

Composite matched filter output partitioning

Robert J. Marks II, James A. Ritcey, Les E. Atlas, and Kwan F. Cheung

A common pattern recognition problem is finding a library element closest, in some sense, to a given reception. In many scenarios, optimal detection requires N matched filters for N library elements. Since N can often be quite large, there is a need for suboptimal techniques that base their decisions on a reduced number of filters. The use of composite matched filters (CMFs) (also called synthetic discriminant functions or linear combination filters) is one technique to achieve this reduction. For two level CMF outputs, the reduction is from N to $\log_2 N$ matched filters. Previously, the coefficients of the CMF output were restricted to positive values—often 0 and 1. We refer to such filters as binary CMFs. An alternative approach is to use -1 and $+1$ for filter coefficients. This alternative filter will be called a bipolar CMF. This paper demonstrates how the extension from a binary to a bipolar CMF greatly improves the detection performance while still maintaining the reduced computational requirements of the binary CMF. Furthermore, the bipolar CMF is invariant to scale: multiplying the input by a positive constant gives the same processor output. This desirable behavior does not exist for the binary CMF.

I. Introduction

Matched filters are commonly used in the design of pattern recognition systems that decide for the presence of one of N library elements. For many scenarios, N matched filters are required. Since N can often be quite large, there is a need for suboptimal techniques that base their decision on a reduced number of filters. One such technique is the use of composite matched filters or CMF's¹⁻⁶ (also referred to as linear combination filters and synthetic discriminant functions). Each CMF is a linear combination of library elements. These filters have been simulated^{5,6} and implemented¹ with some success. In most previous work, the coefficients at the CMF output have been restricted to positive values, often 0 and 1. We will refer to the 0,1 restricted filter as the binary CMF. An alternative approach is to use -1 and $+1$ for filter coefficients. We will call this filter a bipolar CMF. The primary purpose of this paper is to demonstrate how the extension from a binary to a bipolar CMF greatly enhances the detection performance while maintaining reduced complexity, compared with the optimal matched filter

bank. Empirical results show that the bipolar CMF is superior to the binary case even when the processing is inexact.

II. Preliminaries

In this section we briefly review the CMF formulation. The notation follows that of Marks and Atlas.⁵

Let $\{\mathbf{f}_n | n = 0, \dots, N-1\}$ denote N library elements of length L . We form the $L \times N$ library matrix

$$\mathbf{F} = [\mathbf{f}_0 : \dots : \mathbf{f}_{N-1}]$$

and define the $N \times N$ library correlation matrix $\mathbf{R}_F = \mathbf{F}^T \mathbf{F}$. The nm th element of \mathbf{R}_F is $\mathbf{f}_n^T \mathbf{f}_m$ where the superscript T denotes transposition. The $L \times P$ CMF matrix is defined as

$$\mathbf{S} = [\mathbf{s}_1 : \dots : \mathbf{s}_p], \quad (1)$$

where \mathbf{s}_i is the i th CMF. Here $P = \log_2 N$ because, at most, $\log_2 N$ bits are required to specify N distinct binary numbers. Given an $\mathbf{f} \in \{\mathbf{f}_0, \dots, \mathbf{f}_{N-1}\}$, we make a classification decision based on the output $\mathbf{S}^T \mathbf{f}$. As was done in Refs. 1 and 4, we define

$$\mathbf{S} = \mathbf{F} \mathbf{R}_F^{-1} \hat{\Delta}^T,$$

where $\hat{\Delta}$ is the $P \times N$ matrix of ones and zeros formed from the binary representation of the integers $\{0, \dots, N-1\}$. For example, if $N = 4$ ($P = 2$), we have

$$\hat{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ \uparrow & \uparrow & \uparrow & \uparrow \\ 0 & 1 & 2 & 3 \end{bmatrix}. \quad (2)$$

The authors are with University of Washington, Interactive Systems Design Laboratory, Seattle, Washington 98195.

Received 4 October 1986.

0003-6935/87/112274-05\$02.00/0.

