Signal-to-Interference-plus-Noise Ratio Estimation for Wireless Communication Systems: Methods and Analysis

Daniel R. Jeske,[1] Ashwin Sampath[2]

[1] *Department of Statistics, University of California, Riverside, California 92521*

[2] *Wireless Terminal Business Unit, Texas Instruments, San Diego, California 92121*

Received 14 May 2002; revised 20 February 2004; accepted 23 March 2004
DOI 10.1002/nav.20022

Abstract: The Signal-to-Interference-plus-Noise Ratio (SINR) is an important metric of wireless communication link quality. SINR estimates have several important applications. These include optimizing the transmit power level for a target quality of service, assisting with handoff decisions and dynamically adapting the data rate for wireless Internet applications. Accurate SINR estimation provides for both a more efficient system and a higher user-perceived quality of service. In this paper, we develop new SINR estimators and compare their mean squared error (MSE) performance. We show that our new estimators dominate estimators that have previously appeared in the literature with respect to MSE. The sequence of transmitted bits in wireless communication systems consists of both pilot bits (which are known both to the transmitter and receiver) and user bits (which are known only by the transmitter). The SINR estimators we consider alternatively depend exclusively on pilot bits, exclusively on user bits, or simultaneously use both pilot and user bits. In addition, we consider estimators that utilize smoothing and feedback mechanisms. Smoothed estimators are motivated by the fact that the interference component of the SINR changes relatively slowly with time, typically with the addition or departure of a user to the system. Feedback estimators are motivated by the fact that receivers typically decode bits correctly with a very high probability, and therefore user bits can be thought of as quasipilot bits. For each estimator discussed, we derive an exact or approximate formula for its MSE. Satterthwaite approximations, noncentral F distributions (singly and doubly) and distribution theory of quadratic forms are the key statistical tools used in developing the MSE formulas. In the case of approximate MSE formulas, we validate their accuracy using simulation techniques. The approximate MSE formulas, of interest in their own right for comparing the quality of the estimators, are also used for optimally combining estimators. In particular, we derive optimal weights for linearly combining an estimator based on pilot bits with an estimator based on user bits. The optimal weights depend on the MSE of the two estimators being combined, and thus the accurate approximate MSE formulas can conveniently be used. The optimal weights also depend on the unknown SINR, and therefore need to be estimated in order to construct a useable combined estimator. The impact on the MSE of the combined estimator due to estimating the weights is examined. © 2004 Wiley Periodicals, Inc. Naval Research Logistics 51: 720–740, 2004.

Correspondence to: D.R. Jeske (daniel.jeske@ucr.edu); A. Sampath (asampath@ti.com)

Keywords: doubly noncentral F distribution; exponential smoothing; feedback loop estimation; optimal combined estimation; signal-to-noise ratio; Satterthwaite approximation

1. INTRODUCTION

In all communication systems, noise generated by circuit components in the receiver is a source of signal corruption. The ratio of the signal power to noise power, called the *signal-to-noise ratio* (SNR), is an important indicator of communication link quality. In wireless communications systems (particularly in mobile cellular communication systems) interference from other users in the system is a more significant source of signal corruption than the noise from circuitry. An adjusted indicator of link quality is the *signal-to-interference-plus-noise ratio* (SINR). Traditionally, a target link quality is characterized by a tolerable *bit error rate* that, in turn, maps to a required SINR. The SINR is then used to determine the transmitter power, which will deliver the target link quality. The SINR is also utilized by various control actions in wireless communication systems including handoff decisions and data rate adaptation algorithms (see, for example, Furuskar, Mazur, Muller, and Olofsson [4] and Viterbi [19]). We set the context for SINR estimation by first outlining the essential elements of a generic wireless communication system depicted in Figure 1.

An information source outputs an analog (e.g., voice) or digital signal (e.g., data) to a source encoder. The source encoder digitizes the signal, if needed, and typically performs bit compression. The resultant bits are passed to a channel encoder that introduces controlled redundancy (e.g., parity bits) to protect against channel errors. The modulator then maps fixed-size subsets of the bits into one of a finite number of amplitude and/or phase combinations that get impressed on a carrier waveform. The resultant carrier waveform is transmitted through the atmosphere (i.e., through a channel) that randomly attenuates the strength of the signal and shifts its phase. At the receiver, the demodulator is a *matched filter* that tries to recover the amplitude and/or the phase of the transmitted carrier waveform by "matching" the received signal to the known combinations of amplitude and/or phase that are being used by the modulator at the transmitter. The goal of the demodulator is to undo the random effects of the channel and predict the transmitted bits. It is at the output of the demodulator that SINR estimation is undertaken.

Generally, wireless transmission is organized by time slots wherein some pilot bits (i.e., overhead bits that are known *a priori* to both the transmitter and the receiver) are sent followed by user bits (i.e., information bits that are known to the transmitter only). Pilot bits allow the receiver to learn the channel attenuation/phase shift and then undo their effect from the user bits, and are also used in maintaining synchronization between the transmitter and the receiver. As the pilot bits are pure overhead, their number relative to the number of user bits in a time slot is small. With an additive white Gaussian noise (AWGN) model for the noise and interference (Proakis [10]), the output of the demodulator (measured in volts) corresponding to the jth bit of the ith time slot is

$$Y_{ij} = a_{ij}\mu_i + \epsilon_{ij}, \qquad i \geq 1, j = 1, \ldots, N,$$

where a_{ij} is -1 or $+1$ depending on whether the bit sent was 0 or 1, respectively (assuming Binary Phase Shift Keying), μ_i is an unknown constant during the ith time slot that is a function of the channel attenuation, ϵ_{ij} are independent and identically distributed Gaussian random variables with zero mean and variance σ_i^2, and N is the number of bits in a time slot. In cases

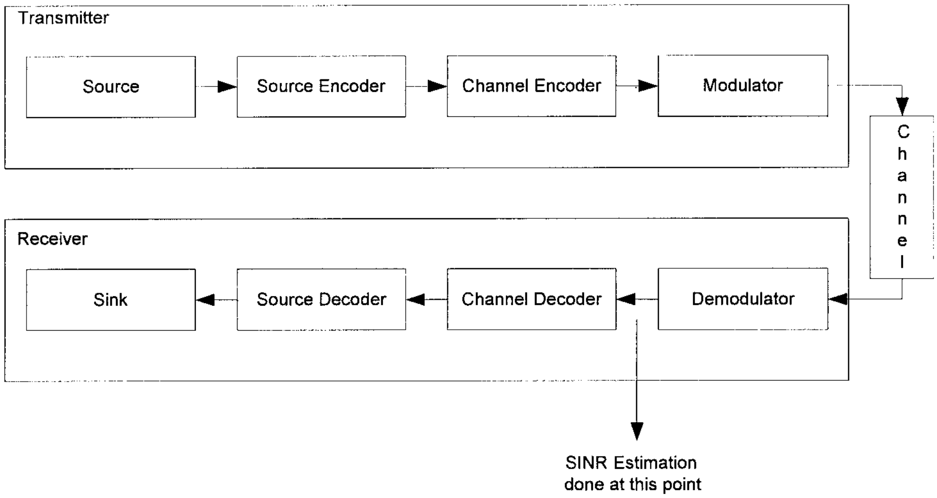

Figure 1. Generic wireless communication system.

of very fast fading, N may need to be small in order for the assumption that μ_i is constant during the time slot to hold. The ϵ_{ij} in the AWGN model represent the effect of noise introduced by the receiver as well as interference from other radio transmissions in the area. The Gaussian assumption for ϵ_{ij} will hold in CDMA systems provided the processing gain is sufficient.

The ratio $\theta_i = \mu_i^2/\sigma_i^2$ is the SINR during the ith time slot and is the parameter that we wish to estimate. For certain applications, such as handoff decisions and slow rate adaptation (e.g., GSM GPRS), SINR is alternatively defined using an average value of μ_i over several adjacent time slots. Estimation methods discussed in Brandao [1], Turkboylari and Stuber [18], and Ramakrishna et al. [11] are more suitable in those contexts than the methods that are discussed here. Applications where our estimation methods apply include power control and fast rate adaptation [e.g., systems such as High Speed Downlink Packet Access (HSDPA), Third Generation Evolution Data Only (3G1x-EVDO), and Third Generation Evolution Data and Voice (3G1x-EVDV)], where it is desired to track the short-term fading as much as possible.

The a_{ij} values in AWGN model are determined by the bit stream, and as such are known by the transmitter. The receiver uses the demodulator output value for each bit to predict a_{ij} and then map that value (-1 or $+1$) to a predicted bit value (0 or 1). The receiver knows the position of pilot bits within a time slot and it also knows that for pilot bits $a_{ij} \equiv 1$. Let n denote the number of pilot bits in a time slot (implying $m = N - n$ user bits), and let P_{ij} and U_{ij} denote the demodulator output values for pilot and user bits, respectively. It follows that, for $i \geq 1$ and $j = 1, \ldots, n$, P_{ij} has a Gaussian distribution with mean μ_i and variance σ_i^2. Moreover, the P_{ij} are mutually independent. The distribution of U_{ij} is a mixture of two Gaussian distributions that each have variance σ_i^2 but have respective means of $-\mu_i$ and μ_i, respectively, since a_{ij} in this case is a random variable taking on each of the values -1 and $+1$ with probability 0.5.

It is clear that the pilot bits provide an opportunity to estimate θ_i. In particular, letting $\bar{P}_i = \sum_{j=1}^{n} P_{ij}/n$ and $S_i^2 = \sum_{j=1}^{n} (P_{ij} - \bar{P}_i)^2/(n - 1)$, a "plug-in" (PI) estimator of θ_i based on pilot bits only, is $\hat{\theta}_i^{PI} = \bar{P}_i^2/S_i^2$. On the other hand, it is less obvious how to use the user bits to estimate θ_i since, for example, the expected value of U_{ij} is zero, not μ_i. While $\hat{\theta}_i^{PI}$ is an intuitive estimator, its drawback is that n is typically small with the consequence being that the mean squared error (MSE) of $\hat{\theta}_i^{PI}$ is too large for most applications. In Section 2 we consider estimation of θ_i based exclusively on pilot bits and show that $\hat{\theta}_i^{PI}$ is dominated in terms of MSE by its bias-corrected version. In Section 3 we consider estimation of θ_i based on user bits. We first review an ad hoc approach for using user bits that exists in the literature and then develop a novel estimator that uses the $\hat{a}_{ij}$ predictions as part of a feedback loop to improve upon $\hat{\theta}_i^{PI}$. In Section 4 we develop further improvements by combining estimators of θ_i that are based on pilot and user bits, respectively. In particular, we derive weights to use in the context of forming a weighted average of the two estimators. In Section 5 we discuss the use of variance-smoothing techniques as an alternative to effectively increase the number of bits available for estimating the SINR. Utilizing both variance-smoothing and combining, we present a very precise SINR estimator. We use MSE as the criteria for evaluating the alternative estimators and derive approximate expressions for the MSE for nearly all of the estimators proposed in this paper. Simulations are used to validate the assumptions underlying the approximation arguments. An overview of the simulation study is provided in an Appendix.

