Wide-Band System Identification Using Multiple Tones With Allpass Filters and Square-Law Detectors

Dennis R. Morgan, *Senior Member, IEEE,* and Christi K. Madsen, *Senior Member, IEEE*

Abstract—**This paper introduces a method of wide-band system identification that uses a uniformly spaced set of tones with allpass filters and square-law detectors to obtain estimates of amplitude and phase at the associated discrete frequencies. The bandwidth of the detectors is on the order of the tone spacing and so can be much smaller than the overall signal bandwidth, thereby making feasible the use of low-cost analog RF integrated circuits and/or pure digital signal processing with all of its inherent advantages. Applications of this technique include equalization and compensation for fiber optic and wide-band cellular communication systems.**

Index Terms—**Allpass filters, square-law detectors, tones, wideband system identification.**

I. INTRODUCTION

WITH ever increasing data rates, the bandwidth of transmission systems is stressing the capabilities of signal processing technology to keep pace. Indeed, with optical fiber transmission, data rates of 40 Gb/s and higher are being contemplated. So, it is a challenge just to realize circuits for demodulating the data. Therefore, the task of equalization and other compensation is largely relegated to analog circuits. This is a distinct disadvantage since digital signal processing, with all of its inherent advantages and compatibility with computer control, cannot then be applied.

Even though analog circuits are used for signal preprocessing, there is still the possibility of digital measurement and control, which would greatly enhance performance and reliability, as well as reduce cost. The task of channel measurement to obtain amplitude and phase over a set of frequencies is a classical system identification problem, where a known training signal at the far end serves as a reference. With extremely wide-band signals, it is not feasible to directly apply digital signal processing for this purpose because analog–digital (A/D) converters with sufficient bandwidth and resolution do not exist. However, if the signal bandwidth can somehow be reduced before sampling, then the feasibility of digital signal processing with all of its advantages becomes an option.

In this paper, we elaborate and generalize a method of wideband system identification that has been suggested for optical applications [1], whereby a uniformly spaced set of tones is used

to indirectly obtain estimates of amplitude and phase at the associated discrete frequencies. The technique uses allpass filters followed by square-law detectors to obtain relatively low-bandwidth signals that can be digitized and processed to obtain the information. The bandwidth of the detectors is on the order of the tone spacing and so can be much smaller than the overall signal bandwidth.

One example where this technique could be usefully applied is in 40 Gb/s fiber optic systems, where it is necessary to identify the channel frequency response for purposes of equalization and polarization mode dispersion (PMD) compensation. In this case, direct digitization of the wide-band signal is not feasible with today's technology. However, making use of low-cost ring-waveguide optical allpass filters, a parallel set of relatively low-bandwidth digitizable signals can indirectly provide the necessary wide-band channel information.

Another application example is obtaining channel state information for wide-band wireless communication systems. In this case, direct digitization is possible, but expensive. Here, the technique of this paper could provide an alternative for low-cost terminals, where mass-produced analog allpass filters are used in conjunction with a slower low-cost A/D converter.

II. BASIC TECHNIQUE

Fig. 1 shows a block diagram of the multi-tone wide-band system identification technique. The system to be identified is denoted by H, and we intend to obtain measurements of its amplitude and phase response over a set of L discrete frequencies $f_1, f_2, \cdots, f_L$ that are applied as a multi-tone input signal $x(t)$. The system output $y(t)$, which is contaminated by noise $n(t)$, is applied to a set of M allpass filters, denoted by $G_m(\omega)$, $m = 1, 2, \cdots, M$, which can either be realized in a fixed parallel bank or as a single adjustable unit to obtain sequential measurements. The allpass filter outputs (in parallel or serial) are then square-law detected to obtain the continuous-time signals $u_m(t)$, which are then sampled and synchronously complex-demodulated [upper output path of Fig. 1(a)] to obtain a set of M complex numbers v_m, $m = 1, 2, \cdots, M$. Alternatively, low-cost analog demodulator chips could be employed first, followed by digitization [lower output path of Fig. 1(a)]. In what follows, the set of complex numbers v_m will be used to estimate the system amplitude and phase over the set of L frequencies [Fig. 1(b)].

We note that such an undertaking could conceivably be accomplished using a bank of narrow-band coherent tone detectors. However, the necessary phase matching would inevitably be very difficult to achieve and maintain with analog circuits.
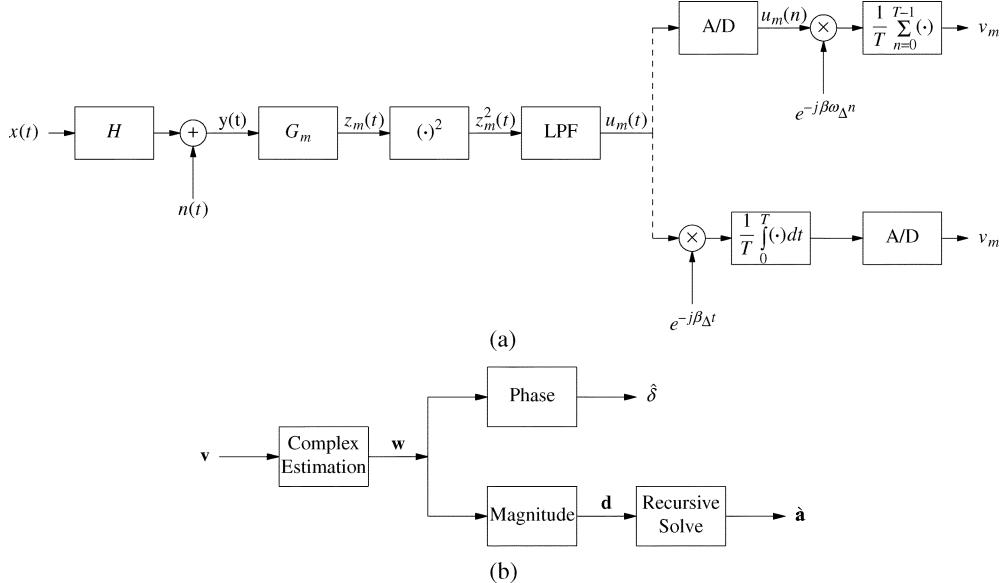
Fig. 1.   Block diagram of the multi-tone wide-band system identification technique. (a) Measurement of mth datum. (b) Estimation of relative phase $\hat{\boldsymbol{\delta}}$ and normalized amplitude $\hat{\mathbf{a}}$.

It is for this reason that we prefer to encode the desired information into *difference* frequency terms that can be reliably processed to obtain accurate estimates.

The input to the system H is defined as the multi-tone complex

$$x(t) = \sum_{l=1}^{L} \cos(\omega_l t), \qquad l = 1, 2, \cdots, L \qquad (1)$$

where $\omega_l \equiv 2\pi f_l$. Here, for convenience we assume that the tones are all unity amplitude and zero reference phase. However, if it is preferable to generate tones at different amplitudes and phases (e.g., the framing sequence in a synchronous optical transmission system), then that information (which is known to the designer) can be subsumed into the unknown system H.

In the absence of noise, the output of H is written as

$$y(t) = \sum_{l=1}^{L} A_l \cos(\omega_l t + \phi_l) \qquad (2)$$

where A_l and ϕ_l are, respectively, the unknown amplitude and phase response of H at frequency f_l, and $\omega_l \equiv 2\pi f_l$.

The allpass filter frequency responses $G_m(\omega)$ take on discrete sets of values at the tone frequencies, denoted as

$$G_{m,l} \equiv G_m(\omega_l). \qquad (3)$$

Nominally, the magnitude of $G_m(\omega)$ is unity; however, we allow for a more general representation here to accommodate actual realizations. Thus, in the noiseless case, the output of the mth allpass filter is expressed

$$z_m(t) = \sum_{l=1}^{L} A_l \Re \left[G_{m,l} e^{j(\omega_l t + \phi_l)} \right] \qquad (4)$$

where $\Re$. denotes the real part.

We will assume here that the tones are uniformly spaced at the difference frequency

$$f_\Delta \equiv f_{l+1} - f_l, \qquad l = 1, 2, \cdots, L - 1. \qquad (5)$$

We then square the allpass filter outputs and retain only baseband terms at low multiples of f_Δ. We will focus on the dc and fundamental tone difference frequency terms, and calculate

$$
\begin{aligned}
u_m(t) &= \text{LPF}\left\{ z_m^2(t) \right\} \\
&= \frac{1}{2} \sum_{l=1}^{L} \sum_{k=1}^{L} A_l A_k \Re \left[G_{m,l} G_{m,k}^* e^{j(\omega_l - \omega_k)t + j(\phi_l - \phi_k)} \right] \\
&= \frac{1}{2} \sum_{l=1}^{L} A_l^2 |G_{m,l}|^2 \\
&\quad + \sum_{l=1}^{L-1} A_l A_{l+1} \Re \left[G_{m,l}^* G_{m,l+1} e^{j(\omega_\Delta t + \delta_l)} \right] \\
&\quad + 2\omega_\Delta, 3\omega_\Delta, \cdots \text{terms} \qquad (6)
\end{aligned}
$$

where LPF denotes a zonal low-pass filter that only retains difference frequency components of the square-law process and

$$\delta_l \equiv \phi_{l+1} - \phi_l, \qquad l = 1, 2, \cdots, L - 1 \qquad (7)$$

are differential channel phases. (Here, we have made the usual assumption that the sum frequencies are much higher than the difference frequencies so that they are effectively filtered out by the zonal filter.) The first term (dc) in (6) contains information on the unknown amplitudes A_l and the second term (ω_Δ) contains information on both A_l and differential phases δ_l. We assume that it is only necessary to identify H up to an arbitrary constant phase term, hence, it is sufficient to obtain the δ_l's rather than the actual channel phases ϕ_l. More will be said later about the A_l's.

The $u_m(t)$ signals are now sampled at times $t = \beta, 2\beta, \cdots$ (assuming normalized system bandwidth), where β is the downsampling ratio, i.e., the ratio of system bandwidth to detector bandwidth, and synchronously demodulated at the tone difference frequency to obtain

$$v_m = \frac{1}{T} \sum_{n=0}^{T-1} u_m(n) e^{-j\beta\omega_\Delta n} \qquad (8a)$$

where T is the number of samples averaged [see upper output path of Fig. 1(a)]. Note that if the filters have not been running for some time before $n = 0$, then it would be necessary to discard a number of samples before (8a) is calculated in order to let the transient response settle out.

Alternatively, one could use existing low-cost analog RF integrated circuits for synchronous demodulation as depicted in the lower output path of Fig. 1(a), whereby

$$v_m = \frac{1}{T} \int_0^T u_m(t) e^{-j\omega_\Delta t} dt. \tag{8b}$$

In this case, the A/D converter bandwidth requirement is even further reduced.

Substituting (6) into (8) and assuming that T is large enough to effectively reject the dc and harmonics of ω_Δ, we obtain

$$v_m \approx \sum_{l=1}^{L-1} d_l F_{ml} e^{j\delta_l} \tag{9}$$

where

$$d_l \equiv \frac{1}{2} A_l A_{l+1} \tag{10}$$

and

$$F_{ml} \equiv G_{m,l}^* G_{m,l+1}. \tag{11}$$

It is convenient to express (9) in vector-matrix form to obtain the compact expression

$$\mathbf{v} = \mathbf{F} \mathbf{D} e^{j\boldsymbol{\delta}} \tag{12}$$

where $\mathbf{v}$ is an $M \times 1$ vector of the synchronous demodulator outputs v_m, $\mathbf{F}$ is an $M \times (L-1)$ matrix with components (11)

$$\mathbf{D} \equiv \text{diag}(d_1, d_2, \cdots, d_{L-1}) \tag{13}$$

is an $(L-1) \times (L-1)$ diagonal matrix, and $\boldsymbol{\delta}$ is an $(L-1) \times 1$ vector of differential phase components δ_l. (The quantity $e^{j\boldsymbol{\delta}}$ is interpreted as a vector with components $e^{j\delta_l}$.) Thus, with this formulation, we now seek a solution for $\boldsymbol{\delta}$, given the measurements $\mathbf{v}$.