© 1987 Optical Society of America.

In general, the i th column of $\hat{\Delta}$ is the binary representation of i . The first row of $\hat{\Delta}$ contains the least significant bit (LSB), while the P th contains the most significant bit (MSB). An unknown input vector is classified by the output vector which is the binary representation of its index; i.e., we have

$$S^T \mathbf{f}_m = \delta_m, \quad (3)$$

where δ_m is the m th column of $\hat{\Delta}$ or the binary representation of $m \in \{0, 1, \dots, N-1\}$.

III. Effect of Input Noise

An exact version of $\mathbf{f} \in \{\mathbf{f}_0, \dots, \mathbf{f}_{N-1}\}$ is never received in practice. A more realistic model assumes we receive $\mathbf{f} + \xi$ where ξ is a vector of length L containing additive noise samples. Replacing \mathbf{f}_m by $\mathbf{f}_m + \xi$ in Eq. (3) produces an output $\delta_m + \eta$ rather than δ_m , where $\eta = S^T \xi$. When ξ is zero mean with covariance matrix $\mathbf{R}_\xi = E \xi \xi^T$, the output noise vector η has zero mean components with covariance matrix

$$\mathbf{R}_\eta = E \eta \eta^T = S^T \mathbf{R}_\xi S, \quad (4)$$

where E denotes expectation. If ξ is jointly Gaussian, $N(\mathbf{O}, \mathbf{R}_\xi)$, then η is Gaussian $N(\mathbf{O}, \mathbf{R}_\eta)$. These second-order properties totally determine $f_\eta(\mathbf{x})$, the probability density function (pdf) of η (Ref. 7)

$$f_\eta(\mathbf{x}) = \{(2\pi)^P |\mathbf{R}_\eta|\}^{-1/2} \exp[-1/2 \mathbf{x}^T \mathbf{R}_\eta^{-1} \mathbf{x}]. \quad (5)$$

Even if ξ is non-Gaussian, under loose conditions, η will be approximately Gaussian distributed due to the central limit theorem. Therefore, we will use the Gaussian model in our analysis.

IV. Decoding Procedure

We take a hypothesis testing viewpoint in which the decision is among N hypotheses $\{H_m\}_{m=0}^{N-1}$. Under H_m , a vector (pattern) $\mathbf{f}_m + \xi$ is received and the CMF outputs a vector

$$H_m : \psi = S^T(\mathbf{f}_m + \xi) = \delta_m + \eta; \quad m = 0, 1, \dots, N-1. \quad (6)$$

We consider two techniques for deciding among the N hypotheses. The first of these deciding procedures was suggested previously¹⁻³ and is computationally straightforward. It is carried out by simple thresholding of the elements of ψ . The second technique is optimal in the sense of minimizing the probability of a decoding error P_E . A decoding error occurs when, for $n = m$, we decide for H_n given that H_m is actually true. Equivalently, we maximize the probability of a correct decision. Thus,

$$P_E = 1 - \sum_{m=0}^{N-1} Pr\{H_m | H_m\} Pr\{H_m\},$$

where $Pr\{H_n | H_m\}$ is the probability of deciding for H_n given that H_m is true and $Pr\{H_m\}$ is the prior probability of H_m . This second approach leads to a computationally intensive postprocessing decoding procedure. As will be shown, minor modification of our $\hat{\Delta}$ matrix leads to a decoding procedure that is both computationally

attractive and optimal in the sense of minimizing of P_E .

V. Thresholding

In the absence of input noise, the components of the CMF output vector ψ are traditionally binary, 0 or 1. The thresholding decoder simply clips the components of ψ at $1/2$. Namely, if ψ_i is a component of ψ , a decoding vector \mathbf{b} with components b_i is formed such that

$$b_i = \begin{cases} 0; & \psi_i \leq 1/2, \\ 1; & \psi_i > 1/2. \end{cases}$$

Since \mathbf{b} has binary components and is of length $P = \log_2 N$, the 2^P possible \mathbf{b} vectors span the range of binary representation of $0, 1, \dots, N-1$. Thus, \mathbf{b} represents the desired classification decision. The procedure is computationally attractive because it acts independently on each of the $P = \log_2 N$ components of \mathbf{b} . The following section will demonstrate that this technique is suboptimal.