2. ESTIMATORS USING PILOT BITS

We first consider the MSE properties of $\hat{\theta}_i^{PI}$, some of which have been detailed in Thomas [16]. First, note that $n\bar{P}_i^2/S_i^2$ has a noncentral F distribution (see, for example, Searle [15], Section 2.4i) with numerator and denominator degrees of freedom equal to 1 and $n - 1$ respectively, and a noncentrality parameter equal to $\lambda_i = n\theta_i/2$. Using the formulas in Searle [15] for the mean and variance of the noncentral F distribution, it follows that

$$MSE(\hat{\theta}_i^{PI}) = \frac{2(n-1)^2}{n^2(n-3)}\left[\frac{(1+n\theta_i)^2}{(n-3)(n-5)} + \frac{1+2n\theta_i}{n-5}\right] + \left[\frac{n-1}{n-3}\left(\frac{1}{n}+\theta_i\right) - \theta_i\right]^2. \quad (1)$$

The estimator $\hat{\theta}_i^{PI}$ is a scaled version of the maximum likelihood estimator of SINR derived by Thomas [16]. A bias-corrected (BC) estimator of θ_i can be formed by linearly transforming $\hat{\theta}_i^{PI}$. Indeed,

$$\hat{\theta}_i^{BC} = \frac{\bar{P}_i^2}{\frac{n-1}{n-3}S_i^2} - \frac{1}{n}$$

removes a variable bias by adjusting the divisor used when estimating σ_i^2 and removes a constant bias by subtracting $1/n$. It follows that the MSE of $\hat{\theta}_i^{BC}$ is simply its variance and thus

$$MSE(\hat{\theta}_i^{BC}) = \frac{2(n-3)}{n^2}\left[\frac{(1+n\theta_i)^2}{(n-3)(n-5)} + \frac{1+n\theta_i}{n-5}\right]. \quad (2)$$

The Lehman-Scheffe theorem (see, for example, Graybill [6], Theorem 2.7.7) implies that $\hat{\theta}_i^{BC}$ is the uniformly minimum variance unbiased estimator (UMVUE) of θ_i when inference is restricted to using pilot bits from the ith time slot. The UMVUE property of $\hat{\theta}_i^{BC}$ was noted in Rukhin [13] in a different application context.

Figure 2 shows the Root Mean Squared Error (RMSE) of $\hat{\theta}_i^{PI}$ and $\hat{\theta}_i^{BC}$ for θ_i in the range -2 dB to 10 dB [SINR is generally expressed in decibels (dB), defined as $10\log_{10}(\theta_i)$] labeled as PI (A) and BC (A), respectively, for the case $n = 8$. The range -2 dB to 10 dB is a practical

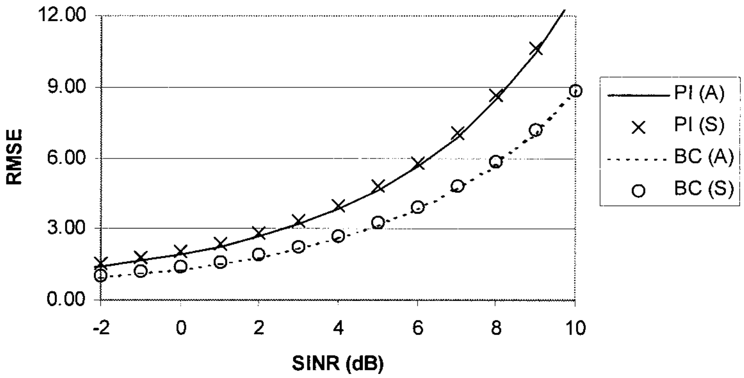

Figure 2. RMSE of pilot-based estimators.

range of SINR values for most wireless communication systems (see, for example, Viterbi [19]). The improvement offered by $\hat{\theta}_i^{BC}$ compared to $\hat{\theta}_i^{PI}$ is quite evident. Also shown in Figure 2 are the RMSE values for $\hat{\theta}_i^{PI}$ and $\hat{\theta}_i^{BC}$ obtained by simulating 50,000 values of each estimator. These values are labeled as PI (S) and BC (S), respectively. Although the simulated RMSE values in Figure 2 are not necessary, since the formulas in (2) and (3) are exact, the good match between the simulated and exact RMSE values suggests that 50,000 will be a sufficient sample size to validate the approximate RMSE formulas in what follows.

3. ESTIMATORS USING USER BITS

3.1. Absolute Value Estimators

Typically there are far fewer pilot bits than there are user bits. As we have previously mentioned, the difficulty with using user bits is that the expected value of U_{ij} is zero instead of μ_i. Hence, it is not possible to estimate θ_i using the U_{ij} values directly. An ad-hoc approach for using user bits, first proposed by Gilchriest [5], is to estimate θ_i by computing either $\hat{\theta}_i^{PI}$ or $\hat{\theta}_i^{BC}$ using $Z_{ij} = |U_{ij}|$ in place of P_{ij}.

The ad hoc approach for using the user bits can be heuristically motivated as follows. Suppose θ_i is large, implying μ_i is large relative to σ_i. Given that $a_{ij} = -1$, $U_{ij} < 0$ with a high probability. It follows that $Z_{ij} \approx -U_{ij}$. Since the distribution of U_{ij}, given $a_{ij} = -1$, is Gaussian with mean $-\mu_i$ and variance σ_i^2, it follows that the conditional distribution of Z_{ij}, given $a_{ij} = -1$, is approximately Gaussian with mean μ_i and variance σ_i^2. Similarly, given that $a_{ij} = 1$, $U_{ij} > 0$ with a high probability and thus $Z_{ij} \approx U_{ij}$. Since the distribution of U_{ij}, given $a_{ij} = 1$, is Gaussian with mean μ_i and variance σ_i^2, the conditional distribution of Z_{ij}, given $a_{ij} = 1$, is approximately Gaussian with mean μ_i and variance σ_i^2. Since the conditional distributions of Z_{ij}, given $a_{ij} = 1$ and given $a_{ij} = -1$, are the same and there are no other possible values for a_{ij}, it follows the unconditional distribution for Z_{ij} is approximately Gaussian with mean μ_i and variance σ_i^2. The ad hoc approach for estimating θ_i relies on the fact that the Z_{ij} values are roughly stochastically equivalent to P_{ij} values. Clearly, for small or even intermediate values of θ_i the Z_{ij} values will have a substantially different distribution than P_{ij} values, and the ad-hoc approach estimators will not have good MSE properties.

Define $\hat{\theta}_i^{PI-Z} = \bar{Z}_i^2/T_i^2$, where $\bar{Z}_i = \sum_{j=1}^{m} Z_{ij}/m$, $T_i^2 = \sum_{j=1}^{m} (Z_{ij} - \bar{Z}_i)^2/(m - 1)$. Layland [9] gives large m approximations for the mean and variance of $\hat{\theta}_i^{PI-Z}$. Let $\Phi(\cdot)$ denote the cumulative distribution function of the Gaussian distribution that has mean 0 and variance 1. The mean and variance of Z_{ij} are $\mu_{Z,i} = \sigma_i\sqrt{2/\pi}e^{-\mu_i/(2\sigma_i^2)} + 2\mu_i\Phi(\mu_i/\sigma_i) - \mu_i$ and $\sigma^2_{Z,i} = \mu_i^2 + \sigma_i^2 - \mu^2_{Z,i}$ (see, for example, Johnson, Kotz, and Balakrishnan [8], pp. 453–454). We use the central limit theorem to approximate the distribution of $\sqrt{m}\bar{Z}_i/\sigma_{Z,i}$ by a Gaussian distribution with mean $\sqrt{m\theta_{Z,i}}$ and variance 1, where $\theta_{Z,i} = \mu^2_{Z,i}/\sigma^2_{Z,i}$. It follows that $V_i^Z = m\bar{Z}_i^2/\sigma^2_{Z,i}$ has an approximate noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter $\lambda_i^Z = m\theta_{Z,i}/2$. We use a Satterthwaite [14] chi-square approximation for the distribution of $W_i^Z = (m-1)T_i^2/\sigma^2_{Z,i}$, writing $W_i^Z \sim g_i\chi^2_{\eta_i}$, where $g_i \equiv g(\theta_i)$ and $\eta_i \equiv \eta(\theta_i)$ are the scale and degree-of-freedom constants developed in the Appendix A.

