}(\mathbf{w}_{MSINR}) = E_0(\mathbf{v}_0^H\mathbf{R}_{I+n}^{-1}\mathbf{v}_0)$ which is the output SINR of the ideal MSINR filter. Thus, near optimum performance of the $\widehat{\mathbf{w}}_{SMI-MSINR}$ filter requires that the *SINR ratio* $\rho_{\mathcal{S}}$ defined as

$$\rho_{\mathcal{S}} \triangleq \mathcal{S}(\widehat{\mathbf{w}}_{SMI-MSINR})/\mathcal{S}(\mathbf{w}_{MSINR}) \tag{23}$$

is close to its maximum value of unity ($\rho_{\mathcal{S}} \leq 1$).

Given the number of antenna elements M, our goal is to derive conditions on the data record size N necessary for the random variables $\rho_{\mathcal{S}}$ and $\rho_{\mathcal{V}}$ to be sufficiently close to 1 in the following probabilistic sense:

$$Pr[\rho_{\mathcal{V}} > 1-\zeta] \geq \epsilon \quad \text{for any given } \zeta \in (0,1) \text{ and } \epsilon \in (0,1) \tag{24}$$

and

$$Pr[\rho_{\mathcal{S}} > 1-\zeta] \geq \epsilon \quad \text{for any given } \zeta \in (0,1) \text{ and } \epsilon \in (0,1). \tag{25}$$

The performance measures in (24) and (25) assign a minimum confidence ϵ to the neighborhood $(1-\zeta, 1]$ of the optimum performance point ($\rho_{\mathcal{V},opt} = \rho_{\mathcal{S},opt} = 1$). The following theorem identifies the performance level that can be reached by the SMI-MVDR and SMI-MSINR filter for a given data record size, as well as the least number of samples that guarantees (24) and (25).

Theorem 1 *(i) For any given* $\zeta \in (0,1)$ *and* $N \geq M(L+N_p)$ *we have*

$$Pr[\rho_{\mathcal{V}} > 1-\zeta] \simeq 1 - Q\left(\frac{\zeta N - M(L+N_p)+2}{\sqrt{N\zeta(1-\zeta)}}\right) \tag{26}$$

and

$$Pr[\rho_S > 1-\zeta] \simeq 1 - Q\left(\frac{\zeta N - M(L+N_p)+2}{\sqrt{N\zeta(1-\zeta)}}\right). \tag{27}$$

(ii) For any given $\zeta \in (0,1)$ *and* $\epsilon \in (0,1)$*, the least number of samples that guarantees* $Pr[\rho_V > 1-\zeta] \geq \epsilon$ *or* $Pr[\rho_S > 1-\zeta] \geq \epsilon$ *can be approximated by*

$$N^* = \begin{cases} \max\left\{M(L+N_p), \left\lceil \frac{(1-\zeta)\gamma^2+2[M(L+N_p)-2]-\sqrt{\gamma^2(1-\zeta)[4M(L+N_p)-8+(1-\zeta)\gamma^2]}}{2\zeta} \right\rceil\right\} & \text{if } \epsilon < 0.5 \\ \max\left\{M(L+N_p), \left\lceil \frac{(1-\zeta)\gamma^2+2[M(L+N_p)-2]+\sqrt{\gamma^2(1-\zeta)[4M(L+N_p)-8+(1-\zeta)\gamma^2]}}{2\zeta} \right\rceil\right\} & \text{if } \epsilon \geq 0.5 \end{cases} \tag{28}$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$, $\gamma \triangleq Q^{-1}(1-\epsilon)$, $\lceil x \rceil$ *denotes the smallest integer larger than or equal to* x*, and* M*,* L *and* N_p *denote the number of antenna elements, the system processing gain, and the number of resolvable paths, respectively.*

Proof: We can show that ρ_V and ρ_S are identically distributed according to a Beta distribution [3] [11] with parameters $N - M(L+N_p) + 2$ and $M(L+N_p) - 1$. We note that the distribution of ρ_S was used in [12] to assess the data record size requirements of the spatial SMI-MVDR beamformer by *visual inspection* of the Beta function itself. In [3], the distribution of ρ_S was used to evaluate the data record size that is necessary for the spatial SMI-MVDR beamformer to make the *expected value* of ρ_S greater than a given constant.

For simplicity in presentation, we denote comprehensively both ρ_V and ρ_S by ρ. The probability density function (pdf) of ρ is given by

$$f_\rho(x) = \frac{1}{B\{N - M(L+N_p)+2,\ M(L+N_p)-1\}} x^{N-M(L+N_p)+1}(1-x)^{M(L+N_p)-2},\ x \in [0,1] \tag{29}$$

where $B\{\cdot,\cdot\}$ is the complete Beta function[6]. Then, $Pr[\rho > 1-\zeta] = \int_{1-\zeta}^1 f_\rho(x)dx$. But [9]

$$\int_{1-\zeta}^1 f_\rho(x)dx = 1 - \sum_{i=N-M(L+N_p)+2}^{N} \binom{N}{i} (1-\zeta)^i \zeta^{N-i}. \tag{30}$$

The right hand side of (30) is the probability that a binomially distributed random variable is less than $N - M(L+N_p) + 2$. Using the Gaussian approximation [9] of the binomial cumulative distribution function we can express (24) or (25) as

$$1 - Q\left(\frac{\zeta N - M(L+N_p)+2}{\sqrt{N\zeta(1-\zeta)}}\right) \geq \epsilon \tag{31}$$

[6]The complete Beta function is defined as $B(x,y) \triangleq \int_0^1 t^{x-1}(1-t)^{y-1}dt$ [9],[10].

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-y^2/2} dy$. Solving (31) with respect to N we find that the least number of samples (greater than $M(L + N_p)$) that guarantees a minimum performance confidence level in the neighborhood of the optimal point as described by (24) and (25) is approximately equal to N^* given by (28). ■

For the special case of $\epsilon = 0.5$ we have $\gamma = 0$. In this case, (28) implies that the least number of samples to achieve $Pr[\rho_{\mathcal{V}} > 1-\zeta] \geq 0.5$ or $Pr[\rho_{\mathcal{S}} > 1-\zeta] \geq 0.5$ is approximately $N^* = [M(L + N_p) - 2]/\zeta$. If, in addition, $\zeta = 0.5$ then $N^* = 2[M(L + N_p) - 2]$. In plain words, this example implies that to be $\epsilon = 50\%$ confident that the output variance of the SMI-MVDR filter (or the output SINR of the SMI-MSINR filter) will be within $\zeta = 50\%$ (-3dB) of the ideal MVDR (or MSINR) filter performance we need to evaluate the SMI estimators using at least $N^* = 2[M(L+N_p)-2]$ data samples. We note that $1-Pr[\rho_{\mathcal{V}} > x]$ and $1 - Pr[\rho_{\mathcal{S}} > x]$ are merely the cumulative distribution functions (cdfs) of the random variables $\rho_{\mathcal{V}}$ and $\rho_{\mathcal{S}}$, respectively. Use of the cdfs allows the evaluation of any statistic of the corresponding variables (e.g. the mean and beyond). As such, the work in [3] appears as a special case of our developments in Theorem 1 (we recall that [3] evaluates the data record size that is necessary for the spatial SMI-MVDR beamformer to achieve *on average (mean)* a given percentage of the optimum performance). In Table I we evaluate the data record size N^* needed by the SMI-MVDR filter and the SMI-MSINR filter to operate with confidence ϵ within ζ of the optimum performance for different values of ϵ and ζ. A DS/CDMA system with a processing gain equal to $L = 31$ is assumed. The number of paths is set $N_p = 3$ and the receiver antenna array utilizes $M = 5$ antenna elements. We see, for example, that for $\epsilon = 0.75$ and $\zeta = 0.25$ we have $N^* = 683$. In other words, SMI estimation of the MVDR filter based on 683 samples guarantees that 75% of the time the performance of the filter estimate will be at most 25% (-6dB) worse than the performance of the ideal filter. In Fig. 1(a) we plot comprehensively the probability $Pr[\rho > 1 - \zeta]$ as a function of ζ and the data record size N, while in Fig. 1(b) we show the value of N^* as a function of ϵ and ζ.

We emphasize that the domain of the random variable $\rho_{\mathcal{V}}$ is different from the domain

of ρ_S since ρ_V and ρ_S are based on different input statistics; the former is the ratio of the output variance of the ideal MVDR filter over the output variance of the (estimated) SMI-MVDR filter, while the latter is the ratio of the output SINR of the (estimated) SMI-SINR filter over the output SINR of the ideal MSINR filter. The fact, however, that they are identically distributed allows us to treat them uniformly with the understanding that the probabilities in the left hand side of (26) and (27) (which correspond to different events) are approximately equal to the *same value* that is given by the right hand side of (26) and (27). Similarly, (28) identifies the approximate number of data required to have the *same level* of confidence ϵ that the distance of the *different* performance measures ρ_V and ρ_S will fall in a neighborhood of radius ζ of the optimal performance point $\rho_{V,opt} = \rho_{S,opt} = 1$.

The data record size N^* in (28) guarantees that $Pr[\rho_V > 1-\zeta] \geq \epsilon$ and $Pr[\rho_S > 1-\zeta] \geq \epsilon$ hold true for any given ζ and ϵ. N^* was obtained by solving the approximate expressions given by (26) and (27) with respect to the data record size. The approximate expressions (26) and (27) were obtained by using the Gaussian approximation of the binomial cumulative distribution under the assumption that $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_{I+n}$ are complex Wishart distributed. To examine both the effect of this assumption on the distribution of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_{I+n}$ as well as the accuracy of the Gaussian approximation used, we compare the approximate with the exact (obtained numerically) probabilities $Pr[\rho_V > 1-\zeta] \geq \epsilon$ and $Pr[\rho_S > 1-\zeta] \geq \epsilon$, respectively. In Fig. 2 we plot the aforementioned quantities as a function of the data-record-size N for $\zeta = 0.2$. As in Fig. 1, an antenna array consisting of $M = 5$ elements is assumed, while the system processing gain L is equal to 31 and $N_p = 3$. The accuracy of the complex Wishart assumption and the Gaussian approximation is evident. It may be interesting to note that the probability density function (pdf) of ρ_V and ρ_S depends only on N, M, L and N_p and it does not depend on the covariance matrix $\mathbf{R}$ or $\mathbf{R}_{I+n}$. This observation implies that (28) is valid regardless of the actual environment, i.e. the number of users, the signal powers, the path coefficients, the path delays, the directions of arrival, etc.

In digital communications systems the ultimate performance measure is the bit-error-rate

(BER), that is the probability of producing an incorrect decision on the information bit b_0, $Pr[\hat{b}_0 \neq b_0]$. In the following, we examine how the performance measures in (24) and (25), which are based on the filter output variance and the filter output SINR, respectively, translate into BER terms. Under the assumption that the received vector is Gaussian distributed (the same assumption that led to the Wishart distribution of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_{I+n}$), the BER $P_e(\mathbf{w})$ exhibited by an arbitrary linear filter $\mathbf{w}$ can be expressed as follows[7]:

$$P_e(\mathbf{w}) \simeq Q\left(\sqrt{\mathcal{S}(\mathbf{w})}\right) \tag{32}$$

where $\mathcal{S}(\cdot)$ is given by

$$\mathcal{S}(\mathbf{w}) \triangleq \frac{E_0|\mathbf{w}^H\mathbf{v}_0|^2}{\mathbf{w}^H\mathbf{R}_{I+n}\mathbf{w}}. \tag{33}$$

Theorem 2 and Corollary 1 presented below identify the relationship between the BER performance and the performance measures in (24) and (25). The proofs are given in the Appendix.

Theorem 2 *Let the ideal MVDR filter have a BER region of operation given by* $(10^{-\lambda}, 10^{-\nu})$ *for some* $\lambda > \nu$. *If for a given* $\zeta \in (0,1)$ *and* $\epsilon \in (0,1)$

$$Pr[\rho_{\mathcal{V}} > 1-\zeta] \geq \epsilon, \tag{34}$$

then

$$P\left[\frac{\log P_e(\widehat{\mathbf{w}}_{SMI-MVDR}) - \beta}{\log P_e(\mathbf{w}_{MVDR})} > \alpha\right] \geq \epsilon \tag{35}$$

where

$$\alpha = \frac{10^{\mu}\left(\sqrt{\frac{(1-\zeta)[Q^{-1}(10^{\mu})]^2}{1+\zeta[Q^{-1}(10^{\mu})]^2}}\right)^3 \exp\left\{-\frac{(1-\zeta)[Q^{-1}(10^{\mu})]^2}{2(1+\zeta[Q^{-1}(10^{\mu})]^2)}\right\}}{(1-\zeta)F_{\zeta}(10^{\mu})[Q^{-1}(10^{\mu})]^3 \exp\left\{-[Q^{-1}(10^{\mu})]^2/2\right\}}, \tag{36}$$

$$\beta = \log F_{\zeta}(10^{\mu}) - \alpha\mu, \tag{37}$$

$$\mu = \frac{1}{2}(\lambda+\nu), \text{ and} \tag{38}$$

$$F_{\zeta}(x) \triangleq Q\left(\sqrt{\frac{(1-\zeta)[Q^{-1}(x)]^2}{1+\zeta[Q^{-1}(x)]^2}}\right). \tag{39}$$

■

Corollary 1 *Let the ideal MSINR filter have a BER region of operation given by* $(10^{-\lambda}, 10^{-\nu})$ *for some* $\lambda > \nu$. *If for a given* $\zeta \in (0,1)$ *and* $\epsilon \in (0,1)$

$$Pr[\rho_{\mathcal{S}} > 1-\zeta] \geq \epsilon, \tag{40}$$

[7]We recall that the data bits are BPSK modulated.

then

$$P\left[\frac{\log P_e(\widehat{\mathbf{w}}_{SMI-MSINR}) - \beta}{\log P_e(\mathbf{w}_{MSINR})} > \alpha\right] \geq \epsilon \tag{41}$$

where

$$\alpha = \frac{10^{\mu}\sqrt{1-\zeta}\exp\{\frac{\zeta}{2}[Q^{-1}(10^{\mu})]^2\}}{F_{*\zeta}(10^{\mu})}, \tag{42}$$

$$\beta = \log F_{*\zeta}(10^{\mu}) - \alpha\mu, \tag{43}$$

$$\mu = \frac{1}{2}(\lambda+\nu), \text{ and} \tag{44}$$

$$F_{*\zeta}(x) \triangleq Q\left(\sqrt{1-\zeta}\, Q^{-1}(x)\right). \tag{45}$$

■

Expression (35) shows that if the data-record-size is chosen to be greater than N^* in (28), then the *exponent*, $\log P_e(\widehat{\mathbf{w}}_{SMI-MVDR})$, of the BER performance of the estimated MVDR receiver will be with probability at least ϵ in the region $\left(\log P_e(\mathbf{w}_{MVDR}),\ \alpha \log P_e(\mathbf{w}_{MVDR}) + \beta\right)$ where the scalars α and β are given by (36), (37). Similarly, expression (41) shows that if the data-record-size is chosen to be greater than N^* in (28), then the *exponent*, $\log P_e(\widehat{\mathbf{w}}_{SMI-MSINR})$, of the BER performance of the estimated MSINR receiver will be with probability at least ϵ in the region $\left(\log P_e(\mathbf{w}_{MSINR}),\ \alpha \log P_e(\mathbf{w}_{MSINR}) + \beta\right)$, where the scalars α and β are given by (42), (43). We note that expressions (35) and (41) are valid regardless of the actual environment, i.e. the number of users, the signal powers, the path coefficients, the path delays, the directions of arrival, etc.

In Table II we show the region $\left(\log P_e(\mathbf{w}_{MVDR}), \alpha \log P_e(\mathbf{w}_{MVDR}) + \beta\right)$ and $\left(\log P_e(\mathbf{w}_{MSINR}),\ \alpha \log P_e(\mathbf{w}_{MSINR}) + \beta\right)$ that the exponent of the BER performance of the SMI-MVDR and the SMI-MSINR receiver, respectively, lies with confidence at least ϵ, for different values of ζ and ϵ. As in prior studies, $L = 31$, $M = 5$, $N_p = 3$. We see, for example, that in the case of the SMI-MVDR filter, 683 samples guarantee that 75% of the time the output variance will be at most 25% $(-6dB)$ worse than the output variance of the ideal filter. Thus, 683 samples guarantee that 75% of the time the exponent $\log P_e(\widehat{\mathbf{w}}_{SMI-MVDR})$ will be in the interval $\left(\log P_e(\mathbf{w}_{MVDR}), 0.16 \log(P_e(\mathbf{w}_{MVDR})) - 0.7\right)$. Similarly, in the case of the SMI-MSINR

receiver 683 samples guarantee that 75% of the time the exponent $\log P_e(\widehat{\mathbf{w}}_{SMI-MSINR})$ will be in the interval $\left(\log P_e(\mathbf{w}_{MSINR}), 0.78\log(P_e(\mathbf{w}_{MSINR})) - 0.1\right)$.