A. Voronoi Partitioning

A less computationally attractive, but optimal, procedure is that suggested by detection theory.⁸ In the absence of noise, the vectors $\psi = \delta_m$; $m = 0, \dots, N-1$ correspond to the vertices of a P -dimensional hypercube. Conditional on H_m being true and in the presence of input noise $\xi \sim N(\mathbf{O}, \mathbf{R}_\xi)$, we find $\psi \sim N(\delta_m, \mathbf{R}_\eta)$, where $\mathbf{R}_\eta = S^T \mathbf{R}_\xi S$. In general even if the elements of ξ are white, the elements of η are correlated; i.e., \mathbf{R}_η is not diagonal. This implies that the density of η is not spherically symmetric. An example for $\mathbf{R}_\xi = \mathbf{I}$ and $P = 2$ is shown in Fig. 1(a). The implication for the optimal decoding procedure is that, because of the element-to-element correlation, any decoder that operates on the elements of ψ separately will be suboptimal.

To derive the optimal decoder we uncorrelate the noise vector η by a coordinate transformation. Since \mathbf{R}_η is a covariance matrix it will in all practical situations be positive definite. Thus, there exists a real orthogonal $P \times P$ matrix \mathbf{L} ($\mathbf{L}\mathbf{L}^T = \mathbf{L}\mathbf{L}^T = \mathbf{I}$) such that $\mathbf{R} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^T$. Here $\mathbf{\Lambda}$ is a diagonal matrix whose i th element, $\lambda_i > 0$, is the i th eigenvalue corresponding to the normalized eigenvector \mathbf{l}_i ; the i th column of \mathbf{L} .⁹ Making the substitution $\eta_w = \mathbf{L}\eta$ gives $\eta_w \sim N(\mathbf{O}, \mathbf{\Lambda})$; i.e., the pdf of η_w is given by

$$f_{\eta_w}(\mathbf{x}) = \prod_{j=1}^P (2\pi\lambda_j)^{-1/2} \exp\left[-\frac{1}{2} x_j^2/\lambda_j\right]. \quad (7)$$

From Eq. (7) we see that the components of η_w are statistically independent. We can define

$$\eta_{iid} = \left[\frac{\eta_w(1)}{\sqrt{\lambda_1}}, \dots, \frac{\eta_w(P)}{\sqrt{\lambda_P}} \right]^T,$$

where $\eta_w(j)$ is the j th component of η_w . The components of η_{iid} are independent and identically distributed (iid).

We have just described the first stage of the optimal decoder. Given ψ we form hypotheses: $\tilde{H}_m : \psi_w = \mathbf{L}\psi$

