Soft Input Channel Estimation for Turbo Equalization

Seongwook Song, Andrew C. Singer, and Koeng-Mo Sung, *Member, IEEE*

***Abstract*—In this paper, we consider soft decision directed channel estimation for turbo equalization. To take advantage of soft information provided by the decoder, a minimum mean square error linear channel estimator is derived under an uncorrelated channel tap model, and a soft input recursive least squares algorithm is also developed by modifying the cost function of the conventional recursive least squares algorithm. The performance of the proposed channel estimators are analyzed in terms of mean square identification error (MSIE) for stationary channels. Simulation results for both time-invariant and time-varying frequency-selective Rayleigh fading channels are also presented.**

***Index Terms*—Channel estimation, soft input, turbo equalization.**

I. Introduction

SINCE the introduction of turbo equalization in [1], a number of researchers have considered practical implementations of this scheme, e.g., [2]–[4], as well as other methods for incorporating soft information from decoding into a host of additional communications and signal processing tasks, e.g., [5]–[7]. The original turbo equalization approach performs extremely well when the channel is known at the receiver, whereas a number of difficulties arise when there is not sufficient channel information known. As a separate task from equalization and decoding, a training sequence can be inserted into the data frame to estimate the channel at the expense of overall system throughput. Adaptive methods for linear equalization (LE) and decision feedback equalization (DFE) can exploit soft statistics from the decoder and can be incorporated directly into an adaptive algorithm to update the equalizer coefficients [4]. Another approach to this problem is to consider the channel estimation problem as simply one additional signal processing task to be performed jointly with the data equalization and decoding tasks. For example, a statistical approach using the sequential Monte Carlo algorithm was taken under the condition that prior distributions of the unknowns are available to the receiver, and alternatives were also discussed for the case when the prior distributions are unknown in [8] and [9]. In [10] and [11], soft statistics from the soft input/soft output (SISO) detector are applied to a channel estimator with decision delay for reduction of error propagation. In [12], the channel is jointly estimated with the data using an expanded trellis and periodic pilot symbols. A method for joint channel estimation and equalization is considered in [13] by combining channel uncertainty into the path metric for a BCJR-like algorithm.

There has also been work on joint decoding and channel estimation, in which hard decisions from the decoder are used to refine a channel estimate, in the context of orthogonal frequency division multiplexing (OFDM) transmission [14], [15]. In [16], an iterative channel estimator exploiting soft decisions after error-correction decoding was investigated for flat fading channels, and, in [17], the multipass channel estimator using both pilot symbols and soft outputs of the decoder was employed for frequency-selective channels, i.e., coded multiuser wideband code-division multiple-access wireless uplink channels. Channel estimators that operate in batch mode may be susceptible to tracking channel variations, if the soft statistics from the detector are not reliable, in particular for a rapidly time-varying channel. Meanwhile, channel estimates can be refined for relatively slowly time varying channels through averaging and exploiting the error-correction code.

In this paper, the turbo equalization setting in [16] and [17] is considered for frequency-selective Rayleigh fading channels. We propose a soft input Kalman channel estimator, which is the best linear unbiased channel estimator for the case of interest. The soft input Kalman channel estimator is derived by restructuring the channel estimation problem as one of Kalman state estimation reflecting the soft information from the decoding process into the statistical description of the channel. Additionally, the average Kalman channel estimator presented in [13] can be interpreted as an approximation to the algorithm proposed here. A soft input recursive least squares (RLS) channel estimator is also derived through a modified cost function and weighting factor taking into account the soft statistics.

Analysis is presented in terms of mean square identification error (MSIE) for stationary channels. To facilitate analytical performance comparisons with the conventional hard decision-based algorithm, the widely used Gaussian distribution assumption [18], [19] for the soft output of the SISO decoder is adopted. The rest of this paper is organized as follows: In Section II, the system model used in this paper is presented. In Section III, a soft input Kalman channel estimator is derived as the linear minimum mean square error channel estimator. A soft input weighted recursive least squares (WRLS) channel estimator is developed in Section IV. Analysis of these algorithms is presented in Section V. In Section VI, performance comparisons

Manuscript received April 28, 2003; revised October 1, 2003. This work was supported in part by the Department of the Navy, Office of Navel Research, under Grant N00014-01-1-0117 and the National Science Foundation under Grant CCR-0092598 (CAREER). The associate editor coordinating the review of this paper and approving it for publication was Prof. Tulay Adali.

S. Song was with the Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Urbana, IL 61801 USA. He is now with Samsung Electronics, Suwon, Gyeonggi 442-600, Korea (e-mail: seongwook.song@samsung.com).

A. C. Singer is with the Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Urbana, IL 61801 USA (e-mail: acsinger@uiuc.edu).

K.-M. Sung is with the Seoul National University, Seoul, Korea (e-mail: kmsung@acoustics.snu.ac.kr).

Digital Object Identifier 10.1109/TSP.2004.834270

are made through computer simulations. Finally, we summarize our results in Section VII.

II. System Overview

Consider a baseband transmission system with the source bit stream $\{s[n]\}$ encoded by a rate R_c binary convolutional code with memory L_c bits. The coded bits are interleaved to reduce the influence of error bursts from the equalizer on the channel decoder. The interleaved code bits $\{a[n]\}$ are mapped to M-ary signal constellation $\mathcal{A}$ and transmitted over a frequency selective Rayleigh fading channel. Usually, the overall modulator-channel-demodulator is modeled with a time-varying finite-length impulse response filter $h[n,k]$ of length L, and then, the baud-sampled received signal $r[n]$ is represented as

$$r[n] = \sum_{k=0}^{L-1} h[n,k]b[n-k] + w_0[n], \quad n \geq 0 \tag{1}$$

where $b[n] \in \mathcal{A}$ denotes the transmitted symbols, and $w_0[n]$ is additive complex white Gaussian noise with zero mean and circular symmetric variance N_0. A typical model used in wireless communications is that each channel tap, corresponding to a different multipath arrival, is an uncorrelated wide sense stationary random process, yielding the standard uncorrelated scattering (WSSUS) model [20]. We will exploit this model later for channel estimation.

III. Soft Input Kalman Channel Estimator

Conventional approaches to channel estimation consider pilot symbols (training) inserted in a data frame. In turbo equalization, soft outputs of the previous iterations are available as well and can also be used for channel estimation. If the decoder output were quantized into hard decisions, as is often done, then the system can be prone to a variety of error propagation effects that also arise in a decision feedback equalizer (DFE) [21]. Instead of using hard decisions, we present a channel estimator that can take advantage of the soft outputs available from the equalizer and the decoder.

From knowledge of the signal constellation, the log likelihood ratio (LLR) of each data bit obtained from the SISO decoder can be converted to the probability mass function (p.m.f.) represented by a length M real vector $\mathcal{P}_n \triangleq (p_0[n], p_1[n], \ldots, p_{M-1}[n])^T$, where $p_m[n] = P\{b[n] = b_m\}, b_m \in \mathcal{A}$. Given $\mathcal{P}_n$, the mean and variance of each received symbol at each sampling time are expressed as $\bar{b}[n] \triangleq E_{\mathcal{P}}[b[n]]$, and $\sigma_b^2[n] \triangleq E_{\mathcal{P}}[|b[n]|^2] - |\bar{b}[n]|^2$, where $E_{\mathcal{P}}[\cdot]$ denotes the expectation using the sequence of p.m.f.'s derived by $\{\mathcal{P}_n\}$ in addition to the statistics of the noise sequences $\mathbf{v}[n]$ and $w_0[n]$, and those of $\mathbf{h}[0]$, as in (4). In order to exploit the soft outputs of the SISO decoder, we treat the transmitted signal $b[n]$ as a stochastic signal, rather than, as is often done in channel estimation, treating it as a deterministic one. Then, given $\mathcal{P}_n$, the transmitted signal can be decomposed into two parts:

$$b[n] = \bar{b}[n] + \tilde{b}[n] \tag{2}$$

where $\tilde{b}[n]$ is the assumed random variation about the mean $\bar{b}[n]$. Note that $\bar{b}[n]$ is deterministic, given $\mathcal{P}_n$, whereas $\tilde{b}[n]$ is a sequence of zero mean random variables with variance $\sigma_b^2[n]$. Considering $b[n]$ as a random process uncorrelated with other symbols under the assumption of the efficient interleaving, it is given that $E_{\mathcal{P}}[\tilde{b}[n+k]\tilde{b}^*[n]] = \sigma_b^2[n]\delta[k]$. With the above decomposition of $b[n]$ taken into consideration, the channel estimation problem formulation can be written as follows:

$$\mathbf{b}[n] = \bar{\mathbf{b}}[n] + \tilde{\mathbf{b}}[n] \tag{3}$$

$$\mathbf{h}[n+1] = \mathbf{F}\mathbf{h}[n] + \mathbf{v}[n] \tag{4}$$

$$r[n] = \mathbf{h}^H[n]\mathbf{b}[n] + w_0[n] \tag{5}$$

where $\mathbf{h}[n] = (h[n,0], h[n,1], \ldots, h[n,L-1])^H$, $\mathbf{b}[n] = (b[n], b[n-1], \ldots, b[n-L+1])^T$, $\bar{\mathbf{b}}[n] = (\bar{b}[n], \bar{b}[n-1], \ldots, \bar{b}[n-L+1])^T$, $\tilde{\mathbf{b}}[n] = (\tilde{b}[n], \tilde{b}[n-1], \ldots, \tilde{b}[n-L+1])^T$, $\mathbf{v}[n]$ is zero mean complex white Gaussian noise with $E[\mathbf{v}[n+k]\mathbf{v}^H[n]] = \mathbf{Q}_v\delta[k]$, and $\mathbf{F}$ denotes an $L \times L$ state transition matrix for the channel $\mathbf{h}[n]$. This is simply the conventional state space model, with (3) introduced to account for the soft information. Since the transmitted vector $\mathbf{b}[n]$ is not known inside the SISO receiver, it is perhaps more meaningful to rewrite (3)–(5) as

$$\mathbf{h}[n+1] = \mathbf{F}\mathbf{h}[n] + \mathbf{v}[n] \tag{6}$$

$$r[n] = \mathbf{h}^H[n]\bar{\mathbf{b}}[n] + g[n] \tag{7}$$

where

$$g[n] = \mathbf{h}^H[n]\tilde{\mathbf{b}}[n] + w_0[n]. \tag{8}$$

In (7), the observation is expressed in terms of the means of the data $\bar{\mathbf{b}}[n]$, replacing the hard decision $\mathbf{b}[n]$, and the uncertain component $\tilde{\mathbf{b}}[n]$ in the data is absorbed into the newly defined noise term $g[n]$. In addition, the signal component $\mathbf{h}^H[n]\bar{\mathbf{b}}[n]$ and the noise component $g[n]$ are uncorrelated because $\tilde{b}[n]$ and $w_0[n]$ have zero mean and are uncorrelated with $\mathbf{v}[n]$, and $\bar{b}[n]$ is deterministic, given $\mathcal{P}_n$. This setup leads to the familiar Kalman filter, except that the observation noise $g[n]$ has a nonlinear noise component containing the multiplication of two random variables of $\mathbf{h}[n]$ and $\tilde{\mathbf{b}}[n]$, with $E_{\mathcal{P}}[g[n]] = 0$. The autocorrelation of $g[n]$, given the soft statistics $\{\mathcal{P}_n\}$, is found as

$$\begin{aligned} E_{\mathcal{P}}&[g[n+k]g^*[n]] \\ &= E_{\mathcal{P}}[\{\mathbf{h}^H[n+k]\tilde{\mathbf{b}}[n+k] + w_0[n+k]\} \\ &\quad \times \{\mathbf{h}^H[n]\tilde{\mathbf{b}}[n] + w_0[n]\}^*] \qquad (9) \\ &= \mathrm{tr}\{\mathbf{R}_h[n+k,n]\mathbf{K}_{\tilde{b}}[n+k,n]\} + N_0\delta[k] \qquad (10) \end{aligned}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace operator, $\mathbf{R}_h[n+k,n] = E[\mathbf{h}[n+k]\mathbf{h}^H[n]]$, and $\mathbf{K}_{\tilde{b}}[n+k,n] = E_{\mathcal{P}}[\tilde{\mathbf{b}}[n+k]\tilde{\mathbf{b}}^H[n]]$ with entries

$$[\mathbf{K}_{\tilde{b}}[n+k,n]]_{i,j} = \begin{cases} \sigma_b^2[n-k-i+1], & \text{if } i-j=k \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $1 \leq i,j \leq L$, and $[\mathbf{A}]_{i,j}$ denotes the ith row and jth column element of $\mathbf{A}$.

Since $g[n]$ is neither white nor Gaussian in general, the Kalman filter is not an optimal minimum mean square error (MMSE) estimator. However, in the case where $g[n]$ is white, the Kalman filter is known to be the best linear unbiased estimator [22]. We can show that $g[n]$ is indeed white based on the WSSUS multipath fading channel model. In this case, the transition matrix $\mathbf{F}$ becomes a diagonal matrix, as does $\mathbf{R}_h[n+k,n]$. With the aid of these two additional conditions and with (11), it can be readily shown that $\mathrm{tr}\{\mathbf{R}_h[n+k,n]\mathbf{K}_{\tilde{b}}[n+k,n]\} = \mathrm{tr}\{\mathbf{F}^k\mathbf{R}_h[n,n]\mathbf{K}_{\tilde{b}}[n+k,n]\} = 0$ for $k \neq 0$. Considering $\mathbf{h}[n]$ as wide sense stationary, $\mathbf{R}_h[0] \triangleq \mathbf{R}_h[0,0] = \mathbf{R}_h[n,n]$ can be used in place of $\mathbf{R}_h[n,n]$. Correspondingly, (10) can be simplified as

$$E_{\mathcal{P}}[g[n+k]g^*[n]] = \mathrm{tr}\{\mathbf{Q}_g[n]\}\delta[k] \tag{12}$$

where

$$\mathbf{Q}_g[n] \triangleq \mathbf{R}_h[0]\mathbf{K}_{\tilde{b}}[n,n] + \frac{1}{L}N_0\mathbf{I}. \tag{13}$$

Each matrix above can be found by the following two relations:

$$\mathbf{K}_{\tilde{b}}[n,n] = \mathrm{diag}\big(\sigma_{\tilde{b}}^2[n], \sigma_{\tilde{b}}^2[n-1], \ldots, \sigma_{\tilde{b}}^2[n-L+1]\big) \tag{14}$$

$$\mathbf{R}_h[0] = \mathbf{F}\mathbf{R}_h[0]\mathbf{F}^H + \mathbf{Q}_v. \tag{15}$$

Therefore, $\mathbf{K}_{\tilde{b}}[n,n]$ is obtained from the soft statistics $\tilde{b}[n]$, and the term $\mathbf{R}_h[0]$ can be computed directly from the state space model. Particularly for the WSSUS channel model, it can be found that $[\mathbf{R}_h[0]]_{i,i} = ([\mathbf{Q}_v]_{i,i})/(1-[\mathbf{F}]_{i,i}^2)$ because both $\mathbf{F}$ and $\mathbf{Q}_v$ are diagonal matrices. As a result, $q[n] \triangleq E_{\mathcal{P}}[g[n]g^*[n]]$, and the variance of uncorrelated noise $g[n]$ can be computed as

$$\begin{aligned} q[n] &= \mathrm{tr}\big\{\mathrm{diag}\big(\sigma_{\tilde{b}}^2[n], \sigma_{\tilde{b}}^2[n-1], \ldots \\ &\qquad \sigma_{\tilde{b}}^2[n-L+1]\big)\,\mathbf{R}_h[0]\big\} + N_0 \qquad (16)\\ &= \sum_{k=0}^{L-1}[\mathbf{R}_h[0]]_{k,k}\sigma_{\tilde{b}}^2[n-k] + N_0. \qquad (17)\end{aligned}$$

Taking information on the noise statistics into account, the following recursion produces the best linear channel estimate in a minimum mean square error sense:

$$\hat{\mathbf{h}}[n\,|\,n-1] = \mathbf{F}\hat{\mathbf{h}}[n-1\,|\,n-1] \tag{18}$$

$$e[n\,|\,n-1] = r[n] - \hat{\mathbf{h}}^H[n\,|\,n-1]\bar{\mathbf{b}}[n] \tag{19}$$

$$q[n] = \sum_{k=0}^{L-1}[\mathbf{R}_h[0]]_{k,k}\sigma_{\tilde{b}}^2[n-k] + N_0 \tag{20}$$

$$\mathbf{k}[n] = \frac{\mathbf{P}[n\,|\,n-1]\bar{\mathbf{b}}[n]}{q[n] + \bar{\mathbf{b}}^H[n]\mathbf{P}[n\,|\,n-1]\bar{\mathbf{b}}[n]} \tag{21}$$

$$\hat{\mathbf{h}}[n\,|\,n] = \hat{\mathbf{h}}[n\,|\,n-1] + \mathbf{k}[n]e^*[n\,|\,n-1] \tag{22}$$

$$\mathbf{P}[n+1\,|\,n] = \mathbf{F}(\mathbf{I} - \mathbf{k}[n]\bar{\mathbf{b}}^H[n])\mathbf{P}[n\,|\,n-1]\mathbf{F}^H + \mathbf{Q}_v[n] \tag{23}$$

where $\hat{\mathbf{h}}[n\,|\,k]$ denotes the linear minimum mean square error estimate of $\mathbf{h}[n]$, given $r[0], r[1], \ldots, r[k]$, and $\mathbf{P}[n\,|\,n-1] \triangleq E_{\mathcal{P}}[\boldsymbol{\epsilon}[n\,|\,n-1]\boldsymbol{\epsilon}^H[n\,|\,n-1]]$ with $\boldsymbol{\epsilon}[n\,|\,n-1] \triangleq \mathbf{h}[n] - \hat{\mathbf{h}}[n\,|\,n-1]$, and $\mathbf{k}[n]$ is the Kalman gain vector. We call this the soft input Kalman channel estimator. When a training sequence is applied, i.e., $\sigma_{\tilde{b}}^2[n] = 0,\ n = 1, \ldots, K$, the right-hand side of (20) becomes N_0, which makes the soft input Kalman channel estimator perform the same operations as the conventional hard-decision Kalman estimator. Otherwise, if $\sigma_{\tilde{b}}^2[n] \neq 0$, the right-hand side of (20) has a value larger than N_0, which lessens the Kalman gain in (21). Subsequently, the innovations containing less reliable data are weighed less than others in the linear channel estimator. Additional computation for the soft input Kalman estimator arises from (20) and is mainly due to L multiplications and additions. As a result, the soft input Kalman channel estimator has almost the same numerical complexity as that of a conventional hard decision Kalman channel estimator.