We note that while the left end-point, $\log P_e(\mathbf{w}_{MVDR})$ or $\log P_e(\mathbf{w}_{MSINR})$, of the SMI-MVDR or the SMI-MSINR performance range attains identical value, the right-end value of the performance range of the SMI-MVDR receiver is larger than the corresponding value of the SMI-MSINR receiver. This observation suggests the superiority of the SMI-MSINR filter relative to the SMI-MVDR. Field engineers have also observed this behavior in the past and occassionally they identify it as the superiority of the SMI-"signal absent" estimator relative to the performance of the SMI-"signal present" estimator. The superiority of the SMI-MSINR filter is quantified in [13] where the pdfs of the output SINR and the induced BER of the SMI-MVDR and SMI-MSINR filter estimators are evaluated and a simple "estimate-and-subtract" algorithm is proposed for desired signal cancellation prior to filter estimation.

IV. Conclusions

We investigated the data-record-size requirements of SMI-type adaptive algorithms to meet a given performance objective in joint S-T interference suppressing DS/CDMA signal detection. We considered the SMI-MVDR and the SMI-MSINR filter estimator and we adopted the following figures of merit. For the SMI-MVDR estimator the figure of merit is the ratio between the output variance of the ideal filter (MVDR) and the estimated filter (SMI-MVDR); for the SMI-MSINR estimator the figure of merit is the ratio of the output SINR of the estimated filter (SMI-MSINR) over the output SINR of the ideal filter (MSINR). For both cases, closed form expressions were derived that provide either the data record size that is necessary to achieve a given performance confidence level in a neighborhood of the optimal performance point or the performance level that can be reached for a given data record size. This was achieved by utilizing close approximations of the involved probability density functions. We emphasize that the derived expressions do not depend on the actual

value of the ideal performance and, thus, knowledge of this value is not required. In addition, the functional relationship between the BER and the SINR exhibited by a linear receiver under the Gaussian input assumption enabled us to translate the data record size requirements of the signal detection algorithms into BER terms. The practical significance of the derived expressions lies in the fact that the expressions are functions only of the number of antenna elements, the system spreading gain, and the number of multipaths, while they are independent of the ideal input covariance matrix which is not known in most realistic applications.

Appendix

A Proof of Theorem 2

It is straightforward to verify that the output variance $\mathcal{V}(\mathbf{w})$ and the output SINR $\mathcal{S}(\mathbf{w})$ of an arbitrary linear filter $\mathbf{w}$ for *user 0* (distortionless in the vector direction $\mathbf{v}_0$) are related through the expression

Thus,
$$\mathcal{V}(\mathbf{w}) = E_0\left(1 + \frac{1}{\mathcal{S}(\mathbf{w})}\right). \tag{46}$$

$$\mathcal{V}(\widehat{\mathbf{w}}_{SMI\text{-}MVDR}) = E_0\left(1 + \frac{1}{\mathcal{S}(\widehat{\mathbf{w}}_{SMI\text{-}MVDR})}\right) \quad \text{and} \quad \mathcal{V}(\mathbf{w}_{MVDR}) = E_0\left(1 + \frac{1}{\mathcal{S}(\mathbf{w}_{MVDR})}\right). \tag{47}$$

Therefore, the inequality $\rho_{\mathcal{V}} > 1 - \zeta$ is equivalent to

$$\mathcal{S}(\widehat{\mathbf{w}}_{SMI\text{-}MVDR}) > \frac{(1-\zeta)\mathcal{S}(\mathbf{w}_{MVDR})}{1 + \zeta\mathcal{S}(\mathbf{w}_{MVDR})}, \tag{48}$$

which, in turn, is equivalent to

$$P_e(\widehat{\mathbf{w}}_{SMI\text{-}MVDR}) < Q\left(\sqrt{\frac{(1-\zeta)\mathcal{S}(\widehat{\mathbf{w}}_{SMI\text{-}MVDR})}{1 + \zeta\mathcal{S}(\widehat{\mathbf{w}}_{SMI\text{-}MVDR})}}\right) \tag{49}$$

since $Q(\sqrt{x})$ is a decreasing function of x. Solving $P_e(\mathbf{w}_{MVDR}) = Q\left(\sqrt{\mathcal{S}(\mathbf{w}_{MVDR})}\right)$ with respect to $\mathcal{S}(\mathbf{w}_{MVDR})$ for values of $P_e(\mathbf{w}_{MVDR})$ less than 0.5, we obtain

$$\mathcal{S}(\mathbf{w}_{MVDR}) = \left[Q^{-1}(P_e(\mathbf{w}_{MVDR}))\right]^2 \tag{50}$$

where $Q^{-1}(\cdot)$ is the inverse function of $Q(\cdot)$. Using (50) in (49) we see that (34) is equivalent to

$$Pr\left[P_e(\widehat{\mathbf{w}}_{SMI-MVDR}) < Q\left(\sqrt{\frac{(1-\zeta)(Q^{-1}(P_e(\mathbf{w}_{MVDR})))^2}{1+\zeta(Q^{-1}(P_e(\mathbf{w}_{MVDR})))^2}}\right)\right] \geq \epsilon. \tag{51}$$

(51) can be written in a more compact form as follows:

$$Pr\left[P_e(\widehat{\mathbf{w}}_{SMI-MVDR}) < F_\zeta(P_e(\mathbf{w}_{MVDR}))\right] \geq \epsilon \tag{52}$$

where

$$F_\zeta(x) \triangleq Q\left(\sqrt{\frac{(1-\zeta)(Q^{-1}(x))^2}{1+\zeta(Q^{-1}(x))^2}}\right). \tag{53}$$

The function $F_\zeta(x)$ in (53) is non-linear. However, $\log F_\zeta(x)$ can be approximated closely by a linear function of $\log x$, i.e. $\log F_\zeta(x) \simeq \alpha \log x + \beta$. For a BER region of operation of the ideal MVDR receiver of the form $(10^{-\lambda}, 10^{-\nu})$ for some $\lambda > \nu$, the parameters α and β in the above approximation can be chosen as the coefficients of the first order Taylor series expansion of the function $\log F_\zeta(10^x)$ about $\mu \triangleq \frac{1}{2}(\lambda+\nu)$:

$$\alpha = \frac{10^\mu \left(\sqrt{\frac{(1-\zeta)[Q^{-1}(10^\mu)]^2}{1+\zeta[Q^{-1}(10^\mu)]^2}}\right)^3 \exp\left\{-\frac{(1-\zeta)[Q^{-1}(10^\mu)]^2}{2(1+\zeta[Q^{-1}(10^\mu)]^2)}\right\}}{(1-\zeta)F_\zeta(10^\mu)[Q^{-1}(10^\mu)]^3 \exp\left\{-[Q^{-1}(10^\mu)]^2/2\right\}}, \tag{54}$$

$$\beta = \log F_\zeta(10^\mu) - \alpha\mu. \tag{55}$$

Thus, $F_\zeta(x)$ can be approximated by

$$\widetilde{F}_\zeta(x) \triangleq 10^\beta x^\alpha. \tag{56}$$

In Fig. 3 we examine the accuracy of the approximation of $F_\zeta(x)$ by (56). In Fig. 3(a) we plot $F_\zeta(x)$ along with its approximation $\widetilde{F}_\zeta(x)$, while in Fig. 3(b) we plot the relative error of the approximation $|F_\zeta(x) - \widetilde{F}_\zeta(x)|/F_\zeta(x)$ for two different values of ζ, $\zeta = 0.2$ and $\zeta = 0.4$. The operating region of the receiver is assumed to be $[10^{-3}, 10^{-1}]$ which implies that $\mu = -2$. The accuracy of the approximation $\widetilde{F}_\zeta(x)$ is evident. Finally, replacing $F_\zeta(x)$ by $\widetilde{F}_\zeta(x)$, (52) becomes

$$Pr\left[\frac{\log P_e(\widehat{\mathbf{w}}_{SMI-MVDR}) - \beta}{\log P_e(\mathbf{w}_{MVDR})} > \alpha\right] \geq \epsilon. \tag{57}$$

■

B Proof of Corollary 1

Following the same approach as in the proof of Theorem 2, we see that $Pr[\rho_S > 1-\zeta] \geq \epsilon$ is equivalent to

$$Pr\left[P_e(\widehat{\mathbf{w}}_{SMI-MSINR}) < Q\left(\sqrt{1-\zeta}\, Q^{-1}\left(P_e(\mathbf{w}_{MSINR})\right)\right)\right] \geq \epsilon. \tag{58}$$

If we define

$$F_{*\zeta}(x) \triangleq Q\left(\sqrt{1-\zeta}\, Q^{-1}(x)\right), \tag{59}$$

we can express (58) in a more compact form as

$$Pr\left[P_e(\widehat{\mathbf{w}}_{SMI-MSINR}) < F_{*\zeta}\left(P_e(\mathbf{w}_{MSINR})\right)\right] \geq \epsilon. \tag{60}$$

The logarithm of the non-linear function $F_{*\zeta}(x)$ can be approximated closely by a linear function of $\log x$, i.e. $\log F_{*\zeta}(x) \simeq \alpha \log x + \beta$. For a BER region of operation of the ideal MSINR receiver of the form $(10^{-\lambda}, 10^{-\nu})$ for some $\lambda > \nu$, the parameters α and β in the above approximation can be chosen as the coefficients of the first order Taylor series expansion of the function $\log F_{*\zeta}(10^x)$ about $\mu = \frac{1}{2}(\lambda+\nu)$:

$$\alpha = \frac{10^{\mu}\sqrt{1-\zeta}\exp\{\frac{\zeta}{2}[Q^{-1}(10^{\mu})]^2\}}{F_{*\zeta}(10^{\mu})}, \tag{61}$$

$$\beta = \log F_{\zeta}(10^{\mu}) - \alpha\mu. \tag{62}$$

Thus, $F_{*\zeta}(x)$ can be approximated by

$$\widetilde{F}_{*\zeta}(x) \triangleq 10^{\beta}x^{\alpha}. \tag{63}$$

In Fig. 4 we examine the accuracy of the approximation of $F_{*\zeta}(x)$ by $\widetilde{F}_{*\zeta}(x)$. In Fig. 4(a) we plot $F_{*\zeta}(x)$ along with its approximation $\widetilde{F}_{*\zeta}(x)$, while in Fig. 4(b) we plot the relative error of the approximation $|F_{*\zeta}(x) - \widetilde{F}_{*\zeta}(x)|/F_{*\zeta}(x)$ for two different values of ζ, $\zeta = 0.2$ and $\zeta = 0.4$. The operating region of the receiver is assumed to be $[10^{-3}, 10^{-1}]$ which implies that $\mu = -2$. The accuracy of the approximation $\widetilde{F}_{*\zeta}(x)$ is evident. Finally, replacing $F_{*\zeta}(x)$ by $\widetilde{F}_{*\zeta}(x)$, (60) becomes

$$P\left[\frac{\log P_e(\widehat{\mathbf{w}}_{SMI-MSINR}) - \beta}{\log P_e(\mathbf{w}_{MSINR})} > \alpha\right] \geq \epsilon. \tag{64}$$

■

References

[1] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406–1415, Sept. 1999.

[2] M. L. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944-960, July 1995.

[3] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, pp. 853-863, Nov. 1974.

[4] W. Y. Tan, "On the distribution of the sample covariance matrix from a mixture of normal densities," *South African Stat. J.*, vol. 12, pp. 47-55, 1978.

[5] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Annals Math. Stat.*, vol 34, pp. 152-177, Mar. 1963.

[6] R. L. Dykstra, "Establishing the positive definiteness of the sample covariance matrix," *Ann. Math. Stat.*, vol. 41, pp. 2153-2154, 1970.

[7] A. O. Steinhardt, "Adaptive multisensor detection and estimation," in *Adaptive Radar Detection and Estimation* , S. Haykin and A. O. Steinhardt, Eds. New York: Wiley, 1992, ch. 3.

[8] C. D. Richmond, "PDF's, confidence regions, and relevant statistics for a class of sample covariance-based array processors," *IEEE Trans. Signal Processing*, vol. 44, pp. 1779-1793, July 1996.

[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* New York, NY: Dover, 1970.

[10] R. N. MacDonough and A. D. Whalen, *"Detection of Signals in Noise."* London, UK:Academic Press, 1995.

[11] R. C. Hanumara, "An alternate derivation of the distribution of the conditioned signal-to-noise ratio," *IEEE Trans. Ant. Prop.*, vol. 34, pp. 463-464, Mar. 1986.

[12] D. M. Boroson, "Sample size considerations for adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 16, pp. 446-451, Jul. 1980.

[13] I. N. Psaromiligkos and S. N. Batalama, "Recursive AV and MVDR filter estimation for maximum SINR adaptive space-time processing," *IEEE Trans. Commun.*, to appear.

TABLE I
DATA-RECORD-SIZE REQUIREMENTS OF SMI-MVDR AND SMI-MSINR SIGNAL DETECTION ALGORITHMS

Space-time product $M \times (L + N_p)$	ζ	ϵ	Data record size N^*
$5 \times (31 + 3)$	0.25 $(-6dB)$	0.75	683
$5 \times (31 + 3)$	0.25 $(-6dB)$	0.9	712
$5 \times (31 + 3)$	0.1 $(-10dB)$	0.9	1793

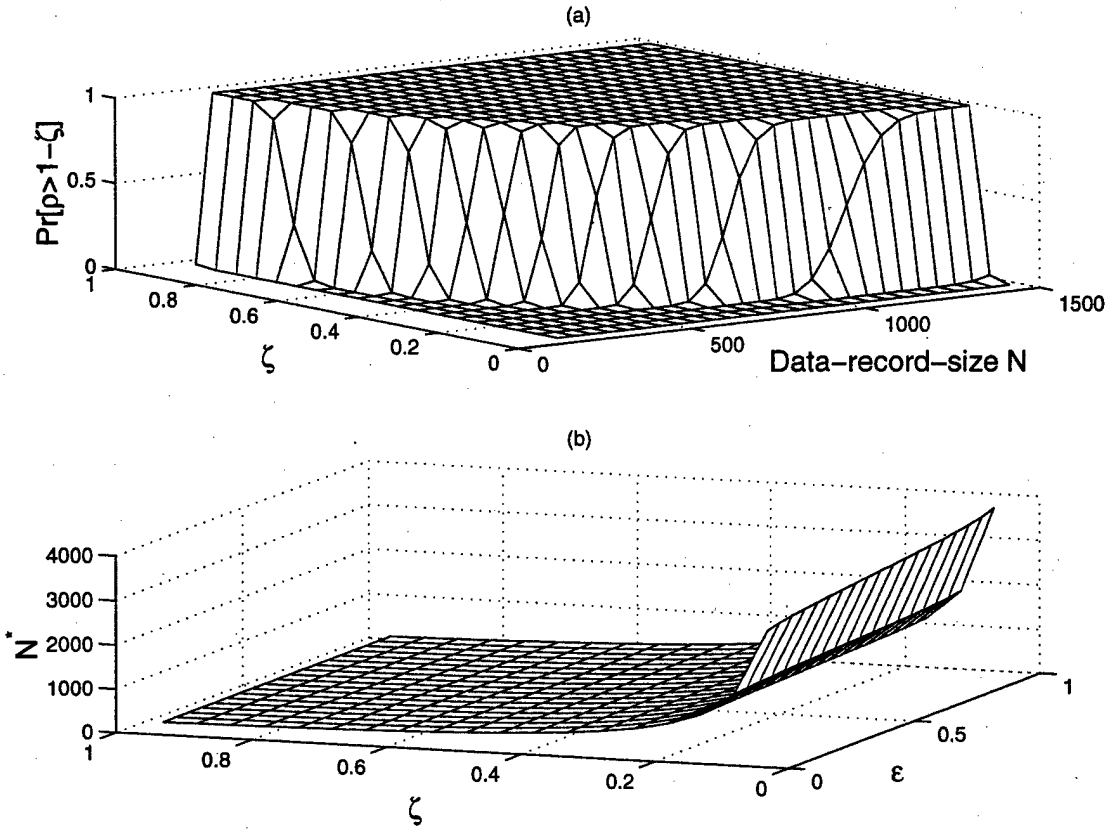


Figure 1: (a) $Pr[\rho > 1 - \zeta]$ as evaluated by Theorem 1, Part (i). (b) Data-record-size requirement N^* as given by Theorem 1, Part (ii).