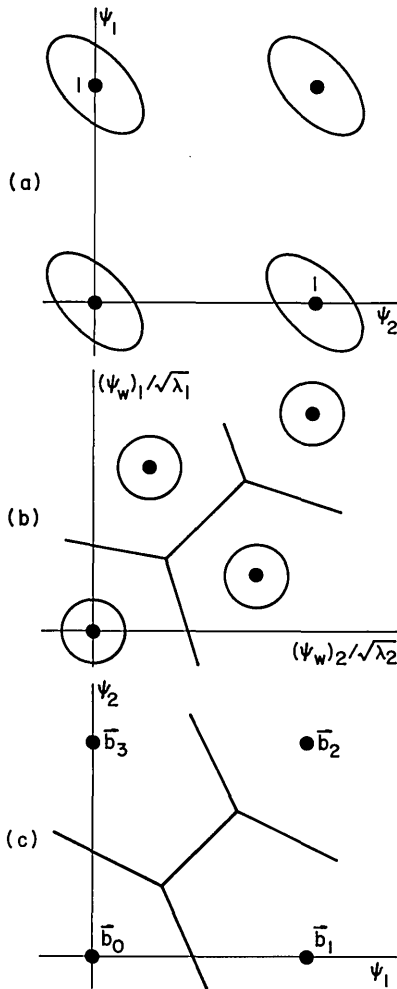


Fig. 1. Illustration of Voronoi partitioning for binary CMF output for $P = 2$. (a) Each of the four library elements with colored Gaussian noise. (b) The axes are scaled so that the noise about each point is iid. The plane is partitioned into four regions. Each region contains those points closest to the library point contained in that region. The partition boundary is mapped back to its original coordinates in (c). This partitioning then leads to maximum detection probability.

$= L\delta_m + \eta_w$; $m = 0, \dots, N-1$. Note that by uncorrelating or whitening the noise we have distorted the hypercube. However, optimal decoding of the hypothesis test \hat{H}_m can operate on the components of ψ_w separately. Intuitively, this is because the noise vector η_w has independent components and the value of $\eta_w(j)$ provides no information about $\eta_w(i)$, $i \neq j$. The effect of distorting the hypercube is depicted in Fig. 1(b). Essentially, while $\psi \sim N(\delta_m, R_\eta)$ under H_m we have $\psi_w \sim N(L\delta_m, \Delta)$ after the linear transformation L . Once we whiten the noise (and assuming equally likely hypotheses $Pr\{H_m\} = 1/N$ for all $m = 0, 1, \dots, N-1$), the optimal decoder or decision procedure decides for H_m when the Euclidean distance between ψ_w and $L\delta_m$ ($|\psi_w - L\delta_m|$), is minimum. This is known as minimum distance decoding.¹⁰ Since any possible observation vector ψ is closest to only one vertex, $L\delta_m$, the decoder induces the partition shown in Fig. 1(b). This process

is known as Voronoi partitioning.¹¹ In practice, we could carry out the procedure in our original observation space (on ψ rather than ψ_w) by first stretching the j th axis by $\sqrt{\lambda_j}$ and then transforming by L . This process is illustrated for our example in Fig. 1(c). The Voronoi partition minimizes the probability of a decoding error. Nonequally likely input patterns can be handled by a straightforward modification, although the Voronoi partition is more difficult to determine.

Although this technique is optimum, the use of Voronoi partitioning clearly requires extensive post-CMF processing relative to simple thresholding. Since, in principle, we need to compare our observation $\psi_w = L\psi$ to each of the $N = 2^P$ library element locations $L\delta_m$, the complexity of the postprocessing is of the order of that of a conventional matched filter. This defeats the purpose of the CMF.

B. Alteration of the CMF Matrix

With a small alteration of the CMF matrix S , we develop a system in which optimal decoding (Voronoi partitioning) reduces to simple thresholding. Thus, given that we reduce our number of matched filters from N to $\log_2 N$, a simple decodable system exists which is optimal. We consider only the following simplified situation:

(a)

$$(a) \xi \sim N(\mathbf{0}, \sigma_\xi^2 \mathbf{I}). \quad (8)$$

The input noise is iid. As before we can generally lift the Gaussian assumption by an appeal to the central limit theorem.

(b) The library elements are orthonormal; namely,

$$R_F = FF^T = \mathbf{I} \quad (9)$$

(c) The library elements $\mathbf{f}_0, \dots, \mathbf{f}_{n-1}$ are equally likely.

The basis of our alteration is to modify the CMF matrix $S = FR_F^{-1}\hat{\Delta}^T$ by replacing every 0 in $\hat{\Delta}$ by -1 . Denote this new bipolar matrix by Δ . For example, Eq. (2) becomes

$$\Delta = \begin{bmatrix} -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}. \quad (10)$$

There are two main advantages to this change:

(1) Using ± 1 s, we threshold at 0 rather than $1/2$ as before. Although for optical processors this will require bipolar encoding techniques,^{12,13} the resulting CMF classification performance increases with the input scale parameter. That is, for all $A > 1$, the performance of the bipolar system to the input $A\mathbf{f}_m + \xi$ is uniformly better than that due to $\mathbf{f}_m + \xi$ since we threshold at 0. Previously, we set our threshold at $1/2$ (when $A = 1$) and the classification performance will decrease when $A > 1$. In detection theory parlance, the bipolar decoding procedure is uniformly most powerful with respect to an unknown $A > 0$. We would not change our decoding procedure even if we knew the value of A .