If we assume that $\bar{Z}_i$ and T_i^2 are independently distributed, then it would follow that $m\bar{Z}_i^2/T_i^2$ has an approximate noncentral F distribution with numerator and denominator degrees of freedom equal to 1 and η_i, respectively, and noncentrality parameter equal to λ_i^Z. In general, the asymmetry of the distribution of Z_{ij} prevents $\bar{Z}_i$ and T_i^2 from being uncorrelated (see, for example, Randles and Wolfe [12], p. 24, Corollary 1.3.33), and therefore they will not be independent. We do note that for large θ_i, $\bar{Z}_i$, and T_i^2 will be approximately independent since the distribution of Z_{ij} is then approximately Gaussian. We absorb the error in assuming $\bar{Z}_i$ and T_i^2 are independent as part of the approximate noncentral F distribution we use for $m\bar{Z}_i^2/T_i^2$. It follows that

$$MSE(\hat{\theta}_i^{PI-Z}) \doteq \frac{2\eta_i^2}{m^2(\eta_i-2)}\left[\frac{(1+m\theta_{Z,i})^2}{(\eta_i-2)(\eta_i-4)} + \frac{1+2m\theta_{Z,i}}{\eta_i-4}\right] + \left[\frac{\eta_i}{\eta_i-2}\left(\frac{1}{m}+\theta_{Z,i}\right) - \theta_i\right]^2. \tag{3}$$

It follows from results in Appendix A that as θ_i gets large, $\eta_i \to m-1$ and $\theta_{Z,i} \to \theta_i$. Thus, the squared bias term in (3) suggests a bias corrected estimator of the form $\hat{\theta}_i^{BC-Z} = (m-3)\hat{\theta}_i^{PI-Z}/(m-1) - 1/m$, which is

$$MSE(\hat{\theta}_i^{BC-Z}) \doteq \frac{2(m-3)^2\eta_i^2}{m^2(m-1)^2(\eta_i-2)}\left[\frac{(1+m\theta_{Z,i})^2}{(\eta_i-2)(\eta_i-4)} + \frac{1+2m\theta_{Z,i}}{\eta_i-4}\right] + \left[\frac{m-3}{m-1}\frac{\eta_i}{\eta_i-2}\left(\frac{1}{m}+\theta_{Z,i}\right) - \frac{1}{m} - \theta_i\right]^2. \tag{4}$$

Figure 3 shows the analytic approximations for the RMSE of $\hat{\theta}_i^{PI-Z}$ and $\hat{\theta}_i^{BC-Z}$, labeled as PI-Z (A) and BC-Z (A), respectively, for the case $m = 20$. The curves, labeled PI-Z (S) and BC-Z (S), are the respective RMSE curves obtained via simulation. The nonmonotone nature of the curves in Figure 3 over the interval -2 dB to 6 dB is explained by the fact that initially the decreasing squared bias component of MSE offsets the increasing variance component, but eventually the bias becomes near zero and the MSE becomes essentially equal to the variance. From Figure 3 we conclude that the analytic RMSE approximations (3) and (4) for $\hat{\theta}_i^{PI-Z}$ and $\hat{\theta}_i^{BC-Z}$, respectively, are quite good when the SINR is greater than or equal to 5 dB. Moreover, by comparing the ordinate scales of Figures 2 and 3, we see that the RMSE values of both $\hat{\theta}_i^{PI-Z}$ and $\hat{\theta}_i^{BC-Z}$ are considerably smaller than the RMSE values of $\hat{\theta}_i^{BC}$ for the companion case $n = 8$.

3.2. Feedback Estimators

In this approach, we take advantage of the fact that the receivers (in typical operation) make correct predictions of the a_{ij} values with a high probability. Let $\hat{a}_{ij}$ denote the receiver's

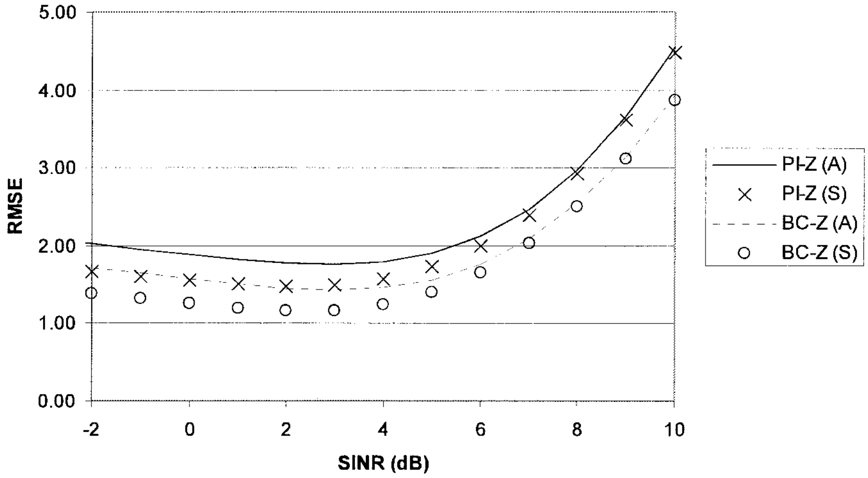

Figure 3. RMSE of absolute value estimators.

predicted value of a_{ij}. Using a likelihood ratio test, the value of $\hat{a}_{ij}$ is $+1$ and -1 when U_{ij} is greater and less than zero, respectively. Define $D_{ij} = \hat{a}_{ij}U_{ij} = a^*_{ij}\mu_i + \epsilon^*_{ij}$, where $a^*_{ij} = \hat{a}_{ij}a_{ij}$ and $\epsilon^*_{ij} = \hat{a}_{ij}\epsilon_{ij}$. Since both $\hat{a}_{ij}$ and a_{ij} are $\{-1, +1\}$ random variables, a^*_{ij} is also a $\{-1, +1\}$ random variable and is equal to $+1$ if and only if $\hat{a}_{ij} = a_{ij}$. It is straightforward to show that $\Pr(a^*_{ij} = 1) \equiv p(\theta_i) = \Phi(\sqrt{\theta_i})$.

We now derive the conditional distribution of ϵ^*_{ij}, given a^*_{ij}. First note that D_{ij} is nonnegative. Thus, conditional on a^*_{ij} we must have $\epsilon^*_{ij} > -a^*_{ij}\mu_i$. Consider first the case where $a^*_{ij} = 1$. We have

$$\Pr(\epsilon^*_{ij} < x | a^*_{ij} = 1) = \frac{\Pr(\epsilon^*_{ij} < x, \hat{a}_{ij} = -1, a_{ij} = -1) + \Pr(\epsilon^*_{ij} < x, \hat{a}_{ij} = 1, a_{ij} = 1)}{\Pr(\hat{a}_{ij} = -1, a_{ij} = -1) + \Pr(\hat{a}_{ij} = 1, a_{ij} = 1)}$$

Substituting $U_{ij} < 0$ ($U_{ij} > 0$) for $\hat{a}_{ij} = -1$ ($\hat{a}_{ij} = 1$), and using $U_{ij} = a_{ij}\mu_i + \epsilon_{ij}$ leads to

$$\Pr(\epsilon^*_{ij} < x | a^*_{ij} = 1) = \frac{\Pr(-x < \epsilon_{ij} < \mu_i, a_{ij} = -1) + \Pr(-\mu_i < \epsilon_{ij} < x, a_{ij} = 1)}{\Pr(\epsilon_{ij} < \mu_i, a_{ij} = -1) + \Pr(\epsilon_{ij} > -\mu_i, a_{ij} = 1)}$$

$$= \frac{\Phi\left(\dfrac{x}{\sigma_i}\right) - [1 - \Phi(\sqrt{\theta_i})]}{\Phi(\sqrt{\theta_i})}, \qquad x \geq -\mu_i.$$

It follows that the conditional distribution of ϵ^*_{ij}, given $a^*_{ij} = 1$, is a zero mean, σ_i^2 variance Gaussian distribution truncated at the point $-\mu_i$. In a similar way, it can be shown that

$$\Pr(\epsilon^*_{ij} < x | a^*_{ij} = -1) = \frac{\Phi\left(\dfrac{x}{\sigma_i}\right) - \Phi(\sqrt{\theta_i})}{1 - \Phi(\sqrt{\theta_i})}, \qquad x \geq \mu_i,$$

and thus the conditional distribution of ϵ^*_{ij}, given $a^*_{ij} = -1$ is a zero mean, σ_i^2 variance Gaussian distribution truncated at the point μ_i. For $|\mu_i| > 2$ the conditional distributions of ϵ^*_{ij}, given a^*_{ij}, will each be close to a zero mean σ_i^2 variance Gaussian distribution. Let $\boldsymbol{D}_i$ be the $m \times 1$ vector of $\{D_{ij}\}_{j=1}^m$ values and $\boldsymbol{a}^*_i$ be the $m \times 1$ vector of $\{a^*_{ij}\}_{j=1}^m$ values. Conditional on $\boldsymbol{a}^*_i$, we approximate the distribution of $\boldsymbol{D}_i$ as multivariate Gaussian with mean vector $\boldsymbol{a}^*_i\mu_i$ and variance–covariance matrix $\sigma_i^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the $m \times m$ identity matrix. We utilize this approximation in what follows.

Define $V_i^F = m\bar{D}_i^2/\sigma_i^2$. Noting that $V_i^F = \boldsymbol{D}_i'\boldsymbol{A}\boldsymbol{D}_i$, where $\boldsymbol{A} = \boldsymbol{J}/(m\sigma_i^2)$ and $\boldsymbol{J}$ is an $m \times m$ matrix of ones, it follows from Searle (see [15], p. 57, Theorem 2) that, conditional on $\boldsymbol{a}^*_i$, V_i^F has an approximate noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter $\lambda_{1i} = \theta_i(\sum_{j=1}^m a^*_{ij})^2/(2m)$. Next let $U_i^2 = \sum_{j=1}^m (D_{ij} - \bar{D}_i)^2/(m-1)$ and define $W_i^F = (m-1)U_i^2/\sigma_i^2$. Noting that $W_i^F = \boldsymbol{D}_i'\boldsymbol{B}\boldsymbol{D}_i$, where $\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{J}/m)/\sigma_i^2$, it similarly follows that, conditional on $\boldsymbol{a}^*_i$, W_i^F has an approximate noncentral chi-square distribution with $m-1$ degrees of freedom and noncentrality parameter $\lambda_{2i} = \theta_i \sum_{j=1}^m (a^*_{ij} - \bar{a}^*_i)^2/2$. Moreover, since $\boldsymbol{AB} = \underset{\sim}{0}$, V_i^F and W_i^F are independently distributed (see, for example, Searle [15], p. 59, Theorem 3). Define $\hat{\theta}_i^{PI-F} = \bar{D}_i^2/U_i^2$. It follows that, conditional on $\boldsymbol{a}^*_i$,

$$F_i^F = \frac{V_i^F}{\left(\dfrac{W_i^F}{m-1}\right)} = m\hat{\theta}_i^{PI-F}$$

has an approximate doubly noncentral F distribution (see, for example, Tiku [17]) with numerator and denominator degrees of freedom equal to 1 and $m-1$, respectively, and numerator and denominator noncentrality parameters equal to λ_{1i} and λ_{2i}, respectively. Using results from Tiku [17], we have

$$E(\hat{\theta}_i^{PI-F}|\underset{\sim}{a}^*_i) \doteq \frac{1}{m}\frac{m-1}{m-3}\frac{1+2\lambda_{1i}}{1+\left(\dfrac{2\lambda_{2i}}{m-1}\right)}$$

$$Var(\hat{\theta}_i^{PI-F}|\underset{\sim}{a}^*_i) \doteq \frac{2}{m^2}\frac{(m-1)^2}{m-3}\left[\frac{(1+2\lambda_{1i})^2}{(m-3)(m-5)} + \frac{1+4\lambda_{1i}}{m-5}\right]\left(1+\frac{2\lambda_{2i}}{m-1}\right)^{-2}.$$