In general, noise and other effects will perturb the measurements (12). Therefore, a least-squares type of solution is desired, which would involve the minimization of $\|\mathbf{v} - \mathbf{F}\mathbf{D}e^{j\boldsymbol{\delta}}\|$ over all possible values of the complex vector $\mathbf{D}e^{j\boldsymbol{\delta}}$, which contains the unknown relative amplitudes A_l and differential phases δ_l. The minimum is reached using the generalized inverse of $\mathbf{F}$, so that we have

$$\hat{\mathbf{D}} e^{j\hat{\boldsymbol{\delta}}} = \mathbf{w} \tag{14}$$

where

$$\mathbf{w} \equiv (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \mathbf{v} \tag{15}$$

and superscript H denotes (Hermitian) complex conjugate. Taking the phase then gives the differential phase estimate vector

$$\hat{\boldsymbol{\delta}} = \text{arg}(\mathbf{w}). \tag{16}$$

Note that for the generalized inverse in (15), the number of measurements M that form the vector $\mathbf{v}$ must always be greater than or equal to the $L-1$ dimensionality of $\hat{\boldsymbol{\delta}}$. In fact, the more measurements that are made, the better, at least theoretically. For example, with $L = 3$ tones and $M = 3$ measurements, $\mathbf{F}$ is of the form

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{bmatrix} = \begin{bmatrix} G_{1,1}^* G_{1,2} & G_{1,2}^* G_{1,3} \\ G_{2,1}^* G_{2,2} & G_{2,2}^* G_{2,3} \\ G_{3,1}^* G_{3,2} & G_{3,2}^* G_{3,3} \end{bmatrix}$$

and (15) is an overdetermined solution. In order to achieve accurate solutions, it is desirable that the matrix $\mathbf{F}$ be well conditioned. Intuitively, this can be ensured by making the allpass filters very sharp and tuning their crossover frequencies, i.e., the frequencies at which the phase of G_m is π radians (180 degrees), equal to the tone frequencies. In addition, we can add one or two additional allpass filters tuned below and above the tone frequencies. Thus, we can achieve well conditioned solutions of the $L-1$ differential phases using up to $M = L+2$ measurements. In practice, realizing high-Q analog allpass filters may be difficult, so there will generally be some overlap of the allpass responses that will tend to degrade the conditioning. We will have more to say about this later in the simulations and theoretical studies.

Also note that in obtaining an unweighted least-squares solution, we have tacitly assumed that the perturbations of each component of $\mathbf{v}$, due to noise and numerical precision, are independent and identically distributed (i.i.d.). As will be shown later, the noise effects are in fact strongly correlated, and so we can contemplate improving the estimation using weighted least-squares techniques. Another refinement would be possible if there were some other means of independently obtaining amplitude estimates, in which case the phase estimates might be improved. These advanced techniques will be discussed in further detail later in Section V.

We now turn to the estimation of the amplitudes $A_1, A_2, \cdots, A_L$. Taking the magnitude of $\mathbf{w}$, we obtain estimates for the components of $\mathbf{D}$

$$\hat{d}_l = \frac{1}{2} \hat{A}_l \hat{A}_{l+1} = |w_l| \tag{17}$$

which can be recursively solved, giving

$$\hat{A}_l = \frac{2|w_{l-1}|}{\hat{A}_{l-1}}, \qquad l = 2, 3, \cdots, L. \tag{18}$$

Note that we have L unknowns and only $L-1$ equations. Therefore, some other information must be gained for the solution.

One way to obtain additional information is to measure the dc value of the sampled square-law detector outputs

$$P_m = \frac{1}{T} \sum_{n=1}^T u_m(n). \tag{19}$$

Here, without loss of generality, we assume that $|G_{m,l}| = 1$, since the A_l's can always be rescaled. Thus, from (6), we see that in the noiseless case each P_m is of value

$$P = \frac{1}{2} \sum_{l=1}^L A_l^2 \tag{20}$$

which is recognized as the sum of the tone powers. In order to gain maximal noise immunity, we will want to average over the M values of (19). Here, we will assume that P is measured over a long enough time so that it is a perfect estimate of the total sinusoidal power. Thus, the additional information of (20) could enable us to solve (18).

First, let us define the normalized variables

$$\hat{a}_l \equiv \begin{cases} \hat{A}_{l+1}\hat{A}_1, & l \text{ odd} \\ \hat{A}_{l+1}/\hat{A}_1, & l \text{ even} \end{cases}, \qquad l = 0, 1, \cdots, L-1. \quad (21)$$

(Note that for convenience we have implicitly defined $\hat{a}_0 = 1$.) Substituting (18) into (21) obtains the recursion

$$\hat{a}_l = \frac{2|w_l|}{\hat{a}_{l-1}}, \qquad l = 1, 2, \cdots, L-1 \quad (22)$$

whereby we have normalized out the effect of $\hat{A}_1$. Next, using (21) we transform the $\hat{a}_l$'s back into the desired variables

$$\hat{A}_l = \begin{cases} \hat{a}_{l-1}\hat{A}_1, & l \text{ odd} \\ \hat{a}_{l-1}/\hat{A}_1, & l \text{ even} \end{cases}, \qquad l = 2, 3, \cdots, L \quad (23)$$

where $\hat{A}_1$ is now the only unknown. Substituting the above into (20) for $A_l = \hat{A}_l$, we obtain the quadratic equation

$$2P = \hat{A}_1^2 \sum_{l \text{ even}} \hat{a}_l^2 + \frac{1}{\hat{A}_1^2} \sum_{l \text{ odd}} \hat{a}_l^2. \quad (24)$$

Solving this equation for A_1^2 gives

$$\hat{A}_1^2 = \frac{P \pm \sqrt{P^2 - \sum\limits_{l \text{ even}} \hat{a}_l^2 \cdot \sum\limits_{l \text{ odd}} \hat{a}_l^2}}{\sum\limits_{l \text{ even}} \hat{a}_l^2}. \quad (25)$$

Thus, (23) and (25) together specify the amplitude estimates $\hat{A}_l$, $l = 1, 2, \cdots, L$, up to the $\pm$ uncertainty in (25). For example, with $L = 3$ and the true values $(A_1, A_2, A_3) = (2, 3, 1)$, we have total power $P = 7$ and, ideally in the absence of noise, $(\hat{a}_0, \hat{a}_1, \hat{a}_2) = (1, 6, 0.5)$. From (23) and (25), we then calculate two possible solutions, $(\hat{A}_1, \hat{A}_2, \hat{A}_3) = (2.6833, 2.2361, 1.3416)$ and $(2,3,1)$. The second solution is obviously the correct one, but the first also satisfies the power constraint (20). Also note that for both solutions, $\hat{A}_1 \hat{A}_2 = 6$ and $\hat{A}_2 \hat{A}_3 = 3$, thereby presenting the same cross-term information (17) and hence are indistinguishable on that basis.

It is not obvious how to easily resolve the sign indeterminacy in (25). Requiring only relative amplitudes $\hat{A}_l / \hat{A}_1$ does not help. It is possible to use the higher-order terms in (6) to gain additional information on the A_l's (as well as the ϕ_l's) but this will complicate the signal processing and require a higher sampling rate, which is undesirable since the whole idea is to reduce the bandwidth. Finally, one could assume some kind of calibration procedure whereby, say, A_1 could be initially determined, and then slow changes tracked by virtue of having established the sign in (25). In any case, we shall assume for the rest of this paper that such means have been taken so that we only have to consider estimation of the normalized amplitude estimates $\hat{a}_l$, which are specified in (22) as a function of the measured variables w_l.

III. EFFECTS OF NOISE

We will now assess the effects of additive noise on the parameter estimates. We assume that the allpass filter outputs $z_m(t)$ contain zero-mean wide-band Gaussian noise with covariance matrix $\mathbf{R}_z$. The nature of this covariance depends on whether the allpass filters are implemented in a fixed parallel filter bank or as a single tunable filter that obtains sequential measurements. In the first case, the noise, which is assumed to originate in the channel signal $y(t)$ as shown in Fig. 1(a), will be almost completely correlated at the allpass filter outputs by the following argument. Recall that for good numerical conditioning and noise immunity, it is desirable to use high-Q allpass filters so that the phase transition bands do not significantly overlap. Therefore, each allpass filter will have zero phase response over most of the frequency band, only making a brief excursion to π at its crossover frequency. This implies that the noise components will be almost completely correlated across the filter bank, so that the output covariance matrix can be expressed as

$$\mathbf{R}_z \approx \sigma^2 \mathbf{U} \quad (26)$$

where

$$\mathbf{U} \equiv \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (27)$$

is a matrix of all 1's. On the other hand, for a serial implementation, the noise will not be correlated at all between channels since the measurements are taken at different times, so that $\mathbf{R}_z = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the indentity matrix.

In either case, we must determine the effects of the noise on the phase and amplitude estimates, respectively, (16) and (22). To this end, we must first determine the covariance at the output of the square-law detector and then at the synchronous demodulator output. For generality, we assume the parallel case for the rest of this paper; the sequential case is similarly handled by simply zeroing the off-diagonal covariance terms.

A. Square-Law Detector Output

Let us assume for the purpose of this analysis that the allpass filters are ideal with unit amplitude response and phase response $\theta_m(\omega)$

$$G_m(\omega) = e^{j\theta_m(\omega)}. \quad (28)$$

Then, the allpass filter outputs (4) can be rewritten as

$$z_m(t) = \sum_{l=1}^{L} A_l \cos[\omega_l t + \phi_l + \theta_m(\omega_l)] + n_m(t) \quad (29)$$

where $n_m(t)$ is the noise at the output of the mth allpass filter due to the previously discussed wide-band Gaussian input noise.

The square-law detector consists of a squaring device followed by a zonal low-pass filter that rejects carrier terms and only retains relatively low-bandwidth difference frequency components. It will be convenient to separate the square-law

detector output $u_m(t)$ into the various mixing components. We first write

$$z_m^2(t) = z_m^{\text{S}\times\text{S}}(t) + 2z_m^{\text{S}\times\text{N}}(t) + z_m^{\text{N}\times\text{N}}(t) \tag{30}$$

where the mixing terms are defined as

$$z_m^{\text{S}\times\text{S}}(t) \equiv \left[\sum_{l=1}^{L} A_l \cos\left[\omega_l t + \phi_l + \theta_m(\omega_l)\right] \right]^2 \tag{31a}$$

$$z_m^{\text{S}\times\text{N}}(t) \equiv \left[\sum_{l=1}^{L} A_l \cos\left[\omega_l t + \phi_l + \theta_m(\omega_l)\right] \right] n_m(t) \tag{31b}$$

$$z_m^{\text{N}\times\text{N}}(t) \equiv n_m^2(t) \tag{31c}$$

and S and N denote, respectively, signal and noise. After the low-pass zonal filter is applied to (30), the surviving components are written as

$$u_m(t) = u_m^{\text{S}\times\text{S}}(t) + 2u_m^{\text{S}\times\text{N}}(t) + u_m^{\text{N}\times\text{N}}(t). \tag{32}$$

Note that the $u_m^{\text{S}\times\text{S}}$ term is the desired noiseless output expressed in (6). Appendix A shows that $u_m^{\text{S}\times\text{N}}(t)$ and $u_m^{\text{N}\times\text{N}}(t)$ are mutually uncorrelated approximately white random processes, with interchannel cross-correlation function approximately given by

$$R_{u_m u_n}^{\text{S}\times\text{N}}(\tau) \approx \frac{\sigma^2}{4} \sum_{l=1}^{L} A_l^2 \Re\left[J_{mn}(\omega_l, \omega_l, \omega_\Delta, 1) \right.$$
$$\left. + J_{nm}(\omega_l, \omega_l, -\omega_\Delta, 1) \right] \delta(\tau) \tag{33}$$

where

$$J_{mn}(\omega_l, \omega_k, \omega_\Delta, T)$$
$$\equiv \frac{\omega_l - \omega_m + j\alpha}{\omega_l - \omega_m - j\alpha} \cdot \frac{\omega_k - \omega_n - j\alpha}{\omega_k - \omega_n + j\alpha}$$
$$\times \left\{ \frac{1}{\beta T} \cdot \frac{\omega_l - \omega_\Delta - \omega_m - j\alpha}{\omega_l - \omega_\Delta - \omega_m + j\alpha} \cdot \frac{\omega_l - \omega_\Delta - \omega_n + j\alpha}{\omega_l - \omega_\Delta - \omega_n - j\alpha} \right.$$
$$- \frac{2\alpha}{(\beta T)^2} \cdot \frac{\omega_m - \omega_n}{\omega_m - \omega_n - j2\alpha}$$
$$\times \left[\frac{1 - e^{-j\beta T(\omega_l - \omega_\Delta - \omega_n - j\alpha)}}{(\omega_l - \omega_\Delta - \omega_n - j\alpha)^2} \right.$$
$$\left. \left. + \frac{1 - e^{+j\beta T(\omega_l - \omega_\Delta - \omega_m + j\alpha)}}{(\omega_l - \omega_\Delta - \omega_m + j\alpha)^2} \right] \right\} \tag{34}$$

$$\alpha \equiv 2\frac{1-r}{1+r} \tag{35}$$

r is the allpass filter pole radius, and β is the ratio of the system bandwidth to the low-pass filter bandwidth. [Note that the generality of (34) will be more fully utilized in the next section.] Also, from Appendix A, we have

$$R_{u_m u_n}^{\text{N}\times\text{N}}(\tau) \approx \frac{2(1 - 2f_\Delta)\sigma^4}{\beta} \delta(\tau). \tag{36}$$

Note that the above expressions are approximations, where the $\delta(\tau)$ symbolically implies a uniform spectrum and should not be taken literally.