(2) The bipolar matrix Δ has the property that

$$\Delta\Delta^T = N\mathbf{I}. \quad (11)$$

That is, the rows of Δ are orthogonal.

Next we examine P_E for the bipolar system. Our CMF matrix is

$$S = F\Delta^T.$$

Using Eq. (9), the covariance matrix of the output noise η is now

$$R_\eta = E\eta\eta = EF\Delta^T\xi\xi^T\Delta F^T = N\sigma_\xi^2 I.$$

The output noise is now uncorrelated and, by the Gaussian assumption, statistically independent. As such, no distortion is required prior to Voronoi partitioning. The optimal decision boundaries are simply hyperplanes aligned with the coordinate axes. This is illustrated for our $N = 4$ example in Fig. 2.

Because the boundaries are so aligned, evaluation of the probability of error is straightforward. We evaluate P_E by considering only the error events associated with a desired output of $\mathbf{f}_{N-1} = [1 \ 1 \dots 1]^T$. Then

$$P_E = 1 - \Pr\{\psi_1 > 0, \psi_2 > 0, \dots, \psi_P > 0 | H_{N-1}\}.$$

Since the random vector ψ has the Gaussian distribution $N(\mathbf{f}_{N-1}, \sigma_\xi^2 I)$ under H_{N-1} , we find

$$P_E = 1 - \prod_{j=1}^P \Pr\{\psi_j > 0 | H_{N-1}\}.$$

Since

$$\Pr\{\psi_j > 0 | H_{N-1}\} = \int_0^\infty \exp[-(\psi_j - 1)^2 / 2N\sigma_\xi^2] \frac{d\psi_j}{\sqrt{2\pi N\sigma_\xi^2}},$$

we find that

$$P_E = 1 - \{\text{erfc}[-(N\sigma_\xi^2)^{1/2}]\}^P, \quad (12)$$

where

$$\text{erfc}(x) = \int_x^\infty \exp(-t^2/2) dt / \sqrt{2\pi}.$$

Plots of $P_C = 1 - P_E$, the probability of detection or correct classification (in percent) are shown in Fig. 3 for various values of σ_ξ and $P = \log_2 N$.

VI. Effect of a Noisy Processor

In Ref. 5 we discussed the problems associated with implementation on a noisy optical processor. By a noisy processor we mean that the ideal CMF matrix S given by Eq. (1) is replaced by $S + \tilde{S}$, where \tilde{S} is a $P \times N$ matrix of noise samples. The components of \tilde{S} are modeled by iid Gaussian random variables with zero mean and a common variance σ_s^2 . Optimality of the bipolar CMF is no longer guaranteed when the processing is inexact. An exact performance analysis is hampered by the non-Gaussian distribution of the output vector ψ in the presence of both input and processor noise. We will demonstrate empirically that the bipolar CMF outperforms the binary processor. The comparison is carried out by a Monte Carlo simulation.

In the simulation we use library elements $f_{nm} = \delta_{n-m}$; $n = 0, 1, \dots, N-1 = 15$ such that the input noise vector ξ is taken to be $N(\mathbf{0}, \sigma_\xi^2 I)$. The processor matrix noise \tilde{S} is taken to be $N(\mathbf{0}, \sigma_s^2 I)$. Plots of P_C vs σ_ξ are given in Fig. 4 corresponding to $\sigma_s = 0.1$ and 0.2 . Each

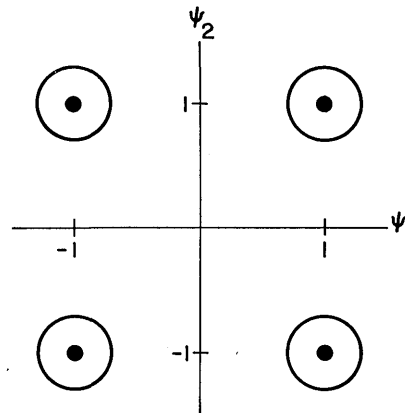


Fig. 2. Illustration for $P = 2$ that, under certain conditions, bipolar CMFs perform optimally with zero thresholding.

