Bayesian Parameter Estimation Using Single-Bit Dithered Quantization

Georg Zeitler, Student Member, IEEE, Gerhard Kramer, Fellow, IEEE, and Andrew C. Singer, Fellow, IEEE

Abstract—The Bayesian parameter estimation problem using a single-bit dithered quantizer is considered. This problem arises, e.g., for channel estimation under low-precision analog-to-digital conversion at the receiver. Based on the Bayesian Cramér-Rao lower bound, bounds on the mean squared error are derived that hold for all dither strategies with strictly causal adaptive processing of the quantizer output sequence. In particular, any estimator using the binary quantizer output sequence is asymptotically (in the sequence length) at least $10 \log_{10}(\pi/2) \approx 1.96 \text{dB}$ worse than the minimum mean squared error estimator using continuous observations, for any dither strategy. Moreover, dither strategies are designed that are shown by simulation to closely approach the derived lower bounds.

EDICS: SSP-PARE, SPC-CEST, SPC-QUAN

I. INTRODUCTION

Analog-to-digital converters (ADCs) form an integral part of modern digital communication systems. However, with increasing bitrates, analog-to-digital (A/D) conversion becomes power-hungry, costly, and time-critical [1]–[3], especially at converter resolutions of 6-12 bits commonly employed today. Examples of such high-speed links include optical transceivers with electronic dispersion compensation for single-mode and multi-mode fiber [4], [5], and chip-to-chip serial links [6].

One approach to reduce the power consumption of ADCs operating at high speeds is to reduce their precision, since it becomes impractical to implement high resolution ADCs at any power budget as ADC speeds increase. In [7], the authors consider the additive white Gaussian noise (AWGN) channel and optimize the ADC and the transmit alphabet to maximize the mutual information between the channel input and the quantizer output, for ADC precisions of 1-3 bits. A related problem is studied in [8], where low-precision ADCs are designed to maximize the information rate over intersymbol-interference (ISI) channels with AWGN; other work on low-precision ADC design for the ISI channel with AWGN includes [9], where the ADC is optimized to minimize the bit error rate (BER) of the link, and [10] where the relative sampling phases of a time-interleaved ADC [11], [12] are chosen to maximize mutual information.

Common to all this work is the assumption of perfect channel state information (CSI) at the receiver, and an important question is how to reliably estimate the channel under low-precision output quantization. In order to study the limit of very coarse quantization we consider a quantizer with only two levels. For such estimation problems, dithered quantizers turn out to be very useful [13]. For example, for a Gaussian prior on the channel coefficients, using the linear minimum mean squared error (MMSE) estimator of the channel as a dither signal is shown to work well in practice [14]. In this paper, we also assume the channel to be estimated to be random, i.e., in contrast to [13] we consider a Bayesian parameter estimation problem as in [14]. For such a single-bit dithered quantizer and a Gaussian prior, we first derive lower bounds on the mean squared error (MSE). In particular, we show that the MSE of any dither strategy is asymptotically (in the quantizer-output sequence length) at least $10 \log_{10}(\pi/2) \approx 1.96 \text{dB}$ worse than the MSE of the MMSE estimator based on unquantized observations. We also design dither strategies that closely approach the derived bounds. Single-bit dithered quantization is also considered in [15] to estimate a deterministic signal parameter, and in [16] the authors characterize optimal additive noise (dither) for parameter estimation based on quantized measurements by formulating a Cramér-Rao lower bound (CRLB) minimization problem. However, in contrast to our work, the dither signal in [16] is not adapted depending on the quantizer output.

Throughout the paper, random variables appear in capital letters, while their realizations are assigned lower case letters. Sequences are denoted by $x^n = \{x_i, x_{i+1}, \ldots, x_j\}$, and we write $x^n$ for the sequence $\{x_1, x_2, \ldots, x_n\}$. Expectation is denoted by $E[\cdot]$, and $\delta(x) = 1$ if $x = 0$; otherwise, $\delta(x) = 0$.

This paper is organized as follows. In Section II, we describe the system model. Lower bounds on the MSE are derived in Section III. Dithering strategies and simulation results are presented in Section V and Section VI, respectively. Concluding remarks appear in Section VII.

We remark that the algorithms we present can also be used for unknown parameters by assigning some prior over the parameter space. However, several of our bounds are based on Gaussian priors and may not apply more generally.
II. SYSTEM MODEL

Consider transmission over the real-valued discrete-time channel with ISI and AWGN, so that the channel output at time $t$ is

$$Y_t = \sum_{\ell=1}^{L} H_\ell X_{t+1-\ell} + N_t, \quad t = 1, 2, \ldots,$$

where the channel of length $L$ has independent random coefficients $H_\ell \sim \mathcal{N}(0, \sigma_H^2)$, the channel input $X_t \in \mathcal{X}$ is real and discrete, and the additive noise $N_t \sim \mathcal{N}(0, \sigma^2)$ satisfies $E[N_t N_{t'}] = \sigma^2 \delta(t - t')$. By assuming the length $L$ of the channel is known at the receiver and by employing a simple periodic training sequence with period $L$, the channel estimation problem for the channel (1) can be decomposed into $L$ parallel estimation problems [14], one for each $H_\ell$, $\ell = 1, 2, \ldots, L$. More precisely, let

$$X_t = \begin{cases} 1 & \text{if } (t-1) \mod L = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$Y_t = H_1+(t-1) \mod L + N_t,$$

so that the estimation problem consists of $L$ independent parallel estimation problems, as illustrated in Fig. 1. Therefore, we only consider one of those $L$ estimation problems in the sequel.