Note that $\mathbf{R}_h[0]$ can also be recursively computed using the expectation of the previous estimate $\hat{\mathbf{h}}[n\,|\,n-1]$ and soft statistics by

$$\begin{aligned}\mathbf{R}_h[0] &= E_{\mathcal{P}}[(\hat{\mathbf{h}}[n\,|\,n-1] + \boldsymbol{\epsilon}[n\,|\,n-1]) \\ &\qquad \times (\hat{\mathbf{h}}[n\,|\,n-1] + \boldsymbol{\epsilon}[n\,|\,n-1])^H] \qquad (24)\\ &= E_{\mathcal{P}}[\hat{\mathbf{h}}[n\,|\,n-1]\hat{\mathbf{h}}^H[n\,|\,n-1]] + \mathbf{P}[n\,|\,n-1] \qquad (25)\end{aligned}$$

where, in the second step, the orthogonality condition $E_{\mathcal{P}}[\boldsymbol{\epsilon}[n\,|\,n-1]\hat{\mathbf{h}}^H[n\,|\,n-1]] = 0$ is applied. By replacing $E_{\mathcal{P}}[\hat{\mathbf{h}}[n\,|\,n-1]\hat{\mathbf{h}}^H[n\,|\,n-1]]$ with a current estimate $\hat{\mathbf{h}}[n\,|\,n-1]\hat{\mathbf{h}}^H[n\,|\,n-1]$, an approximation to (25) can be obtained

$$\mathbf{R}_h[0] \approx \hat{\mathbf{h}}[n\,|\,n-1]\hat{\mathbf{h}}^H[n\,|\,n-1] + \mathbf{P}[n\,|\,n-1] \tag{26}$$

which yields the average Kalman channel estimator of [13].[1]

IV. Soft Input Weighted RLS

If the WSSUS model is not deemed appropriate or not warranted, then the RLS algorithm can also be used as a channel estimator. Application of the RLS algorithm with soft information is discussed in [23]. Here, we propose a soft input weighted RLS algorithm to use the soft statistics more effectively. In [26] and [27], modified least squares cost functions of the desired parameter were effectively used to develop the new adaptive algorithms and to relate the algorithms to the Kalman filter by choosing suitable state space models. In this section, the conventional RLS problem is reformulated to reflect the effect of confidence in the data using a weighting factor such that the following error is recursively minimized:

$$J(\hat{\mathbf{h}}[n]) = \sum_{k=0}^{n}\lambda^{n-k}\frac{|r[k] - \hat{\mathbf{h}}^H[n]\bar{\mathbf{b}}[k]|^2}{q[k]} \tag{27}$$

where $q[k]$ is the noise variance from the previous section, and λ denotes a forgetting factor. The RLS algorithm uses the fading memory parameter λ to track changes in the channel statistics. Similarly, λ together with $q[n]$ in (27) determine the relative

[1]Singularity of the matrix in (26) does not make significant performance loss of the algorithm, unlike the singularity of the correlation matrix of the regressor vector because the matrix in (26) is used only to estimate the signal power within the received signal and, therefore, does not require any type of matrix inversion.

weighting of recent data. In many RLS applications, the noise is modeled as stationary, whereas in this case, the noise variance changes at every instance and is determined by the confidence in the data as well as any additive white Gaussian noise, as in (16). Under the assumption that the channel estimate and error covariance matrix $\mathbf{P}[n]$ of the RLS algorithm are relatively close to those from the soft input Kalman channel estimate, $q[n]$ can be approximated with help of (26) as

$$\hat{q}[n] = \mathrm{tr}\Big\{\mathrm{diag}\big(\sigma_b^2[n], \sigma_b^2[n-1], \ldots, \sigma_b^2[n-L+1]\big) \times \big(\hat{\mathbf{h}}[n-1]\hat{\mathbf{h}}^H[n-1] + \mathbf{P}[n-1]\big)\Big\} + N_0. \quad (28)$$

With aid of (28), the soft input WRLS algorithm is summarized as

$$e[n] = r[n] - \hat{\mathbf{h}}^H[n-1]\bar{\mathbf{b}}[n] \quad (29)$$

$$\hat{q}[n] = \mathrm{tr}\Big\{\mathrm{diag}\big(\sigma_b^2[n], \ldots, \sigma_b^2[n-L+1]\big) \times \big(\hat{\mathbf{h}}[n-1]\hat{\mathbf{h}}^H[n-1] + \mathbf{P}[n-1]\big)\Big\} + N_0 \quad (30)$$

$$\mathbf{k}[n] = \frac{\mathbf{P}[n-1]\bar{\mathbf{b}}[n]}{\lambda\hat{q}[n] + \bar{\mathbf{b}}^H[n]\mathbf{P}[n-1]\bar{\mathbf{b}}[n]} \quad (31)$$

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \mathbf{k}[n]e^*[n] \quad (32)$$

$$\mathbf{P}[n] = \frac{1}{\lambda}(\mathbf{I} - \mathbf{k}[n]\bar{\mathbf{b}}^H[n])\mathbf{P}[n-1]. \quad (33)$$

As the variables of the RLS algorithm were shown to have one-to-one correspondence with those of the Kalman algorithm in [25]–[27], it is possible to interpret the soft input WRLS algorithm as the soft input Kalman algorithm for a special case when the state space model is expressed by $\mathbf{F} = \lambda^{-1/2}\mathbf{I}$, $\mathbf{Q}_v = 0$, and $E[g[n]g[n+k]] = \hat{q}[n]\delta[k]$ in (6) and (7).

V. Convergence Analysis

In this section, channel estimation algorithms with either hard decisions or soft decisions as inputs are analyzed in terms of the convergence rate of the MSIE. To make the analysis tractable, we consider a stationary channel for which the state space model is described by $\mathbf{F} = \mathbf{I}$ and $\mathbf{Q}_v = 0$ with the initial channel covariance matrix $\mathbf{R}_h[0]$. To compare the MSIE of the soft input Kalman algorithm with those of other algorithms, we introduce the assumption that the sequence of priors is a stationary random process and its p.d.f. is symmetric about the real and imaginary axes. Consequently, the convergence rate of the MSIE of the hard decision-based Kalman algorithm can be expressed by the second-order statistics such as a correlation coefficient between the true data and hard decisions, whereas that of the soft input Kalman algorithm depends on the higher order statistics, which makes it difficult to obtain its closed-form expression. Therefore, we derive upper and lower bounds of the convergence rate of the MSIE of the soft input Kalman algorithm for a special case of M-PSK signaling scheme instead. As an example, three types of the Kalman channel estimators are tested for binary phase shift keying (BPSK) signaling, and the performance of each algorithm is compared with the results from analysis.

For the case of a stationary channel, given received data of length n, the channel estimation problem using either hard decisions or soft decisions can be reshaped into matrix form by suitable choices of matrix $\mathbf{A}_n$, observations $\mathbf{y}_n$, and channel parameter $\mathbf{x}$ in

$$\mathbf{y}_n = \mathbf{A}_n\mathbf{x} + \mathbf{z}_n \quad (34)$$

where $\mathbf{z}_n$ denotes additive noise with $E[\mathbf{z}_n\mathbf{z}_n^H] = \mathbf{Q}_n$, and $\mathbf{Q}_n$ is positive definite and is uncorrelated with $\mathbf{x}$. For the channel estimation problem, $\mathbf{A}_n$ is an $n \times L$ convolution matrix with either soft decisions or hard decisions as its entries. Using this linear model, the MSIE for each algorithm in the stationary case can be modeled through appropriate selection of $\mathbf{A}_n$ and $\mathbf{Q}_n$. For the model in (34), the resulting Kalman filter would be a recursive implementation of the linear minimum mean square error (LMMSE) solution. Therefore, the resulting optimal filter and corresponding mean square error are summarized as

$$\hat{\mathbf{x}} = \mathbf{T}_n\mathbf{y}_n, \quad (35)$$

$$\mathbf{T}_n = \mathbf{R}_x\mathbf{A}_n^H(\mathbf{A}_n\mathbf{R}_x\mathbf{A}_n + \mathbf{Q}_n)^{-1} = \big(\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n + \mathbf{R}_x\big)^{-1}\mathbf{Q}_n^{-1}\mathbf{A}_n^H \quad (36)$$

$$\mathbf{e}_n = \mathbf{x} - \hat{\mathbf{x}} \quad (37)$$

$$E\left[\mathbf{e}_n\mathbf{e}_n^H \,\middle|\, \mathbf{A}_n, \mathbf{y}_n, \mathbf{Q}_n, \mathbf{R}_x\right] = \big(\mathbf{R}_x^{-1} + \mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n\big)^{-1}. \quad (38)$$

By setting $\mathbf{x} = \mathbf{h}$, $\mathbf{y}_n = (r[0], r[1], \ldots, r[n-1])^T$, and $\mathbf{R}_x = \mathbf{R}_h[0]$, we can examine the behavior of the stationary channel estimation algorithm. For this stationary setup, the MSIE of each of the adaptive algorithms considered here eventually converges to 0 as $n \rightarrow \infty$, but the rates of convergence may be different. The following figure of merit is introduced to compare the convergence rate of these algorithms.