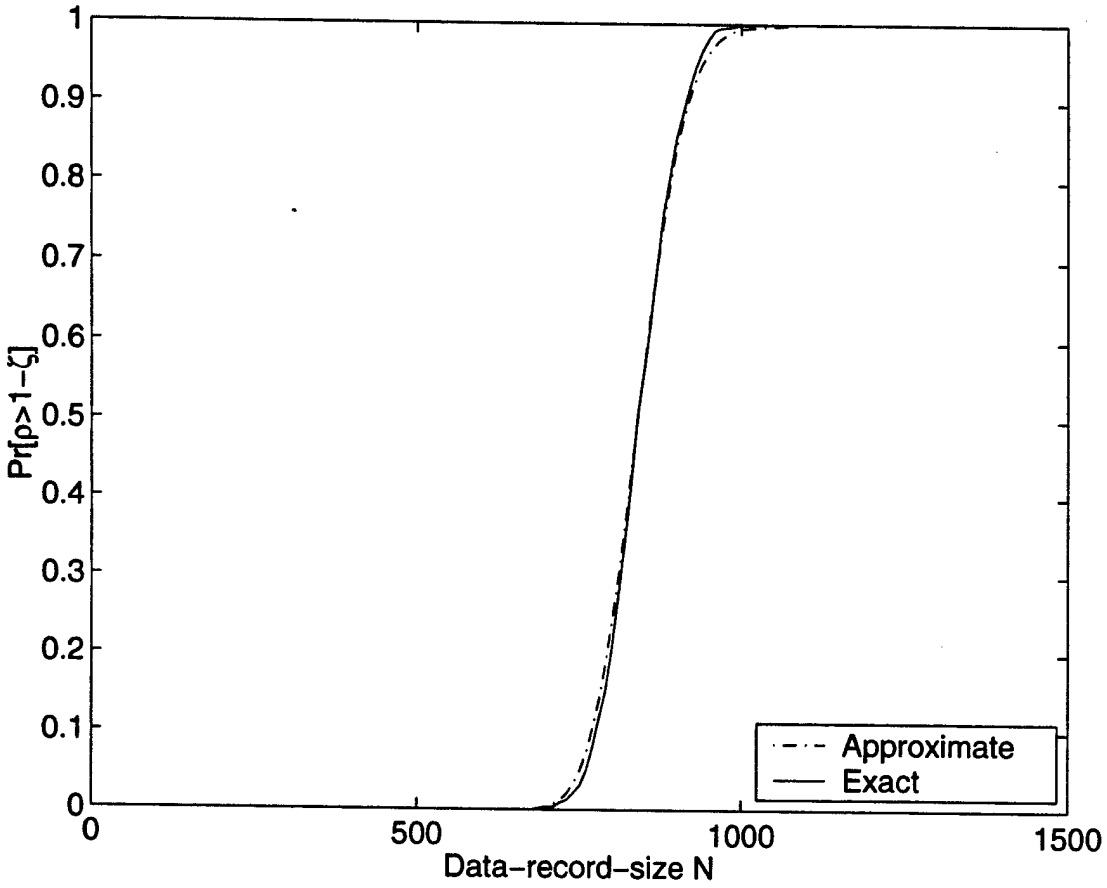


Figure 2: Exact and approximate $Pr[\rho > 1 - \zeta]$ (given by (26), (27)).

TABLE II
BIT-ERROR-RATE PERFORMANCE OF SMI-MVDR AND SMI-MSINR
SIGNAL DETECTION ALGORITHMS

Space-time product $M \times (L+N_p)$	ζ	ϵ	Data record size N^*	$\log P_e(\widehat{\mathbf{w}}_{SMI-MVDR})$ or $\log P_e(\widehat{\mathbf{w}}_{SMI-MSINR})$
SMI-MVDR				
$5 \times (31+3)$	0.25 $(-6dB)$	0.75	683	$(\log P_e(\mathbf{w}_{SMI-MVDR}), 0.16 \log(P_e(\mathbf{w}_{SMI-MVDR})) - 0.7)$
$5 \times (31+3)$	0.25 $(-6dB)$	0.9	712	$(\log P_e(\mathbf{w}_{SMI-MVDR}), 0.16 \log(P_e(\mathbf{w}_{SMI-MVDR})) - 0.7)$
$5 \times (31+3)$	0.1 $(-10dB)$	0.9	1793	$(\log P_e(\mathbf{w}_{SMI-MVDR}), 0.41 \log(P_e(\mathbf{w}_{SMI-MVDR})) - 0.6)$
SMI-MSINR				
$5 \times (31+3)$	0.25 $(-6dB)$	0.75	683	$(\log P_e(\mathbf{w}_{SMI-MSINR}), 0.78 \log(P_e(\mathbf{w}_{SMI-MSINR})) - 0.1)$
$5 \times (31+3)$	0.25 $(-6dB)$	0.9	712	$(\log P_e(\mathbf{w}_{SMI-MSINR}), 0.78 \log(P_e(\mathbf{w}_{SMI-MSINR})) - 0.1)$
$5 \times (31+3)$	0.1 $(-10dB)$	0.9	1793	$(\log P_e(\mathbf{w}_{SMI-MSINR}), 0.91 \log(P_e(\mathbf{w}_{SMI-MSINR})) - 0.04)$

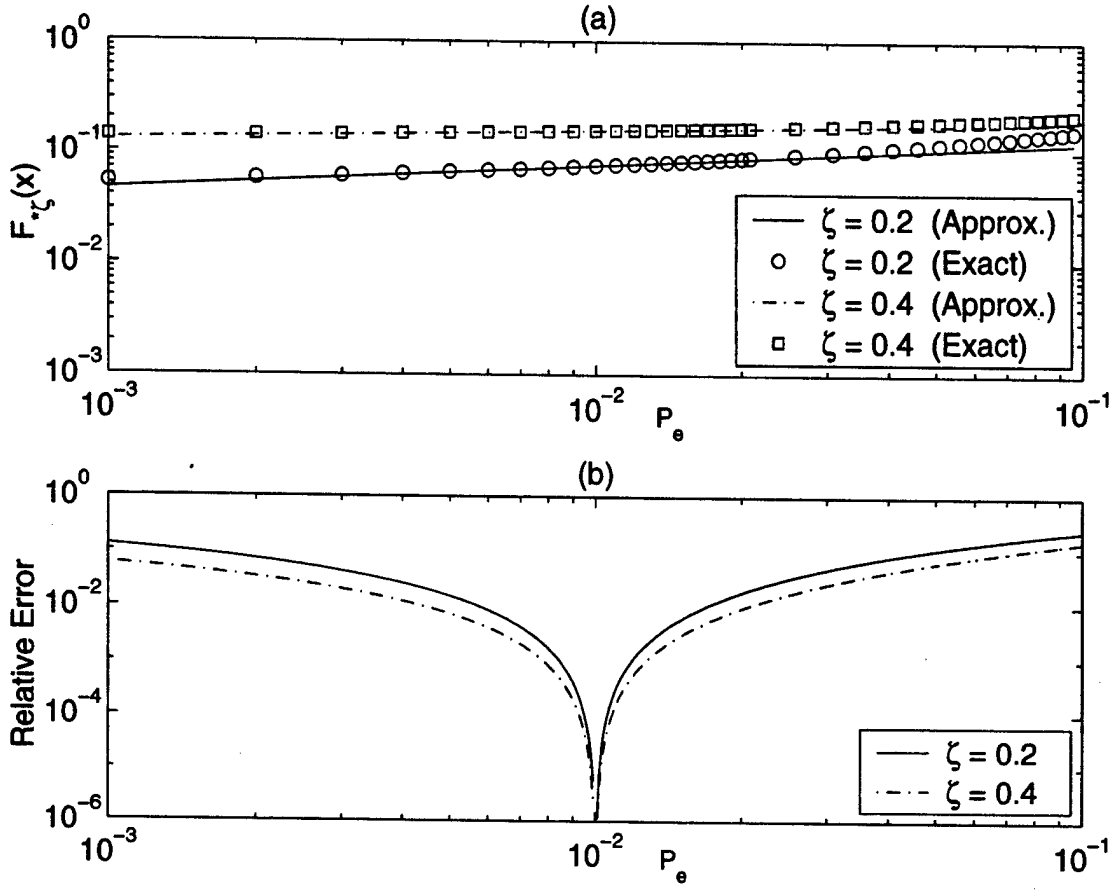


Figure 3: (a) Approximation of $F_\zeta(x)$ by $\widetilde{F}_\zeta(x)$. (b) Relative error of the approximation.

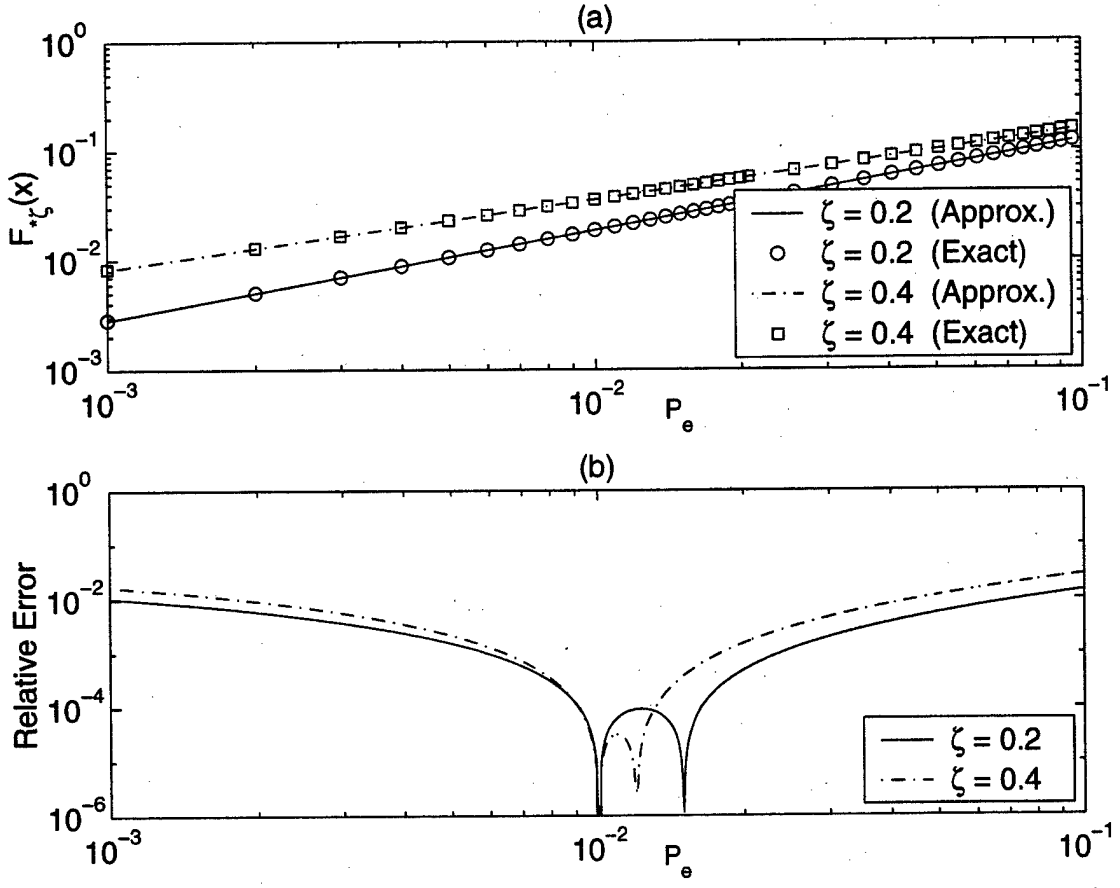


Figure 4: (a) Approximation of $F_{*\zeta}(x)$ by $\widetilde{F}_{*\zeta}(x)$. (b) Relative error of the approximation.

Performance Analysis of Doubly Optimal CDMA Spreading Codes with Odd Length*

George N. Karystinos and Dimitris A. Pados

Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA

ABSTRACT

A doubly optimal binary signature set is a set of binary spreading sequences that can be used for code division multiplexing purposes and exhibits minimum total-squared-correlation (TSC) and minimum maximum-squared-correlation (MSC) at the same time. In this article, we focus on such sets with signatures of odd length and we derive closed-form expressions for the signature cross-correlation matrix, its eigenvalues, and its inverse. Then, we derive analytic expressions for *(i)* the bit-error-rate (BER) upon decorrelating processing, *(ii)* the maximum achievable signal-to-interference-plus-noise (SINR) ratio upon minimum-mean-square-error (MMSE) filtering, and *(iii)* the total asymptotic efficiency of the system. We find that doubly optimal sets with signature length of the form $4m+1$, $m = 1, 2, \ldots$, are in all respects superior to doubly optimal sets with signature length of the form $4m-1$ (the latter class includes the familiar Gold sets as a small proper subset). "$4m+1$" sets perform practically at the single-user-bound (SUB) after decorrelating or MMSE processing (not true for "$4m-1$" sets). The total asymptotic efficiency of "$4m+1$" sets is lower bounded by $\frac{2}{e}$ for any system user load. The corresponding lower bound for "$4m-1$" sets is zero.

Keywords: Binary sequences, code division multiaccess, codes, Gold codes, signal design, spread spectrum communication, Welch bound.

1. INTRODUCTION AND BACKGROUND

In code division multiplexing systems, such as the present wireless direct-sequence code-division-multiple-access (DS-CDMA) technology, distinct data streams are modulated by individual signature patterns and sent over a common in time and frequency channel. The success or, more general, performance of the multiplexing scheme (or multiple-access scheme depending on the point of view) depends on the specific assigned set of signatures. Two critical signature set metrics are the total-squared-correlation (TSC) and the maximum-squared-correlation (MSC) among signatures.[1–11] Intuitively, both metric values need to be kept low.

In Ref. 8, we derived new lower bounds on the TSC of *binary* antipodal signature sets for all possible combinations of number of data streams/users K and signature length/processing gain L and we proved the tightness of the new bounds for all K, L except *(i)* $L = K \equiv 1 \pmod 4$, *(ii)* $L = K \equiv 2 \pmod 4$, *(iii)* $L+1 = K \equiv 2 \pmod 4$, and *(iv)* $K+1 = L \equiv 2 \pmod 4$. Ding, Golin, and Kløve[9] established the tightness of these bounds also under Cases *(ii)*-*(iv)*.† All proofs of tightness in Ref. 8, as well as in Ref. 9, are by construction and present us with simple algorithms for the design of minimum-TSC optimum binary signature sets based on Hadamard matrix transformations. The familiar Rademacher-Walsh-Hadamard code sets[12,13] with $L = 2^m$, $m = 1, 2, \ldots$, and $K \leq L$ and their extension to $L = 4m$, $m = 1, 2, \ldots$, and $K \leq L$, the m-sequence sets[14] with $L = 2^m - 1$, $m = 3, 4, \ldots$, and $K \leq L$, and the Gold sets[15] with $L = 2^m - 1$, $m = 5, 6, \ldots$, and $K \leq L + 2$,

*This work was supported in part by the National Science Foundation under Grant ECS-0073660 and the U.S. Air Force Office of Scientific Research under Grant F49620-01-1-0176.

Further author information: (Send correspondence to D.A.P.)

G.N.K.: E-mail: cary@eng.buffalo.edu, Telephone: 1 716 645 2422 ext. 2181

D.A.P.: E-mail: pados@eng.buffalo.edu, Telephone: 1 716 645 3115 ext. 2134

Address: Department of Electrical Engineering, 332 Bonner Hall, State University of New York at Buffalo, Buffalo, NY 14260 USA.

†The case $K = L \equiv 1 \pmod 4$ remains open. Ding, Golin, and Kløve[9] showed that our bound in Ref. 8 is tight for $K = L = 5$ or 13, but not for $K = L = 9$. What happens when $K = L \geq 17$ is an open question.

can all be seen as specific output examples of the design procedure in Ref. 8, 9. In Ref. 11 we were able to show that for "underloaded systems" ($K \leq L$) the minimum-TSC optimal designs in Ref. 8, 9 have, in fact, minimum maximum-squared-correlation (MSC) at the same time. We call, therefore, these sets doubly optimal: optimal both in the minimum-TSC and minimum-MSC sense.