Since a^*_{ij} is a $\{-1, +1\}$ random variable, it is easy to verify that λ_{1i} and λ_{2i} can equivalently be expressed as $\lambda_{1i} = \theta_i(2N_i - m)^2/(2m)$ and $\lambda_{2i} = 2\theta_i N_i(m - N_i)/m$, where N_i is the number of $\{a^*_{ij}\}_{j=1}^m$ values that are equal to $+1$. Consequently, the unconditional mean of $\hat{\theta}_i^{PI-F}$ can be approximated by

$$E(\hat{\theta}_i^{PI-F}) \doteq \frac{1}{m}\frac{m-1}{m-3}E\left\{\frac{1+2\lambda_{1i}}{1+2\lambda_{2i}/(m-1)}\right\}, \tag{5}$$

where the expectation on the right-hand side is with respect to the distribution of N_i which is binomial with parameters m and $p(\theta_i)$. Similarly, the unconditional variance of $\hat{\theta}_i^{PI-F}$ can be approximated by

$$Var(\hat{\theta}_i^{PI-F}) \doteq \frac{2}{m^2}\frac{(m-1)^2}{m-3} E\left\{\left[\frac{(1+2\lambda_{1i})^2}{(m-3)(m-5)} + \frac{1+4\lambda_{1i}}{m-5}\right]\left(1+\frac{2\lambda_{2i}}{m-1}\right)^{-2}\right\} + \frac{(m-1)^2}{m^2(m-3)^2} Var\left\{\frac{1+2\lambda_{1i}}{1+2\lambda_{2i}/(m-1)}\right\}, \quad (6)$$

where both the mean and variance operators on the right-hand side are with respect to the distribution of N_i. Equations (5) and (6) together provide the components needed to evaluate the MSE of $\hat{\theta}_i^{PI-F}$.

An approximate bias correction for $\hat{\theta}_i^{PI-F}$ can be motivated from the special case where $p(\theta_i)$ is close to unity and therefore N_i is stochastically close to m. In the special case where $N_i \equiv m$, λ_{1i} and λ_{2i} reduce to $m\theta_i/2$ and zero, respectively. It follows from (5) that

$$E(\hat{\theta}_i^{PI-F}) \doteq \frac{1}{m}\frac{m-1}{m-3}(1+m\theta_i), \qquad \text{for } p(\theta_i) \doteq 1,$$

which suggests an approximate bias-corrected version of $\hat{\theta}_i^{PI-F}$ is $\hat{\theta}_i^{BC-F} = (m-3)\hat{\theta}_i^{PI-F}/(m-1) - 1/m$. It follows from (5) and (6) that

$$E(\hat{\theta}_i^{BC-F}) \doteq \frac{1}{m}\left[E\left\{\frac{1+2\lambda_{1i}}{1+\left(\frac{2\lambda_{2i}}{m-1}\right)}\right\} - 1\right], \quad (7)$$

$$Var(\hat{\theta}_i^{BC-F}) \doteq \frac{2(m-3)}{m^2} E\left\{\left[\frac{(1+2\lambda_{2i})^2}{(m-3)(m-5)} + \frac{1+4\lambda_{2i}}{m-5}\right]\left(1+\frac{2\lambda_{2i}}{m-1}\right)^{-2}\right\} + \frac{1}{m^2} Var\left\{\frac{1+2\lambda_{1i}}{1+\left(\frac{2\lambda_{2i}}{m-1}\right)}\right\}. \quad (8)$$

Equations (7) and (8) together provide the components needed to evaluate the MSE of $\hat{\theta}_i^{BC-F}$. Figure 4 shows the analytic approximations (for the case $m = 20$) of the RMSE of $\hat{\theta}_i^{PI-F}$ and $\hat{\theta}_i^{BC-F}$, denoted as PI-F (A) and BC-F (A), respectively. The simulation values of RMSE for $\hat{\theta}_i^{PI-F}$ and $\hat{\theta}_i^{BC-F}$ are not shown as they virtually coincide with the analytic approximations. Also shown in Figure 4 is the previously discussed analytic approximation for the RMSE of $\hat{\theta}_i^{BC-Z}$, denoted as BC-Z (A). We conclude from Figure 4 that, for SINR values less than 4 dB, both $\hat{\theta}_i^{PI-F}$ and $\hat{\theta}_i^{BC-F}$ have smaller RMSE than $\hat{\theta}_i^{BC-Z}$, a consequence of the fact that significant bias exists in $\hat{\theta}_i^{BC-Z}$ when SINR is less than 4 dB. For SINR greater than 4 dB, the absolute value estimator is (slightly) better than both of the feedback estimators, reflecting the fact that the bias in $\hat{\theta}_i^{BC-Z}$ dissipates more quickly than the effect of incorrect bit decisions in $\hat{\theta}_i^{PI-F}$ and $\hat{\theta}_i^{BC-F}$.

4. ESTIMATORS USING PILOT AND USER BITS

The model for the demodulator output discussed in Section 1 implies that an estimator of θ_i based on pilot bits will be statistically independent of an estimator of θ_i based on user bits. The

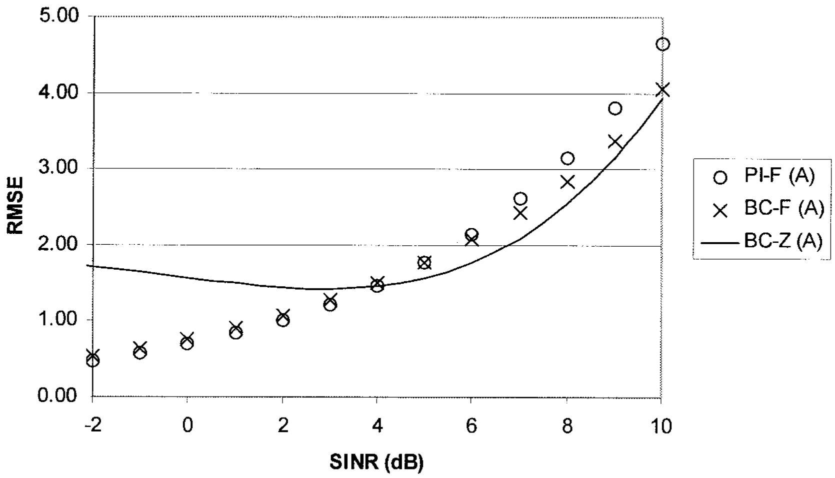

Figure 4. RMSE of feedback estimators.

following lemma shows how to linearly combine two estimators, based on pilot bits and user bits respectively, to yield a new estimator that has smaller MSE.

LEMMA 1: Suppose T_1 and T_2 are independent estimators of a parameter δ with biases B_1 and B_2, respectively, and variances σ_1^2 and σ_2^2, respectively. Consider the class C of estimators of δ which have the form $T(\alpha_1, \alpha_2) = \alpha_1 T_1 + \alpha_2 T_2$, where the weights α_1 and α_2 are arbitrary real numbers. The weights of the estimator in C that has the smallest MSE are

$$\alpha_i^{opt} = \frac{\left(\dfrac{\delta + B_i}{\sigma_i}\right)^2}{\left(1 + \dfrac{B_i}{\delta}\right)\left[1 + \left(\dfrac{\delta + B_1}{\sigma_1}\right)^2 + \left(\dfrac{\delta + B_2}{\sigma_2}\right)^2\right]} \qquad (i = 1, 2),$$

and the resulting MSE of $T(\alpha_1^{opt}, \alpha_2^{opt})$ is

$$MSE[T(\alpha_1^{opt}, \alpha_2^{opt})] = \frac{\delta^2}{1 + \left(\dfrac{\delta + B_1}{\sigma_1}\right)^2 + \left(\dfrac{\delta + B_2}{\sigma_2}\right)^2}.$$

PROOF OF LEMMA 1: The MSE of an arbitrary estimator of the form $T(\alpha_1, \alpha_2) = \alpha_1 T_1 + \alpha_2 T_2$ is $MSE(T) = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + [\alpha_1(\delta + B_1) + \alpha_2(\delta + B_2) - \delta]^2$. It follows that $\partial\{MSE(T)\}/\partial\alpha_i = 2\alpha_i\sigma_i^2 + 2[\alpha_1(\delta + B_1) + \alpha_2(\delta + B_2) - \delta](\delta + B_i)$, for $i = 1, 2$. Setting the two partial derivative equations equal to zero and solving for α_1 and α_2 gives the optimal weights and substituting them into the expression for $MSE(T)$ gives the minimum MSE value.

The class of estimators defined in Lemma 1 allow arbitrary weights on T_1 and T_2. If instead we impose the constraint that $\alpha_1 + \alpha_2 = 1$, the estimator $T(\alpha_1, \alpha_2) = \alpha_1 T_1 + (1 - \alpha_1)T_2$

becomes a weighted average of T_1 and T_2. If both T_1 and T_2 are unbiased then $T(\alpha_1, \alpha_2)$ will also be unbiased.

LEMMA 2: Suppose T_1 and T_2 are independent estimators of a parameter δ with biases B_1 and B_2, respectively, and variances σ_1^2 and σ_2^2, respectively. Consider the class C^* of estimators of δ which has the form $T(\alpha) = \alpha T_1 + (1 - \alpha)T_2$, where $0 \leq \alpha \leq 1$. The weight α of the estimator in C^* that has the smallest MSE is

$$\alpha^{opt} = \frac{MSE_2 - B_1B_2}{MSE_1 + MSE_2 - 2B_1B_2},$$

where $MSE_i = \sigma_i^2 + B_i^2$, and the resulting MSE of $T(\alpha^{opt})$ is

$$MSE[T(\alpha^{opt})] = \frac{MSE_1MSE_2 - B_1^2B_2^2}{MSE_1 + MSE_2 - 2B_1B_2}.$$

The proof of Lemma 2 is similar to the proof of Lemma 1 and is thus omitted.