Definition 1: Let $\mathrm{MSIE}[n] \triangleq E[\mathbf{e}_n^H\mathbf{e}_n \,|\, \mathbf{A}_n, \mathbf{y}_n, \mathbf{Q}_n, \mathbf{R}_x]$. For an algorithm with $\lim_{n\rightarrow\infty} \mathrm{MSIE}[n] = 0$, the convergence rate is defined as

$$\kappa \triangleq \lim_{n\rightarrow\infty} n\mathrm{MSIE}[n] \quad (39)$$

when the limit exists.

Using this definition, κ can be computed for the estimator of (35)–(38) as

$$\kappa = \lim_{n\rightarrow\infty} n\mathrm{tr}\Big\{\big(\mathbf{R}_x^{-1} + \mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n\big)^{-1}\Big\} \quad (40)$$

$$= \lim_{n\rightarrow\infty} \mathrm{tr}\left\{\left(\frac{1}{n}\mathbf{R}_x^{-1} + \frac{1}{n}\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n\right)^{-1}\right\} \quad (41)$$

$$= \mathrm{tr}\left\{\left(\lim_{n\rightarrow\infty}\frac{1}{n}\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n\right)^{-1}\right\}. \quad (42)$$

Here, we assume positive definiteness of $\mathbf{R}_x^{-1}$ to guarantee existence of the inverse matrix. The following theorem summarizes the MSIE for the case analogous to the soft input Kalman channel estimator and that for the trained Kalman channel estimator.

Theorem 1: Define $\mathcal{S}_n = (\mathcal{P}_n, \mathcal{P}_{n-1}, \ldots, \mathcal{P}_{n-L+1})$. If the stochastic process $\mathcal{S}_n$ is a stationary Markov process, then the convergence rate of the soft input Kalman algorithm κ_s and that of the trained Kalman algorithm κ_t are given by

$$\kappa_s = \text{tr}\{\mathbf{R}_s^{-1}\} \tag{43}$$

$$\kappa_t = \text{tr}\{\mathbf{R}_t^{-1}\} \tag{44}$$

where $\mathbf{R}_s = E[(\bar{\mathbf{b}}[n]\bar{\mathbf{b}}^H[n])/(N_0 + \text{tr}\{\mathbf{K}_{\tilde{b}}[n]\mathbf{R}_h[0]\})]$, and $\mathbf{R}_t = E[(\mathbf{b}[n]\mathbf{b}^H[n])/(N_0)]$.

Proof: By setting the ith row of $\mathbf{A}_n$ equal to $\bar{\mathbf{b}}^H[i]$, and $\mathbf{Q}_n = \text{diag}(q[0], q[1], \ldots, q[n-1])$, where $q[n] = N_0 + \text{tr}\{\mathbf{K}_{\tilde{b}}[n,n]\mathbf{R}_h[0]\}$ in (16), the estimator of (35)–(38) characterizes the soft input Kalman channel estimator and the corresponding MSIE for a stationary channel. Then, it follows that $\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n = \sum_{i=0}^{n-1}(\bar{\mathbf{b}}[i]\bar{\mathbf{b}}^H[i])/(q[i])$ and that $\bar{\mathbf{b}}[i]$ and $(\bar{\mathbf{b}}[i]\bar{\mathbf{b}}^H[i])/(q[i])$ are continuous and bounded functions of $\mathcal{S}_n$, because $q[i] \geq N_0 > 0$, and both $\bar{\mathbf{b}}[n]$ and $q[i]$ are functions of $\mathcal{S}_n$. Since $\mathcal{S}_n$ is a stationary Markov process, the ergodic theorem (see [29, Sec. 6.3]) can be applied. As a result, $(1/n)\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n \xrightarrow{p} E[(\bar{\mathbf{b}}[i]\bar{\mathbf{b}}^H[i])/(q[i])]$ as $n \to \infty$, where $\xrightarrow{p}$ denotes the convergence in probability. When the data is error free, $q[i] = N_0$ and $\bar{b}[i] = b[i]$ for all i, which leads to $(1/n)\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n = (1/n)\sum_{i=0}^{n-1}(\mathbf{b}[i]\mathbf{b}^H[i]/N_0)$, which converges to $\mathbf{R}_t$ in probability. Substituting $(1/n)\mathbf{A}_n^H\mathbf{Q}_n^{-1}\mathbf{A}_n$ by $\mathbf{R}_s$ or $\mathbf{R}_t$ in (42) completes the proof. ■

For the hard decision-based Kalman channel estimator, the matrix $\mathbf{A}_n$ used to construct the LMMSE filter $\mathbf{T}_n$ is not known to receiver; however, $\hat{\mathbf{A}}_n$ is used instead based on hard decisions from the decoder, which may in general contain errors. It turns out in this case that there exists some residual MSIE, even for arbitrarily many observations, unless the symbol error rate is zero. Let $\hat{\mathbf{T}}_n$ be the LMMSE filter constructed from hard decisions $\hat{b}[i]$. Then, the MSIE of the hard decision-based Kalman channel estimator can be modeled in the following theorem.

Theorem 2: Assume $E[\hat{\mathbf{b}}[n]\hat{\mathbf{b}}^H[n]] = E[\mathbf{b}[n]\mathbf{b}^H[n]]$, and $E[\hat{\mathbf{b}}[n]\mathbf{b}^H[n]] = \rho E[\mathbf{b}[n]\mathbf{b}^H[n]]$, where $\hat{\mathbf{b}}[n] = (\hat{b}[n], \hat{b}[n-1], \ldots, \hat{b}[n-L+1])^T$. The MSIE of the hard decision based Kalman algorithm $\text{MSIE}_h[n]$ is given as follows:
For $\rho \neq 1$,

$$\text{MSIE}_h[n] \xrightarrow{p} |1-\rho|^2\text{tr}\{\mathbf{R}_h[0]\}. \tag{45}$$

For $\rho = 1$

$$\text{MSIE}_h[n] \to 0 \text{ as } n \to \infty, \quad \text{and} \quad \kappa_h = \kappa_t.$$

Proof: The estimation error given hard decisions can be expressed in term of $\hat{\mathbf{T}}_n$ as $\mathbf{e}_n = (\hat{\mathbf{T}}_n\mathbf{A}_n - \mathbf{I})\mathbf{x} - \hat{\mathbf{T}}_n\mathbf{z}$, and the MSIE can be found by computing the trace of $E[\mathbf{e}_n\mathbf{e}_n^H \mid \hat{\mathbf{A}}_n, \mathbf{y}_n, \mathbf{Q}_n, \mathbf{R}_x] = (\hat{\mathbf{T}}_n\mathbf{A}_n - \mathbf{I})\mathbf{R}_x(\hat{\mathbf{T}}_n\mathbf{A}_n - \mathbf{I})^H + \hat{\mathbf{T}}_n\mathbf{Q}_n\hat{\mathbf{T}}_n^H$. Noting that $\mathbf{Q}_n = N_0\mathbf{I}$, the first term of the right side is computed using $\hat{\mathbf{T}}_n\mathbf{A}_n = ((1/n)\mathbf{R}_x^{-1} + (1/n)(\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n/N_0))^{-1}((1/n)(\hat{\mathbf{A}}_n^H\mathbf{A}_n/N_0)) \xrightarrow{p} (E[(\hat{\mathbf{b}}[n]\hat{\mathbf{b}}^H[n]/N_0)])^{-1}E[(\hat{\mathbf{b}}[n]\mathbf{b}^H[n]/N_0)] = (\mathbf{R}_t)^{-1}\rho\mathbf{R}_t = \rho\mathbf{I}$. Hence, $(\hat{\mathbf{T}}_n\mathbf{A}_n - \mathbf{I})\mathbf{R}_x(\hat{\mathbf{T}}_n\mathbf{A}_n - \mathbf{I})^H \xrightarrow{p} |1-\rho|^2\mathbf{R}_h[0]$. The second term of the right side is can be expressed as

$$\begin{aligned}
&\lim_{n\to\infty} \hat{\mathbf{T}}_n\mathbf{Q}_n\hat{\mathbf{T}}_n^H \\
&= \lim_{n\to\infty} \frac{1}{n}\left(\frac{1}{n}\mathbf{R}_x^{-1} + \frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\right)^{-1} \\
&\quad \times \frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\left(\frac{1}{n}\mathbf{R}_x^{-1} + \frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\right)^{-1} &(46)\\
&= \left(\lim_{n\to\infty}\frac{1}{n}\right)\left(\lim_{n\to\infty}\left(\frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\right)^{-1}\right. \\
&\quad \left. \times \left(\frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\right)\left(\frac{1}{n}\frac{\hat{\mathbf{A}}_n^H\hat{\mathbf{A}}_n}{N_0}\right)^{-1}\right) &(47)\\
&= 0 \cdot (\mathbf{R}_d)^{-1} = 0. &(48)
\end{aligned}$$