B. Synchronous Demodulator Output

The synchronous demodulator, which is a linear processing device, can be considered to act independently on each of the above mixing components, producing an output

$$v_m = v_m^{\text{S}\times\text{S}} + 2v_m^{\text{S}\times\text{N}} + v_m^{\text{N}\times\text{N}}. \tag{37}$$

It is shown in Appendix A that these components are mutually uncorrelated and that the *Hermitian* covariance matrix of $\mathbf{v}$ can be written as

$$\mathbf{R}_v \equiv E\left[(\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^H\right] = 4\mathbf{R}_v^{\text{S}\times\text{N}} + \mathbf{R}_v^{\text{N}\times\text{N}} \tag{38}$$

where $\bar{\mathbf{v}} \equiv E\{\mathbf{v}\}$ is the mean, defined in the noiseless case by (12), H denotes (Hermitian) complex conjugate, and the components of $\mathbf{R}_v^{\text{S}\times\text{N}}$ and $\mathbf{R}_v^{\text{N}\times\text{N}}$ are given by

$$R_{v_m v_n}^{\text{S}\times\text{N}} \approx \frac{\sigma^2}{4} \sum_{l=1}^{L} A_l^2 [J_{mn}(\omega_l, \omega_l, \omega_\Delta, T) + J_{nm}(\omega_l, \omega_l, -\omega_\Delta, T)] \tag{39}$$

and

$$R_{v_m v_n}^{\text{N}\times\text{N}} \approx \frac{2(1 - 2f_\Delta)\sigma^4}{\beta T} \tag{40}$$

where J_{mn} is defined in (34). Note that $R_{u_m u_n}^{\text{S}\times\text{N}}$ in (33) is equivalent to the real part of $R_{v_m v_n}^{\text{S}\times\text{N}}$ in (39) for the special case $T = 1$. This relates to the effect of the synchronous demodulator, which is described fully in Appendix A. A similar relationship holds between $R_{u_m u_n}^{\text{N}\times\text{N}}$ in (36) and $R_{v_m v_n}^{\text{N}\times\text{N}}$ in (40).

The complex representation of $\mathbf{v}$ is rather artificial here, being a mathematical artifice rather than arising from some physical process. Therefore, $\mathbf{v}$ is in general not a so-called *proper* complex random vector because the usual condition of zero correlation when the complex conjugate transpose (H) in (38) is replaced by transpose (T) does not hold [2]–[4]. Accordingly, we also need to define what we shall call, following [4], the *complementary* covariance matrix, which is expressed in Appendix A as

$$\mathbf{C}_v \equiv E\left[(\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^T\right] = 4\mathbf{C}_v^{\text{S}\times\text{N}} \tag{41}$$

where

$$C_{v_m v_n}^{\text{S}\times\text{N}} \approx \frac{\sigma^2}{4} \sum_{l=3}^{L} A_l A_{l-2} e^{j(\phi_l - \phi_{l-2})}$$
$$\times [J_{mn}(\omega_l, \omega_{l-2}, \omega_\Delta, T) + J_{nm}(\omega_l, \omega_{l-2}, \omega_\Delta, T)]. \tag{42}$$

Note that in comparison with (38), (41) has no complex conjugation on the second factor and the N × N term is absent. Also note that (42) differs from (39) in that the sum starts at $l = 3$ instead of $l = 1$, $A_l A_{l-2}$ replaces A_l^2, a differential phase term is added, the second argument of the J_{mn} and J_{nm} terms is ω_{l-2} instead of ω_l, and ω_Δ does not change sign between J_{mn} and J_{nm}.

It can be easily shown that if (33) and (36) were autocorrelations of strictly white signals, instead of only approximately so, the complementary term (41) would be zero when the synchronous demodulation (8) was performed over an integer number of cycles. In fact, this is not so, as the above theory shows and which is confirmed later by simulation. Therefore, the complementary term is nonnegligible and its inclusion is necessary to obtain the correct estimate statistics. This lesson

was painfully learned during the arduous process of developing this theory.

C. Phase Estimates

We now go on to use these covariances for calculating the covariance of the estimated phases (16). First, we note that if z is a complex variable with phase angle ϕ, then (see Appendix B)

$$d\phi = \Im\left(\frac{dz}{z}\right) \tag{43}$$

where $\Im(\cdot)$ denotes imaginary part. We now apply this to perturbations of (16). Note that in the noiseless case, substituting (12) into (15) gives

$$\mathbf{w}_0 = \mathbf{D}e^{j\boldsymbol{\delta}} \tag{44}$$

where $\mathbf{D}$ is defined in (13). Taking differentials about $\mathbf{w} = \mathbf{w}_0$ in (16) and using (43) and (15) then obtains

$$\begin{aligned} d\hat{\boldsymbol{\delta}} &= \Im\left[\mathbf{D}^{-1}\mathrm{diag}(e^{-j\boldsymbol{\delta}})d\mathbf{w}\right] \\ &= \Im(\mathbf{Q}d\mathbf{v}) \end{aligned} \tag{45}$$

where

$$\mathbf{Q} \equiv \mathbf{D}^{-1}\mathrm{diag}(e^{-j\boldsymbol{\delta}})(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H. \tag{46}$$

Assuming small noise perturbations, we then calculate the covariance of $\hat{\boldsymbol{\delta}}$, giving

$$\begin{aligned} \mathbf{R}_\delta &\equiv E[d\hat{\boldsymbol{\delta}}\, d\hat{\boldsymbol{\delta}}^T] \\ &= \frac{1}{2}\Re(\mathbf{Q}\mathbf{R_v}\mathbf{Q}^H - \mathbf{Q}\mathbf{C_v}\mathbf{Q}^T) \end{aligned} \tag{47}$$

where $\mathbf{R_v}$ and $\mathbf{C_v}$ are, respectively, the Hermitian covariance matrix and symmetric complementary covariance matrices of $\mathbf{v}$ given by (38) and (41).

D. Amplitude Estimates

Next, we compute the covariance of the normalized amplitude estimates (22), for which we will first need to obtain a closed-form solution. By taking the natural log, we can rewrite (22) as a set of linear equations in vector form

$$\ln|2\mathbf{w}| = \mathbf{T}\ln\hat{\mathbf{a}} \tag{48}$$

where $|\cdot|$ denotes component-wise magnitude

$$\mathbf{T} \equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \tag{49}$$

is an $(L-1) \times (L-1)$ matrix, and

$$\hat{\mathbf{a}} \equiv [\hat{a}_1 \quad \hat{a}_2 \quad \cdots \quad \hat{a}_{L-1}]^T \tag{50}$$

is an $(L-1) \times 1$ vector. Solving (48) for $\hat{\mathbf{a}}$ gives

$$\hat{\mathbf{a}} = \exp\left(\mathbf{T}^{-1}\ln|2\mathbf{w}|\right) \tag{51}$$

where (52), shown at the bottom of page, expresses $\mathbf{T}^{-1}$.

Now we need to use the fact that for any matrix $\mathbf{A}$ and vector function

$$\mathbf{y} = \exp(\mathbf{A}\ln\mathbf{x}), \tag{53}$$

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \mathrm{diag}(\mathbf{y})\mathbf{A}\mathrm{diag}^{-1}(\mathbf{x}) \tag{54}$$

which can be easily established by showing that

$$\frac{\partial y_i}{\partial x_j} = \frac{y_i a_{ij}}{x_j} \tag{55}$$

and collecting the components. Applying this to (51) gives

$$d\hat{\mathbf{a}} = \mathrm{diag}(\hat{\mathbf{a}})\mathbf{T}^{-1}\mathrm{diag}^{-1}(|\mathbf{w}|)\, d|\mathbf{w}|. \tag{56}$$

Next, we need to determine the differential of $|\mathbf{w}|$ with respect to $\mathbf{w}$. For any complex variable z with phase angle ϕ, we have (see Appendix B)

$$d|z| = \Re(e^{-j\phi}dz). \tag{57}$$

Applying this to (56) for $\mathbf{w} = \mathbf{w}_0 = \mathbf{D}e^{j\boldsymbol{\delta}}$ (whereby $\hat{\mathbf{a}} = \mathbf{a}$) and using (15), we obtain finally

$$\begin{aligned} d\hat{\mathbf{a}} &= \mathrm{diag}(\mathbf{a})\mathbf{T}^{-1}\mathbf{D}^{-1}\Re\left[\mathrm{diag}(e^{-j\boldsymbol{\delta}})d\mathbf{w}\right] \\ &= \Re(\mathbf{P}d\mathbf{v}) \end{aligned} \tag{58}$$

where

$$\mathbf{P} \equiv \mathrm{diag}(\mathbf{a})\mathbf{T}^{-1}\mathbf{D}^{-1}\mathrm{diag}(e^{-j\boldsymbol{\delta}})(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H. \tag{59}$$

With the above result, we can now write the covariance matrix of $\hat{\mathbf{a}}$

$$\begin{aligned} \mathbf{R}_a &\equiv E\left[(d\hat{\mathbf{a}}\, d\hat{\mathbf{a}}^T\right] \\ &= \frac{1}{2}\Re(\mathbf{P}\mathbf{R_v}\mathbf{P}^H + \mathbf{P}\mathbf{C_v}\mathbf{P}^T). \end{aligned} \tag{60}$$

E. Cross Estimates

Finally, we note that the phase and amplitude estimates are coupled and we shall compute their cross-covariance. From (45) and (58), we have

$$\begin{aligned} \mathbf{R}_{\delta a} &\equiv E\left[(d\hat{\boldsymbol{\delta}}\, d\hat{\mathbf{a}}^T\right] \\ &= \frac{1}{2}\Im(\mathbf{Q}\mathbf{R_v}\mathbf{P}^H + \mathbf{Q}\mathbf{C_v}\mathbf{P}^T). \end{aligned} \tag{61}$$

$$\mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ (-1)^{L-2} & (-1)^{L-3} & (-1)^{L-4} & (-1)^{L-5} & \cdots & -1 & 1 \end{bmatrix} \tag{52}$$
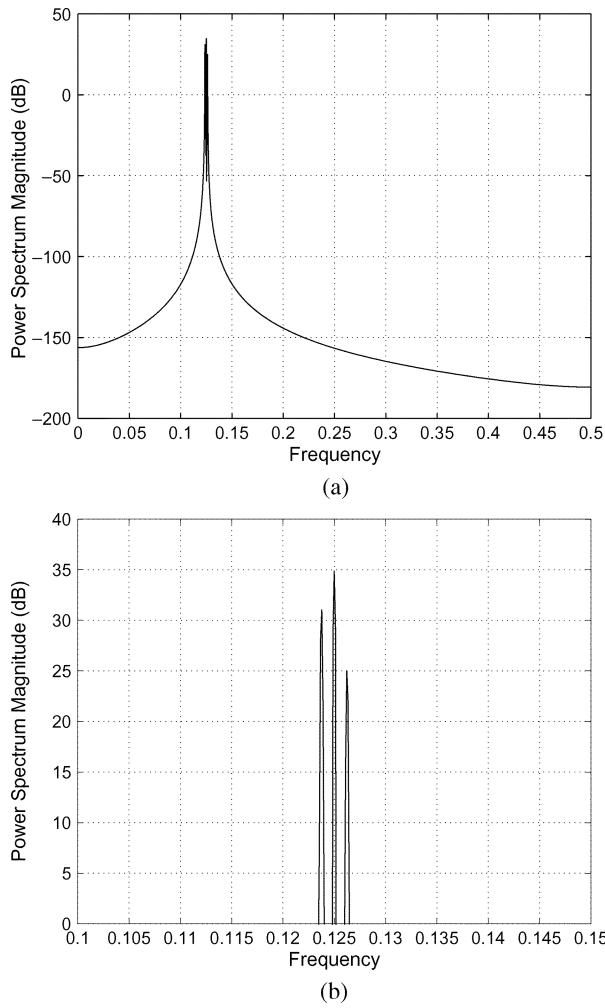
Fig. 2. Power spectrum of 3-tone signal used in simulations. (a) Full band. (b) Zoom in at center frequency.