COROLLARY 1: If one of the estimators, say T_1, in Lemma 2 is unbiased, then

$$\alpha^{opt} = \frac{MSE_2}{MSE_1 + MSE_2} = \frac{MSE_1^{-1}}{MSE_1^{-1} + MSE_2^{-1}}$$

and

$$MSE[T(\alpha^{opt})] = \frac{MSE_1MSE_2}{MSE_1 + MSE_2}.$$

Note that the optimal weight in Corollary 1 can equivalently be expressed as $\alpha^{opt} = \sigma_1^{-2}/(\sigma_1^{-2} + MSE_2^{-1})$, and when both estimators are unbiased, $\alpha^{opt} = \sigma_1^{-2}/(\sigma_1^{-2} + \sigma_2^{-1})$, which is a well-known result (see, for example, Christensen [2], Theorem 2.7.1).

We now apply Corollary 1 using $\hat{\theta}_i^{BC}$ and $\hat{\theta}_i^{BC-Z}$ as T_1 and T_2, respectively. The condition $B_1 = 0$ is met. Define

$$\hat{\theta}_i^{C1} = \alpha^{opt}(\theta_i)\hat{\theta}_i^{BC} + [1 - \alpha^{opt}(\theta_i)]\hat{\theta}_i^{BC-Z},$$

where

$$\alpha^{opt}(\theta_i) = \frac{MSE(\hat{\theta}_i^{BC-Z})}{MSE(\hat{\theta}_i^{BC}) + MSE(\hat{\theta}_i^{BC-Z})}$$

with $MSE(\hat{\theta}_i^{BC})$ and $MSE(\hat{\theta}_i^{BC-Z})$ given by (2) and (4), respectively. It follows from Corollary 1 that

$$MSE(\hat{\theta}_i^{C1}) = \frac{MSE(\hat{\theta}_i^{BC})MSE(\hat{\theta}_i^{BC-Z})}{MSE(\hat{\theta}_i^{BC}) + MSE(\hat{\theta}_i^{BC-Z})}.$$

An alternative way to combine $\hat{\theta}_i^{BC}$ and $\hat{\theta}_i^{BC-Z}$ is to use Lemma 1 to form the estimator $\hat{\theta}_i^{C2} = \alpha_1^{opt}(\theta_i)\hat{\theta}_i^{BC} + \alpha_2^{opt}(\theta_i)\hat{\theta}_i^{BC-Z}$. Since $\hat{\theta}_i^{BC}$ is unbiased, $B_1 = 0$. The second term in (4) is B_2^2, and thus

$$\alpha_1^{opt}(\theta_i) = \frac{\dfrac{\theta_i^2}{Var(\hat{\theta}_i^{BC})}}{1 + \dfrac{\theta_i^2}{Var(\hat{\theta}_i^{BC})} + \dfrac{(\theta_i + B_2)^2}{Var(\hat{\theta}_i^{BC-Z})}},$$

$$\alpha_2^{opt}(\theta_i) = \frac{\dfrac{(\theta_i + B_2)^2}{Var(\hat{\theta}_i^{BC-Z})}}{\left(1 + \dfrac{B_2}{\theta_i}\right)\left[1 + \dfrac{\theta_i^2}{Var(\hat{\theta}_i^{BC})} + \dfrac{(\theta_i + B_2)^2}{Var(\hat{\theta}_i^{BC-Z})}\right]}$$

and the MSE of $\hat{\theta}_i^{C2}$ is

$$MSE(\hat{\theta}_i^{C2}) = \frac{\theta_i^2}{1 + \dfrac{\theta_i^2}{Var(\hat{\theta}_i^{BC})} + \dfrac{(\theta_i + B_2)^2}{Var(\hat{\theta}_i^{BC-Z})}}.$$

Since both $\hat{\theta}_i^{C1}$ and $\hat{\theta}_i^{C2}$ depend on θ_i through the optimal weights, neither is directly useable. A practical approach is to estimate the optimal weights using an unbiased estimator such as $\hat{\theta}_i^{BC}$, for example. We thus define

$$\hat{\theta}_i^{EC1} = \alpha^{opt}(\hat{\theta}_i^{BC})\hat{\theta}_i^{BC} + [1 - \alpha^{opt}(\hat{\theta}_i^{BC})]\hat{\theta}_i^{BC-Z},$$

$$\hat{\theta}_i^{EC2} = \alpha_1^{opt}(\hat{\theta}_i^{BC})\hat{\theta}_i^{BC} + \alpha_2^{opt}(\hat{\theta}_i^{BC})\hat{\theta}_i^{BC-Z}.$$

Analytic approximations for the MSE of $\hat{\theta}_i^{EC1}$ and $\hat{\theta}_i^{EC2}$ are largely intractable, but not particularly necessary since we will not be defining other estimators that depend on these expressions.

Figure 5 shows the optimal constrained (solid curves) and unconstrained weights (dashed curves), as a function of SINR, for the combined estimators obtained from $\hat{\theta}_i^{BC}$ and $\hat{\theta}_i^{BC-Z}$ for the case of $n = 8$ and $m = 20$. In the case of constrained weights, Figure 5 shows that for small SINR (where the bias in $\hat{\theta}_i^{BC-Z}$ is significant) most of the weight is given to $\hat{\theta}_i^{BC}$. On the other hand, for intermediate to large values of SINR, where the bias in $\hat{\theta}_i^{BC-Z}$ has become small, most of the weight is given to $\hat{\theta}_i^{BC-Z}$ since it is based on a larger sample size. In the case of unconstrained weights, there is no crossover point. For small SINR values, both unconstrained weights are small which makes the combined estimator relatively small. As the SINR increases, the weight on $\hat{\theta}_i^{BC-Z}$ increases more rapidly than the weight on $\hat{\theta}_i^{BC}$.

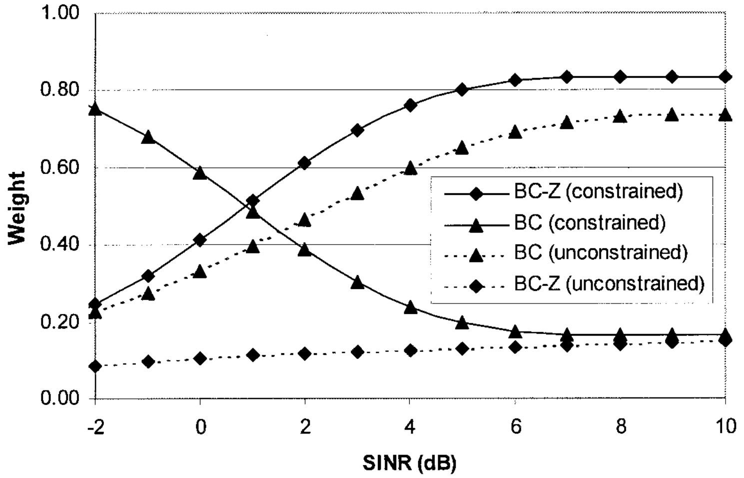

Figure 5. Ideal constrained and unconstrained weights for combining the BC and BC-Z estimators.

Figure 6 shows the simulated RMSE values of $\hat{\theta}_i^{C1}$ and $\hat{\theta}_i^{C2}$ labeled as C1 (S) and C2 (S), and the analytic RMSE approximations labeled as C1 (A) and C2 (A), respectively. The approximate values match the simulated values quite well. By comparing the ordinate scales of Figures 2 and 3 to the same in Figure 6, it is clear that the combined estimator offers significant reductions in RMSE compared to the individual estimators, $\hat{\theta}_i^{BC}$ and $\hat{\theta}_i^{BC-Z}$.

Figure 7 shows the effect of estimating the optimal weights by contrasting the simulated RMSE values of $\hat{\theta}_i^{C1}$ and $\hat{\theta}_i^{C2}$ with the simulated RMSE values of $\hat{\theta}_i^{EC1}$ and $\hat{\theta}_i^{EC2}$, which are labeled as EC1 (S) and EC2 (S), respectively. It is evident from Figure 7 that the cost of estimating the unconstrained weights is appreciably more than the cost of estimating the constrained weights. In fact, although the RMSE of $\hat{\theta}_i^{C2}$ is significantly smaller than the RMSE of $\hat{\theta}_i^{C1}$, the RMSE of $\hat{\theta}_i^{EC2}$ and $\hat{\theta}_i^{EC1}$ are virtually the same. Once again, comparing the ordinate scales of Figures 2 and 3 to the same in Figure 7 demonstrates that the cost of estimating the weights does not diminish the value of using combined estimators.

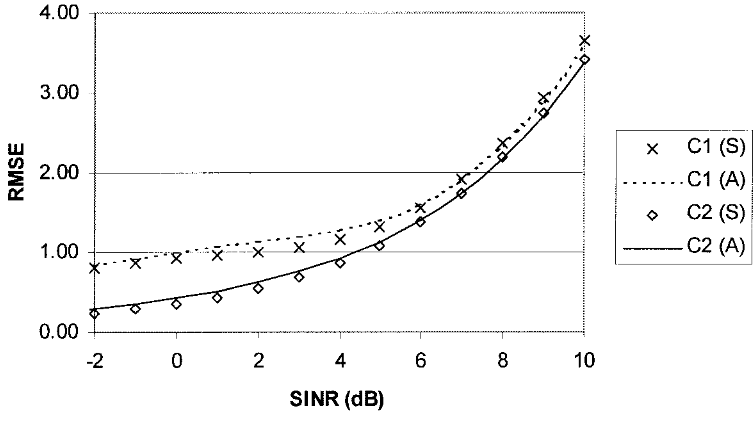

Figure 6. RMSE of combined BC and BC-Z estimators (known weights).

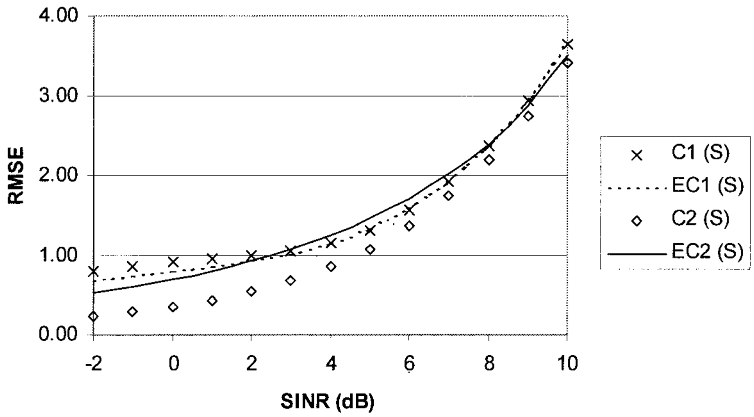

Figure 7. RMSE of the combined BC and BC-Z estimators.