As a result, $\text{MSIE}_h[n] \xrightarrow{p} |1-\rho|^2\text{tr}\{\mathbf{R}_h[0]\}$. If $\rho = 1$, the first term disappears, and κ_h exists and is equal to κ_t. ■

Remark: The assumption that $E[\hat{\mathbf{b}}[n]\hat{\mathbf{b}}^H[n]] = E[\mathbf{b}[n]\mathbf{b}^H[n]]$ can be justified when the distribution of $\hat{\mathbf{b}}[n]$ is identical to that of $\mathbf{b}[n]$, which can be obtained when a fair detection rule is used. When $\mathbf{b}^H[n]$ is zero mean and white, $E[\hat{\mathbf{b}}[n]\mathbf{b}^H[n]]$ is a diagonal matrix, and $\rho = (E[\hat{b}[n]b^*[n]]/E[b[n]b^*[n]])$. For example, ρ is equal to the bit error rate (BER) for BPSK.

We can now examine the relative convergence behavior of each of the algorithms considered by investigating $\mathbf{R}_t$, $\mathbf{R}_s$, and ρ. Here, comparisons are made under the assumption that $\mathcal{P}_n$ is i.i.d., which is reasonable for a practical system. It follows that $\mathbf{R}_t = q_t\mathbf{I}$ for some positive real-valued q_t, and, additionally, it can be shown that $\mathbf{R}_s = q_s\mathbf{I}$ for some q_s under reasonable assumptions on the soft statistics.

Assumption 1: Let $\text{Re}(x)$ and $\text{Imag}(x)$ denote the real and imaginary parts, respectively, of the complex variable x. The random processes $\text{Re}(\bar{b}[n])$ and $\text{Imag}(\bar{b}[n])$ are i.i.d. with an even distribution function, i.e., $P\{\text{Re}(\bar{b}[n]) = a\} = P\{\text{Re}(\bar{b}[n]) = -a\}$, for any real a.

Assumption 1 can be justified for most practical signaling schemes (e.g., QAM, M-PSK). For example, in BPSK, the distribution of the LLR of the decoder output at time n is assumed given by $P\{\text{LLR}[n] \mid b[n] = 1\} = \mathcal{N}((\sigma^2/2), \sigma^2)$, and $P\{\text{LLR}[n] \mid b[n] = -1\} = \mathcal{N}(-(\sigma^2/2), \sigma^2)$ for some real σ, where $\text{LLR}[n] \triangleq \log(p_1[n]/p_0[n])$. When $b[n]$ is transmitted with equal probability, we have that $P\{\text{LLR}[n]\} = 0.5\mathcal{N}((\sigma^2/2), \sigma^2) + 0.5\mathcal{N}(-(\sigma^2/2), \sigma^2)$, and this can be parameterized by $p_0[n]$ with $p_0[n] = (1/1 + \exp(\text{LLR}[n]))$, which indicates that $\mathcal{P}_n = (p_0[n], p_1[n])^T$ is a stationary random process. The p.d.f. of $\bar{b}[n]$ is an even function because $P\{\bar{b}[n] = a\} = P\{\text{LLR}[n] = \tanh^{-1}(a)\} = P\{\text{LLR}[n] = -\tanh^{-1}(a)\} = P\{\bar{b}[n] = -a\}$ using $\bar{b}[n] = 1 - 2p_0[n] =$

TABLE I
ASYMPTOTIC MSIE FOR KALMAN CHANNEL ESTIMATION

	κ	$MSIE(n) \sim$
Trained	Lq_t^{-1}	$L(nq_t)^{-1}$
Hard decision based	$N.A.$	$\|1-\rho\|^2 + L(nq_t)^{-1}$
Soft input	Lq_s^{-1}	$L(nq_s)^{-1}$

$\tanh((\text{LLR}[n]/2))$, and $\text{LLR}[n] = 2\tanh^{-1}(\bar{b}[n])$. The task is to show $\mathbf{R}_s = q_s\mathbf{I}$ for some q_s; equivalently

$$[\mathbf{R}_s]_{i,j} = E\left[\frac{\bar{b}[n-i+1]\bar{b}_j^*[n-j+1]}{N_0 + \text{tr}\{\mathbf{K}_{\tilde{b}}[n,n]\mathbf{R}_h[0]\}}\right] \quad (49)$$

$$= \begin{cases} q_s, & \text{if } i=j \\ 0, & \text{otherwise} \end{cases} \quad (50)$$

where $1 \leq i,j \leq L$. When $i=j$, the result is trivial. For the case $i \neq j$, we will use the following lemma.

Lemma 1: Let x and y be i.i.d. real-valued random variables with $E[|x|] < \infty$ and $E[|y|] < \infty$. If the p.d.f. for each is an even function, and $w(x,y)$ is a positive, even function of both x and y, then it follows that

$$E\left[\frac{xy}{w}\right] = 0. \quad (51)$$

Proof: Since $w(x,y)$ is even in x, $(x/w(x,y))$ is an odd function in x. Since the p.d.f. of x is even, $E[(x/w(x,y)) \mid y] = 0$. Considering that $E[(xy/w)] = E[yE[(x/w) \mid y]]$, the result follows immediately. ■

From Assumption 1, the real and imaginary parts of $\bar{b}[n-i+1]$ and $\bar{b}[n-j+1]$ are i.i.d., and the denominator of (49)—$N_0 + \text{tr}\{\mathbf{K}_{\tilde{b}}[n,n]\mathbf{R}_h[0]\}$—is an even function with regard to the real and imaginary parts of $\bar{b}[n-i+1]$ and $\bar{b}[n-j+1]$ by $\mathbf{K}_{\tilde{b}}[n,n] = \text{diag}(E[|b[n]|^2] - |\bar{b}[n]|^2, \ldots, E[|b[n-L+1]|^2] - |\bar{b}[n-L+1]|^2)$. Therefore, it is possible to apply Lemma 1 by choosing the real and imaginary parts of $\bar{b}[n-i+1]$ and $\bar{b}[n-j+1]$ in (49) as x and y, which proves that the off-diagonal elements of $\mathbf{R}_s$ are zero.

For sufficiently large n, the asymptotic analysis provides that the MSIE of each algorithm can be approximated as shown in Table I.

Consider the case of M-PSK signaling, and, without loss of generality, let $\mathbf{R}_h = \sigma_h^2 \text{diag}(\nu_0, \nu_1, \ldots, \nu_{L-1})$ for some σ_h, where $\sum_{k=0}^{L-1} \nu_k = 1$. Since the magnitude of each symbol is identical for M-PSK, i.e., $E_{\mathcal{P}}[|b[n]|^2] = |b[n]|^2 = \sigma_s^2$ regardless of $\mathcal{P}$, it follows that $\sigma_{\tilde{b}}^2 = \sigma_s^2 - |\bar{b}[n]|^2$. This provides that $q_s = E[(\bar{b}^2[n]/N_0 + \sigma_h^2\sigma_s^2 - \sum_{k=0}^{L-1} \nu_k |\bar{b}[n]|^2)]$, which, since $\bar{b}[n]$ is i.i.d., can be written as

$$q_s = E\left[\frac{\sum_{k=0}^{L-1} \nu_k |\bar{b}[n-k]|^2}{N_0 + \sigma_h^2\sigma_s^2 - \sigma_h^2 \sum_{k=0}^{L-1} \nu_k |\bar{b}[n-k]|^2}\right]. \quad (52)$$

Therefore, $q_s = E[(\eta/(N_0 + \sigma_h^2\sigma_s^2) - \sigma_h^2\eta)]$, where $\eta = \sum_{k=0}^{L-1} \nu_k |\bar{b}[n-k]|^2$. Hence, the dependence of the MSIE on the soft statistics can be reflected in a single variable η. Nevertheless, it is still a difficult task to evaluate q_s even for the BPSK case because η is a nonlinear function of $\bar{b}[n], \bar{b}[n-1], \ldots, \bar{b}[n-k+1]$. However, it is possible to obtain upper and lower bounds of q_s using Lemma 3 below.

The following Lemma is useful to obtain the necessary performance bound.