Fig. 3. Phase response of allpass filter for crossover frequency 0.125 and various pole radii r.

IV. SIMULATION MODEL

A. Description

We now describe the discrete-time model used in the simulations. For simplicity, we chose as an example $L = 3$ tones with amplitudes $(A_1, A_2, A_3) = (1, 1.5, 0.5)$ and phases $(\phi_1, \phi_2, \phi_3) = (30°, 60°, 135°)$ at the receiver input [$y(t)$ in Fig. 1]. The sinusoids are represented in discrete time by nominally eight samples per cycle, so that their normalized frequencies are centered at 0.125 (1/8). The tone spacing is chosen as 1/100 of the center frequency so that $f_\Delta = 0.00125$, with specific frequencies given by $(f_1, f_2, f_3) = (0.12375, 0.12500, 0.12625)$. Fig. 2 shows the power spectrum of this 3-tone signal (8192 points). To this signal, we add white Gaussian noise samples of standard deviation σ.

The allpass filters were modeled using the standard second-order discrete-time formulation

$$G_m(z) = \frac{z^{-2} - 2r\cos\nu_m z^{-1} + r^2}{r^2 z^{-2} - 2r\cos\nu_m z^{-1} + 1} \qquad (62)$$

where $r < 1$ is the pole radius and ν_m is the angular crossover frequency, at which the phase of G_m is π radians (180 degrees).
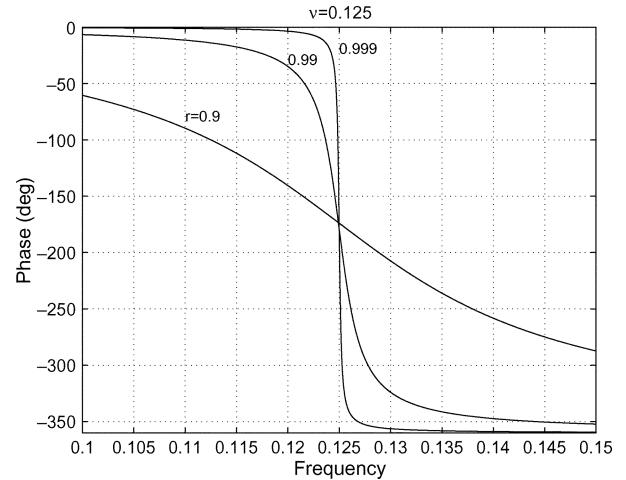
In the simulations, up to $M = 4$ filters are used, where ν_1, ν_2, and ν_3 are set to the angular tone frequencies (for best conditioning, as previously discussed) and we select $\nu_4 = \pi/2$, which is well outside of the tone bandwidth. (The only purpose of the fourth allpass filter instead of no filter is to achieve delay match with the other three.) Fig. 3 plots the phase response of G_2 (center frequency, $\nu_2 = f_2 = 0.125$) for several values of r. As can be seen, the response sharpens as r becomes closer to 1. However, for pole radius r close to 1, the transient response also becomes longer, taking approximately $1/(1 - r)$ samples to die out to the $1/e$ level (e.g., 1000 samples for $r = 0.999$). Therefore, we have made a provision to discard a certain number of samples before synchronous demodulation.

The allpass filter outputs are squared and decimated by the factor D to simulate the square-law detector low-pass filter outputs u_m. (At this point, there is no need to sample the signals as shown in Fig. 1 because the simulated signals are already discrete-time.) The standard Matlab decimation routine involves first low-pass filtering to limit aliasing and then downsampling to retain only one out of every D samples. Thus, the bandwidth is reduced by the factor $\beta = D$, relating to the previously defined analog low-pass filtering associated with the square-law detector. For the simulation results presented here, we have selected $D = 128$ as a reasonable compromise between numerical efficiency and controlling aliasing effects due to nonideal downsampling filters.

Finally, T samples are accumulated as in the upper output path of Fig. 1(a) to produce output measurements v_m, which are then used to compute the differential phase estimates per (16) and normalized amplitude estimates per (22). We have chosen $T = 50$ for this simulation so that the synchronous demodulator bandwidth is $1/50 = 0.02$, being sufficiently selective to isolate and recover the decimated difference frequency component at $Df_\Delta = 128 \times 0.00125 = 0.16$. This choice was also motivated by the desire to minimize windowing effects, since then $TDf_\Delta = 50 \times 0.16 = 8$ encompasses an integer number of cycles at the decimated difference frequency.

In the simulation, $(n_0 - 1)$ samples are discarded before the synchronous demodulation calculation is started in order to let

the transient response of the allpass filters stabilize, as previously discussed. Here, $r = 0.999$ is the largest value considered, which corresponds to a transient time constant of about 1000 samples. Given that we have selected $DT = 128 \times 50 = 6400$ undecimated samples for synchronous demodulation, we have chosen a total number of $N = 8192$ input samples so that the initial 1792 samples will provide for about two time constants. Thus, discarding $1792/D = 1792/128 = 14$ decimated samples, we start the synchronous demodulation at $n_0 = 15$ in all cases.

In all simulations, the starting phase of the sinusoids is varied over an ensemble of realizations. This is accomplished by interjecting a delay between each batch of N samples. Since all of the desired information is at the difference frequency f_Δ [2nd term of (6)], we span its starting phase over $i = 1, 2, \cdots, I$ iterations by choosing the interblock delay for the ith iteration as the integer nearest to $i/(If_\Delta)$. (Recall that $f_\Delta = 0.00125$ here.) So, for example, with $I = 20$, we have 20 realizations with difference frequency starting phases uniformly distributed over $(0°, 360°)$ in $18°$ increments.

B. Results and Discussion

The power spectra at the output of the center frequency square law detector (u_2) are shown by the solid traces in Fig. 4 for $r = 0.999$ and several values of σ (8192 samples, overlapped 128-point Hanning window). Also plotted are the spectra of the various mixing components: $u_2^{S \times S}$ (dashed), $u_2^{N \times N}$ (dotted), and $u_2^{S \times N}$ (dashed-dot). The difference frequency spectral component is the peak centered at $Df_\Delta = 0.16$. We also see peaks at dc and $2Df_\Delta$. The relative strength of the component spectra depends on the input noise power: for $\sigma = 0.1$ [Fig. 4(a)], the S × S term dominates, followed by the S × N term, with negligible N × N contribution; for $\sigma = 1$ [Fig. 4(b)], the S × N and N × N terms are roughly comparable, but are still considerably lower than the peak of the S × S term at the difference frequency; for $\sigma = 10$ [Fig. 4(c)], the N × N term dominates. The theoretical amplitude of the difference frequency signal component is calculated from the second term of (6) for this example as 1.0492. Therefore, the 64-point (8192/128) power spectrum of the S × S component should be $64/2 \times (1.0492)^2 = 35.23$ (15.5 dB) at the difference frequency 0.16, in agreement with the figure. The theoretical S × N and N × N power spectra are approximately white with power levels shown in Table I, which were calculated from (33) and (36), and are seen to be in rough agreement with the levels shown in Fig. 4.

Fig. 5 shows the calculated estimates $\hat{\delta}$ and $\hat{a}$ for $I = 20$ iterations. The three traces in each plot are for number of measurements $M = 2$ (solid), $M = 3$ (dashed), and $M = 4$ (dotted). For the noiseless case ($\sigma = 0$) [Fig. 5(a)], the estimate variation is solely due to slightly different aliasing effects as the starting phases are varied, as previously discussed. We also see some estimate bias from the true values (horizontal lines), again due to the aliasing. The noiseless case should be considered as a baseline for assessing noise effects since the nonideal behavior represents simulation artifacts and does not reflect fundamental performance limitations—the actual wide-band input signal is analog, not digital. Fig. 5(b) and (c) shows how the estimates are affected by the addition of noise ($\sigma = 1, 2$). We see that
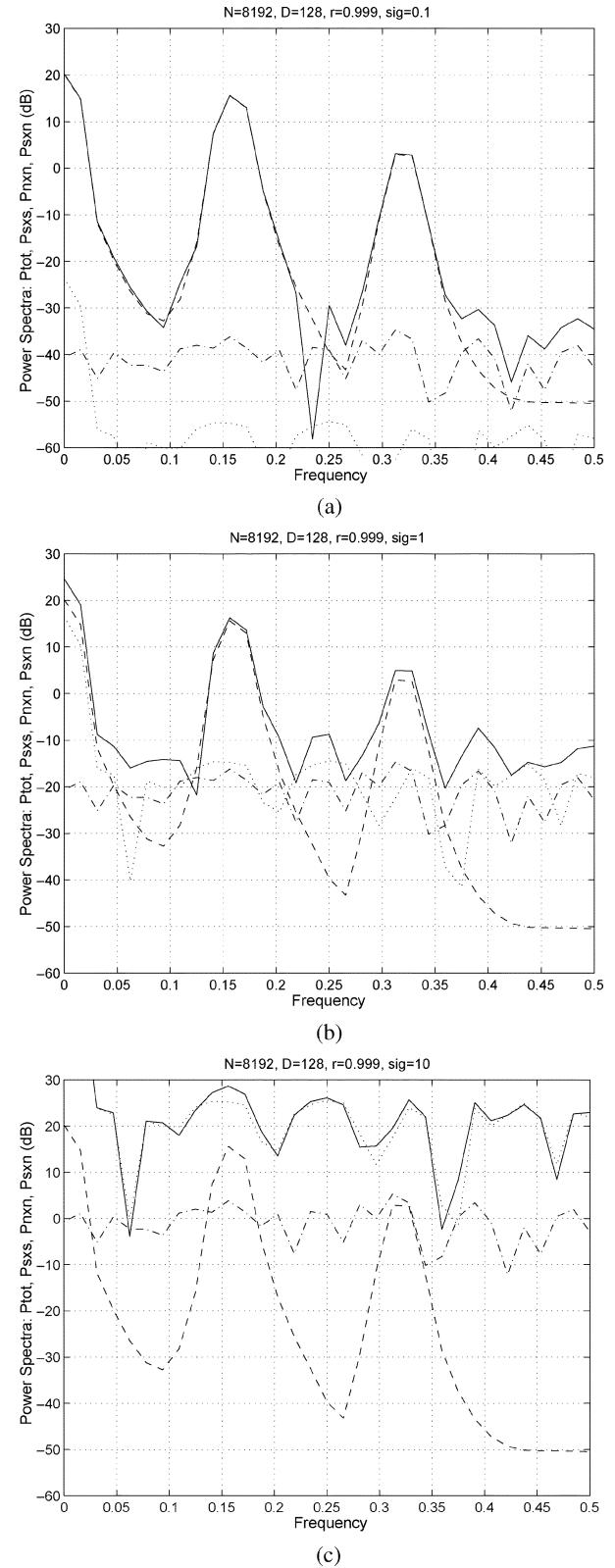

Fig. 4. Power spectra of square-law detector output for $r = 0.999$, showing total spectrum (solid) and components from S × S (dashed), N × N (dotted), and S × N (dashed-dot) terms. (a) $\sigma = 0.1$. (b) $\sigma = 1$. (c) $\sigma = 10$.

the variation of the parameter estimates is increased from their noiseless values.