5. SMOOTHED VARIANCE ESTIMATORS

5.1. Pilot-Based Estimator

The AWGN model for the demodulator output discussed in Section 1 shows that the variance of the noise factor, σ_i^2, varies from time slot to time slot. It has been recognized that σ_i^2 is a slowly varying function of i (see Viterbi [19]). The average noise plus interference component, for example, would only change with the addition/departure of a call in a CDMA system, or when a discernible "rearrangement" of users has taken place due to mobility. The slowly varying nature of σ_i^2 suggests using exponential weighting of the within time slot estimates of the noise variance. In particular, we could consider an estimator of σ_i^2 of the form

$$\hat{\sigma}_i^2 = (1 - r)^{i-1} S_1^2 + \sum_{k=2}^{i} r(1 - r)^{i-k} S_k^2,$$

where $0 < r \leq 1$. An alternative recursive form of $\hat{\sigma}_i^2$ is $(i \geq 2)$

$$\hat{\sigma}_i^2 = r S_i^2 + (1 - r)\hat{\sigma}_{i-1}^2.$$

Note that $r = 1$ delivers $\hat{\sigma}_i^2 = S_i^2$. Higuchi [7] smoothes variance estimates calculated using both pilot and user bits. Analogous to $\hat{\theta}_i^{PI}$, we could consider an estimator of θ_i of the form $\hat{\theta}_i^{SV} = \bar{P}_i^2/\hat{\sigma}_i^2$, which, of course, reduces to $\hat{\theta}_i^{PI}$ when $r = 1$. Furthermore, we could consider scaled versions of $\hat{\theta}_i^{SV}$ to reduce bias.

The MSE properties of $\hat{\theta}_i^{SV}$ depend on the nature of the $\{\sigma_k^2\}$ sequence. An interesting class of models for the $\{\sigma_k^2\}$ sequence would be an appropriate family of stationary stochastic processes (e.g., ARMA models). The random nature of $\{\sigma_k^2\}$ implies the $\{\theta_k\}$ sequence is also random. The inference problem shifts from estimating the parameter θ_i to predicting the realized value of the random variable θ_i. The MSE of $\hat{\theta}_i^{SV}$ would then be evaluated as $E(\hat{\theta}_i^{SV} - \theta_i)^2$, where the expectation is with respect to the joint distribution of $\hat{\theta}_i^{SV}$ and θ_i. We defer the evaluation of the MSE of $\hat{\theta}_i^{SV}$ in the context of stochastic models for $\{\sigma_k^2\}$ to a future paper and consider here only the MSE properties of $\hat{\theta}_i^{SV}$ in the special case where $\sigma_k^2 \equiv \sigma^2$. The MSE results for the special case where $\sigma_k^2 \equiv \sigma^2$ are approximations to the case where $\{\sigma_k^2\}$ is a slowly varying sequence that has a mean equal to σ^2.

Define $W = (n - 1)\hat{\sigma}_i^2/\sigma^2 = \sum_{k=1}^{i} w_k(n - 1)S_k^2/\sigma^2$, where $w_1 = (1 - r)^{i-1}$ and $w_k = r(1 - r)^{i-k}$, for $2 \le k \le i$. Since the terms $(n - 1)S_k^2/\sigma^2$ are independent and identically distributed chi-square random variables with $n - 1$ degrees of freedom, it follows that the mean and variance of W are $\mu_W = n - 1$ and $\sigma_W^2 = 2(n - 1)r/(2 - r)$, respectively, where for σ_W^2 we have used the fact that $\sum_{k=1}^{i} w_k^2$ converges to $r/(2 - r)$ as $i \to \infty$.

We shall approximate the distribution of W using a Satterthwaite approximation and write $W \sim h\chi_\nu^2$, where $h = \sigma_W^2/(2\mu_W)$ and $\nu = 2\mu_W^2/\sigma_W^2$. Making the appropriate substitutions, we find $\nu = (n - 1)(2 - r)/r$ and $h = r/(2 - r)$. Next let $V_i = n\bar{P}_i^2/\sigma^2$. It follows that V_i has a noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter λ_i. It is well known that $\bar{P}_i$ (and hence V_i) is independently distributed of S_i^2, and it is clearly independent of $S_1^2, \ldots, S_{i-1}^2$. Thus, V_i and W are distributed independently and consequently

$$F = \frac{V_i}{\dfrac{(W/h)}{\nu}} = \frac{n\bar{P}_i^2}{\hat{\sigma}_i^2}$$

has an approximate noncentral F distribution with numerator and denominator degrees of freedom equal to 1 and ν, respectively, and noncentrality parameter equal to λ_i. It follows that

$$MSE(\hat{\theta}_i^{SV}) \doteq \frac{2\nu^2}{n^2(\nu - 2)}\left[\frac{(1 + n\theta_i)^2}{(\nu - 2)(\nu - 4)} + \frac{1 + 2n\theta_i}{\nu - 4}\right] + \left[\frac{\nu}{\nu - 2}\left(\frac{1}{n} + \theta_i\right) - \theta_i\right]^2. \quad (9)$$

The squared bias term in (9) suggests a bias corrected estimator of the form $\hat{\theta}_i^{BCSV} = (\nu - 2)\hat{\theta}_i^{SV}/\nu - (1/n)$ from which it follows that

$$MSE(\hat{\theta}_i^{BCSV}) \doteq \frac{2(\nu - 2)}{n^2}\left[\frac{(1 + n\theta_i)^2}{(\nu - 2)(\nu - 4)} + \frac{1 + 2n\theta_i}{\nu - 4}\right]. \quad (10)$$

We note that in the special case where $r = 1$, then $\nu = n - 1$ and (9) and (10) reduce to (1) and (2), respectively.

Figure 8 shows the simulated RMSE values of $\hat{\theta}_i^{BCSV}$, labeled as BCSV (S), for the case where $n = 8$ and $r = 0.1$. Approximate RMSE values for $\hat{\theta}_i^{BCSV}$ using (10) agree very closely with the simulated values shown in Figure 8. We also note that there is very little difference in the RMSE values of $\hat{\theta}_i^{SV}$ and $\hat{\theta}_i^{BCSV}$, suggesting that bias correction after variance-smoothing has a small effect.

5.2. Absolute Value Estimator

Let $U^2 = \sum_{k=1}^{i} w_k T_k^2$ and define $\hat{\theta}_i^{SV-Z} = \bar{Z}_i^2/U^2$. We now derive an approximation for the MSE of $\hat{\theta}_i^{SV-Z}$. Recalling from Section 3.1 the definition $W_k^Z = (m - 1)T_k^2/\sigma_{Z,i}^2$, let

$$W^Z = \frac{(m - 1)U^2}{\sigma^2} = \sum_{k=1}^{i} \frac{w_k\sigma_{Z,k}^2 W_k^Z}{\sigma^2}.$$

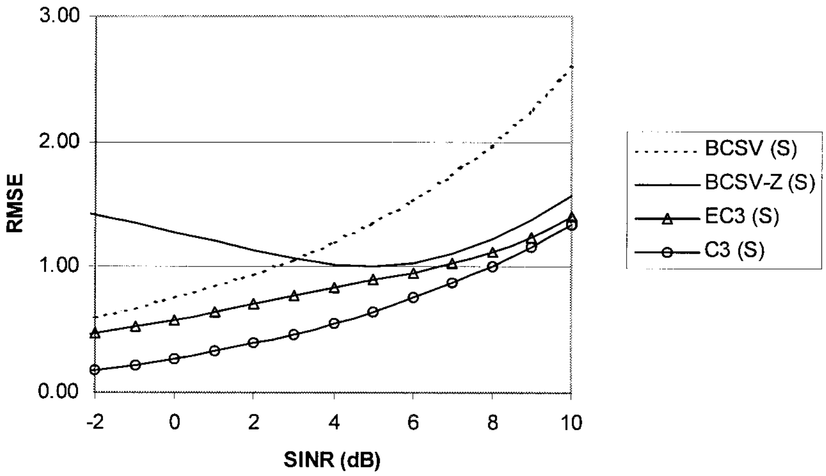

Figure 8. RMSE of variance-smoothed estimators BCSV and BCSV-Z and of the unconstrained combined estimator based on BCSV and BCSV-Z.

As in Section 3.1, we approximate the distribution of W_k^Z as $g_k\chi^2_{\eta_k}$, where $g_k \equiv g(\theta_k)$ and $\eta_k \equiv \eta(\theta_k)$ are the scale and degree-of-freedom parameters of a Satterthwaite approximation. Explicit expressions for both g_k and ν_k were previously mentioned as appearing in Appendix A. It follows that approximate expressions for the mean and variance of W^Z are thus

$$E(W^Z) \doteq (m-1)\sum_{k=1}^{i} w_k \frac{\sigma^2_{Z,k}}{\sigma^2}$$

$$Var(W^Z) \doteq 2(m-1)\sum_{k=1}^{i} w_k^2 \left(\frac{\sigma^2_{Z,k}}{\sigma^2}\right)^2 g_k.$$

As noted in Section 3.1, $\sigma^2_{Z,k} \to \sigma^2$ and $g_k \to m-1$ as θ_k gets large. Consequently, for large θ_k, $E(W^Z) \doteq m-1$ and $Var(W^Z) \doteq 2(m-1)^2 r/(2-r)$. A Satterthwaite approximation based on the large θ_k approximations for the mean and variance of W^Z implies $W^Z \overset{\sim}{\sim} h^Z\chi^2_{\nu^Z}$, where $h^Z = r/(2-r)$ and $\nu^Z = (m-1)(2-r)/r$ from which it follows that

$$F^Z = \frac{V_i^Z}{\frac{W^Z}{(h^Z\nu^Z)}} = m\hat{\theta}_i^{SV-Z}$$

has an approximate noncentral F distribution with numerator and denominator degrees of freedom equal to 1 and ν^Z, respectively, and noncentrality parameter equal to λ_i^Z. An approximate expressions for the MSE of $\hat{\theta}_i^{SV-Z}$ is thus

$$MSE(\hat{\theta}_i^{SV-Z}) \doteq \frac{2(\nu^Z)^2}{m^2(\nu^Z-2)}\left[\frac{(1+m\theta_{Z,i})^2}{(\nu^Z-2)(\nu^Z-4)} + \frac{1+2m\theta_{Z,i}}{\nu^Z-4}\right] + \left[\frac{\nu^Z}{\nu^Z-2}\left(\frac{1}{m}+\theta_{Z,i}\right) - \theta_i\right]^2. \tag{11}$$