Lemma 2: Let $X_1, X_2, \ldots, X_L$ be i.i.d. random variables and $\sum_{k=1}^{L} c_k = 1$. If $f(x)$ is a convex function of x, then the following inequality holds:

$$E\left[f\left(\sum_{j=1}^{L} c_j X_j\right)\right] \leq E[f(X_1)]. \quad (53)$$

Proof: Using the Jensen's inequality for a convex function $f(x)$, $f(\sum_{j=1}^{L} c_j X_j) \leq \sum_{j=1}^{L} c_j f(X_j)$, it follows that $f(\sum_{j=1}^{L} c_j X_j) \leq \sum_{j=1}^{L} c_j f(X_j)$. Since $X_1, X_2, \ldots, X_L$ are i.i.d.

$$E\left[f\left(\sum_{j=1}^{L} c_j X_j\right)\right] \leq \sum_{j=1}^{L} c_j E[f(X_j)] = E[f(X_1)]. \quad (54)$$

■

Lemma 3: The following inequality with regard to q_s holds for M-PSK signaling

$$q_0 \leq q_s \leq q_1 \quad (55)$$

where

$$q_0 = \frac{E[|\bar{b}[n]|^2]}{(N_0 + \sigma_h^2\sigma_s^2) - \sigma_h^2 E[|\bar{b}[n]|^2]}$$

and

$$q_1 = E\left[\frac{|\bar{b}[n]|^2}{(N_0 + \sigma_h^2\sigma_s^2) - \sigma_h^2 |\bar{b}[n]|^2}\right].$$

Proof: First, let $f(x) = (x/(N_0 + \sigma_h^2\sigma_s^2) - \sigma_h^2 x)$, which is a convex function of x for $0 \leq x \leq \sigma_s^2$. Then, q_s can be rewritten as $q_s = E[f(\eta)]$, with $\eta = \sum_{k=0}^{L-1} \nu_k |\bar{b}[n-k]|^2$ and $\sum_{k=0}^{L-1} \nu_k = 1$. If we take $X_i = |\bar{b}[n-i]|^2$ and $c_i = \nu_i$ in Lemma 2, then it results in $q_s = E[f(\sum_{k=0}^{L-1} \nu_k |\bar{b}[n-k]|^2)] \leq E[f(|\bar{b}[n]|^2)] = q_1$. The lower bound can be obtained readily using Jensen's inequality. Since the function $f(x)$ is a convex function, $q_s = E[f(\eta)] \geq f(E[\eta]) = f(E[|\bar{b}[n]|^2])$, which completes the proof. ■

Note that q_0 and q_1 can be computed by taking an expectation over a single variable $\bar{b}[n]$, whereas q_s requires the integration over the multiple variables $\bar{b}[n], \bar{b}[n-1], \ldots, \bar{b}[n-k+1]$ in η. When M-PSK signaling is used, the following inequality with regard to κ_s then holds for sufficiently large n:

$$Lq_1^{-1} \leq \kappa_s \leq Lq_0^{-1}. \quad (56)$$

From Theorems 1 and 2, we see that the trained Kalman channel estimator outperforms both the soft input and hard de-

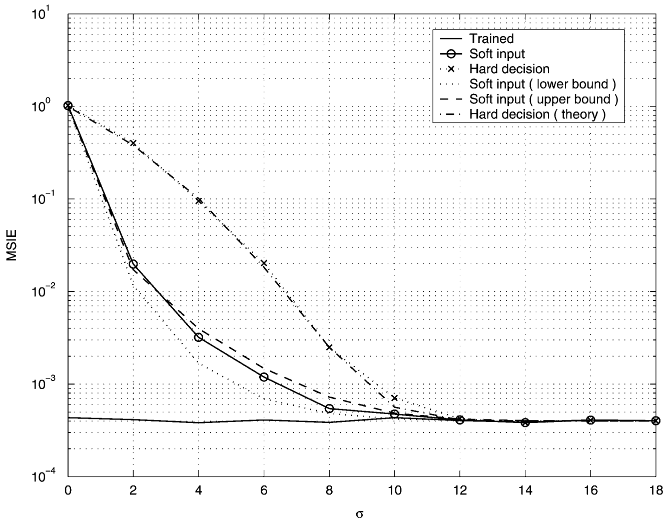


Fig. 1. MSIE performance comparison under Gaussian assumption $L = 4$, SNR $= 10$ dB.

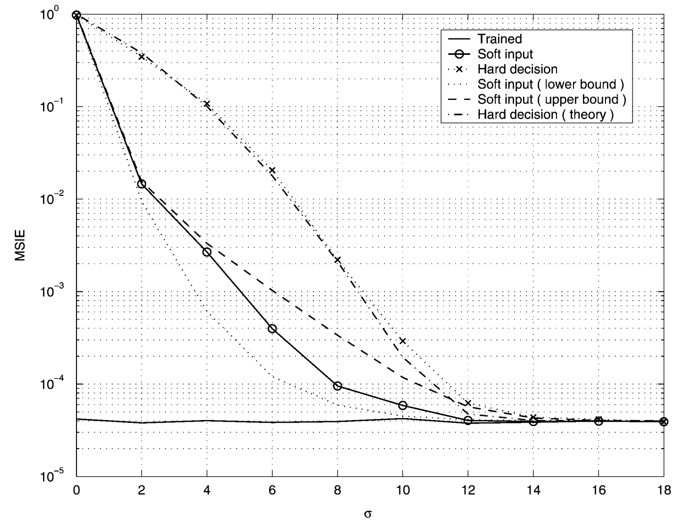


Fig. 2. MSIE performance comparison under Gaussian assumption $L = 4$, SNR $= 20$ dB.

cision-based Kalman algorithms since $\kappa_t \geq \kappa_s$ and, due to the residual $\mathrm{MSIE}_h[n]$ for $\rho \neq 1$, for sufficiently large n. Performance comparisons between the hard decision-based Kalman algorithm and the soft input Kalman algorithm can be further explained by considering the distribution of the log likelihood $\mathrm{LLR}[n]$ in order to link ρ, q_t, and q_s for the hard decision-based Kalman and soft input Kalman algorithms, respectively. The following example shows the performance gain achieved by the soft input Kalman based on the assumption that the LLRs of the decoded bits are described by the widely used Gaussian distribution LLR $\sim \mathcal{N}(\pm(\sigma^2/2), \sigma^2)$ [18], [19].

Example: Consider BPSK signaling, for which $q_t = 1$, and $\rho = 1 - 2p_e$ can be computed using the probability of error $p_e = Q(\sigma^2/2)$, where $Q(x) = \int_x^\infty e^{-(\tau^2/2)}\, d\tau$. In the simulations, 1000 BPSK symbols were generated with Gaussian distributed LLRs of each symbol and passed through a 4-tap, equal power, fading channel, i.e., $\mathbf{R}_h = (1/L)\mathbf{I}$. Each performance curve was obtained by averaging the MSIE of 100 different realizations of the channel. Figs. 1 and 2 show performance curves of the different algorithms obtained from theoretical calculation and simulations for 10 and 20 dB, where SNR $\triangleq (E[|\mathbf{h}^H[n]\mathbf{b}[n]|^2]/N_0)$. The MSIE of the hard decision-based Kalman channel estimator follows closely the theoretical performance curve, and the performance of the soft input Kalman channel estimator is well bounded by the upper and lower limits as computed.

VI. Simulations and Results

Each of the algorithms are tested in the communication system in Fig. 3 and compared in terms of their overall BER. The performance should be verified as a part of an overall communication system because the LLRs, which were assumed accurate in Section V, tend to be noisy due to, among other reasons, errors in channel knowledge in the SISO detector.

Soft information can serve two purposes, namely, 1) refining the channel estimate with extra data other than training sequence for time-invariant channels and 2) improving tracking capability

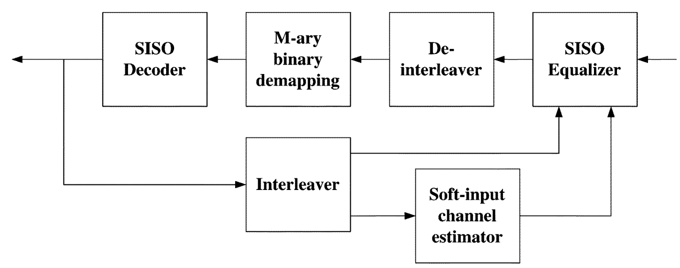


Fig. 3. Receiver structure with a decision-directed channel estimator.

by reducing error propagation due to the relatively large detection amplitude errors of hard decisions for dynamic channels. Therefore, simulations are performed under the two different scenarios, i.e., for time-invariant and time-varying channels.