In this paper we focus exclusively on binary doubly optimal sets with signatures of odd length (m-sequence and Gold sets fall under this study) and we attempt to evaluate theoretically their multiaccess behavior. In view of the results in Ref. 8, 9, we recognize the need to identify two separate cases: sets with signature length of the form $4m+1$, $m = 1, 2, \ldots$, and sets with signature length of the form $4m-1$, $m = 1, 2, \ldots$ (m-sequence and Gold sets fall now under the latter case). For both cases we derive analytic expressions for *(i)* the probability of error (bit-error-rate or BER) of the decorrelating detector, *(ii)* the mean square error (MSE) and output signal-to-interference-plus-noise ratio (SINR) of the minimum-mean-square-error (MMSE) linear filter, and *(iii)* the total asymptotic efficiency of the system. Direct analytic comparisons weigh heavily in favor of "$4m+1$" code sets. Decorrelation of "$4m+1$" codes leads to almost single-user-bound (SUB) BER performance. This is not so for "$4m-1$" codes. "$4m+1$" codes attain higher SINR at the output of the MMSE filter for smaller processing gain (therefore, smaller occupied bandwidth) than "$4m-1$" codes. "$4m+1$" codes achieve higher total asymptotic efficiency than "$4m-1$" codes and guarantee a lower bound of $\frac{2}{e}$ for any user load $0 < \frac{K}{L} \leq 1$. In contrast, the corresponding bound for "$4m-1$" codes is zero.

The presentation of this material is organized as follows. In Section 2, after we introduce pertinent notation and definitions, we derive a few essential properties of the doubly optimal signature sets under consideration. The performance of code division multiplexing systems that utilize such spreading codes is examined theoretically in Section 3. In Section 4, we present a few numerical examples and simulation studies to illustrate and interpret the theoretical developments of Section 3. Some concluding remarks are given in Section 5.

2. NOTATION, DEFINITIONS, AND A FEW DERIVATIONS

We consider a binary antipodal signature set with K normalized signatures $\mathbf{s}_k \in \left\{\pm\frac{1}{\sqrt{L}}\right\}^L$, $k = 1, 2, \ldots, K$, where L is the DS-CDMA system processing gain; we represent this set by the $L \times K$ signature matrix $\mathbf{S} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K]$. The *signature cross-correlation matrix* $\mathbf{R}$ is defined as the matrix that contains all cross-correlations (inner products) between the signatures in the set:

$$\mathbf{R} \triangleq \mathbf{S}^T\mathbf{S}. \tag{1}$$

Below, we derive some useful expressions for $\mathbf{R}$, its eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_K$, and its inverse $\mathbf{R}^{-1}$ when $\mathbf{S}$ is an underloaded ($K \leq L$) doubly optimal set of odd processing gain.[11] We examine the two cases $L = 4m - 1$ and $L = 4m + 1$, $m = 1, 2, \ldots$, separately.

1) $L = 4m - 1$

From Ref. 8, 9, we can show that for a doubly optimal signature matrix $\mathbf{S}$ with processing gain of the form $L = 4m - 1$, $m = 1, 2, \ldots$,

$$\mathbf{R} = \frac{L+1}{L}\mathbf{I}_K - \mathbf{s}\mathbf{s}^T \tag{2}$$

where the vector $\mathbf{s} \in \left\{\pm\frac{1}{\sqrt{L}}\right\}^K$ depends on[‡] the specific signature matrix $\mathbf{S}$ and $\mathbf{I}_K$ denotes the $K \times K$ identity matrix. Then, the K ordered eigenvalues of $\mathbf{R}$ given in the form of a $K \times 1$ vector $\boldsymbol{\lambda}$ are as follows:

$$\boldsymbol{\lambda} = \left[\frac{L+1}{L}, \ldots, \frac{L+1}{L}, \frac{L-K+1}{L}\right]^T. \tag{3}$$

The inverse of $\mathbf{R}$ can be obtained from (2) using the Matrix Inversion Lemma[16] (Woodbury's identity):

$$\mathbf{R}^{-1} = \frac{L}{L+1}\mathbf{I}_K + \frac{L}{L+1}\mathbf{I}_K\mathbf{s}\left(1 - \mathbf{s}^T\frac{L}{L+1}\mathbf{I}_K\mathbf{s}\right)^{-1}\mathbf{s}^T\frac{L}{L+1}\mathbf{I}_K = \frac{L}{L+1}\mathbf{I}_K + \frac{L^2}{(L+1)(L-K+1)}\mathbf{s}\mathbf{s}^T. \tag{4}$$

[‡]For example, if $\mathbf{S}$ is a Gold signature set ($L = 2^m - 1$, $m = 5, 6, \ldots$),[15] then $\mathbf{s} = \frac{1}{\sqrt{L}}[1, 1, \ldots, 1]^T$.

2) $L = 4m + 1$

For a doubly optimal signature matrix $\mathbf{S}$ with processing gain of the form $L = 4m + 1$, $m = 1, 2, \ldots$, we can show using the findings in Ref. 8, 9 that

$$\mathbf{R} = \frac{L-1}{L}\mathbf{I}_K + \mathbf{s}\mathbf{s}^T \tag{5}$$

where the vector $\mathbf{s} \in \left\{\pm\frac{1}{\sqrt{L}}\right\}^K$ depends on the specific signature matrix $\mathbf{S}$. The $K \times 1$ vector $\boldsymbol{\lambda}$ with the ordered eigenvalues of $\mathbf{R}$ is

$$\boldsymbol{\lambda} = \left[\frac{L+K-1}{L}, \frac{L-1}{L}, \ldots, \frac{L-1}{L}\right]^T. \tag{6}$$

The inverse of $\mathbf{R}$ is obtained from (5) using the Matrix Inversion Lemma:

$$\mathbf{R}^{-1} = \frac{L}{L-1}\mathbf{I}_K - \frac{L}{L-1}\mathbf{I}_K\mathbf{s}\left(1 + \mathbf{s}^T\frac{L}{L-1}\mathbf{I}_K\mathbf{s}\right)^{-1}\mathbf{s}^T\frac{L}{L-1}\mathbf{I}_K = \frac{L}{L-1}\mathbf{I}_K - \frac{L^2}{(L-1)(L+K-1)}\mathbf{s}\mathbf{s}^T. \tag{7}$$

3. ODD-LENGTH DOUBLY OPTIMAL CODES: THEORETICAL PERFORMANCE ANALYSIS

We consider a synchronous DS-CDMA system with processing gain L where $K \leq L$ users transmit binary information symbols, via phase-shift-keying for example, over a common in time and frequency additive white Gaussian noise (AWGN) channel. Each user k, $k = 1, 2, \ldots, K$, is assigned a spreading code $\mathbf{s}_k$ which is one column of the $L \times K$ signature matrix $\mathbf{S}$. After carrier demodulation, conventional chip-matched filtering and chip-rate sampling, the baseband received signal vector over one information bit period takes the familiar form

$$\mathbf{r} = \sum_{k=1}^{K} b_k\sqrt{E_k}\mathbf{s}_k + \mathbf{n} \tag{8}$$

where $b_k \in \{\pm 1\}$ is the kth user information bit, E_k denotes received energy per bit, and $\mathbf{n}$ represents the filtered noise vector. If $\frac{N_0}{2}$ is the power spectral density of the underlying AWGN process, then $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}_{L\times 1}, \frac{N_0}{2}\mathbf{I}_L\right)$ and $\gamma_k \triangleq \frac{2E_k}{N_0}$ is the kth user signal-to-noise ratio (SNR) per bit. We define the $K \times K$ diagonal matrix that consists of the K user energies $\mathbf{E} \triangleq \mathrm{diag}\,(E_1, E_2, \ldots, E_K)$ and, similarly, the diagonal matrix $\boldsymbol{\Gamma} \triangleq \mathrm{diag}\,(\gamma_1, \gamma_2, \ldots, \gamma_K)$ that contains the K user SNRs. In the sequel, we evaluate theoretically the performance of this code division multiplexing system when doubly optimal spreading codes of odd processing gain are used.

3.1. Decorrelating Receivers

The decorrelating linear filter (or decorrelator)[17,18] output is the ML *linear estimate* of the information bit b_k.[19] The kth user decorrelator is given by

$$\mathbf{w}_k^{(\mathrm{DEC})} = \mathbf{S}\left[\mathbf{R}^{-1}\right]_{k\text{th column}} \tag{9}$$

where $\mathbf{R}$ is the signature cross-correlation matrix[§]. It is known that the BER of the decorrelator followed by a sign detector is

$$P_{e_k}^{(\mathrm{DEC})} = Q\left(\sqrt{\frac{\gamma_k}{[\mathbf{R}^{-1}]_{k,k}}}\right) \tag{10}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$. Therefore, the BER performance of the kth decorrelator is determined by the kth diagonal element of the inverse signature cross-correlation matrix $\mathbf{R}^{-1}$ (and the kth user SNR of course). Below, we consider the use of odd-length doubly optimal spreading codes[8,9,11] and we examine their effect on the decorrelator BER performance. We need to study the two cases $L = 4m - 1$ and $L = 4m + 1$, $m = 1, 2, \ldots$, separately.

[§]In our notation, if $\mathbf{A}$ is an $m \times n$ matrix, then $[\mathbf{A}]_{k\text{th column}}$, $k = 1, 2, \ldots, n$, is the $m \times 1$ vector defined by the kth column of $\mathbf{A}$. $[\mathbf{A}]_{j,k}$, $j = 1, 2, \ldots, m$, $k = 1, 2, \ldots, n$, is the (j, k)th element of $\mathbf{A}$.

1) $L = 4m - 1$
From (4) we obtain:

$$\left[\mathbf{R}^{-1}\right]_{k,k} = \frac{L}{L+1} + \frac{L^2}{(L+1)(L-K+1)}\left(\mathbf{s}[k]\right)^2 = \boxed{1 + \frac{K-1}{(L+1)(L-K+1)}.} \tag{11}$$

2) $L = 4m + 1$
From (7) we obtain:

$$\left[\mathbf{R}^{-1}\right]_{k,k} = \frac{L}{L-1} - \frac{L^2}{(L-1)(L+K-1)}\left(\mathbf{s}[k]\right)^2 = \boxed{1 + \frac{K-1}{(L-1)(L+K-1)}.} \tag{12}$$

We note that the $\left[\mathbf{R}^{-1}\right]_{k,k} = 1$ value in (10) corresponds to single-user-bound (SUB) performance and we recall that for any K, L and signature set $\mathbf{S}$ for which the decorrelator exists, $\left[\mathbf{R}^{-1}\right]_{k,k} \geq 1$. From (11), it is interesting to observe that when we use doubly optimal signatures with $L = 4m-1$, then the coefficient $\left[\mathbf{R}^{-1}\right]_{k,k}$ becomes much greater than 1 when the system is heavily loaded ($\left[\mathbf{R}^{-1}\right]_{k,k} \underset{K=L}{=} 1 + \frac{L-1}{L+1} \underset{K=L\to\infty}{\longrightarrow} 2$). On the other hand, the coefficient $\left[\mathbf{R}^{-1}\right]_{k,k}$ in (12) for $L = 4m + 1$ doubly optimal signatures is always kept close to 1, even for high user loads ($\left[\mathbf{R}^{-1}\right]_{k,k} \underset{K=L}{=} 1 + \frac{1}{2L-1} \underset{K=L\to\infty}{\longrightarrow} 1$). As a result, we conclude that $(4m+1)-$long doubly optimal codes maintain near SUB performance even for heavily loaded systems, while $(4m-1)-$long doubly optimal codes fail (to place this conclusion in the broad context of pertinent past literature, we recall that the Gold codes[15] are in fact a small proper subset of the $(4m-1)-$long doubly optimal codes[8,9,11]). Later, in Section 4, we present for illustration purposes some numerical BER performance plots for the decorrelator when $(4m-1)$ or $(4m+1)-$long doubly optimal codes are utilized.

3.2. MMSE Receivers

The MMSE (or Wiener) filter[20,21] is the linear filter that minimizes the MSE between its output and the *true* information bit. It can be shown that the MMSE filter maximizes the SINR at its output. The kth user MMSE filter is given by

$$\mathbf{w}_k^{(\mathrm{MMSE})} = \mathbf{R}_A^{-1}\mathbf{s}_k \tag{13}$$

where $\mathbf{R}_A$ is the *auto-correlation matrix of the received input vector* $\mathbf{r}$:

$$\mathbf{R}_A = E\left\{\mathbf{r}\mathbf{r}^T\right\} = \mathbf{S}\mathbf{E}\mathbf{S}^T + \frac{N_0}{2}\mathbf{I}_L. \tag{14}$$

It is well known that the MSE of the kth user optimum MMSE filter is

$$\mathrm{MMSE}_k = 1 - E_k\mathbf{s}_k^T\mathbf{R}_A^{-1}\mathbf{s}_k \tag{15}$$

and the kth user maximum output SINR value is

$$\mathrm{SINR}_k = \frac{1}{\mathrm{MMSE}_k} - 1. \tag{16}$$

We now apply the Matrix Inversion Lemma to (14) to obtain:

$$\mathbf{R}_A^{-1} = \frac{2}{N_0}\mathbf{I}_L - \frac{2}{N_0}\mathbf{S}\mathbf{E}^{\frac{1}{2}}\left(\mathbf{I}_K + \frac{2}{N_0}\mathbf{E}^{\frac{1}{2}}\mathbf{S}^T\mathbf{S}\mathbf{E}^{\frac{1}{2}}\right)^{-1}\mathbf{E}^{\frac{1}{2}}\mathbf{S}^T\frac{2}{N_0}. \tag{17}$$

Therefore,

$$\mathbf{S}^T\mathbf{R}_A^{-1}\mathbf{S} = \frac{2}{N_0}\mathbf{R} - \frac{2}{N_0}\mathbf{R}\mathbf{E}^{\frac{1}{2}}\left(\mathbf{I}_K + \frac{2}{N_0}\mathbf{E}^{\frac{1}{2}}\mathbf{R}\mathbf{E}^{\frac{1}{2}}\right)^{-1}\mathbf{E}^{\frac{1}{2}}\mathbf{R}\frac{2}{N_0}. \tag{18}$$

Next, we use the Matrix Inversion Lemma to calculate the inverse of $\mathbf{E}+\frac{N_0}{2}\mathbf{R}^{-1}$:

$$\left(\mathbf{E}+\frac{N_0}{2}\mathbf{R}^{-1}\right)^{-1}=\frac{2}{N_0}\mathbf{R}-\frac{2}{N_0}\mathbf{R}\mathbf{E}^{\frac{1}{2}}\left(\mathbf{I}_K+\mathbf{E}^{\frac{1}{2}}\frac{2}{N_0}\mathbf{R}\mathbf{E}^{\frac{1}{2}}\right)^{-1}\mathbf{E}^{\frac{1}{2}}\frac{2}{N_0}\mathbf{R}. \tag{19}$$

From (18), (19), we conclude that

$$\mathbf{S}^T\mathbf{R}_A^{-1}\mathbf{S}=\left(\mathbf{E}+\frac{N_0}{2}\mathbf{R}^{-1}\right)^{-1}. \tag{20}$$

We note that $\mathbf{s}_k^T\mathbf{R}_A^{-1}\mathbf{s}_k=\left[\mathbf{S}^T\mathbf{R}_A^{-1}\mathbf{S}\right]_{k,k}$. Then, from (15) and (20) we obtain:

$$\mathrm{MMSE}_k=1-E_k\left[\left(\mathbf{E}+\frac{N_0}{2}\mathbf{R}^{-1}\right)^{-1}\right]_{k,k}=1-\left[\left(\mathbf{I}_K+\frac{N_0}{2}\mathbf{E}^{-\frac{1}{2}}\mathbf{R}^{-1}\mathbf{E}^{-\frac{1}{2}}\right)^{-1}\right]_{k,k}. \tag{21}$$

We recall the definition $\mathbf{\Gamma}\triangleq\mathrm{diag}\left(\gamma_1,\gamma_2,\ldots,\gamma_K\right)=\mathrm{diag}\left(\frac{2E_1}{N_0},\frac{2E_2}{N_0},\ldots,\frac{2E_K}{N_0}\right)=\frac{2}{N_0}\mathbf{E}$. Then,

$$\mathrm{MMSE}_k=1-\left[\left(\mathbf{I}_K+\mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{R}^{-1}\mathbf{\Gamma}^{-\frac{1}{2}}\right)^{-1}\right]_{k,k}. \tag{22}$$

In the following, we consider the possible use of odd-length doubly optimal spreading codes[8, 9, 11] in conjunction with MMSE receivers. As in the previous decorrelator subsection, we examine the two cases $L=4m-1$ and $L=4m+1$, separately.