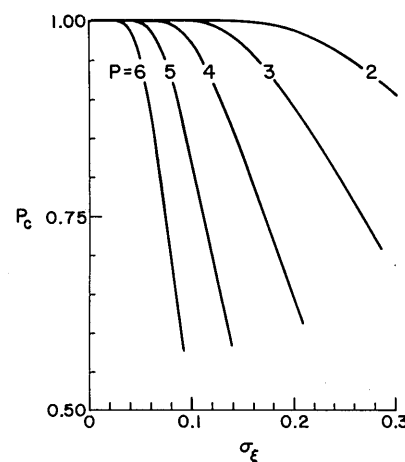


Fig. 3. Percentage of correct decisions vs input noise standard deviation (σ_ξ) for $P = 2$ (top curve), 3, 4, 5, and 6 for the bipolar CMF.

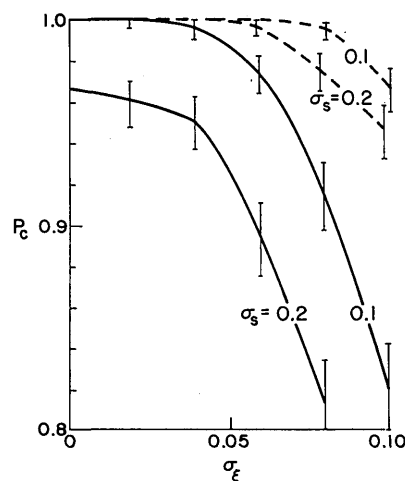


Fig. 4. Monte Carlo simulation of percentage of correct decision vs σ_ξ for matrix noise standard deviations $\sigma_s = 0.1$ and 0.2 . The upper dashed lines are for bipolar and the lower solid plots are for the binary CMF; 90% confidence intervals are shown; $P = 4$.

point is the result of 800 trials—50 for each input. The error bars indicate 90% confidence intervals about each data point and were calculated assuming Bernoulli trials. In both cases we see that the bipolar CMF achieves a greater detection probability than the binary CMF. This is directly related to the bipolar system using optimal discussion processing, at least in the processor noise-free case. As expected, the performance of both systems degrades as σ_s^2 , the processor noise variance, increases.

VII. Conclusion

We have applied the methodology of signal detection theory to develop optimal minimum probability of error CMF processors. By using a CMF matrix with bipolar rather than binary components, the computationally attractive threshold decoder is shown to be identical to the optimal Voronoi partition decoding technique. We have evaluated the probability of correct classification for a noise-free processor and extended the results, by Monte Carlo simulation, to the more realistic noisy processor situation. In all cases, the bipolar system outperforms the binary CMF. Future papers will report on ongoing research which includes:

- (1) a comparison of conventional and composite matched filter error performance;
- (2) error correcting coding techniques for noisy optical processors as in Ref. 5.

This research was supported by the SDIO/IST's Ultra High Speed Computing Program administered through the U.S. Office of Naval Research in conjunc-

tion with the Optical Systems Laboratory at Texas Tech University and was also partially supported by a National Science Foundation Presidential Young Investigator Award and a University of Washington Career Development Award.