The approximate MSE formula for $\hat{\theta}_i^{SV-Z}$ is similar to the MSE formula for $\hat{\theta}_i^{SV}$, the two differences being that m is used in place of n and $\theta_{Z,i}$ is used in place of θ_i. The squared bias term in (11) suggests a bias corrected estimator of the form $\hat{\theta}_i^{BCSV-Z} = (\nu^Z - 2)\hat{\theta}_i^{SV-Z}/\nu^Z - (1/m)$ from which it follows

$$MSE(\hat{\theta}_i^{BCSV-Z}) \doteq \frac{2(\nu^Z - 2)}{m^2}\left[\frac{(1 + m\theta_{Z,i})^2}{(\nu^Z - 2)(\nu^Z - 4)} + \frac{1 + 2m\theta_{Z,i}}{\nu^Z - 4}\right] + (\theta_{Z,i} - \theta_i)^2. \tag{12}$$

Figure 8 includes the simulated RMSE values of $\hat{\theta}_i^{BCSV-Z}$, labeled as BCSV-Z (S), for the case where $m = 20$ and $r = 0.1$. The simulated RMSE values for $\hat{\theta}_i^{SV-Z}$ are very close to the simulated RMSE values of $\hat{\theta}_i^{BCSV-Z}$, and consequently are not shown in Figure 8. In addition, the approximate RMSE values from (11) and (12) match closely with the simulated values and thus are not shown in Figure 8.

5.3. Combined Smoothed Estimators

Estimators based on combining $\hat{\theta}_i^{BCSV}$ and $\hat{\theta}_i^{BCSV-Z}$ can be obtained with constrained or unconstrained weights using the approximate MSE expressions given by (10) and (12), together with the weight formulas given in Section 4. Figure 8 shows the simulated RMSE values of $\hat{\theta}_i^{C3} = \alpha_1^{opt}(\theta_i)\hat{\theta}_i^{BCSV} + \alpha_2^{opt}(\theta_i)\hat{\theta}_i^{BCSV-Z}$, labeled as C3 (S) for the case where $m = 20$ and $r = 0.1$. The analytic approximate RMSE values for $\hat{\theta}_i^{C3}$ are not shown since they match the simulated RMSE values extremely well. We conclude from Figure 8 that the RMSE of $\hat{\theta}_i^{C3}$ is uniformly smaller than the RMSE of any other estimator considered in this paper, and for a majority of the operating region the magnitude of the reduced RMSE is appreciable. Also shown in Figure 8 are the simulated RMSE values of $\hat{\theta}_i^{EC3} = \alpha_1^{opt}(\hat{\theta}_i^{BCSV})\hat{\theta}_i^{BCSV} + \alpha_2^{opt}(\hat{\theta}_i^{BCSV})\hat{\theta}_i^{BCSV-Z}$, where the weights have been estimated using $\hat{\theta}_i^{BCSV}$. The increase in the RMSE due to using estimated weights is quite apparent, however, it is still the case that the RMSE of $\hat{\theta}_i^{EC3}$ is uniformly smaller than any other useable estimator considered in this paper.

For completeness, we note that the RMSE values of the constrained combined estimators $\hat{\theta}_i^{C4} = \alpha^{opt}(\theta_i)\hat{\theta}_i^{BCSV} + [1 - \alpha^{opt}(\theta_i)]\hat{\theta}_i^{BCSV-Z}$ and $\hat{\theta}_i^{EC4} = \alpha^{opt}(\hat{\theta}_i^{BCSV})\hat{\theta}_i^{BCSV} + [1 - \alpha^{opt}(\hat{\theta}_i^{BCSV})]\hat{\theta}_i^{BCSV-Z}$ are very close to each other and to the RMSE of $\hat{\theta}_i^{EC3}$. Thus, when estimated weights must be used, it does not matter much whether estimated unconstrained weights or estimated constrained weights are used.

6. SUMMARY

Several important control actions in wireless communication systems (e.g., transmit power level determination, handoff decisions, data rate adaptation algorithms) depend on an accurate SINR estimator. In this paper, we have developed SINR estimators using some new approaches that yield estimators with significantly smaller RMSE than traditional estimators. Estimators based on pilot bits only or user bits only were considered first and then estimators that are linear combinations of these two types of estimators were discussed. It was discussed that for estimating the interference variance (the denominator of the SINR), long-term averaging via exponential smoothing is viable and dramatically reduces the RMSE of SINR estimators.

Approximate analytic expressions for the RMSE of the estimators were derived and their accuracy was validated using simulation methods. The analytic expressions serve two purposes. First, improved estimators based on bias correction and/or variance reduction become apparent

Table 1. Summary of the estimators considered in this paper.

Estimator	Section	Description	Remark
$\hat{\theta}^{PI}$	2	Plug-in estimator based on pilot bits	Has the highest RMSE for the sample sizes illustrated in the paper
$\hat{\theta}^{BC}$	2	Bias-corrected version of $\hat{\theta}^{PI}$	Dominates $\hat{\theta}^{PI}$ by reducing bias
$\hat{\theta}^{PI-Z}$	3.1	Plug-in estimator based on user bits	Dominates $\hat{\theta}^{BC}$ for the sample sizes illustrated in the paper
$\hat{\theta}^{BC-Z}$	3.1	Bias-corrected version of $\hat{\theta}^{PI-Z}$	Dominates $\hat{\theta}^{PI-Z}$ by reducing bias
$\hat{\theta}^{PI-F}$	3.2	Plug-in feedback estimator based on pilot bits	Dominates $\hat{\theta}^{BC-Z}$ for small to mid-range SINR values
$\hat{\theta}^{BC-F}$	3.2	Bias-corrected version of $\hat{\theta}^{PI-F}$	Reduces RMSE of $\hat{\theta}^{PI-F}$ for large SINR values
$\hat{\theta}^{C1}$	4	Constrained (weights sum to unity) linear combination of $\hat{\theta}^{BC}$ and $\hat{\theta}^{BC-Z}$	Significant reduction in RMSE by combining, but it assumes perfect knowledge of weights
$\hat{\theta}^{EC1}$	4	Variation of $\hat{\theta}^{C1}$ where the weights are estimated using $\hat{\theta}^{BC}$	Not much penalty in RMSE for using estimated weights
$\hat{\theta}^{C2}$	4	Unconstrained (weights need not sum to unity) linear combination of $\hat{\theta}^{BC}$ and $\hat{\theta}^{BC-Z}$	Significantly smaller RMSE than $\hat{\theta}^{C1}$ for small values of SINR
$\hat{\theta}^{EC2}$	4	Variation of $\hat{\theta}^{C2}$ where the weights are estimated using $\hat{\theta}^{BC}$	Higher cost for using estimated weights than is the case for $\hat{\theta}^{C1}$
$\hat{\theta}^{SV}$	5.1	Variation of $\hat{\theta}^{PI}$ but with use of smoothed variance estimator	Removes the small sample size limitation associated with $\hat{\theta}^{PI}$ and $\hat{\theta}^{BC}$; RMSE is comparable to $\hat{\theta}^{EC1}$ and $\hat{\theta}^{EC2}$
$\hat{\theta}^{BCSV}$	5.1	Bias-corrected version of $\hat{\theta}^{SV}$	Not much additional reduction in RMSE with bias correction after smoothing
$\hat{\theta}^{SV-Z}$	5.2	Variation of $\hat{\theta}^{PI-Z}$ but with use of smoothed variance estimator	Significantly worse than $\hat{\theta}^{BCSV}$ for small to midrange SINR; significantly better than $\hat{\theta}^{BCSV}$ for large SINR
$\hat{\theta}^{BCSV-Z}$	5.2	Bias-corrected version of $\hat{\theta}^{SV-Z}$	Not much additional reduction in RMSE with bias correction after smoothing
$\hat{\theta}^{C3}$	5.3	Unconstrained linear combination of $\hat{\theta}^{BCSV}$ and $\hat{\theta}^{BCSV-Z}$	Smallest RMSE of all the estimators considered in this paper, but weights depend on unknown SINR
$\hat{\theta}^{EC3}$	5.3	Variation of $\hat{\theta}^{C3}$ where weights are estimated using $\hat{\theta}^{BCSV}$	RMSE of $\hat{\theta}^{C3}$ remains the smallest of all the estimators considered in this paper, even when the weights are estimated
$\hat{\theta}^{C4}$	5.3	Constrained linear combination of $\hat{\theta}^{BCSV}$ and $\hat{\theta}^{BCSV-Z}$	RMSE values are comparable to those of $\hat{\theta}^{C3}$
$\hat{\theta}^{EC4}$	5.3	Variation of $\hat{\theta}^{C4}$ where weights are estimated using $\hat{\theta}^{BCSV}$	RMSE values are comparable to those of $\hat{\theta}^{EC3}$

from the expressions and, second, optimal weights for combining pilot bit and user bit based estimators can be evaluated. Table 1 provides a summary and review of all the different estimators considered in this paper. A brief comment about the RMSE performance of each estimator is also provided. In general, our results show that the best estimators are obtained through optimal combining. Using estimated weights in place of the unknown optimal weights predictably increases the RMSE, but not to the extent of diminishing the value of combining.