The above simulation was repeated for various values of allpass filter parameter r using $I = 2000$ iterations and the resulting statistics are compiled in Table II. Since the 2000 iter-

TABLE I
THEORETICAL SQUARE-LAW DETECTOR OUTPUT SPECTRAL LEVELS OF
CENTER FREQUENCY CHANNEL

σ	$S_{u_2 u_2}^{\text{S} \times \text{N}}$ (dB)	$S_{u_2 u_2}^{\text{N} \times \text{N}}$ (dB)
0.1	−38.6	−58.1
1	−18.6	−18.1
10	1.4	21.9

ations involve different aliasing and noise realizations, the statistical significance of the values in the table are relatively accurate to the level of $1/\sqrt{2000} = 0.0224$, i.e., about 2%. We see that the mean of δ_1 and δ_2 is negatively biased by about 1° in all cases, being about the same with ($\sigma = 1, 2$) and without ($\sigma = 0$) noise. As previously discussed, these effects are due to a small amount of aliasing in the discrete-time simulation and the fact that they do not significantly change in the presence of noise means that they are probably negligible. The mean normalized amplitude estimates are also similarly biased with and without noise. In contrast to the means, the standard deviations do all increase in the presence of noise, as expected. They also increase systematically as r is reduced. This is due to the fact that with less sharp allpass filters, there is some correlation between the response at different tone frequencies which tends to degrade the conditioning of the matrix $\mathbf{F}$ defined in (11), thereby enhancing perturbations due to aliasing and extrinsic noise. In most cases, the bias and standard deviation decrease somewhat as M increases, which is understandable because redundant measurements are being utilized in the overdetermined case ($M > 2$). Finally, we observe the estimate correlations, which are not meaningful for the noiseless case but take on definite values in the presence of extrinsic noise. This relates to the correlation between the allpass filter outputs, as extensively discussed in Section III and Appendix A.

We now compare the simulation results of Table II with theoretical calculations. We will concentrate on the standard deviations and correlations in the presence of noise, which are the most meaningful measures. Table III compares calculations of the theoretical exact [5] and approximate values derived for the differential phase estimates in Section III using the results of Appendix A with the simulation results in Table II, all for $\sigma = 1$. For $r = 0.999$ and $r = 0.99$ we see that there is close agreement between the exact and approximate theoretical values, which are moreover in reasonable agreement with the simulation results. For $r = 0.9$, the approximation breaks down because the filter responses are too broad. Also, we see a wider gap between the exact theoretical and simulation results. This is due to the poorly conditioned $\mathbf{F}$ matrix, as previously discussed. Similar calculations are presented in Table IV for the amplitude estimates, where again we see good agreement except for the problematic case $r = 0.9$.

V. ADVANCED ESTIMATION TECHNIQUES

A. Side Amplitude Information

We have so far considered joint estimation of the amplitude and phase, which is captured in the least-squares estimate (15). If there were some independent means of obtaining amplitude
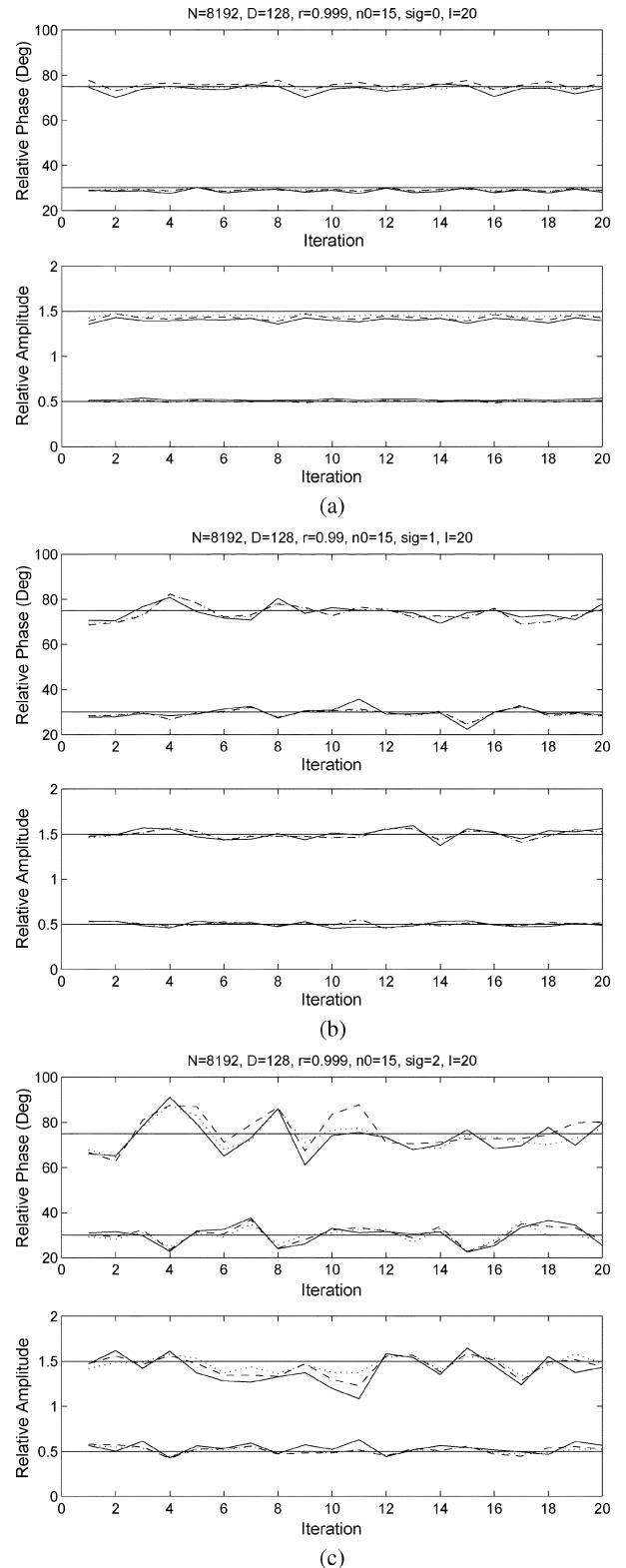

Fig. 5. Calculated estimates $\hat{\boldsymbol{\delta}}$ and $\hat{\mathbf{a}}$ for discrete-time simulation with $M = 2$ (solid), $M = 3$ (dashed), and $M = 4$ (dotted) measurements; $r = 0.999$. (a) $\sigma = 0$. (b) $\sigma = 1$. (c) $\sigma = 2$.

measurements, one could consider using this information to possibly refine the phase estimates. Ideally, we would then know $\mathbf{D}$ in (12) exactly. In that case, instead of minimizing $\|\mathbf{v} - \mathbf{FD}e^{j\boldsymbol{\delta}}\|$ over all values of the complex vector $\mathbf{D}e^{j\boldsymbol{\delta}}$, we would minimize only over values of the phase vector $\boldsymbol{\delta}$. This is equivalent to minimizing $\|\mathbf{v} - \mathbf{FDz}\|$ over all values of a complex vector $\mathbf{z}$ with

TABLE II
MEAN, STANDARD DEVIATION, AND CORRELATION OF ESTIMATION ERRORS FOR VARIOUS VALUES OF r, σ, AND M. (δ_1, δ_2 STATISTICS IN DEGREES; $\hat{a}_1$, $\hat{a}_2$ STATISTICS RELATIVE TO $A_1 = 1$)

r	σ	M	$\hat{\delta}_1$		$\hat{\delta}_2$		Corr $\hat{\delta}_1, \hat{\delta}_2$	$\hat{a}_1$		$\hat{a}_2$		Corr $\hat{a}_1, \hat{a}_2$
			Bias	Std	Bias	Std		Bias	Std	Bias	Std	
0.999	0	2	−1.3173	0.8562	−1.1812	1.7274	0.1630	−0.0992	0.0218	0.0188	0.0088	0.0299
		3	−0.8931	0.5856	0.7552	1.3183	−0.0014	−0.0736	0.0223	0.0003	0.0101	−0.3256
		4	−0.8167	0.6092	−0.5139	0.6931	−0.0814	−0.0497	0.0163	0.0044	0.0082	−0.6149
0.99	0	2	−0.8848	1.3076	−0.8852	1.3429	−0.1163	−0.0107	0.0340	0.0003	0.0171	−0.7359
		3	−0.8848	0.3854	−0.8847	1.4251	0.8120	−0.0110	0.0100	0.0001	0.0099	0.6899
		4	−0.8848	0.3057	−0.8848	1.0005	0.7157	−0.0111	0.0080	0.0000	0.0071	0.5114
0.9	0	2	−0.8847	1.4308	−0.8854	1.3608	−0.3921	−0.0106	0.0371	0.0004	0.0200	−0.8440
		3	−0.8848	0.4466	−0.8848	1.4610	0.7911	−0.0110	0.0115	0.0001	0.0097	0.6889
		4	−0.8839	2.2977	−0.8838	6.2931	−0.6809	−0.0099	0.0595	0.0049	0.0707	−0.8228
0.999	1	2	−1.2608	1.8261	−1.1939	3.2821	−0.0075	−0.0973	0.0457	0.0194	0.0286	−0.3527
		3	−0.8447	1.4746	0.7678	3.0978	0.0168	−0.0716	0.0408	0.0007	0.0258	−0.2612
		4	−0.7891	1.4147	−0.5303	2.3691	−0.1548	−0.0477	0.0361	0.0044	0.0230	−0.4540
0.99	1	2	−0.8612	2.4792	−0.8922	2.9632	−0.1911	−0.0072	0.0640	0.0005	0.0356	−0.6800
		3	−0.8629	1.6605	−0.9230	2.9660	−0.0172	−0.0084	0.0429	0.0003	0.0289	−0.3717
		4	−0.8675	1.6289	−0.9235	2.7446	−0.0928	−0.0085	0.0421	0.0002	0.0278	−0.4418
0.9	1	2	−0.8903	3.0093	−0.9071	4.4739	−0.4833	−0.0070	0.0781	0.0023	0.0557	−0.7605
		3	−0.8625	2.5304	−0.9549	4.6718	−0.4176	−0.0078	0.0658	0.0018	0.0519	−0.6751
		4	−0.7974	3.5763	−1.0711	8.7302	−0.6279	−0.0066	0.0916	0.0076	0.0941	−0.7535
0.999	2	2	−1.1557	4.9568	−1.2052	5.8744	−0.0250	−0.0919	0.1195	0.0225	0.0646	−0.6058
		3	−0.7706	3.3551	0.8034	6.8706	0.2321	−0.0681	0.0854	0.0032	0.0551	−0.1327
		4	−0.7566	2.6820	−0.5516	4.7125	−0.1194	−0.0448	0.0679	0.0055	0.0446	−0.4163
0.99	2	2	−0.8175	5.5799	−0.9045	5.5244	−0.1848	0.0019	0.1454	0.0033	0.0728	−0.7349
		3	−0.8334	3.7218	−0.9761	6.3836	0.1095	−0.0033	0.0961	0.0026	0.0596	−0.3226
		4	−0.8490	3.6023	−0.9847	6.0350	0.0497	−0.0038	0.0925	0.0022	0.0578	−0.3786
0.9	2	2	−0.8363	7.6548	−1.1531	12.1491	−0.5037	0.0060	0.1943	0.0192	0.1499	−0.7750
		3	−0.7651	6.6377	−1.2906	12.9218	−0.4671	0.0020	0.1709	0.0185	0.1454	−0.7242
		4	−0.5897	7.6668	−1.7918	21.0690	−0.5942	0.0040	0.1972	0.0439	0.2161	−0.7506