1) $L=4m-1$

From (22) and (4) we are able to obtain the following closed-form analytic expression for the kth user MMSE as a function of the system parameters L, K, and γ_1, γ_2, ..., γ_K when any doubly optimal signature set of length $L=4m-1$, $m=1,2,\ldots$, is used (we mention again that the familiar Gold sets fall exactly under this class):

$$\mathrm{MMSE}_k=1-\frac{\gamma_k}{\frac{L}{L+1}+\gamma_k}\left(1-\frac{1}{\left(\frac{L}{L+1}+\gamma_k\right)\left[\frac{(L+1)(L-K+1)}{L}+\sum_{j=1}^{K}\frac{1}{\frac{L}{L+1}+\gamma_j}\right]}\right). \tag{23}$$

The detailed derivation (proof) of (23) is omitted due to lack of space. Since $\mathrm{SINR}_k=\frac{1}{\mathrm{MMSE}_k}-1$ (cf. (16)), we conclude that

$$\boxed{\mathrm{SINR}_k=\frac{1}{1-\frac{\gamma_k}{\frac{L}{L+1}+\gamma_k}\left(1-\frac{1}{\left(\frac{L}{L+1}+\gamma_k\right)\left[\frac{(L+1)(L-K+1)}{L}+\sum_{j=1}^{K}\frac{1}{\frac{L}{L+1}+\gamma_j}\right]}\right)}-1.} \tag{24}$$

2) $L=4m+1$

Similarly, from (22) and (7) we obtain the following closed-form expression for doubly optimal sets of signatures with length $L=4m+1$, $m=1,2,\ldots$. The derivation is omitted.

$$\mathrm{MMSE}_k=1-\frac{\gamma_k}{\frac{L}{L-1}+\gamma_k}\left(1+\frac{1}{\left(\frac{L}{L-1}+\gamma_k\right)\left[\frac{(L-1)(L+K-1)}{L}-\sum_{j=1}^{K}\frac{1}{\frac{L}{L-1}+\gamma_j}\right]}\right). \tag{25}$$

We conclude that (cf. (16)),

$$\boxed{\mathrm{SINR}_k=\frac{1}{1-\frac{\gamma_k}{\frac{L}{L-1}+\gamma_k}\left(1+\frac{1}{\left(\frac{L}{L-1}+\gamma_k\right)\left[\frac{(L-1)(L+K-1)}{L}-\sum_{j=1}^{K}\frac{1}{\frac{L}{L-1}+\gamma_j}\right]}\right)}-1.} \tag{26}$$

The analytic expressions derived in this section can be used for the direct evaluation/comparison of the SINR (or MMSE) performance of odd-length doubly optimal spreading codes. As a case study, we proved the following proposition that holds under "perfect power-control." The proof is omitted due to lack of space.

PROPOSITION 1. *Consider all doubly optimal spreading codes*[8,9,11] *of length* $L_1 = 4m-1$ *and* $L_2 = 4m+1$, $m = 1,2,\ldots$. *Assume that* $\gamma_k = \gamma$, $k = 1,2,\ldots,K$, *and consider MMSE filter receivers. Then,* $\mathrm{SINR}^{(L_2)} > \mathrm{SINR}^{(L_1)}$ *(or, equivalently,* $\mathrm{MMSE}^{(L_2)} < \mathrm{MMSE}^{(L_1)}$*) if and only if* $\boxed{L_2 > L_1 - \frac{\gamma}{\gamma+1}(K-2).}$ □

Proposition 1 identifies the minimum processing gain L_2 of the form *4m+1* that is required to achieve better SINR (or MMSE) performance than a given processing gain L_1 of the form *4m-1*. For example, set $K = 30$ and $\gamma = 10$dB. If $L_1 = 63$, then for any $L_2 > 37.5$ (i.e. $L_2 = 41, 45, 49, 53, \ldots$) the corresponding doubly optimal codes outperform in SINR (or MMSE) all 63-long codes (that, we recall, include the 63-long Gold codes). In general, Proposition 1 implies that *(4m+1)−long spreading codes can achieve a higher output SINR than (4m−1)−long codes and simultaneously occupy a significantly smaller bandwidth.*

In Section 4, we present a few numerical studies that illustrate the maximum SINR performance of odd-length doubly optimal spreading codes derived above.

3.3. Total Asymptotic Efficiency

The total asymptotic efficiency[22,23] of a DS-CDMA system equals the determinant of the signature cross-correlation matrix $|\mathbf{R}|$. Since $|\mathbf{R}| = \prod_{k=1}^{K} \lambda_k$, where λ_k, $k = 1,2,\ldots,K$, are the eigenvalues of $\mathbf{R}$, we can use our eigenvalue findings in (3) and (6) to obtain closed-form expressions for the total asymptotic efficiency of CDMA systems that utilize odd-length doubly optimal spreading codes:

$$|\mathbf{R}| = \begin{cases} \frac{(L+1)^{K-1}(L-K+1)}{L^K}, & L = 4m-1, \quad m = 1,2,\ldots, \\ \frac{(L-1)^{K-1}(L+K-1)}{L^K}, & L = 4m+1, \quad m = 1,2,\ldots. \end{cases} \tag{27}$$

In the following proposition we show that DS-CDMA systems with $(4m+1)-$long doubly optimal spreading codes have *strictly greater* total asymptotic efficiency than systems with $(4m-1)-$long codes. The proof is omitted due to lack of space.

PROPOSITION 2. *Consider code division multiplexing systems with doubly optimal signature sets of length* $L_1 = 4m-1$ *and* $L_2 = 4m+1$, $m = 1,2,\ldots$, *and* $K > 1$ *users. Then, the total asymptotic efficiency of* L_2-*type systems is strictly greater than* L_1-*type systems:* $|\mathbf{R}_{L_2}| > |\mathbf{R}_{L_1}|$. □

It may be even more interesting to examine the asymptotic behavior of the total asymptotic efficiency $|\mathbf{R}|$ as the processing gain and the number of users increase ($L \to \infty$ and $K \to \infty$) while their ratio (user load) $\frac{K}{L}$ is kept constant. Below, we establish formally that $(4m+1)-$long codes are significantly superior to $(4m-1)-$long codes in terms of the asymptotic behavior of $|\mathbf{R}|$ for the same user load $\alpha \triangleq \frac{K}{L}$. We present our theoretical developments in the form of the following proposition whose proof is omitted.

PROPOSITION 3.

(a) *Let* $\mathbf{R}_1$ *be the signature cross-correlation matrix of a doubly optimal spreading code set with processing gain of the form* $L = 4m-1$, $m = 1,2,\ldots$, *and user load* $\alpha \in (0,1]$.

(i) Fix the user load α. *Then,* $|\mathbf{R}_1|$ *is a monotonically decreasing function of* L, $L \geq \frac{1}{\alpha}$. *For* $L = \frac{1}{\alpha}$, $|\mathbf{R}_1| = 1$ *and*

$$\boxed{\lim_{L\to\infty} |\mathbf{R}_1| = (1-\alpha)e^{\alpha}.} \tag{28}$$

(ii) Fix the processing gain L. *Then,* $|\mathbf{R}_1|$ *is a monotonically decreasing function of* α, $\frac{1}{L} \leq \alpha \leq 1$, *and*

$$|\mathbf{R}_1| = \begin{cases} 1, & \alpha = \frac{1}{L} \\ \frac{1}{L}\left(1+\frac{1}{L}\right)^{L-1}, & \alpha = 1. \end{cases} \tag{29}$$

(b) *Let* $\mathbf{R}_2$ *be the signature cross-correlation matrix of a doubly optimal spreading code set with processing gain of the form* $L = 4m+1$, $m = 1, 2, \ldots$, *and user load* $\alpha \in (0, 1]$.

(i) Fix the user load α. *Then,* $|\mathbf{R}_2|$ *is a monotonically decreasing function of* L, $L \geq \frac{1}{\alpha}$. *For* $L = \frac{1}{\alpha}$, $|\mathbf{R}_2| = 1$ *and*

$$\boxed{\lim_{L\to\infty} |\mathbf{R}_2| = (1+\alpha)e^{-\alpha}.} \tag{30}$$

(ii) Fix the processing gain L. *Then,* $|\mathbf{R}_2|$ *is a monotonically decreasing function of* α, $\frac{1}{L} \leq \alpha \leq 1$, *and*

$$|\mathbf{R}_2| = \begin{cases} 1, & \alpha = \frac{1}{L} \\ \left(2 - \frac{1}{L}\right)\left(1 - \frac{1}{L}\right)^{L-1}, & \alpha = 1. \end{cases} \tag{31}$$

□

For example, according to Proposition 3, if we fix the user load α to 1 (fully loaded system) and we let the processing gain L increase to infinity, the total asymptotic efficiency $|\mathbf{R}_1|$ of $(4m-1)-$long codes converges to 0 while the total asymptotic efficiency $|\mathbf{R}_2|$ of $(4m+1)-$long codes converges to $\frac{2}{e}$. In fact, since $(1-\alpha)e^{\alpha} < (1+\alpha)e^{-\alpha}$ $\forall \alpha \in (0, 1]$, Proposition 3 shows that $(4m+1)-$long codes achieve higher total asymptotic efficiency than $(4m-1)-$long codes as $L \to \infty$ for *any* load $\alpha \in (0, 1]$. This conclusion is showcased below in the form of a corollary.

COROLLARY 1. *Consider two doubly optimal sets with processing gain of the form* $L_1 = 4m-1$ *and* $L_2 = 4m+1$, $m = 1, 2, \ldots$, *and signature cross-correlation matrices* $\mathbf{R}_1$ *and* $\mathbf{R}_2$, *respectively. If the two sets have the same fixed user load* $\alpha \in (0, 1]$, *then*

$$\lim_{L_1\to\infty} |\mathbf{R}_1| = (1-\alpha)e^{\alpha} < (1+\alpha)e^{-\alpha} = \lim_{L_2\to\infty} |\mathbf{R}_2|. \tag{32}$$

□

In the context of doubly optimal sets with finite (non-asymptotic) length L and load $\alpha \in (0, 1]$, Proposition 3 offers the following conclusion. While in general the total asymptotic efficiency $|\mathbf{R}|$ of a signature set is between 0 and 1 (in fact, these numbers are tight bounds for $|\mathbf{R}|$ when the processing gain is of the form $L = 4m-1$ (cf. Proposition 3)), $(4m+1)-$long doubly optimal spreading codes guarantee a minimum total asymptotic efficiency $|\mathbf{R}| = \frac{2}{e} \simeq .73$ for *any* processing gain $L = 4m+1$ and *any* user load $\alpha \in (0, 1]$. We present this finding below as a second corollary to Proposition 3.

COROLLARY 2. *The total asymptotic efficiency* $|\mathbf{R}|$ *of doubly optimal sets with processing gain of the form* $L = 4m+1$, $m = 1, 2, \ldots$, *is lower bounded by* $\frac{2}{e}$:

$$\boxed{\frac{2}{e} < |\mathbf{R}| \leq 1.} \tag{33}$$

□

4. NUMERICAL EXAMPLES AND SIMULATION STUDIES

In this section we present a few numerical examples that may help us visualize the theoretical developments of Section 3. The analysis of Section 3 applies to any doubly optimal spreading code set with odd processing gain. Here we choose to compare the performance of 31-long and 63-long doubly optimal Gold codes to the performance of 33-long and 65-long doubly optimal codes.[8, 9, 11]

In Fig. 1 we consider underloaded-by-one doubly optimal sets, that is *(a)* a Gold set with $L_1 = 31$ and $K_1 = 30$, *(b)* a doubly optimal set with $L_2 = 33$ and $K_2 = 32$, *(c)* a Gold set with $L_1' = 63$ and $K_1' = 62$, and *(d)* a doubly optimal set with $L_2' = 65$ and $K_2' = 64$. We plot the BER curve of the single-user decorrelating detector as a function of the received (input) SNR γ_k of the user of interest k directly from (10), (11), and (12). The single-user bound (SUB) is also included for reference purposes. The loads $\frac{32}{33}$ and $\frac{64}{65}$ of the "$4m+1$" codes are higher than the corresponding loads $\frac{30}{31}$ and $\frac{62}{63}$ of the "$4m-1$" (Gold) codes. Still, the decorrelating detector

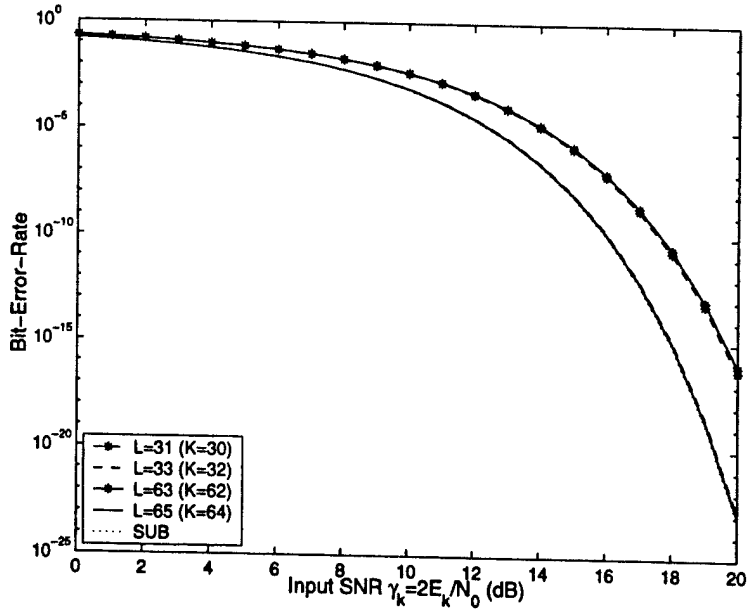


Figure 1. BER of single-user decorrelating detector versus input SNR for 31 (Gold), 33, 63 (Gold), and 65-long doubly optimal sets with 30, 32, 62, and 64 users, respectively.

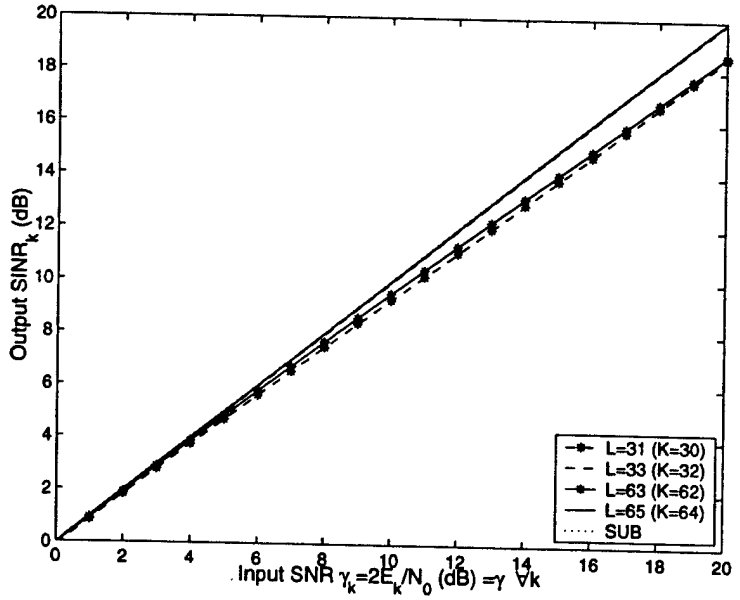


Figure 2. Output SINR of single-user MMSE filter versus input SNR for 31 (Gold), 33, 63 (Gold), and 65-long doubly optimal sets with 30, 32, 62, and 64 users, respectively. Perfect user power control is assumed.

performs practically at the SUB for the former but not for the latter. We conclude that we achieve almost SUB performance with a linear decorrelator if the processing gain is of the form $4m+1$ and the signature set is doubly optimal. This is not true, however, for $(4m-1)-$long doubly optimal spreading codes (that include the Gold sets).