References

1. B. Braunecker, R. Hauck, and A. W. Lohmann, "Optical Character Recognition Based on Nonredundant Correlation Measurements," *Appl. Opt.* **18**, 2746 (1979).
2. B. V. K. V. Kumar, "Efficient Approach to Designing Linear Combination Filters," *Appl. Opt.* **22**, 1445 (1983).
3. D. Casasent, "Unified Synthetic Discriminant Function Computational Formulation," *Appl. Opt.* **23**, 1620 (1984).
4. H. J. Caulfield and R. S. Putnam, "Fault Tolerance and Self-Healing in Optical Systolic Array Processors," *Opt. Eng.* **24**, 65 (1985).
5. R. J. Marks II and L. E. Atlas, "Composite Matched Filtering with Error Correction," *Opt. Lett.* **12**, 135 (1987).
6. J. Riggins and S. Butler, "Simulation of Synthetic Discriminant Function Optical Implementation," *Opt. Eng.* **23**, 721 (1984).
7. A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1984).
8. H. L. VanTrees, *Detection, Estimation and Modulation Theory, Part 1* (Wiley, New York, 1968).
9. R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics* (Macmillan, New York, 1970).
10. B. P. Lathi, *Modern Digital and Analog Communication Systems* (Holt, Rinehart & Winston, New York, 1983).
11. D. T. Lee, "Two-Dimensional Voronoi Diagrams in the L Metric," *J. Assoc. Comput. Mach.* **27**, 604 (1980).
12. R. P. Bocker, S. R. Clayton, and K. Bromley, "Electrooptical Matrix Multiplication Using the Twos Complement Arithmetic for Improved Accuracy," *Appl. Opt.* **22**, 2019 (1983).
13. D. P. Casasent and C. Perles, "Bipolar Biasing in High Accuracy Optical Linear Algebra Processors," *Appl. Opt.* **25**, 1033 (1986).

Meetings continued from page 2259

1988 January

- 15-20 Conf. & Exhibition on Electronic Imaging, Los Angeles
SPSE, 7003 Kilworth La., Springfield, VA 22151
- 24-28 **Optical Fiber Communication Conf., New Orleans**
OSA Mtgs. Dept., 1816 Jefferson Pl., NW, Wash., DC 20036.
- 24-29 Australian Optical Soc. Mtg., Sydney *P. Hariharan, CSIRO Div. of Applied Physics, P.O. Box 218, Lindfield, NSW 2070, Australia*
- 25-27 Outlook for Compact & Video Disc Systems & Applications Mtg., Key Biscayne *Inst. for Graphic Communication, Inc., 375 Commonwealth Ave., Boston, MA 02115*
- 27-29 35th Ann. Western Spectroscopy Assoc. Conf., Pacific Grove *H. Lafferty, P.O. Box 50, Springfield, OR 97477*
- 27-29 Optical Fiber Sensors Conf., New Orleans *OSA Mtgs. Dept., 1816 Jefferson Pl., NW, Wash., DC 20036*
- 31-5 Feb. Medical Imaging Mtg., Newport Beach *SPIE, P.O. Box 10, Bellingham, WA 98827*

February

- 28-4 Mar. Microlithography Conf., Santa Clara *SPIE, P.O. Box 10, Bellingham, WA 98827*

March

- 13-18 Optics & Optoelectronics Symp. Southeast, Orlando
SPIE, P.O. Box 10, Bellingham, WA 98827
- 29-31 Optical Storage of Documents & Images Mtg., Wash., DC
TOC, P.O. Box 14817, San Francisco, CA 94114

April

- 5-8 Materials Research Soc. Mtg., Reno *MRS, 9800 McKnight Rd., Ste. 327, Pittsburgh, PA 15237*
- 10-15 4th Int. Congr. on Advances in Non-Impact Printing Technologies, New Orleans
SPSE, 7003 Kilworth La., Springfield, VA 22151
- 12-14 **Optical Interference Coatings Top. Mtg., Tucson**
OSA Mtgs. Dept., 1816 Jefferson Pl., NW, Wash., DC 20036
- 25-29 **Conf. Lasers & Electro-Optics, Anaheim** *OSA Mtgs. Dept., 1816 Jefferson Pl., NW, Wash., DC 20036*

May

- 22-27 41st Ann. Conf., Arlington *SPSE, 7003 Kilworth La., Springfield, VA 22151*

continued on page 2282