TABLE III
EXACT, APPROXIMATE, AND SIMULATED PHASE ESTIMATE STATISTICS (IN DEGREES) FOR VARIOUS VALUES OF r, AND M; $\sigma = 1$

r	Method	Standard Deviation $\hat{\delta}_1$			Standard Deviation $\hat{\delta}_2$			Correlation $\hat{\delta}_1, \hat{\delta}_2$		
		$M = 2$	$M = 3$	$M = 4$	$M = 2$	$M = 3$	$M = 4$	$M = 2$	$M = 3$	$M = 4$
0.999	Exact	1.5392	1.3148	1.2448	2.6824	2.6501	2.2215	−0.0530	0.0175	−0.1704
	Approximation	1.5398	1.3144	1.2443	2.6808	2.6499	2.2205	−0.0535	0.0175	−0.1707
	Simulation	1.8261	1.4746	1.4147	3.2821	3.0978	2.3691	−0.0075	0.0168	−0.1548
0.99	Exact	2.0312	1.5538	1.5376	2.5885	2.5492	2.5010	−0.2383	−0.0969	−0.1280
	Approximation	2.0353	1.5578	1.5421	2.6049	2.5587	2.5112	−0.2410	−0.1022	−0.1327
	Simulation	2.4792	1.6605	1.6289	2.9632	2.9660	2.7446	−0.1911	−0.0172	−0.0928
0.9	Exact	2.4072	2.2409	2.3805	3.7360	3.8464	5.1175	−0.4580	−0.4315	−0.5157
	Approximation	5.1301	5.0409	5.2078	9.8099	9.8189	10.5735	−0.7137	−0.7134	−0.7176
	Simulation	3.0093	2.5304	3.5763	4.4739	4.6718	8.7302	−0.4833	−0.4176	−0.6279

TABLE IV
EXACT, APPROXIMATE, AND SIMULATED AMPLITUDE ESTIMATE STATISTICS (RELATIVE TO $A_1 = 1$) FOR VARIOUS VALUES OF r, AND M; $\sigma = 1$

r	Method	Standard Deviation $\hat{a}_1$			Standard Deviation $\hat{a}_2$			Correlation $\hat{a}_1, \hat{a}_2$		
		$M = 2$	$M = 3$	$M = 4$	$M = 2$	$M = 3$	$M = 4$	$M = 2$	$M = 3$	$M = 4$
0.999	Exact	0.0404	0.0344	0.0323	0.0250	0.0228	0.0207	−0.3800	−0.2252	−0.3955
	Approximation	0.0404	0.0344	0.0323	0.0250	0.0228	0.0207	−0.3803	−0.2241	−0.3951
	Simulation	0.0457	0.0408	0.0361	0.0286	0.0258	0.0230	−0.3527	−0.2612	−0.4540
0.99	Exact	0.0532	0.0407	0.0403	0.0296	0.0245	0.0244	−0.6482	−0.4355	−0.4563
	Approximation	0.0533	0.0408	0.0404	0.0297	0.0246	0.0245	−0.6474	−0.4371	−0.4578
	Simulation	0.0640	0.0429	0.0421	0.0356	0.0289	0.0278	−0.6800	−0.3717	−0.4418
0.9	Exact	0.0630	0.0587	0.0622	0.0446	0.0440	0.0568	−0.7309	−0.6932	−0.7081
	Approximation	0.1239	0.1215	0.1255	0.1072	0.1066	0.1145	−0.7972	−0.7930	−0.7936
	Simulation	0.0781	0.0658	0.0916	0.0557	0.0519	0.0941	−0.7605	−0.6751	−0.7535

the constraint that the components of $\mathbf{z}$, z_m, all have unit magnitude: $|z_m| = 1$, $m = 1, 2, \cdots, M$. However, this is a difficult problem to solve because using Lagrange multipliers to enforce the constraints generally leads to a set of nonlinear simultaneous equations. We have instead obtained a gradient solution that recursively minimizes

$$J = \|\mathbf{v} - \mathbf{F}\mathbf{D}e^{j\boldsymbol{\delta}}\| \qquad (63)$$

over $\boldsymbol{\delta}$. Taking the complex gradient gives, after some matrix algebra

$$\frac{dJ}{d\boldsymbol{\delta}} = -2\Im\left\{\operatorname{diag}\left[\mathbf{DF}^H(\mathbf{v} - \mathbf{DF}^H e^{j\boldsymbol{\delta}})\right]e^{j\boldsymbol{\delta}}\right\}. \quad (64)$$

Using the negative of this to adjust $\boldsymbol{\delta}$, we obtain the recursion

$$\boldsymbol{\delta}_{p+1} = \boldsymbol{\delta}_p + \mu\Im\left\{\operatorname{diag}\left[\mathbf{DF}^H(\mathbf{v} - \mathbf{DF}^H e^{j\boldsymbol{\delta}_p})\right]e^{j\boldsymbol{\delta}_p}\right\} \quad (65)$$

where μ is a step size chosen to achieve fast convergence while maintaining stability.

We modified the simulation to implement the above constrained estimation algorithm, assuming perfect *a priori* information for the amplitudes A_1, A_2, A_3 (i.e., perfect knowledge of $\mathbf{D}$). For the same parameter values used in Section IV, rapid and stable convergence was achieved in less than ten steps for $\mu = 0.5$. However, the noise-impaired differential phase estimates did not significantly improve. This result held even if we artificially fixed the mean value of $\mathbf{v}$ to its ideal value (12), thereby eliminating all aliasing effects in the simulation. We therefore conclude that, at least for this example, side information on the amplitudes is not very useful.

B. Weighted Least-Squares/Maximum Likelihood

So far in this paper, we have used a least-squares estimator for the unknown parameters. However, given the results of Section III, we know that the noise is not i.i.d. Therefore, we could possibly improve the estimation quality by incorporating this knowledge in a form of weighted least squares, which is equivalent to a maximum-likelihood solution in the case of Gaussian noise. As we have already observed in Section III-B, the nonproper complex demodulator output requires special treatment using both the Hermitian covariance matrix $\mathbf{R}_v$ and symmetric complementary covariance matrix $\mathbf{C}_v$. Here, we will need to derive a maximum-likelihood estimator of $\mathbf{w} = \mathbf{D}e^{j\boldsymbol{\delta}}$, given the complex synchronous demodulator output vector $\mathbf{v}$. From Appendix C, we can express the maximum likelihood estimator as

$$\mathbf{w}_{\mathrm{ML}} = (\tilde{\mathbf{P}}^{-1} - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\tilde{\mathbf{Q}})^{-1}\left[(\mathbf{F}^H\mathbf{P}^{-1} - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\mathbf{F}^T\mathbf{Q})\mathbf{v}\right.$$
$$\left. + (\mathbf{F}^H\mathbf{Q}^* - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\mathbf{F}^T\mathbf{P}^{-*})\mathbf{v}^*\right] \quad (66)$$

where

$$\mathbf{P} = \mathbf{P}^H \equiv \mathbf{R}_v - \mathbf{C}_v\mathbf{R}_v^{-*}\mathbf{C}_v^* \quad (67)$$
$$\tilde{\mathbf{P}}^{-1} = \tilde{\mathbf{P}}^{-H} \equiv \mathbf{F}^H\mathbf{P}^{-1}\mathbf{F} \quad (68)$$
$$\mathbf{Q} = \mathbf{Q}^T \equiv \mathbf{R}_v^{-*}\mathbf{C}_v^*\mathbf{P}^{-1} \quad (69)$$

and

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{Q}}^T \equiv \mathbf{F}^T\mathbf{Q}\mathbf{F}. \quad (70)$$

Note that in the special case when $\mathbf{F}$ is square ($M = L - 1$), $\mathbf{w}_{\mathrm{ML}} = \mathbf{F}^{-1}\mathbf{v}$. Also, if $\mathbf{v}$ were proper with $\mathbf{C}_v = \mathbf{0}$, then $\mathbf{Q} = \tilde{\mathbf{Q}} = \mathbf{0}$ and (66) reduces to $\mathbf{w}_{\mathrm{ML}} = (\mathbf{F}^H\mathbf{R}_v^{-1}\mathbf{F})^{-1}\mathbf{F}^H\mathbf{R}_v^{-1}\mathbf{v}$, which for $\mathbf{R}_v = \mathbf{I}$ further reduces to the least-squares formulation (15).

The simulation was modified to incorporate the maximum likelihood estimator. For the same parameter values used in Sec-

tion IV, the differential phase and relative amplitude estimates were only slightly improved. As the allpass pole radius is decreased from 0.999 to 0.9, the numerical conditioning becomes slightly better than for the unweighted least-squares solution, but not dramatically so. So from this exercise, we conclude that weighted least-squares offers little improvement, at least for the parameter values considered in this paper.

VI. CONCLUSION

In this paper, we have presented a wide-band system identification technique that can be implemented with parallel digital processing, making it attractive for systems that are otherwise only measurable using analog techniques. This method can be implemented using either a simple tunable wide-band analog allpass filter or a bank of fixed allpass filters, followed by square-law detection and digital sampling at a reduced rate. A mathematical basis was derived for the scheme and an analysis was presented for the effects of noise on estimate errors and correlations. A simulation for three tones was presented and results were compared with their theoretical values. A major conclusion of this study is that the allpass filters should be neither too sharp nor too broad: sharp filters would lead to modeling inaccuracies, especially when they have to be calibrated and controlled as tunable analog filters; broad filters have strong coupling between the responses at different tone frequencies and this leads to poor numerical conditioning. For the simulations in this paper we found that a pole radius of about 0.99 ($Q = 100$) provided reasonably accurate and predictable results.

Future work could extend these results to a moderately large number of tones, say 10–30, to see how the technique scales. Also, other filter types that are more representative of hardware for specific applications could be studied using the general theory of this paper. For example, in fiber optic applications, an optical ring resonator with couplers could be modeled as a more general IIR filter, still with allpass-like phase but also incorporating frequency-dependent loss.

APPENDIX A
DERIVATION OF CROSS-CHANNEL CORRELATIONS

In the main text, the square-law mixing terms (30) of each allpass filtered channel are low-pass filtered to produce the detector output components (32), and then are synchronously demodulated to form output components (37). In this appendix, we sketch a theoretical derivation of the cross-correlation between the different channels, before and after synchronous demodulation. A more complete treatment, along with asymptotic exact results, can be found in [5].

As a starting point, we shall represent the wide-band Gaussian noise input $n(t)$, depicted in Fig. 1, as the sum of a large number of K sinusoids with random phase, viz.,

$$n(t) = \frac{\sigma}{\sqrt{K/2}}\sum_{k=1}^{K}\cos(\nu_k t + \psi_k) \quad (A1)$$

where σ^2 is the noise power, the ν_k's are uniformly spaced over the normalized (single-sided) system radial bandwidth from 0 to π, and the ψ_k's are independent random variables, uniformly distributed on $(0, 2\pi)$. This representation assumes that the noise is "white," i.e., uniform, within the system bandwidth, and enables tractable analysis of the mixing terms that result from the nonlinear square-law device. The model does not employ a random amplitude for each sinusoid; however, we maintain that this is not necessary for large K by appealing to the central limit theorem.

With the above model, the noise output of the mth allpass filter output is written

$$n_m(t) = \frac{\sigma}{\sqrt{K/2}} \sum_{k=1}^{K} \cos\left[\nu_k t + \psi_k + \theta_m(\nu_k)\right] \tag{A2}$$

where $\theta_m(\cdot)$ is the phase of the mth allpass filter frequency response, which is assumed here to be ideal with unity magnitude (28). Substituting this into (31b) expresses the $\mathrm{S} \times \mathrm{N}$ cross-term of the square-law detector output as

$$u_m^{\mathrm{S \times N}}(t) = \frac{\sigma}{\sqrt{2K}} \sum_{l=1}^{L} A_l \sum_{k \in \mathcal{K}_l} \cos[(\omega_l - \nu_k)t + \phi_l \\ - \psi_k + \theta_m(\omega_l) - \theta_m(\nu_k)] \tag{A3}$$

where

$$\mathcal{K}_l \equiv \left\{k : \omega_l - \frac{\pi}{\beta} \leq \nu_k \leq \omega_l + \frac{\pi}{\beta}\right\} \tag{A4}$$

is a subset of the K angular frequencies centered around $\omega_l = 2\pi f_l$ that contribute to difference frequency components within the passband of the low-pass filter.