For the following simulations, the rate $1/2$ recursive systematic convolutional (RSC) code with generation polynomials equal to (23, 35) (expressed in octal) is used in an encoder, and each pair of code bits is mapped to one of the QPSK symbols. Interleaved symbols are transmitted over a frequency-selective channel. At the receiver, the LOGMAP algorithm is employed as the SISO detector and the decoder. The decoded output is deinterleaved and fed to the channel estimator. To show the benefit from the soft information used in the channel estimator, we successively tested the turbo equalizer with perfect channel knowledge, the hard decision-directed channel estimate, and the soft decision-based channel estimate. In addition, we tested the receivers equipped with the soft input WRLS algorithm and the conventional hard decision-based RLS algorithm. For convenience, we introduced the following acronyms:

- SKTE: turbo equalizer with the soft input Kalman channel estimator;
- HKTE: turbo equalizer with the hard decision based Kalman channel estimator;
- SWRLS: turbo equalizer with the soft input WRLS channel estimator;
- HRLS: turbo equalizer with the hard decision-based RLS channel estimator.

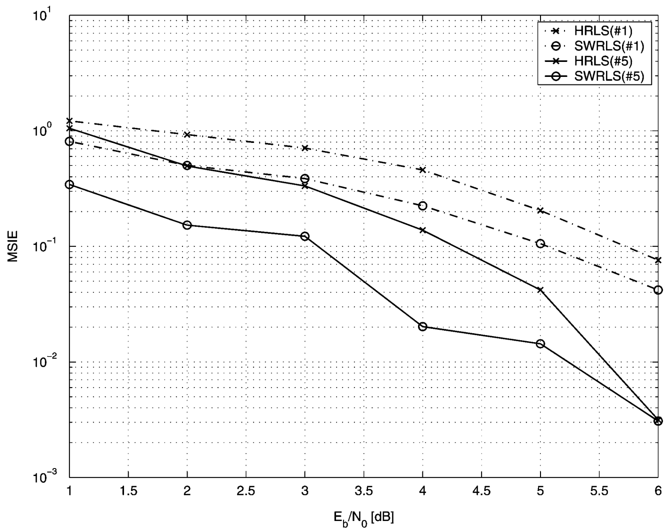


Fig. 4. MSIE performance after the first and the fifth iteration with a time invariant channel.

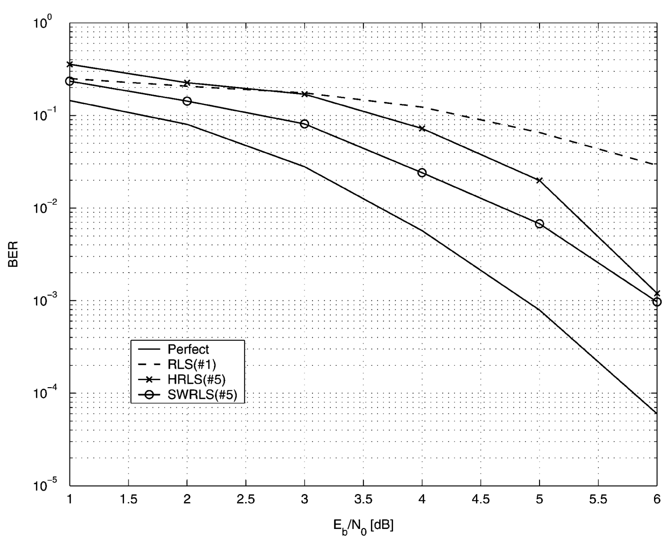


Fig. 5. BER performance after the fifth iteration with a time invariant channel.

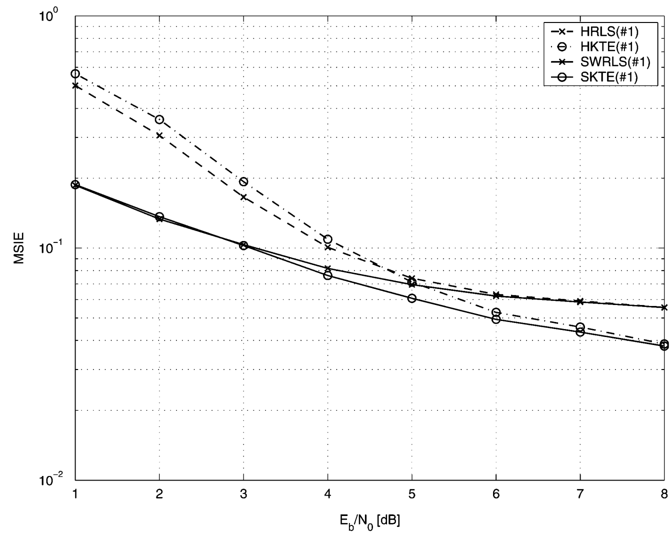


Fig. 6. MSIE performance after the first iteration with a first order AR channel model.

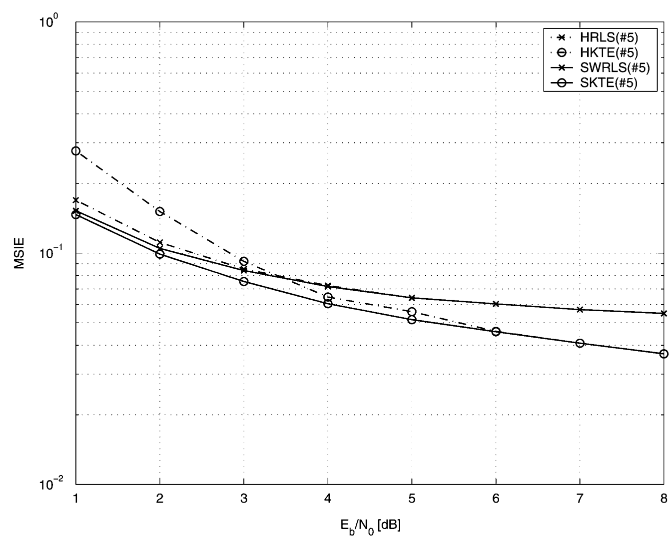


Fig. 7. MSIE performance after the fifth iteration with a first-order AR channel model.

For the first iteration, the fixed channel estimate obtained from a training sequence is used in the SISO detector. Therefore, the BER performance of the first iteration is identical for all receivers, except for the turbo receiver with perfect channel knowledge. Since the channel estimate is unreliable until the first iteration is complete, soft information from the decoder is not used in the equalizer until the first channel estimate has been used. All performance curves are obtained by averaging more than 100 different realizations of the channel specified by each simulation set.

A. Time-Invariant Channel

In the first set of simulations, 500 QPSK symbols are encapsulated into one data frame with ten training symbols up front, which are sent through a three-tap time-invariant channel $\mathbf{h} = (-0.691 - 0.501i, 0.361 + 0.506i, -0.528 - 0.408i)^T$. To see the refining of the channel estimate, we choose a relatively short training period compared with the overall data block length. For the training period, the conventional RLS algorithm is used to estimate the channel and is then switched to either the HRLS algorithms or the SWRLS algorithm. The forgetting factors of these algorithms was set to 0.99. Fig. 4 presents the MSIE performance of the various receivers in terms of E_b/N_0. It is observed that soft input algorithm achieves more than 1 dB E_b/N_0 gain in the terms of MSIE performance compared with the hard decision-based algorithm. Fig. 5 shows the performance gain of the soft input algorithms in terms of BER, which shows that the gain in BER is approximately 1 dB.

B. Time-Varying Channel

For the dynamic situation, a 228×10 block interleaver was used, and ten consecutive frames with 114 QPSK symbols and 26 training symbols (which gives comparable throughput to that of GSM [24]) are transmitted over a two-tap, equal power, Rayleigh fading channel, satisfying the WSSUS channel model. Note that in this set of simulations, the fixed channel

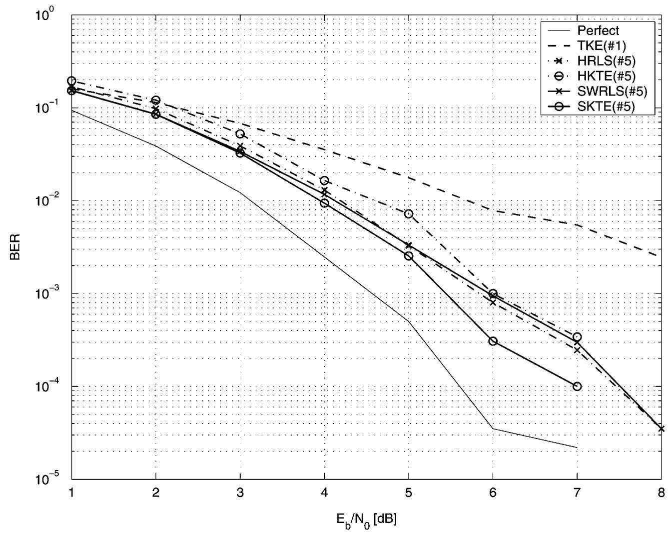


Fig. 8. BER performance after the fifth iteration with a first-order AR channel model.