In Fig. 2, we maintain the same set-up as in Fig. 1 and we use expressions (24) and (26) to plot the *output SINR* of the single-user MMSE filter as a function of the *input SNR* which for simplicity is assumed to be the

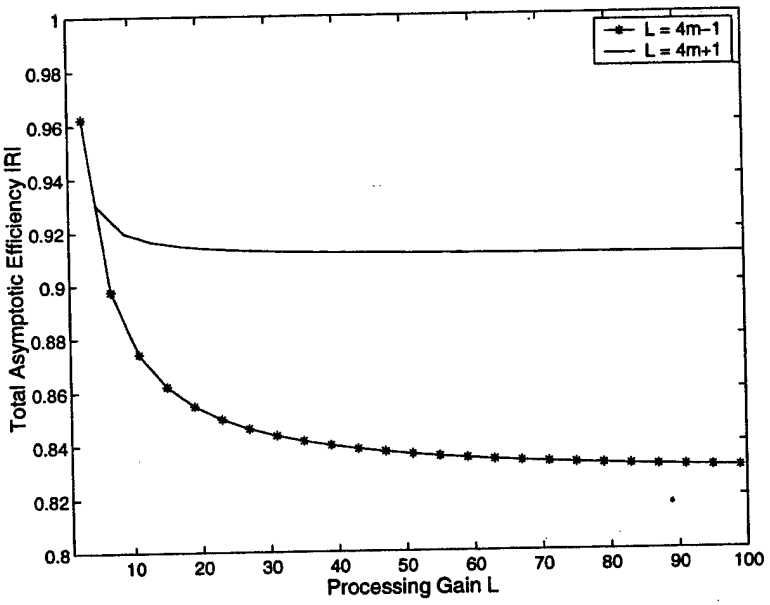


Figure 3. Total asymptotic efficiency versus processing gain for fixed user load $\alpha = 0.5$.

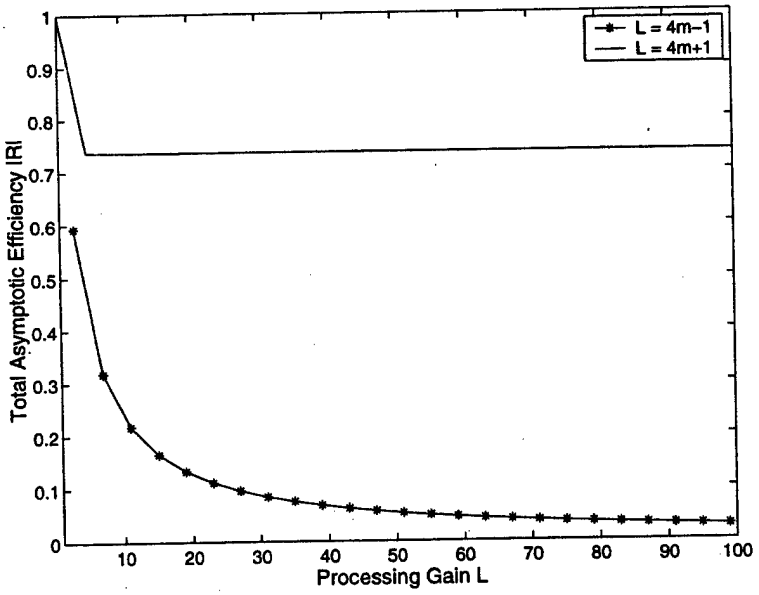


Figure 4. Total asymptotic efficiency versus processing gain for fixed user load $\alpha = 1$.

same for all users (power control). For $(4m+1)-$long codes the SNR "loss" due to multiuser interference is practically non-existent. This is not, however, the case for $(4m-1)-$long (Gold) codes. Notice that for an input SNR of 20dB the output SINR of the MMSE filter for 31 and 63-long doubly optimal (Gold) codes is about 18.6dB, while it is 19.9dB for 33 and 65-long "$4m+1$" doubly optimal codes.

Numerical studies on the total asymptotic efficiency $|\mathbf{R}|$ are presented in Figs. 3 and 4 where we plot $|\mathbf{R}|$ as a function of the processing gain L for a fixed load α for both $(4m-1)-$long and $(4m+1)-$long doubly optimal codes. In Fig. 3, α is set to 0.5 (half-loaded system). In Fig. 4, α is set to 1 (fully loaded system). In both studies, the superiority of $(4m+1)-$long codes is apparent. It is interesting, for example, to observe in Fig. 3

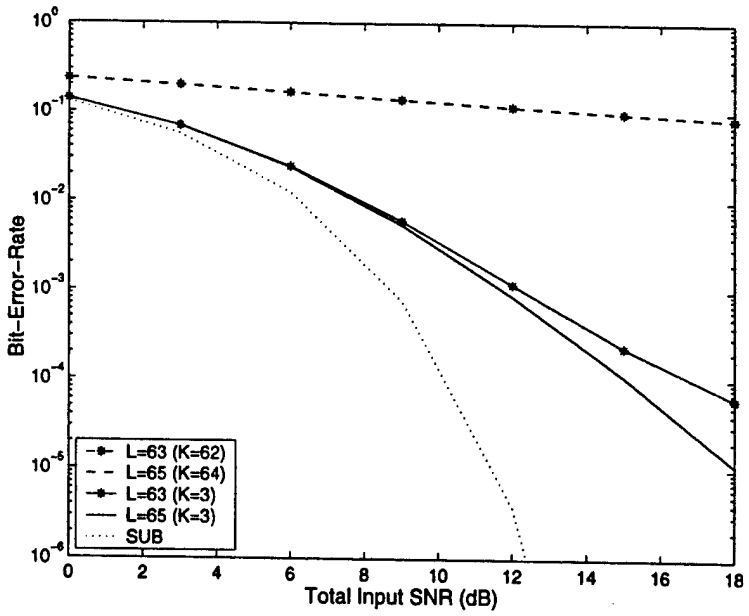


Figure 5. BER of single-user MMSE filter versus total input SNR for 63-long (Gold) and 65-long doubly optimal sets (asynchronous link, three unit-gain paths).

the convergence of $|\mathbf{R}|$ to $\frac{\sqrt{e}}{2} \simeq .82$ and $\frac{3}{2\sqrt{e}} \simeq .91$ for "$4m-1$" and "$4m+1$" doubly optimal codes, respectively, as proved in the previous section. Similarly, we observe in Fig. 4 the quick convergence of $|\mathbf{R}|$ to 0 and $\frac{2}{e} \simeq .73$, respectively.

Finally, a *simulation* study that examines the performance of a 63-long Gold and a 65-long doubly optimal set for *asynchronous* transmissions over multipath channels is presented in Fig. 5. We consider the presence of 3 fixed chip-spaced paths (of equal gain set to 1) for each user signal and we assume independent uniformly distributed chip-spaced time delay for each interfering signal with respect to the user signal of interest. We plot the average BER of the MMSE filter receiver (of length equal to the multipath-extended processing gain $L+3-1=L+2$) as a function of the total received (input) SNR. Here, the SUB curve identifies the performance of one-shot ML detection in complete absence of multiuser interference ($K=1$ user signal). The study indicates that the $(4m-1)-$long Gold set and the $(4m+1)-$long doubly optimal set perform quite similarly for both lightly loaded ($K=3$) and heavily loaded ($K=62$, 64) systems. We notice the rather disappointing performance of both spreading code sets especially when the system is heavily loaded, which -from an optimistic point of view- can be seen as significant room for improvement in signature set design for asynchronous/multipath channels. The exact same study is repeated in Fig. 6 for a *synchronous* link with three unit-gain paths and indicates that a primary reason for the performance deterioration exhibited in Fig. 5 by both code sets is asynchronism (a secondary reason is the few multipaths that are present). We recall that the "$4m+1$" codes of Ref. 8, 9, 11 have not been "optimized" in any way for asynchronous/multipath transmissions. The Gold codes are instead presented in the literature as optimized but under a *cyclic-shift* cross-correlation and auto-correlation measure which can be argued to be inappropriate for asynchronous and multipath, correspondingly, DS-CDMA channels.[¶]

5. SUMMARY AND CONCLUSIONS

This paper presented a theoretical study of the code division multiplexing performance of doubly optimal binary signature sets of odd length. These are sets, we recall, that have minimum both total-squared-correlation (TSC) and maximum-squared-correlation (MSC) at the same time.[8,9,11] We examined separately the doubly optimal

[¶]Gold spreading codes[15] possess preferred *cyclic-shift* cross-correlation and auto-correlation properties. However, they do not guarantee "good" *zero-padded-shift* cross-correlation and auto-correlation properties. The latter measure may represent better the needs of DS-CDMA systems with asynchronous links over single or multipath channels.

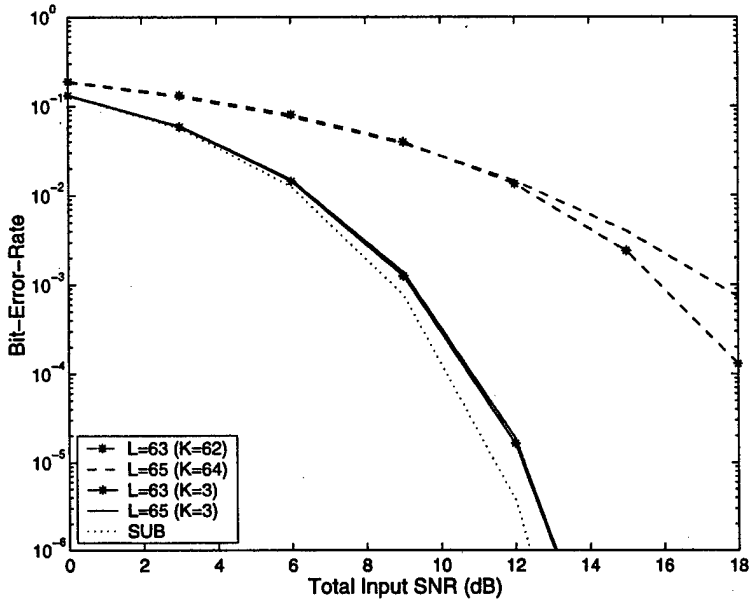


Figure 6. Same study as in Fig. 5 but for a synchronous link (still three unit-gain paths).

sets with signature length of the form $4m-1$ and $4m+1$, $m = 1, 2, \ldots$, and we noted that the familiar Gold sets[15] fall under the "$4m-1$" class of doubly optimal sets.

First, we derived explicit closed-form formulas for the signature cross-correlation matrix of such sets, its eigenvalues, and its inverse. Based on these formulas, we were able to obtain analytic expressions for the BER of decorrelating filters, the (maximum) output SINR of MMSE filters, and the total asymptotic efficiency of synchronous code division multiplexing systems. The calculated across-the-board superiority of the "$4m+1$" signature sets is intriguing. Combined with the indicated similar to Gold sets performance under asynchronous transmissions, this strong proven superiority for (near-)synchronous links of the "$4m+1$" sets questions seriously the status quo and the popularity of Gold sets (or any other "$4m-1$" doubly optimal candidate set to that respect). Regretfully, neither the "$4m+1$" codes nor the Gold codes have what we might call satisfactory performance for heavy-loaded asynchronous transmissions. The "$4m+1$" doubly optimal codes have not been optimized in any way for asynchronous transmissions. The Gold codes are presented as optimized but under a cyclic-shift cross-correlation metric that, apparently, may be criticized as inappropriate.

REFERENCES

1. L. R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inform. Theory* **20**, pp. 397–399, May 1974.
2. M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory* **40**, pp. 1261–1266, July 1994.
3. J. L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II, Methods in Communication, Security, and Computer Sciences*, R. Capocelli, A. D. Santis, and U. Vaccaro, eds., New York: Springer-Verlag, 1993.
4. P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inform. Theory* **45**, pp. 1968–1983, Sept. 1999.
5. P. Cotae, "An algorithm for obtaining welch bound equality sequences for S-CDMA channels," *AEÜ. Int. J. Electron. Commun.* **55**, pp. 95–99, Mar. 2001.
6. S. Ulukus and R. D. Yates, "Iterative construction of optimum signature sequence sets in synchronous CDMA systems," *IEEE Trans. Inform. Theory* **47**, pp. 1989–1998, July 2001.

7. C. Rose, "CDMA codeword optimization: Interference avoidance and convergence via class warfare," *IEEE Trans. Inform. Theory* **47**, pp. 2368–2382, Sept. 2001.
8. G. N. Karystinos and D. A. Pados, "New bounds on the total squared correlation and optimum design of DS-CDMA binary signature sets," *IEEE Trans. Commun.* **51**, pp. 48–51, Jan. 2003.
9. C. Ding, M. Golin, and T. Kløve, "Meeting the Welch and Karystinos-Pados bounds on DS-CDMA binary signature sets," *Designs, Codes and Cryptography*, to appear. (Available on-line at http://www.cs.ust.hk/tcsc/RR/index_5.html).
10. E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence cdma and wideband cdma cellular networks," *IEEE Comm. Magazine* **36**, pp. 48–54, Sept. 1998.
11. G. N. Karystinos and D. A. Pados, "Binary CDMA signature sets with concurrently minimum total-squared-correlation and maximum-squared-correlation," in *Proc. 2003 IEEE Intern. Conf. Commun.*, Anchorage, AK, May 2003.
12. H. Rademacher, "Einige sätze über reihen von allgemeinen orthogonalfunktionen," *Mathematische Annalen* **87**, pp. 112–138, 1922.
13. J. L. Walsh, "A closed set of normal orthogonal functions," *American Journal of Mathematics* **45**, pp. 5–24, Jan. 1923.
14. N. Zierler, "Linear recurring sequences," *Journal of the Society for Industrial and Applied Math.* **7**, pp. 31–48, Mar. 1959.
15. R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. Inform. Theory* **13**, pp. 619–621, Oct. 1967.
16. C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, PA, 2000.
17. K. Schneider, "Optimum detection of code division multiplexed signals," *IEEE Trans. Aerospace Electron. Syst.* **15**, pp. 181–185, Jan. 1979.
18. R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple access channels," *IEEE Trans. Inform. Theory* **35**, pp. 123–136, Jan. 1989.
19. J. G. Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, New York, NY, 1995.
20. Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.* **8**, pp. 683–690, May 1990.
21. U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.* **42**, pp. 3178–3188, Dec. 1994.
22. S. Verdu, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. 24th Annual Allerton Conf. Commun., Control and Computing*, pp. 1025–1034, Allerton, IL, Oct. 1986.
23. D. Parsavand and M. K. Varanasi, "RMS bandwidth constrained signature waveforms that maximize the total capacity of PAM-synchronous CDMA channels," *IEEE Trans. Commun.* **44**, pp. 65–75, Jan. 1996.

Near-ML Multiuser Detection with Linear Filters and Reliability-based Processing

Zhenyu Liu, *Student Member, IEEE,* and Dimitris A. Pados, *Member, IEEE*

***Abstract*— The prohibitive -exponential in the number of users- computational complexity of the maximum likelihood (ML) multiuser detector for direct-sequence code-division-multiple-access (DS/CDMA) communications has fueled an extensive research effort for the development of low complexity multiuser detection alternatives. In this article, we show that we can efficiently and effectively approach the error rate performance of the optimum multiuser detector as follows. We utilize a multiuser zero-forcing or minimum-mean-square-error (MMSE) linear filter as a preprocessor and we establish that the output magnitudes, when properly scaled, provide a reliability measure for each user bit decision. Then, we prepare an ordered reliability-based error search sequence of length linear in the number of users that returns the most likely user bit vector among all visited options. Numerical and simulation studies for moderately loaded systems that permit exact implementation of the optimum detector indicate that the error rate performance of the optimum and the proposed detector are nearly indistinguishable over the whole pre-detection signal-to-noise ratio (SNR) range of practical interest. Similar studies for higher user loads (that prohibit comparisons with the optimum detector) demonstrate error rate performance gains of orders of magnitude in comparison with straight decorrelating or MMSE multiuser detection.**

***Index Terms*— Code-division-multiaccess, decorrelation, maximum likelihood detection, mean square error methods, multiuser channels, reliability, soft decision.**

I. Introduction

In direct-sequence code-division-multiple-access (DS/CDMA) systems, multiple users transmit information bit sequences over a common channel utilizing distinct individually assigned signature waveforms. In the presence of strong interferers and/or high cross-correlation signature waveforms, the performance of the conventional single-user signature matched-filter detector degrades significantly. On the other hand, if the user signatures, energies, and the channel noise power are known at the receiver end, use of the maximum-likelihood (ML) multiuser detector [1] guarantees minimum probability of error system performance. ML optimum multiuser detection, however, comes with computational complexity that grows exponentially with the number of users.

In search of a satisfactory trade-off between performance and complexity, numerous sub-optimum multiuser detectors have been proposed. Two linear multiuser detectors that are of particular interest to our work are the familiar zero-forcing decorrelating detector [2] and the minimum-mean-square-error (MMSE) detector [3]-[6]. Both are known to be near-far resistant [7]. The decorrelator is the maximum likelihood estimator of the energy modulated user bits for additive white Gaussian noise (AWGN) multiple-access channels. The MMSE receiver is the linear filter solution that maximizes the pre-detection signal-to-interference-plus-noise ratio (SINR) of each user.