APPENDIX A: SATTERTHWAITE APPROXIMATION CONSTANTS

The constants $g_i \equiv g(\theta_i)$ and $\eta_i \equiv \eta(\theta_i)$ for the Satterthwaite approximation to the distribution of $W_i^Z = (m-1)T_i^2/\sigma_{Z,i}^2$, which is utilized in Sections 3.1 and 5.2, satisfy the equations $m - 1 = g_i\eta_i$ and $(m-1)^2 Var(T_i^2) = 2g_i^2\eta_i\sigma_{Z,i}^4$. Equivalently, $g_i = (m-1)Var(T_i^2)/2\sigma_{Z,i}^4$ and $\eta_i = 2\sigma_{Z,i}^4/Var(T_i^2)$. We therefore need to evaluate the ratio $Var(T_i^2)/\sigma_{Z,i}^4$. It is well known (see, for example, Cramer [3], Section 27.4) that

$$\frac{Var(T_i^2)}{\sigma_{Z,i}^4} = \frac{(m-1)\mu_4 - (m-3)\mu_2^2}{m(m-1)\sigma_{Z,i}^4},$$

where $\mu_4 = E(Z_{ij} - \mu_{Z,i})^4$ and $\mu_2 = E(Z_{ij} - \mu_{Z,i})^2 = \sigma_{Z,i}^2$. Clearly $\mu_2^2/\sigma_{Z,i}^4 = 1$ and hence

$$\frac{Var(T_i^2)}{\sigma_{Z,i}^4} = \frac{\mu_4}{m\sigma_{Z,i}^4} - \frac{m-3}{m(m-1)}. \tag{13}$$

Expanding μ_4 gives

$$\mu_4 = -3\mu_{Z,i}^4 + 6\mu_{Z,i}^2 E(Z_{ij}^2) - 4\mu_{Z,i}E(Z_{ij}^3) + E(Z_{ij}^4). \tag{14}$$

Clearly,

$$E(Z_{ij}^2) = \mu_i^2 + \sigma_i^2. \tag{15}$$

Writing $Z_{ij}^4 = (a_{ij}\mu_i + \epsilon_{ij})^4$ and expanding the right-hand side, it is straightforward to show that

$$E(Z_{ij}^4) = 3\sigma_i^4 + 6\mu_i^2\sigma_i^2 + \mu_i^4. \tag{16}$$

It remains to find $E(Z_{ij}^3)$ which we can evaluate as

$$E(Z_{ij}^3) = E|a_{ij}\mu_i + \epsilon_{ij}|^3 = E\{|a_{ij}\mu_i + \epsilon_{ij}|^3 | a_{ij} = 1\} \times p + E\{|a_{ij}\mu_i + \epsilon_{ij}|^3 | a_{ij} = -1\} \times (1-p), \tag{17}$$

where $p = \Pr(a_{ij} = 1)$. Let R be a random variable with a Gaussian distribution having mean μ_i and variance σ_i^2. It follows from (16) that

$$E(Z_{ij}^3) = E[|R|^3] \times p + E[|-R|^3] \times (1-p) = E[|R|^3].$$

Evaluating $E|R|^3$ directly gives

$$E(Z_{ij}^3) = (\mu_i^3 + 3\sigma_i^2\mu_i)\left[2\Phi\left(\frac{\mu_i}{\sigma_i}\right) - 1\right] + \frac{1}{\sqrt{2\pi}}e^{-(\mu_i^2/2\sigma_i^2)}(2\sigma_i\mu_i^2 + 4\sigma_i^3). \tag{18}$$

Combining (14)–(18) gives

$$\begin{aligned}\frac{\mu_4}{\sigma_{Z,i}^4} = {} & -3\left(\frac{\mu_{Z,i}}{\sigma_{Z,i}}\right)^4 + 6\left(\frac{\mu_{Z,i}}{\sigma_{Z,i}}\right)^2\left[\left(\frac{\mu_i}{\sigma_{Z,i}}\right)^2 + \left(\frac{\sigma_i}{\sigma_{Z,i}}\right)^2\right] - 4\left(\frac{\mu_{Z,i}}{\sigma_{Z,i}}\right)\left\{\left[\left(\frac{\mu_i}{\sigma_{Z,i}}\right)^3 + 3\left(\frac{\sigma_i}{\sigma_{Z,i}}\right)^2\left(\frac{\mu_i}{\sigma_{Z,i}}\right)\right]\left[2\Phi\left(\frac{\mu_i}{\sigma_i}\right) - 1\right]\right. \\ & \left. + \frac{1}{\sqrt{2\pi}}e^{-(\mu_i^2/2\sigma_i^2)}\left[2\left(\frac{\sigma_i}{\sigma_{Z,i}}\right)\left(\frac{\mu_i}{\sigma_{Z,i}}\right)^2 + 4\left(\frac{\sigma_i}{\sigma_{Z,i}}\right)^3\right]\right\} + 3\left(\frac{\sigma_i}{\sigma_{Z,i}}\right)^4 + 6\left(\frac{\mu_i}{\sigma_{Z,i}}\right)^2\left(\frac{\sigma_i}{\sigma_{Z,i}}\right)^2 + \left(\frac{\mu_i}{\sigma_{Z,i}}\right)^4.\end{aligned}$$

It is easy to verify that the ratios $\mu_{Z,i}/\sigma_{Z,i}$, $\mu_i/\sigma_{Z,i}$, and $\sigma_i/\sigma_{Z,i}$ all depend on μ_i and σ_i through the ratio θ_i. Thus, the ratio $\mu_4/\sigma_{Z,i}^4$ is a function of θ_i. It therefore follows from (13) that $Var(T_i^2)/\sigma_{Z,i}^4$ is a function of θ_i and, consequently, g_i and η_i are as well.

APPENDIX B: SIMULATION DETAILS

Simulations were run to obtain RMSE estimates that could be compared to the analytic RMSE approximations. An overview of the simulation details is provided in this appendix. For each θ_i in the set $\{-2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, each estimator was computed for 50,000 datasets of demodulator outputs and the simulation estimate of the RMSE was obtained from sample RMSE. The sample size of 50,000 was deemed sufficient based on observed accuracy of the simulation results in cases where exact analytic RMSE formulas are available (e.g., $\hat{\theta}^{PI}$ and $\hat{\theta}^{BC}$).

For the estimators based on pilot bits (Section 2), each dataset of demodulator outputs consisted of $n = 8$ P_{ij} observations that were simulated from a Gaussian distribution with mean $\sqrt{\theta_i}$ and variance 1. In the case of the smoothed pilot bit-based estimators (Section 5.1), a warm-up period of 2000 datasets was used to allow the initial value associated with the exponential smoothing of the variance term to dissipate.

For the absolute value estimators based on user bits (Section 3.1), each dataset of demodulator outputs consisted of $m = 20$ U_{ij} observations that were simulated from a symmetric mixture of two Gaussian distributions that have respective means $\pm\sqrt{\theta_i}$ and variance 1. Again, a warm-up period of 2000 datasets was used when computing the smoothed user bit-based absolute value estimator (Section 5.2).

The datasets used for the feedback estimator (Section 3.2) were the same as those used for the absolute value estimator, but in this case the U_{ij} observations were used to obtain user bit predictions ($\hat{a}_{ij}$) rather than the Z_{ij} observations. The user bit predictions were then used to obtain the D_{ij} observations from which the feedback estimator was computed.

For the combined estimators in Section 4, each dataset consist of both $n = 8$ P_{ij} observations (used to compute the estimator based on pilot bits) and $m = 20$ U_{ij} observations (used to compute the estimator based on user bits). For the combined smoothed estimators in Section 5.3, each dataset similarly consisted of both P_{ij} observations and U_{ij} observations, but in addition 2000 preliminary datasets were used as a warm-up period.

ACKNOWLEDGMENTS

We would like to thank Yasuo Amemiya for a helpful question pertaining to the justification for our approximation to the MSE of the feedback estimator. We thank the referees for comments and suggestions that greatly improved the readability of our paper.

REFERENCES

[1] A.L. Brandao, L.B. Lopez, and D.C. McLernon, Co-channel interference estimation for M-ary PSK modulated signals, Wireless Personal Commun 1(1) (1994), 23–32.

[2] R. Christensen, Plane answers to complex questions: The theory of linear models, Springer-Verlag, New York, 1987.

[3] H. Cramer, Mathematical methods of statistics, Princeton University Press, Princeton, NJ, 1946.

[4] A. Furuskar, S. Mazur, F. Muller, and H. Olofsson, EDGE: Enhanced Data Rates for GSM and TDMA/136 Evolution, IEEE Personal Commun Mag (June 1999), 56–66.

[5] C.E. Gilchriest, Signal-to-noise monitoring, JPL Space Programs Summary, 37-27, IV (1964), 169–184.

[6] F.A. Graybill, Theory and application of the linear model, Wadsworth, Belmont, CA, 1976.

[7] K. Higuchi, H. Andoh, K. Okawa, M. Sawahashi, and F. Adachi, Experimental evaluation of combined effect of coherent rake combining and SIR-based fast transmit power control for reverse link of DS-CDMA mobile radio, IEEE J Selected Areas Commun 18(8) (2000), 1526–1535.

[8] N.L. Johnson, S. Kotz, and N. Balakrishnan, Continuous univariate distributions (2nd edition), Wiley, New York, 1995, Vol. 2.

[9] J.W. Layland, On S/N estimation, JPL Space Programs Summary, 37-48, III (1967), 209–212.

[10] J.G. Proakis, Digital communications, McGraw-Hill, New York, 1989.

[11] D. Ramakrishna, N.B. Mardayam, and R. Yates, Subspace based SIR estimation for CDMA cellular systems, IEEE Trans Vehicular Technol 49(5) (2000), 1732–1742.

[12] R.H. Randles and D.A. Wolfe, Introduction to the theory of nonparametric statistics, Wiley, New York, 1979.

[13] A.L. Rukhin, Estimating the noncentrality parameter of a t-distribution, Systems Sci Math Sci 5 (1992), 1–8.

[14] F.E. Satterthwaite, An approximate distribution of estimates of variance components, Biometrika Bull 2 (1946), 110–114.
[15] S.R. Searle, Linear models, Wiley, New York, 1971.
[16] C.M. Thomas, Maximum likelihood estimation of signal-to-noise ratio, Ph.D. dissertation, University of Southern California, Department of Electrical Engineering, Los Angeles, 1967.
[17] M.L. Tiku, Doubly noncentral F distribution—tables and applications, Sel Tables Math Statist 2 (1974), 139–149.
[18] M. Turkboylari and G.L. Stuber, An efficient algorithm for estimating the signal-to-interference ratio in TDMA cellular systems, IEEE Trans Commun 46(6) (1998), 728–731.
[19] A.J. Viterbi, CDMA: Principles of spread spectrum communication, Addison-Wesley, New York, 1995.