We now consider the correlation between the mth and nth channels and calculate (A5), as shown at the bottom of the page,

where we have used the following reasoning to obtain the final result. First, since the ψ_k's are uniformly distributed over $(0, 2\pi)$ and independent over the frequency index k, only cross-terms with $k_1 = k_2 = k$ survive ensemble expectation. Next, taking expectation over time t, only sinusoids with frequency index $l_1 = l_2 = l$ will contribute. Similar calculations can be performed to obtain

$$R_{u_m u_n}^{\mathrm{N \times N}}(\tau) \equiv E\left\{u_m^{\mathrm{N \times N}}(t)u_n^{\mathrm{N \times N}}(t-\tau)\right\} \\ = \frac{\sigma^4}{K^2} \sum_{k=1}^{K} \sum_{l \in \mathcal{L}_k} \cos\left[\theta_m(\nu_k) - \theta_n(\nu_k) - \theta_m(\nu_l)\right. \\ \left. + \theta_n(\nu_l)\right]\delta(\tau) \tag{A6}$$

where

$$\mathcal{L}_k \equiv \left\{l : \max\left(0, \nu_k - \frac{\pi}{\beta}\right) \leq \nu_l \leq \min\left(\pi, \nu_k + \frac{\pi}{\beta}\right)\right\}. \tag{A7}$$

Further approximation of (A5) and (A6) is deferred to the end of this appendix, where it will be conveniently expressed in terms of a more general formulation.

We now derive expressions for the correlation of the synchronous demodulator outputs (37). In order to obtain statistics over an ensemble, we consider multiple realizations over time and, for analytical convenience, center the T-sample synchronous demodulator window at each time instance t. By considering the synchronous demodulation as a linear filter, we obtain (A8) and (A9), shown at the bottom of the page, where

$$S(\omega) \equiv \frac{1}{T} \sum_{n=-\frac{T-1}{2}}^{\frac{T+1}{2}} e^{\pm j\beta(\omega - \omega_\Delta)n} = \frac{\sin\left[\frac{\beta T(\omega - \omega_\Delta)}{2}\right]}{T\sin\left[\frac{\beta(\omega - \omega_\Delta)}{2}\right]}. \tag{A10}$$

$$R_{u_m u_n}^{\mathrm{S \times N}}(\tau) \equiv E\left\{u_m^{\mathrm{S \times N}}(t)u_n^{\mathrm{S \times N}}(t-\tau)\right\}$$

$$= E\left\{\frac{\sigma}{\sqrt{2K}} \sum_{l_1=1}^{L} A_{l_1} \sum_{\nu_{k_1} \in \mathcal{K}_{l_1}} \cos\left[(\omega_{l_1} - \nu_{k_1})t + \phi_{l_1} - \psi_{k_1} + \theta_m(\omega_{l_1}) - \theta_m(\nu_{k_1})\right]\right.$$

$$\left. \times \frac{\sigma}{\sqrt{2K}} \sum_{l_2=1}^{L} A_{l_2} \sum_{\nu_{k_2} \in \mathcal{K}_{l_2}} \cos\left[(\omega_{l_2} - \nu_{k_2})(t-\tau) + \phi_{l_2} - \psi_{k_2} + \theta_n(\omega_{l_2}) - \theta_n(\nu_{k_2})\right]\right\}$$

$$= \frac{\sigma^2}{4K} \sum_{l=1}^{L} A_l^2 \sum_{k \in \mathcal{K}_l} \cos\left[\theta_m(\omega_l) - \theta_n(\omega_l) - \theta_m(\nu_k) + \theta_n(\nu_k)\right]\delta(\tau) \tag{A5}$$

$$v_m^{\mathrm{S \times N}}(t) = \frac{\sigma}{\sqrt{8K}} \sum_{l=1}^{L} A_l \sum_{k \in \mathcal{K}_l} \left\{S(\omega_l - \nu_k)e^{+j[(\omega_l - \nu_k - \omega_\Delta)t + \phi_l - \psi_k + \theta_m(\omega_l) - \theta_m(\nu_k)]}\right.$$

$$\left. + S(\nu_k - \omega_l)e^{-j[(\omega_l - \nu_k + \omega_\Delta)t + \phi_l - \psi_k + \theta_m(\omega_l) - \theta_m(\nu_k)]}\right\} \tag{A8}$$

$$v_m^{\mathrm{N \times N}}(t) = \frac{\sigma^2}{2K} \sum_{k=1}^{K} \sum_{l \in \mathcal{L}_k} \left\{S(\nu_k - \nu_l)e^{+j[(\nu_k - \nu_l - \omega_\Delta)t + \psi_k - \psi_l + \theta_m(\nu_k) - \theta_m(\nu_l)]}\right.$$

$$\left. + S(\nu_l - \nu_k)e^{-j[(\nu_k - \nu_l + \omega_\Delta)t + \psi_k - \psi_l + \theta_m(\nu_k) - \theta_m(\nu_l)]}\right\} \tag{A9}$$

In a similar manner as before, we now calculate the cross-channel correlations as shown in (A11)–(A13), shown at the bottom of the page, and

$$C^{\mathrm{N \times N}}_{v_m v_n} = E\left\{v^{\mathrm{N \times N}}_m(t) v^{\mathrm{N \times N}}_n(t)\right\} = 0. \qquad \text{(A14)}$$

Equations (A11)–(A14) were used for the exact calculations in the main text. However, these are not very efficient to compute when K is a large number, e.g., 8192, especially (A12) which involves a double sum over K. Therefore, we now go on the derive approximate expressions for these quantities that can be quickly computed over many parameter sets. First note that for T moderately large so that the bandwidth of $S(\omega)$ is small compared to the allpass transition bandwidth, (A10) can be approximated as

$$S(\omega) \approx \frac{\sin\left[\dfrac{\beta T(\omega - \omega_\Delta)}{2}\right]}{\dfrac{\beta T(\omega - \omega_\Delta)}{2}} = \mathrm{si}\left[\frac{\beta T(\omega - \omega_\Delta)}{2}\right] \qquad \text{(A15)}$$

where $\mathrm{si}(x) \equiv \sin(x)/x$. Substituting this into (A11) and taking the limit as $K \to \infty$ (whence K/β frequencies are contained in $\mathcal{K}_l$) yields the integral expression (A16), shown at the botttom of the page.

Next, we assume second-order allpass filters of the form (62) and approximate their phase response as [5]

$$\theta_m \approx -\pi - 2\tan^{-1}\left(\frac{\omega - \nu_m}{\alpha}\right) \qquad \text{(A17)}$$

where α is defined in (35) and r is the pole radius. Using this to express the various phase terms in (A16) gives

$$R^{\mathrm{S \times N}}_{v_m v_n} \approx \frac{\sigma^2}{4} \sum_{l=1}^{L} A_l^2 \frac{\omega_l - \nu_m + j\alpha}{\omega_l - \nu_m - j\alpha} \cdot \frac{\omega_l - \nu_n - j\alpha}{\omega_l - \nu_n + j\alpha}$$
$$\times \left[I_{mn}(\omega_l, \omega_\Delta) + I^*_{mn}(\omega_l, -\omega_\Delta)\right] \qquad \text{(A18)}$$

where we have defined

$$I_{mn}(\omega_l, \omega_\Delta) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{si}^2\left[\frac{\beta T(\omega_l - \omega - \omega_\Delta)}{2}\right]$$
$$\times \frac{\omega - \nu_m - j\alpha}{\omega - \nu_m + j\alpha} \cdot \frac{\omega - \nu_n + j\alpha}{\omega - \nu_n - j\alpha} d\omega. \qquad \text{(A19)}$$

The integral I_{mn} can be evaluated using contour integration in the complex plane [5], thereby with (A18) above yielding (39) and (34) in the main text for the special case when $\nu_m = \omega_m$. Similar calculations for the complementary $\mathrm{S \times N}$ cross-correlation (A13) yield (42). For the $\mathrm{N \times N}$ cross-correlation, the total phase is approximately zero over most indexes, and summing the si function as an integral obtains (40).

The square-law detector outputs can be considered as special cases of the above theory for $T = 1$, whereby the integration is approximately limited to the appropriate bandwidth $1/(\beta T)$. Thus, by substituting $T = 1$ into (39) and (40), the cross-correlations (33) and (36) follow as approximations of (A5) and (A6), expressed in terms of (34).

$$R^{\mathrm{S \times N}}_{v_m v_n} \equiv E\left\{v^{\mathrm{S \times N}}_m(t) \left[v^{\mathrm{S \times N}}_n(t)\right]^*\right\}$$
$$= \frac{\sigma^2}{8K} \sum_{l=1}^{L} A_l^2 \sum_{k \in \mathcal{K}_l} \left\{S^2(\omega_l - \nu_k) e^{+j[\theta_m(\omega_l) - \theta_n(\omega_l) - \theta_m(\nu_k) + \theta_n(\nu_k)]}\right.$$
$$\left. + S^2(\nu_k - \omega_l) e^{-j[\theta_m(\omega_l) - \theta_n(\omega_l) - \theta_m(\nu_k) + \theta_n(\nu_k)]}\right\} \qquad \text{(A11)}$$

$$R^{\mathrm{N \times N}}_{v_m v_n} \equiv E\left\{v^{\mathrm{N \times N}}_m(t) \left[v^{\mathrm{N \times N}}_n(t)\right]^*\right\}$$
$$= \frac{\sigma^4}{2K^2} \sum_{k=1}^{K} \sum_{l \in \mathcal{L}_k} \left\{S^2(\nu_k - \nu_l) e^{+j[\theta_m(\nu_k) - \theta_n(\nu_k) - \theta_m(\nu_l) + \theta_n(\nu_l)]}\right.$$
$$\left. + S^2(\nu_l - \nu_k) e^{-j[\theta_m(\nu_k) - \theta_n(\nu_k) - \theta_m(\nu_l) + \theta_n(\nu_l)]}\right\} \qquad \text{(A12)}$$

$$C^{\mathrm{S \times N}}_{v_m v_n} \equiv E\left\{v^{\mathrm{S \times N}}_m(t) v^{\mathrm{S \times N}}_n(t)\right\}$$
$$= \frac{\sigma^2}{8K} \left\{\sum_{l=3}^{L} A_l A_{l-2} \sum_{k \in \mathcal{K}_l} S^2(\omega_l - \nu_k) e^{+j[\phi_l - \phi_{l-2} + \theta_m(\omega_l) - \theta_n(\omega_{l-2}) - \theta_m(\nu_k) + \theta_n(\nu_k)]}\right.$$
$$\left. + \sum_{l=1}^{L-2} A_l A_{l+2} \sum_{k \in \mathcal{K}_l} S^2(\omega_{l+2} - \nu_k) e^{-j[\phi_l - \phi_{l+2} + \theta_m(\omega_l) - \theta_n(\omega_{l+2}) - \theta_m(\nu_k) + \theta_n(\nu_k)]},\right\} \qquad \text{(A13)}$$

$$R^{\mathrm{S \times N}}_{v_m v_n} \approx \frac{\sigma^2}{4} \sum_{l=1}^{L} A_l^2 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{\mathrm{si}^2\left[\frac{\beta T(\omega_l - \omega - \omega_\Delta)}{2}\right] e^{+j[\theta_m(\omega_l) - \theta_n(\omega_l) - \theta_m(\omega) + \theta_n(\omega)]}\right.$$
$$\left. + \mathrm{si}^2\left[\frac{\beta T(\omega_l - \omega + \omega_\Delta)}{2}\right] e^{-j[\theta_m(\omega_l) - \theta_n(\omega_l) - \theta_m(\omega) + \theta_n(\omega)]}\right\} d\omega \qquad \text{(A16)}$$
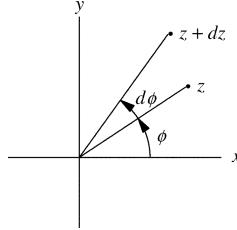
Fig. 6.   Differentials in the complex plane.