Recently, there has been an effort to directly approximate the ML multiuser detector in an efficient and effective manner. Examples of such work are the "greedy" detector in [8] which partitions the likelihood metric function and proceeds with "part-by-part" maximization and the neighbor searching mechanisms in [9], [10]. The "MK-face" detector in [9] searches within a pre-calculated set of neighboring candidates. The gradient search algorithm in [10] happens to coincide with the "K-face" detector of [9].

As a combinatorial optimization problem, ML multiuser detection has certain similarities with ML decoding of binary linear block codes. One of the earliest, well known low-complexity algorithms for sub-ML block-code decoding is Chase's algorithm [11]. The Chase procedure selects a small fixed number of bit decision perturbations based on reliability measurements which are supposed to be part of the demodulator output (usually directly related to the sufficient decision statistic). In this paper, we follow an analogous approach to tackle the ML multiuser detection problem. First, we use a linear multiuser detector (such as the decorrelator or MMSE detector) to obtain an initial multiuser bit decision. The soft output vector, before being discarded, is processed to derive the corresponding reliability measurement vector. Next, as dictated by the reliability vector, a sequence of "error patterns" is generated to "correct" the initial decision vector and produce multiuser bit combination candidates that include the maximum likelihood solution with high probability. Final decision is made by simply choosing the bit combination with the highest likelihood among all candidates. When the length of the error pattern sequence is in the order of the number of users, the proposed algorithm exhibits near-ML performance at remarkably reduced computational complexity compared to the ML-optimum multiuser detector.[1]

We organize the rest of the paper as follows. In Section II, we present our signal model and notation. The near-ML

Manuscript received XXXX; revised XXXX. This work was supported in part by the National Science Foundation under Grant ECS-0073660 and Grant CCR-0219903 and in part by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory, USAF, under agreement number F30602-00-1-0520. This paper was presented in part at the 2001 SPIE Digital Wireless Communication Conference, Orlando, FL, April 2001, and in part at IEEE Milcom 2001, Tysons Corner, VA, October 2001.

The authors are with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: zliu2@eng.buffalo.edu; pados@eng.bufalo.edu).

[1] Another coding theory inspired multiuser detector was presented recently in [12]. In contrast to the ML approximations in [8]-[10] and the work herein, the multiuser detector of [12] is a complete exponential-complexity depth-first tree search implementation of the full ML decision rule based on the lattice decoder developed in [13].

multiuser detection algorithm is developed in Section III. Simulation studies are presented in Section IV and a few final conclusions are drawn in Section V.

II. Signal Model and Notation

For the sake of clarity in presentation, we consider a synchronous DS/CDMA system where K users share a multiple-access AWGN channel with power spectral density σ^2. Our developments can be generalized to cover near-ML block multiuser detection for asynchronous systems by creating error pattern sequences of length proportional to the total number of bits in a block. The general block diagram representation of the receiver that we develop in this work is given in Fig. 1. The received signal is processed by a front end that consists of a block of K signature-matched filters. We denote the ith bit period output of the matched filter bank by $\mathbf{y}(i) \triangleq [y_1(i), y_2(i), \cdots, y_K(i)]^T$ (T denotes the transpose operation) and we write

$$\mathbf{y}(i) = \mathbf{R}\mathbf{E}(i)\mathbf{b}(i) + \mathbf{n}(i), \quad i = 1, 2, \ldots, \tag{1}$$

where $\mathbf{b}(i) \triangleq [b_1(i), b_2(i), \cdots, b_K(i)]^T \in \{\pm 1\}^K$ is the ith period information bit vector, $\mathbf{E}(i) \triangleq \mathrm{diag}\left(\left[\sqrt{E_1(i)}, \sqrt{E_2(i)}, \cdots, \sqrt{E_K(i)}\right]\right)$ is the $K \times K$ user signal amplitude matrix, $\mathbf{R}$ is the $K \times K$ signature cross-correlation matrix and $\mathbf{n}(i) \triangleq [n_1(i), n_2(i), \cdots, n_K(i)]^T$ is the filtered noise vector. The ML optimum multiuser detector [1] is

$$\hat{\mathbf{b}}_{\mathrm{ML}} = \arg \max_{\mathbf{b} \in \{\pm 1\}^K} \left(\Omega(\mathbf{b}, \mathbf{y}) \triangleq 2\mathbf{b}^T\mathbf{E}\mathbf{y} - \mathbf{b}^T\mathbf{E}\mathbf{R}\mathbf{E}\mathbf{b}\right) \tag{2}$$

where for notational simplicity we dropped the index i with the understanding that the decision rule in (2) refers to a given information bit interval.

III. Near-ML Multiuser Detection

We switch now our attention to the linear operator $\mathbf{L}$ and the "Search & Decision" blocks in Fig. 1. The linear operator can be viewed as a multiuser detector that operates on the output of the matched-filter bank $\mathbf{y}$ to produce an initial decision $\hat{\mathbf{b}}^{(0)}$ for the transmitted bit vector $\mathbf{b}$. Alongside the initial decision, we also extract a "reliability" measurement vector $\boldsymbol{\alpha}$. During the final "Search & Decision" stage, an ordered sequence of D distinct error patterns $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \ldots, \mathbf{e}^{(D)} \in \{0,1\}^K$ is generated from the reliability vector $\boldsymbol{\alpha}$ where the positive integer D is a system parameter to be adjusted according to our performance versus complexity requirements. Corresponding to each error pattern $\mathbf{e}^{(d)}$, $d = 1, 2, \ldots, D$, a new tentative decision $\hat{\mathbf{b}}^{(d)}$ is produced. The final decision output becomes

$$\hat{\mathbf{b}}_{\text{Near-ML}} = \arg \max_{\mathbf{b} \in \{\hat{\mathbf{b}}^{(1)}, \hat{\mathbf{b}}^{(2)}, \ldots, \hat{\mathbf{b}}^{(D)}\}} \Omega(\mathbf{b}, \mathbf{y}). \tag{3}$$

As we will see later in Section IV, when $\mathbf{L}$ is the decorrelating or MMSE operator, a small D (for instance of the same order as the number of users/bits to be detected) results in bit-error-rate system performance that is nearly indistinguishable from the maximum likelihood detector in operational environments of practical interest.

A. Initial Decision and Reliability Measurement

We use a linear operator (multiuser detector) $\mathbf{L}_{K \times K}$ to obtain the initial bit decision vector $\hat{\mathbf{b}}^{(0)} = \mathrm{sgn}(\mathbf{L}\mathbf{y})$. To exploit the reliability information contained in the linear operator/filter output, we define for each initial bit decision $\hat{b}_j^{(0)}$ the corresponding soft output

$$\tilde{\alpha}_j \triangleq \boldsymbol{l}_j^T\mathbf{y}, \quad j = 1, 2, \ldots, K, \tag{4}$$

where $\boldsymbol{l}_j^T$ is the jth row of the matrix $\mathbf{L}$. For convenience, we also define $\mathbf{w}_j^T \triangleq \boldsymbol{l}_j^T\mathbf{R}\mathbf{E}$ and substitute (1) in (4) to calculate

$$\tilde{\alpha}_j = \mathbf{w}_j^T\mathbf{b} + \boldsymbol{l}_j^T\mathbf{n} = w_{jj}b_j + \sum_{i=1, i \neq j}^{K} w_{ji}b_i + \boldsymbol{l}_j^T\mathbf{n}, \; j = 1, 2, \ldots, K. \tag{5}$$

The first term on the right-hand side of (5) is the scaled jth transmitted bit of interest. The last term, $\boldsymbol{l}_j^T\mathbf{n} \triangleq n_j$, is zero-mean filtered Gaussian noise with variance $E\{n_j^2\} = \sigma^2\boldsymbol{l}_j^T\mathbf{R}\boldsymbol{l}_j \triangleq \sigma_{n_j}^2$. The middle term $\sum_{i=1, i \neq j}^{K} w_{ji}b_i \triangleq z_j$ contains the information bits transmitted by the other users (multiple-access-interference or MAI) and is a zero-mean random variable with variance $E\{z_j^2\} = \sum_{i=1, i \neq j}^{K} w_{ji}^2 = \|\mathbf{w}_j\|^2 - w_{jj}^2 \triangleq \sigma_{z_j}^2$. We define the *reliability value* α_j of the initial bit decision $\hat{b}_j^{(0)}$ as the log-likelihood ratio $\alpha_j \triangleq \frac{1}{2}\ln\frac{f\left(\tilde{\alpha}_j | b_j = \hat{b}_j^{(0)}\right)}{f\left(\tilde{\alpha}_j | b_j = -\hat{b}_j^{(0)}\right)}$ where $f(\tilde{\alpha}_j | b_j)$ denotes the probability density function of $\tilde{\alpha}_j$ conditioned on b_j. If the filtered MAI variance $E\{z_j^2\}$ is relatively small, the soft output $\tilde{\alpha}_j$ is well approximated [14], [15] by a Gaussian random variable with mean $E\{\tilde{\alpha}_j\} = w_{jj}b_j$ and variance $E\{\tilde{\alpha}_j^2\} = \sigma_{z_j}^2 + \sigma_{n_j}^2$. Then, the reliability value α_j simplifies to a scaled version of $|\tilde{\alpha}_j|$:

$$\alpha_j \triangleq \frac{1}{2}\ln\frac{f\left(\tilde{\alpha}_j | b_j = \hat{b}_j^{(0)}\right)}{f\left(\tilde{\alpha}_j | b_j = -\hat{b}_j^{(0)}\right)} \cong \frac{\tilde{\alpha}_j w_{jj}\hat{b}_j^{(0)}}{\sigma_{z_j}^2 + \sigma_{n_j}^2} = \frac{w_{jj}}{\sigma_{z_j}^2 + \sigma_{n_j}^2}|\tilde{\alpha}_j|. \tag{6}$$

The generated reliability vector $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_K]^T$ is passed to the "Search & Decision" stage along with the hard-limited initial "guess" $\hat{\mathbf{b}}^{(0)}$.

B. Generation of the Error Pattern Sequence

When $\mathbf{L}$ is the decorrelating or MMSE filter operator, we expect intuitively that a small perturbation of the initial vector $\hat{\mathbf{b}}^{(0)}$ may produce the maximum likelihood decision $\hat{\mathbf{b}}_{\mathrm{ML}}$. With this motivation, we generate a sequence of error patterns $\mathbf{e}^{(d)} \in \{0,1\}^K$, $d = 1, 2, \ldots,$ where $e_j^{(d)} = 1$ stands for an error in bit position j. We apply the error sequence $\mathbf{e}^{(d)}$ on the initial bit vector $\hat{\mathbf{b}}^{(0)}$ to create the bit vector sequence $\hat{\mathbf{b}}^{(d)}$ as follows:

$$\hat{b}_j^{(d)} = \hat{b}_j^{(0)} \oplus e_j^{(d)}, \quad j = 1, 2, \ldots, K, \tag{7}$$

where $\oplus$ denotes the error correction operation $b \oplus e = b \cdot (-1)^e$, $b \in \{\pm 1\}$, $e \in \{0,1\}$. Naturally, the error sequence generation criterion will be the likelihood of $\hat{\mathbf{b}} = \hat{\mathbf{b}}^{(0)} \oplus \mathbf{e}$ with observation the soft-output vector $\tilde{\boldsymbol{\alpha}} = \mathbf{L}\mathbf{y}$, $f\left(\tilde{\boldsymbol{\alpha}} | \mathbf{b} = \hat{\mathbf{b}}\right)$. If,

approximately, we treat the soft outputs $\tilde{\alpha}_j$ as conditionally independent random variables, the likelihood function becomes $\prod_{j=1}^{K} f\left(\tilde{\alpha}_j | b_j = \hat{b}_j\right)$ and the log-likelihood ratio is

$$\Lambda(\hat{\mathbf{b}}) \cong \frac{1}{2}\sum_{j=1}^{K} \ln \frac{f\left(\tilde{\alpha}_j | b_j = \hat{b}_j\right)}{f\left(\tilde{\alpha}_j | b_j = -\hat{b}_j\right)} = \sum_{j:\hat{b}_j = \hat{b}_j^{(0)}} \alpha_j - \sum_{j:\hat{b}_j \neq \hat{b}_j^{(0)}} \alpha_j$$

$$= \sum_{j=1}^{K} \alpha_j - 2\sum_{j:e_j=1} \alpha_j. \tag{8}$$

We immediately observe that the first term in (8), $\sum_{j=1}^{K} \alpha_j$, is independent of the choice of the error pattern $\mathbf{e}$. Therefore, the error pattern sequence should be designed in ascending order of the other term $\phi(\mathbf{e}, \boldsymbol{\alpha}) \triangleq \sum_{j:e_j=1} \alpha_j$. For our purposes of near-ML multiuser detection, if all 2^K error patterns $\left\{\mathbf{e}^{(d)}\right\}_{j=1}^{2^K}$ are in ascending order under the $\phi(\cdot, \boldsymbol{\alpha})$ key

$$\phi\left(\mathbf{e}^{(d)}, \boldsymbol{\alpha}\right) \leq \phi\left(\mathbf{e}^{(d+1)}, \boldsymbol{\alpha}\right), \quad d = 1, 2, \ldots, 2^K - 1, \tag{9}$$

we will consider and process only the first D elements of the sequence.

We now derive an algorithm that produces the first D $\phi(\cdot,\cdot)$-ordered error patterns with computational complexity linear in the number of bits, $O(DK)$. The origins of our algorithm are in the "subset-sum problem" where the objective is to find all subsets of a set of numbers that have sum of elements less than or equal to a given number [16], [17]. Suppose that the reliability values α_j, $j = 1, 2, \ldots, K$, are such that $\alpha_{h_1} \leq \alpha_{h_2} \leq \cdots \leq \alpha_{h_K}$. Define the error pattern sequences $\mathcal{E}_p \triangleq \left\{\mathbf{e}^{(p_t)}\right\}$, $p = 0, 1, 2, \ldots, K$, $t = 1, 2, \ldots, 2^p$, where $\left\{\mathbf{e}^{(p_t)}\right\}$ is ordered under the key $\phi(\cdot, \boldsymbol{\alpha})$

$$\phi\left(\mathbf{e}^{(p_t)}, \boldsymbol{\alpha}\right) \leq \phi\left(\mathbf{e}^{(p_{t+1})}, \boldsymbol{\alpha}\right), \quad t = 1, 2, \ldots, 2^p - 1, \tag{10}$$

and satisfies

$$e_h^{(p_t)} = 0 \text{ for every } h \in \{h_{p+1}, \cdots, h_K\} \text{ and } 1 \leq t \leq 2^p. \tag{11}$$

Here, $e_h^{(\cdot)}$ is the hth bit in the error pattern $\mathbf{e}^{(\cdot)}$ and (11) states that $\mathbf{e}^{(p_t)}$ may include errors only in the p lowest reliability positions of the initial bit decision vector $\hat{\mathbf{b}}^{(0)}$. Define also the error pattern sequences $\mathcal{E}'_p \triangleq \left\{\mathbf{e}'^{(p_t)}\right\}$, $p = 0, 1, 2, \ldots, K-1$, $t = 1, 2, \ldots, 2^p$, where $\left\{\mathbf{e}'^{(p_t)}\right\}$ is such that

$$e_h'^{(p_t)} = \begin{cases} 1, & \text{if } h = h_{p+1} \\ e_h^{(p_t)}, & \text{otherwise.} \end{cases} \tag{12}$$

Therefore, each $\mathbf{e}'^{(p_t)}$ contains *exactly* one more error in bit position h_{p+1} than the corresponding error pattern $\mathbf{e}^{(p_t)}$. It follows that $\phi\left(\mathbf{e}'^{(p_t)}, \boldsymbol{\alpha}\right) = \phi\left(\mathbf{e}^{(p_t)}, \boldsymbol{\alpha}\right) + \alpha_{h_{p+1}}$ and $\left\{\mathbf{e}'^{(p_t)}\right\}$ is also ordered under the key $\phi(\cdot, \boldsymbol{\alpha})$.