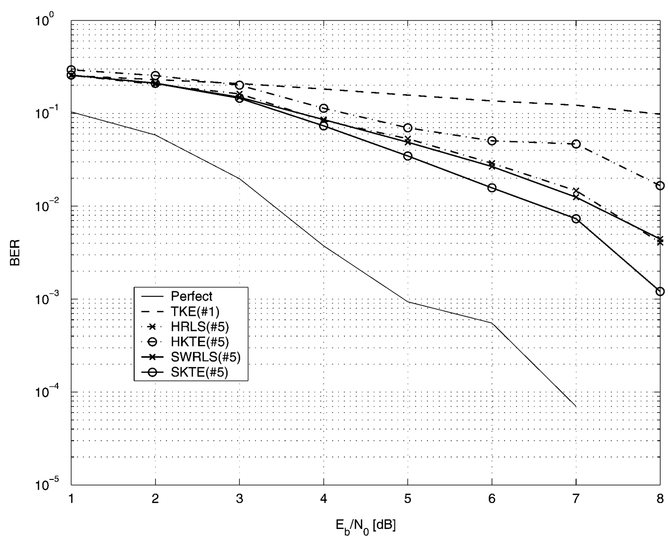


Fig. 9. BER performance comparison after the fifth iteration for a Rayleigh fading channel at $f_d = 0.002$.

estimate obtained from the training sequence is used to compute the metric in the LOGMAP detector in the first iteration; meanwhile, through the following iterations, the time-varying channel estimate from an adaptive algorithm is applied. The proposed algorithms are applied to time-varying channels, which are generated by the state space model with $\mathbf{F} = \lambda^{1/2}\mathbf{I}$ and $\mathbf{Q}_v = (1 - \lambda)\mathbf{I}$, where λ is fixed 0.999. At the receiver, the exact state space model is available and used in the SKTE and the HKTE. Experimental results for the various algorithms are presented in Figs. 6 and 7. Here, it is observed that the SKTE provides performance improvement over other algorithms in terms of the MSIE. In Fig. 8, the BER performance gain of the SKTE becomes larger as E_b/N_0 increases.

The algorithms are also applied to a more realistic setup, in which a Rayleigh fading channel with normalized Doppler spread $f_d = 0.002$ is assumed. For the SKTE and the HKTE, the state-space model is approximated with a first-order autoregressive model with the parameter determined as in [28]. Fig. 9 illustrates the BER performance curves for various algorithms. It is observed that the SKTE achieves a lower BER than other algorithms for the entire tested range of E_b/N_0, and the performance gain continues to increase as E_b/N_0 becomes large.

VII. Conclusion

This paper develops and studies several channel estimation algorithms that can exploit soft statistics available in a turbo equalization-based receiver under various channel conditions. The benefits of channel estimator are verified by convergence analysis and through simulations.

References

[1] C. Douillard *et al.*, “Iterative correction of intersymbol interference: Turbo equalization,” *Eur. Trans. Telecommun.*, vol. 6, pp. 507–511, Sept.–Oct. 1995.

[2] M. Tüchler, A. C. Singer, and R. Koetter, “Minimum mean squared error equalization using *a priori* information,” *IEEE Trans. Signal Processing*, vol. 50, pp. 673–683, Mar. 2002.

[3] M. Tüchler, R. Koetter, and A. C. Singer, “Turbo equalization: Principles and new results,” *IEEE Trans. Commun.*, vol. 50, pp. 754–767, May 2002.

[4] A. Glavieux, C. Laot, and J. Labat, “Turbo equalization over a frequency selective channel,” in *Proc. Int. Symp. Turbo Codes Related Topics*, Sept. 1997, pp. 96–102.

[5] X. Wang and H. Poor, “Iterative (turbo) soft interference cancellation and decoding for coded CDMA,” *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, July 1999.

[6] P. D. Alexander *et al.*, “Iterative multiuser interference reduction: Turbo CDMA,” *IEEE Trans. Commun.*, vol. 47, pp. 1008–1014, July 1999.

[7] B. Lu and X. Wang, “Iterative receivers for multiuser space-time coding systems,” *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, Nov. 2000.

[8] X. Wang and R. Chen, “Blind turbo equalization in Gaussian and impulsive noise,” *IEEE Trans. Vehicular Tech.*, vol. 50, pp. 1092–1105, July 2001.

[9] R. Chen, J. S. Liu, and X. Wang, “Convergence analysis and comparisons of Markov chain Monte Carlo algorithms in digital communications,” *IEEE Trans. Signal Processing*, vol. 50, pp. 255–270, Feb. 2002.

[10] E. Baccarelli, R. Cusani, and S. Galli, “A novel adaptive receiver with enhanced channel tracking capability for TDMA-based mobile radio communication,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1630–1639, Dec. 1998.

[11] E. Baccarelli and S. Galli, “A new approach based on soft statistics to the nonlinear blind-deconvolution of unknown data channels,” *IEEE Trans. Signal Processing*, vol. 49, pp. 1481–1491, July 2001.

[12] L. M. Davis, I. B. Collings, and P. Hoeher, “Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels,” *IEEE Trans. Commun.*, vol. 49, pp. 2106–2114, Dec. 2001.

[13] A. Anastasopoulos and K. M. Chugg, “Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty,” *IEEE Trans. Commun.*, vol. 48, pp. 1638–1648, Oct. 2000.

[14] P. Frenger and A. Svensson, “Decision directed coherent detection in multicarrier systems on Rayleigh fading channels,” *IEEE Trans. Veh. Technol.*, vol. 48, pp. 490–498, Mar. 1999.

[15] V. Mignone and A. Morello, “CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers,” *IEEE Trans. Commun.*, vol. 44, pp. 1144–1151, Sept. 1996.

[16] M. C. Valenti and B. D. Woerner, “Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels,” *IEEE Trans. Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.

[17] J. Thomas and E. Geraniotis, “Soft iterative multisensor multiuser detection in coded dispersive CDMA wireless channels,” *IEEE J. Select. Areas Commun.*, vol. 10, pp. 1334–1351, July 2001.

[18] S. ten Brink, “Convergence of iterative decoding,” *Electron. Lett.*, vol. 35, pp. 806–808, May 1999.

[19] ——, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Trans. Commun.*, vol. 40, pp. 1727–1737, Oct. 2001.

[20] P. A. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 360–393, Dec. 1963.

[21] J. G. Proakis, *Digital Communication*. New York: McGraw-Hill, 1989.

[22] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[23] M. Tüchler, R. Otnes, and A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalization," in *Proc. Int. Contr. Conf.*, vol. 3, 2002, pp. 1858–1862.

[24] R. Steele, *Wireless Communications*, London, U.K.: Pentech, 1992.

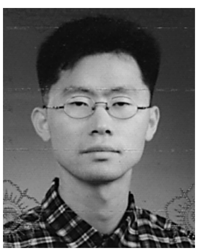
[25] S. Haykin, *Adaptive Filter Theory*, Third ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[26] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Processing*, vol. 45, pp. 1118–1128, May 1997.

[27] A. H. Sayed and T. Kailath, "State-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, pp. 18–60, July 1994.

[28] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 941–950, Feb./Mar./Apr. 1994.

[29] S. R. S. Varadhan, *Probability Theory*. New York: AMSCIMS, 2001.



Seongwook Song received the B.S. and M.S. degrees in electrical engineering from the Seoul National University, Seoul, Korea, in 1997 and 1999, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 2004. He also received the Ph.D. degree in acoustics from the Seoul National University.

Since 2003, he has worked as a Senior Engineer at Samsung Electronics, Suwon, Korea. His research interests are in the general areas of communication systems, advanced signal processing for digital communications, and audio signal processing.

Andrew C. Singer was born in Akron, OH, in 1967. He received the S.B., S.M., and Ph.D. degrees, all in electrical engineering and computer science, from the Massachusetts Institute of Technology (MIT), Cambridge, in 1990, 1992, and 1996, respectively.

Since 1998, he has been with the faculty of the Department of Electrical and Computer Engineering (ECE), University of Illinois, Urbana-Champaign, where he is a Willett Faculty Scholar and an Associate Professor with ECE and a Research Associate Professor with the Coordinated Science Laboratory. In 1996, he was a Postdoctoral Research Affiliate with the Research Laboratory of Electronics, MIT. From 1996 to 1998, he was a Research Scientist with Sanders, A Lockheed Martin Company, Manchester, NH. His research interests include statistical signal processing and communication, universal prediction and data compression, and machine learning.

Dr. Singer was a Hughes Aircraft Masters Fellow and was the recipient of the Harold L. Hazen Memorial Award for excellence in teaching in 1991. In 2000, he received the National Science Foundation CAREER Award, and in 2001, he recieved the Xerox Faculty Research Award. He is currently a member of the MIT Educational Council, Eta Kappa Nu, and Tau Beta Pi.

Koeng-Mo Sung (M'83) was born in Incheon, Korea, in 1947. He received the Dipl.-Ing. degree in communication engineering in 1977 and the Dr.-Ing. degree in acoustics in 1982 from Technische Hochschule Aachen (RWTH Aachen), Aachen, Germany.

He was with the Department of Electronics Engineering, Seoul National University, Seoul, Korea, from 1965 to 1971. He was a research engineer at RWTH Aachen from 1977 to 1983. Since 1983, he has been with Seoul National University, where he is a Professor with the School of Electrical Engineering. His research interests are ultrasonics, musical acoustics, and audio signal processing.

Dr. Sung is a member of the Board of Directors of the Acoustical Society of Korea.