APPENDIX B
DIFFERENTIALS IN COMPLEX PLANE

Consider the differential of the phase function

$$\phi = \arg z = \tan^{-1} \frac{y}{x} \tag{B1}$$

where $z = x + jy$ is a complex variable with real components x and y. Note that ϕ can be considered as a function of z and z^*. The usual representation of the derivative of a complex variable forms

$$2\frac{\partial \phi}{\partial z} = \frac{\partial \phi}{\partial x} - j\frac{\partial \phi}{\partial y}$$
$$= \frac{-y - jx}{x^2 + y^2} = \frac{-jz^*}{|z|^2} = -\frac{j}{z}. \tag{B2}$$

Combining this with a similar expression for the partial with respect to z^* gives the total differential

$$d\phi = 2\Re\left(\frac{\partial \phi}{\partial z}dz\right)$$
$$= \Re\left(-\frac{j}{z}dz\right) = \Im\left(\frac{dz}{z}\right). \tag{B3}$$

This result can also be appreciated by drawing a diagram in the complex plane with $z = |z|e^{j\phi}$ (see Fig. 6), where one immediately sees the special cases

$$d\phi = \frac{dy}{|z|}, \qquad \phi = 0 \tag{B4a}$$

$$d\phi = -\frac{dx}{|z|}, \qquad \phi = \frac{\pi}{2}. \tag{B4b}$$

We also consider differentials of the magnitude

$$|z| = \sqrt{x^2 + y^2}. \tag{B5}$$

Proceeding as above, we calculate

$$2\frac{\partial |z|}{\partial z} = \frac{\partial |z|}{\partial x} - j\frac{\partial |z|}{\partial y}$$
$$= \frac{x - jy}{|z|} = \frac{z^*}{|z|} = e^{-j\phi} \tag{B6}$$

and then obtain

$$d|z| = 2\Re\left(\frac{\partial |z|}{\partial z}dz\right)$$
$$= \Re(e^{-j\phi}dz). \tag{B7}$$

Again, one immediately sees the special cases (Fig. 6)

$$d|z| = dx, \qquad \phi = 0 \tag{B8a}$$

$$d|z| = dy, \qquad \phi = \frac{\pi}{2}. \tag{B8b}$$

APPENDIX C
MAXIMUM LIKELIHOOD ESTIMATION IN GENERAL COMPLEX GAUSSIAN NOISE

In this appendix, we derive the maximum likelihood estimator of a complex $N \times 1$ vector $\mathbf{x}$, given measurements of the complex $M \times 1$ vector $\mathbf{y}$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{C1}$$

where $\mathbf{A}$ is an arbitrary $M \times N$ matrix and $\mathbf{n}$ is an $M \times 1$ vector of zero-mean complex Gaussian noise with Hermitian covariance matrix

$$\mathbf{R} \equiv E\{\mathbf{n}\mathbf{n}^H\} \tag{C2}$$

where superscript H denotes conjugate transpose. For generality, we also define the *complementary* covariance matrix

$$\mathbf{C} \equiv E\{\mathbf{n}\mathbf{n}^T\}. \tag{C3}$$

For so-called *proper* complex random variables, the real and imaginary parts are uncorrelated and of equal variance (circularly symmetric) so that $\mathbf{C} = \mathbf{0}$. However, there are many applications for which this condition does not hold [2]–[4].

From [3] we can write the $2M$-dimensional probability density of $\mathbf{n}$ as

$$f_{\mathbf{N}}(\mathbf{n}) = c\exp\left[-q(\mathbf{n},\mathbf{n}^*)\right] \tag{C4}$$

where c is an unimportant constant and

$$q(\mathbf{n},\mathbf{n}^*) \equiv [\mathbf{n}^H \quad \mathbf{n}^T]\begin{bmatrix} \mathbf{R} & \mathbf{C} \\ \mathbf{C}^* & \mathbf{R}^* \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{n} \\ \mathbf{n}^* \end{bmatrix}$$
$$= [\mathbf{n}^H \quad \mathbf{n}^T]\begin{bmatrix} \mathbf{P}^{-1} & \mathbf{Q}^* \\ \mathbf{Q} & \mathbf{P}^{-*} \end{bmatrix}\begin{bmatrix} \mathbf{n} \\ \mathbf{n}^* \end{bmatrix}$$
$$= 2\mathbf{n}^H\mathbf{P}^{-1}\mathbf{n} + \mathbf{n}^H\mathbf{Q}^*\mathbf{n}^* + \mathbf{n}^T\mathbf{Q}\mathbf{n} \tag{C5}$$

where

$$\mathbf{P} = \mathbf{P}^H \equiv \mathbf{R} - \mathbf{C}\mathbf{R}^{-*}\mathbf{C}^* \tag{C6}$$

is a Hermitian matrix (Shure complement) and

$$\mathbf{Q} = \mathbf{Q}^T \equiv -\mathbf{R}^{-*}\mathbf{C}^*\mathbf{P}^{-1} = (\mathbf{C} - \mathbf{R}\mathbf{C}^{-*}\mathbf{R}^*)^{-1} \tag{C7}$$

is a symmetric matrix. Therefore, considering $\mathbf{x}$ as an unknown parameter vector, we write the probability density of the observations $\mathbf{y}$ as

$$f_{\mathbf{Y}}(\mathbf{y}) = c\exp\left[-q(\mathbf{y} - \mathbf{A}\mathbf{x}, \mathbf{y}^* - \mathbf{A}^*\mathbf{x}^*)\right]. \tag{C8}$$

The maximum-likelihood estimate of $\mathbf{x}$ is then determined by maximizing (C8)

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} f_{\mathbf{Y}}(\mathbf{y}). \tag{C9}$$

Equivalently, we seek to minimize

$$q(\mathbf{y} - \mathbf{A}\mathbf{x}, \mathbf{y}^*\mathbf{A}^*\mathbf{x}^*) = 2(\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{P}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x})$$
$$+ (\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{Q}^*(\mathbf{y} - \mathbf{A}\mathbf{x})^* + (\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{Q}(\mathbf{y} - \mathbf{A}\mathbf{x}). \quad \text{(C10)}$$

Taking derivatives of the components of (C10) with respect to $\mathbf{x}^*$, we have

$$\frac{d}{d\mathbf{x}^*}(\mathbf{y}^H - \mathbf{x}^H\mathbf{A}^H)\mathbf{P}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x})$$
$$= -\mathbf{A}^H\mathbf{P}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}) \quad \text{(C11)}$$

$$\frac{d}{d\mathbf{x}^*}(\mathbf{y}^H - \mathbf{x}^H\mathbf{A}^H)\mathbf{Q}^*(\mathbf{y}^* - \mathbf{A}^*\mathbf{x}^*)$$
$$= -2\mathbf{A}^H\mathbf{Q}^*(\mathbf{y}^* - \mathbf{A}^*\mathbf{x}^*) \quad \text{(C12)}$$

and

$$\frac{d}{d\mathbf{x}^*}(\mathbf{y}^T - \mathbf{x}^T\mathbf{A}^T)\mathbf{Q}(\mathbf{y} - \mathbf{A}\mathbf{x}) = 0. \quad \text{(C13)}$$

(Note that in minimizing such quadratic forms, $\mathbf{x}$ and $\mathbf{x}^*$ can be considered as independent variables [6].) Thus, equating the derivative of (C10) to zero formulates the maximum likelihood estimate $\hat{\mathbf{x}}$

$$\frac{1}{2}\frac{d}{d\hat{\mathbf{x}}^*}q(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}, \mathbf{y}^* - \mathbf{A}^*\hat{\mathbf{x}}^*)$$
$$= -\mathbf{A}^H\mathbf{P}^{-1}(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) - \mathbf{A}^H\mathbf{Q}^*(\mathbf{y}^* - \mathbf{A}^*\hat{\mathbf{x}}^*)$$
$$= 0. \quad \text{(C14)}$$

To obtain $\hat{\mathbf{x}}$, we simultaneously solve the above equation and its complex conjugate, eliminating $\hat{\mathbf{x}}^*$, giving

$$\hat{\mathbf{x}} = (\tilde{\mathbf{P}}^{-1} - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\tilde{\mathbf{Q}})^{-1}\left[(\mathbf{A}^H\mathbf{P}^{-1} - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\mathbf{A}^T\mathbf{Q})\mathbf{y}\right.$$
$$\left. + (\mathbf{A}^H\mathbf{Q}^* - \tilde{\mathbf{Q}}^*\tilde{\mathbf{P}}^*\mathbf{A}^T\mathbf{P}^{-*})\mathbf{y}^*\right] \quad \text{(C15)}$$

where

$$\tilde{\mathbf{P}}^{-1} = \tilde{\mathbf{P}}^{-H} \equiv \mathbf{A}^H\mathbf{P}^{-1}\mathbf{A} \quad \text{(C16)}$$

and

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{Q}}^T \equiv \mathbf{A}^T\mathbf{Q}\mathbf{A}. \quad \text{(C17)}$$

Note that in the special case when $\mathbf{A}$ is square $(N = M)$, $\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{y}$. Also, if $\mathbf{n}$ were proper with $\mathbf{C} = \mathbf{0}$, then $\mathbf{Q} = \tilde{\mathbf{Q}} = \mathbf{0}$ and (C15) reduces to $\hat{\mathbf{x}} = (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{y}$, which for $\mathbf{R} = \mathbf{I}$ further reduces to the usual least-squares solution $\hat{\mathbf{x}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{y}$.

REFERENCES

[1] C. K. Madsen, "A novel optical vector spectral analysis technique," *Eurasip J. Appl. Signal Process.*, vol. 2005, no. 10, pp. 1566–1573, Jul. 1, 2005.

[2] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 7, pp. 1293–1302, Jul. 1993.

[3] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.

[4] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[5] D. R. Morgan, "Variance and correlation of square-law detected allpass channels with bandpass harmonic signals in Gaussian noise," *IEEE Trans. Signal Process.*, to be published.

[6] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng. Pts. F and H*, vol. 130, pp. 11–16, Feb. 1983.

Dennis R. Morgan (S'63–S'68–M'69–SM'92) was born in Cincinnati, OH, on February 19, 1942. He received the B.S. degree from the University of Cincinnati, OH, in 1965, and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1968 and 1970, respectively, all in electrical engineering.

From 1965 to 1984, he was with the General Electric Company, Electronics Laboratory, Syracuse, NY, specializing in the analysis and design of signal processing systems used in radar, sonar, and communications. He is now a Distinguished Member of Technical Staff with Bell Laboratories, Lucent Technologies, where he has been employed since 1984. From 1984 to 1990, he was with the Special Systems Analysis Department, Whippany, NJ, where he was involved in the analysis and development of advanced signal processing techniques associated with communications, array processing, detection estimation, and active noise control. From 1990 to 2002, he was with the Acoustics Research Department, Murray Hill, NJ, where he was engaged in research on adaptive signal processing techniques applied to electroacoustic systems, including adaptive microphones, echo cancellation, talker direction finders, and blind source separation. Since 2002, he has been with the Wireless Research Laboratory, Murray Hill, NJ, where he is involved in research on adaptive signal processing applied to RF and optical communication systems. He has authored numerous journal publications and is coauthor of *Active Noise Control Systems: Algorithms and DSP Implementations* (New York: Wiley, 1996) and *Advances in Network and Acoustic Echo Cancellation* (New York: Springer-Verlag, 2001).

Dr. Morgan served as an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING from 1995 to 2000. He has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2001 to 2004.

Christi K. Madsen (S'84–M'86–SM'96) received the Bachelor's degree from The University of Texas at Austin, the Master's degree from Stanford University, Stanford, CA, and the Ph.D. degree from Rutgers University, Piscataway, NJ, in 1986, 1987, and 1996, respectively, all in electrical engineering.

She joined AT&T Bell Laboratories in 1987 and worked for the submarine systems business unit. After completing her Ph.D., she transferred to the Integrated Photonics Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ. Her research has focused on the application of digital filter and signal processing techniques to optical filters for high-speed, high-capacity optical communication systems. She was promoted to Distinguished Member of the Technical Staff at Bell Laboratories in 2002. In 1998, she invented a class of tunable, multi-stage optical allpass filters that allow any phase response to be approximated and have application in chromatic dispersion compensation and polarization mode dispersion compensation. She holds 24 U.S. patents and has given over 70 technical talks and papers. She is now a Professor at Texas A&M University in College Station, TX.

Dr. Madsen has given a short course on "Optical Filters for WDM Systems: Theory, Technologies, and Applications" at the Optical Filter Conference and was the 2004 General Chair for the Integrated Photonics Research (IPR) Conference. She achieved Fellow ranking in the Optical Society of America in 2003.