At this point, it is important to observe the following. If we merge the two ordered sequences $\mathcal{E}_p = \{\mathbf{e}^{(\cdot)}\}$ and $\mathcal{E}'_p = \{\mathbf{e}'^{(\cdot)}\}$ in increasing order of the same key $\phi(\cdot, \boldsymbol{\alpha})$, then we will obtain $\mathcal{E}_{p+1}$. Therefore, iterative generation of $\mathcal{E}'_p$ from $\mathcal{E}_p$ and merging to form $\mathcal{E}_{p+1}$ for $p = 0, 1, \ldots, K-1$ leads to $\mathcal{E}_K$ where $\mathcal{E}_K$ is the whole set of 2^K error patterns ordered under the key $\phi(\cdot, \boldsymbol{\alpha})$. Since only the first D error patterns in $\mathcal{E}_K$ are needed, we can simplify this process greatly. We observe that the first D elements in $\mathcal{E}_K$ come from the first D elements of $\mathcal{E}_{K-1}$ and the first D elements[2] of $\mathcal{E}'_{K-1}$. Therefore, we can truncate the length of $\mathcal{E}_{K-1}$ to D (which also limits the length of $\mathcal{E}'_{K-1}$ to D). By induction, we conclude that every sequence $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_K$ can be shortened to length D. Considering the involved computational complexity, we need up to D additions to calculate $\phi(\cdot, \boldsymbol{\alpha})$ and another D comparisons to merge $\mathcal{E}_p$ and $\mathcal{E}'_p$ in each step. Therefore, the computational complexity is $O(DK)$.

In summary, the proposed multiuser detection system in Fig. 1 operates as follows:

1) Apply the linear operator/filter $\mathbf{L}$ to the matched-filter bank output $\mathbf{y}$ to obtain the initial decision $\hat{\mathbf{b}}^{(0)}$ and the reliability measures $\{\alpha_j\}$ in (6).
2) Sort $\{\alpha_j\}$ in increasing order.
3) Initialize $p \leftarrow 0$, $\mathcal{E}_0 = \left\{\mathbf{e}^{(1)} = [0\ 0\ \cdots\ 0]\right\}$, $\phi\left(\mathbf{e}^{(1)}, \boldsymbol{\alpha}\right) = 0$.
4) Create $\mathcal{E}'_p = \left\{ \mathbf{e}'^{(i)} = \mathbf{e}^{(i)} + [\underbrace{0\cdots0}_{h_{p+1}-1}\ 1\ \underbrace{0\cdots0}_{K-h_{p+1}}] : \mathbf{e}^{(i)} \in \mathcal{E}_p \right\}$. Calculate $\phi\left(\mathbf{e}'^{(i)}, \boldsymbol{\alpha}\right) = \phi\left(\mathbf{e}^{(i)}, \boldsymbol{\alpha}\right) + \alpha_{h_{p+1}}$.
5) Merge $\mathcal{E}_p$ and $\mathcal{E}'_p$ to form $\mathcal{E}_{p+1}$ in ascending order under key $\phi(\cdot, \boldsymbol{\alpha})$. Truncate the length of $\mathcal{E}_{p+1}$ to D if longer than D.
6) If $p < K-1$, then set $p \leftarrow p+1$ and go back to Step 4. Else go to Step 7.
7) For $d = 1, 2, \ldots, D$ apply $\mathbf{e}^{(d)}$ to the initial decision $\hat{\mathbf{b}}^{(0)}$ as in (7). Produce the final decision $\hat{\mathbf{b}}_{\text{Near-ML}}$ by (3).

The overall system computational complexity can be calculated as follows. The complexity of applying the linear operator $\mathbf{L}$ to the matched-filter output $\mathbf{y}$ is $O\left(K^2\right)$. Sorting the reliability values costs $O\left(K \log_2 K\right)$. As stated earlier, the complexity of generating the ordered sequence $\left\{\mathbf{e}^{(d)}\right\}_{d=1}^{D}$ is $O(DK)$. Finally, $\Omega(\cdot,\cdot)$ is calculated D times with complexity, again, $O(DK)$. Therefore, the complexity of the whole scheme is approximately $O\left(K(K + 2D + \log_2 K)\right)$ per bit period. The cost of the "Near-ML Search & Decide" block in Fig. 1 is barely more than twice the cost of the $\mathbf{Ly}$ operation if D is chosen equal to K (that is, if D is equal to the number of user bits to be detected). As we will see later in Section IV, when the linear operator $\mathbf{L}$ is chosen appropriately a value of D equal to K or -more general- a few times K can be sufficient for the algorithm to maintain near-ML performance for all practical purposes.

C. Selection of the Linear Operator

So far, we have intentionally treated the linear operator $\mathbf{L}$ as arbitrary. Yet, it is important to choose an operator/filter that produces initial decisions with low probability of error such that only a small perturbation is needed to approach the ML

[2]In fact, only the first $\lfloor \frac{D}{2} \rfloor$ elements of $\mathcal{E}'_{K-1}$ can contribute to the first D elements of $\mathcal{E}_K$ but this does not affect our conclusion on the rough order of the involved computational complexity and may complicate our presentation unnecessarily.

optimum decision vector. Conversely, if the initial decisions are poor, then to achieve the same bit-error-rate performance level we will need a larger search length parameter D and that will thereby increase the computational complexity.

Besides the initial performance considerations, the condition that the variance of the MAI term $z_j = \sum_{i=1, i\neq j}^{K} w_{ji} b_i$ in (5) is relatively small should also be fulfilled. This directly affects the accuracy of the Gaussian approximation of the soft output $\tilde{\alpha}_j$ which in turn affects the quality of the reliability measure α_j in (6). There are two well-researched linear filters that can satisfy the quality of initial decision and small MAI variance requirements: the decorrelating and MMSE filters. The decorrelating filter $\mathbf{L}_{\text{DEC}} = \mathbf{R}^{-1}$, leading to zero errors in the absence of noise, has perfect near-far resistance and is the maximum likelihood solution for real-valued "symbol" vectors. If we use $\mathbf{L}_{\text{DEC}} = \mathbf{R}^{-1}$ in (4), (5), we obtain $w_{jj} = \mathbf{E}_{jj}$, $\sigma_{n_j}^2 = \sigma^2 (\mathbf{R}^{-1})_{jj}$, and $\sigma_{z_j}^2 = 0$. Because MAI is completely cancelled, the Gaussian approximation is not needed anymore. The reliability measures in (6) take the form

$$\alpha_j = \frac{\mathbf{E}_{jj}}{\sigma^2 (\mathbf{R}^{-1})_{jj}} |\tilde{\alpha}_j|, \quad j = 1, 2, \ldots, K, \qquad (13)$$

with *no approximation*. The MMSE multiuser filter $\mathbf{L}_{\text{MMSE}} = (\mathbf{R} + \sigma^2 \mathbf{E}^{-2})^{-1}$ minimizes the mean norm square error between its output and the true transmitted information bit vector and maximizes the output SINR per user. However, the filtered MAI term in (5) is not zero: $z_j = \sum_{i=1, i\neq j}^{K} w_{ji} b_i \neq 0$. Still, the MMSE filtered MAI is approximately Gaussian [14],[15] and the reliability measure in (6) maintains its accuracy.

As a final comment, it is clear that technically the soft outputs $\tilde{\alpha}_j$, $j = 1, 2, \ldots, K$, are not independent and their degree of dependency can affect the quality of the log-likelihood ratio approximation in (8) and, in turn, the quality of the generated sequence of error patterns. When the linear operator $\mathbf{L}$ is the decorrelator, it is easy to calculate the correlation coefficient ρ_{ij} between $\tilde{\alpha}_i$ and $\tilde{\alpha}_j$, $i \neq j$: $\rho_{ij} = \frac{E\{\tilde{\alpha}_i \tilde{\alpha}_j\}}{\sqrt{E\{\tilde{\alpha}_i^2\} E\{\tilde{\alpha}_j^2\}}} = \frac{\mathbf{L}_{ij}}{\sqrt{(\text{SNR}_i + \mathbf{L}_{ii})(\text{SNR}_j + \mathbf{L}_{jj})}}$ where $\text{SNR}_i \triangleq \frac{E_i}{\sigma^2}$. In most cases of practical interest, $|\rho_{ij}|$ is less than 10^{-2} and the independence approximation is well justified. In such cases, we found (see for example Section IV) that $D = K$ is sufficient to approximate very closely the performance of the true ML detector. In extreme situations, such as pre-detection SNR values less than 0dB and high signature cross-correlations (greater than 0.5), $|\rho_{ij}|$ can be greater than 0.1 and thus non-negligible. In these cases, a greater D should be used to approach ML performance. For MMSE preprocessors, we can calculate $|\rho_{ij}|$ numerically and, again, adjust D accordingly.

IV. Simulation Studies

We consider the DS/CDMA signal model of Section II with spreading gain 31 and $K = 15$ users with Gold signature code assignments. In Fig. 2, we plot the bit-error-rate (BER) of the user of interest (user 1) under different multiuser detection schemes as a function of its SNR over the $4 - 16$dB range. The SNRs of the 14 interferers are distributed in the $8.5 - 15$dB range with a 0.5dB increment. We use the multiuser MMSE filter as the pre-processor in our proposed system and we fix the length of the error pattern sequence to $D = K = 15$ (hence, the computational complexity per bit interval is $O(DK) = O(K^2)$). For purposes of comparison, we include *(i)* the ML optimum multiuser detector [1] (which serves as a BER lower bound), *(ii)* the stand-alone multiuser MMSE detector[3] [3]-[6], *(iii)* the "greedy" multiuser detector in [8] with different memory parameter settings $L = 1, 2, 3$ and computational complexity per bit interval $O(LK^2)$, *(iv)* the "MK-face" detector [9] with parameter setting $MK = 2 \cdot 15 = 30$, computational complexity per bit interval $O(MK^2)$ and space complexity $O(2^K MK)$, and *(v)* the "gradient search" detector [10] that coincides with the "K-face" algorithm of [9] and has computational complexity $O(K^2)$ per iteration. We observe that in the relatively low SNR region ($4 - 12$dB) the greedy, the MK-face, and the proposed detector have performance very close to the optimum. We recall that the greedy algorithm decides the value of the user information bits sequentially and the order favors high energy bits/users. As the SNR of the "user of interest" increases relative to the other users, bit decisions are made early with the adverse effect of an increase in BER as seen in Fig. 2. For the MK-face detector, the pre-calculated neighboring candidates fail to perturb high SNR user bits and the performance curve degrades to the pre-detector (MMSE in this case). The proposed system maintains "near-ML" performance throughout the range of this study. In Fig. 3, we repeat the study of Fig. 2 and we examine the BER of the proposed detector as a function of the length of the error pattern sequence D at $\text{SNR}_1 = 12$dB and 14dB. The proposed algorithm approaches effectively the ML bound for modest values of D (the number of bits to be detected is $K = 15$; $D = 1$ corresponds to the stand-alone MMSE detector and $D = 2^{15}$ corresponds to the true ML detector).

Finally, in Fig. 4 we increase the number of active users from $K = 15$ to $K = 25$. The SNRs of the 24 interferers are distributed over the $8.5 - 20$dB range with a 0.5dB increment. The performance gain of the proposed "near-ML" algorithm with $D = 4K = 100$ over the MMSE detector is remarkable. True ML detection has complexity proportional to 2^{25} and we cannot afford this computational cost at present. For this reason, we replace the true ML bound that appears in the previous Figs. 2 and 3 by the *single-user bound*. We also cannot afford execution of the "MK-face" algorithm that requires exponential ($2^{25} MK$) storage space. Instead, we present in Fig. 4 the performance curves of the "greedy" ($L = 4$) and "gradient search" ("K-face") algorithms [8]-[10].

V. Conclusions

We described a new multiuser detection algorithm. A decorrelating or MMSE multiuser filter is used as a pre-processor that provides initial decisions and reliability measurements based on which an ordered error pattern sequence of variable length is formed. The error pattern sequence is followed to its

[3] In this study we choose a Gold code assignment that makes the normalized signature cross-correlation between user 1 and the other users equal to $\frac{7}{31}$ or $-\frac{9}{31}$. Notice the remarkable performance difference between the MMSE and the ML detector.

end and the most likely bit vector among all visited options is returned. When the length of the pattern sequence is of the order of the number of bits to be detected, the additional imposed computational cost compared to straight decorrelating or MMSE detection is rather insignificant. Still, in extensive simulation studies for both synchronous and asynchronous links (not reported herein) we saw that the proposed multiuser detection algorithm is able to maintain near-ML bit-error-rate performance over the whole studied SNR range of interest. There is strong resemblance between this scheme and "efficient" decoding algorithms for binary linear block codes [11], [17].

REFERENCES

[1] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 32, pp. 85-96, Jan. 1986.

[2] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123-136, Jan. 1989.

[3] P. B. Rapajic and B. S. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685-697, May 1994.

[4] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, Dec. 1994.

[5] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, pp. 1746-1755, Feb./Mar./Apr. 1995.

[6] C. N. Pateros and G. J. Saulnier, "An adaptive correlator receiver for direct-sequence spread-spectrum communication," *IEEE Trans. Commun.*, vol. 44, pp. 1543-1552, Nov. 1996.

[7] R. Lupas and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496-508, Apr. 1990.

[8] A. AlRustamani and B. R. Vojcic, "A new approach to greedy multiuser detection," *IEEE Trans. Commun.*, vol. 50, pp. 1326-1336, Aug. 2002.

[9] Q. Li and C. N. Georghiades, "On a geometric view of multiuser detection for synchronous DS/CDMA channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2723-2731, Nov. 2000.

[10] J. Hu and R. S. Blum, "A gradient guided search algorithm for multiuser detection," *IEEE Commun. Letters*, vol. 4, pp. 340-342, Nov. 2000.

[11] D. Chase, "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inform. Theory*, vol. 18, pp. 170-182, Jan. 1972.

[12] L. Brunel and J. Boutros, "Lattice decoding for joint detection in direct sequence CDMA systems," *IEEE Trans. Inform. Theory*, submitted Jan. 2000.

[13] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1639-1642, July 1999.

[14] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Commun.*, vol. 43, pp. 858-871, May 1997.

[15] J. Zhang, E. K. P. Chong, and D. N. C. Tse, "Output MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1128-1144, Mar. 2001.

[16] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. New York: McGraw-Hill, 1990.

[17] Y. Wu and D. A. Pados, "An adaptive two-stage algorithm for ML and sub-ML decoding of binary linear block codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 261-269, Jan. 2003.

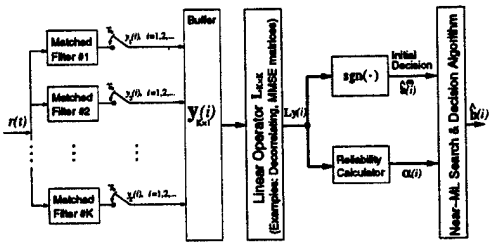


Fig. 1. The proposed "near-ML" multiuser detection architecture.

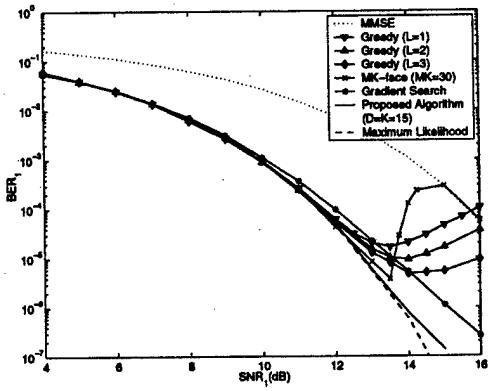


Fig. 2. Bit-error-rate versus SNR for various multiuser detectors ($K = 15$, SNR_i in 8.5 to 15dB range, $i = 2, 3, \cdots, 15$).

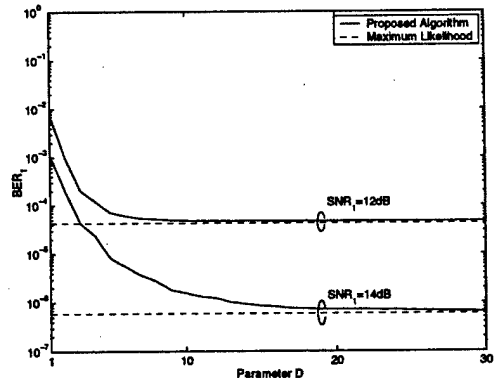


Fig. 3. Bit-error-rate versus length of the error pattern sequence D for the system of Fig. 2 (detection of $K = 15$ user bits).

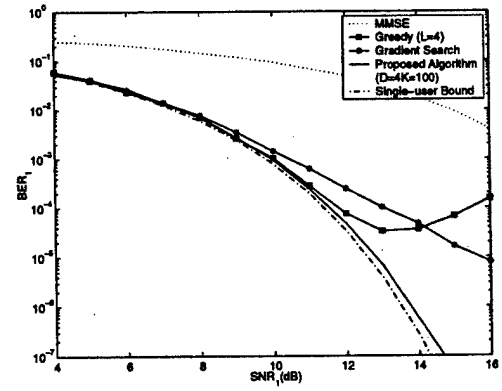


Fig. 4. Bit-error-rate versus SNR performance in the presence of severe MAI ($K = 25$, SNR_i in 8.5 to 20dB range, $i = 2, 3, \cdots, 25$).