Suppose that the real-valued random parameter $H$ is corrupted by AWGN, and the receiver observes

$$Y_i = H + N_i,$$

where $H \sim \mathcal{N}(0, \sigma_H^2)$, $N_i \sim \mathcal{N}(0, \sigma^2)$, and the sequence $\{N_i\}$ is independent and identically distributed (i.i.d.) and independent of $H$. The quantizer output $Z_i$ at time $i$ is

$$Z_i = Q(Y_i - D_i),$$

where $D_i$ is a real-valued dither signal to be designed, and the quantization function $Q: \mathbb{R} \to \{0, 1\}$ satisfies

$$Q(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{if } y \geq 0. \end{cases}$$

The signal-to-noise ratio (SNR) is defined as $\sigma_H^2/\sigma^2$.

It remains to design the dither signal $D_i$. We permit the receiver to use a strictly causal and adaptive dither signal $D_i = \tau_i(Z_i - 1)$, for some function $\tau_i : \{0, 1\}^{i-1} \to \mathbb{R}$. The system is depicted in Fig. 2. Note that dithering increases the system complexity by requiring digital calculations (or look-up tables) and a high-precision digital-to-analog converter (DAC).

III. BAYESIAN CRAMÉR-RAO LOWER BOUNDS

The celebrated CRLB [17], [18], [19, p. 66] provides a lower bound on the variance of any unbiased estimate of a non-random parameter, which is not applicable in our setting due to the random nature of $H$. However, for the random (Bayesian) parameter estimation problem, a similar bound on the MSE known as the Bayesian Cramér-Rao lower bound (BCRLB) [19, p. 72] holds under some mild regularity conditions. We state the BCRLB in the following theorem.

**Theorem 1 (The BCRLB [19]):** Let $\Theta$ be a random variable and $Y \in \mathbb{R}^k$ an observation vector, let $p_{\Theta|Y}$ be the joint density of $\Theta$ and $Y$, and let $\hat{\Theta}(Y)$ be an estimator of $\Theta$. Suppose the following conditions hold:

1. $\frac{\partial p_{\Theta|Y}(\theta, y)}{\partial \theta}$ is absolutely integrable with respect to $\theta$ and $y$.
2. $\frac{\partial^2 p_{\Theta|Y}(\theta, y)}{\partial \theta^2}$ is absolutely integrable with respect to $\theta$ and $y$.
3. The conditional expectation of the error, given $\Theta = \theta$, is

$$B(\theta) = \int_{\mathbb{R}^k} (\hat{\Theta}(y) - \theta)p_{Y|\Theta}(y|\theta)dy.$$

We have

$$\lim_{\theta \rightarrow \pm \infty} B(\theta)p_{\Theta}(\theta) = 0$$

Then the MSE of $\hat{\Theta}(Y)$ satisfies the inequality

$$E[(\Theta - \hat{\Theta}(Y))^2] \geq -E\left[\frac{\partial^2 \ln p_{\Theta|Y}(\Theta, Y)}{\partial \Theta^2}\right]^{-1}. $$

Theorem 1 holds for continuous observation vectors. We next state a version of the BCRLB that holds for discrete...
observations, and therefore applies to the estimation problem considered in this paper.

**Theorem 2 (The BCRLB for discrete observations):** Let \( \Theta \) be a random variable and \( Z^n \in \mathbb{Z}^n \) an observation sequence, where \( \mathbb{Z} \) is a finite set. Let \( p_{\theta}(z) \) be the density of \( \Theta \), let \( P_{Z^n|\Theta}(z^n|\theta) \) denote the probability of \( z^n \) given \( \theta \), and let \( \hat{\theta}(Z^n) \) be an estimator of \( \Theta \). Suppose the following conditions hold:

1) The conditional expectation of the error, given \( \Theta = \theta \), is
   
   \[
   \bar{B}(\theta) = \sum_{z^n} \left[ \hat{\theta}(z^n) - \theta \right] P_{Z^n|\Theta}(z^n|\theta).
   \]

   We have
   
   \[
   \lim_{\theta \to \infty} \bar{B}(\theta)p_{\theta}(\theta) = 0 \quad \text{(11)}
   \]

   \[
   \lim_{\theta \to \infty} B(\theta)p_{\theta}(\theta) = 0. \quad \text{(12)}
   \]

2) The first derivative of \( p_{\theta}(\theta) \) exists and satisfies

   \[
   \lim_{\theta \to \pm \infty} \frac{\partial p_{\theta}(\theta)}{\partial \theta} = 0 \quad \text{(13)}
   \]

Then the MSE of \( \hat{\theta}(z^n) \) satisfies the inequality

\[
E \left[ (\Theta - \hat{\theta}(Z^n))^2 \right] \geq \left\{ -E \left[ \frac{\partial^2 \ln (P_{Z^n|\Theta}(Z^n|\Theta)p_{\Theta}(\Theta))}{\partial \Theta^2} \right] \right\}^{-1}, \quad \text{(14)}
\]

assuming that the right-hand side of (14) exists.

The proof of Theorem 2 is a straightforward modification of the proof of Theorem 1 given in [19]. For the sake of completeness, we give the proof in Appendix A.

**IV. LOWER BOUNDS ON THE MSE USING SINGLE-BIT DITHERED QUANTIZATION**

In this section, performance bounds on the MSE for single-bit dithered quantizers are derived.

**A. The BCRLB for parameter estimation with quantized observations**

We apply Theorem 2 to the problem of estimating \( \hat{h} \) in AWGN with a single-bit adaptively dithered quantizer, and obtain the following result.

**Theorem 3:** Suppose that \( |\hat{h}(z^n)| < \infty \) for any \( z^n \). For any dither signal \( D_i = \tau_i(Z^{i-1}) \), \( i = 1, 2, \ldots, n \), the MSE of \( \hat{h}(Z^n) \) is lower bounded by

\[
E \left[ (H - \hat{h}(Z^n))^2 \right] \geq \frac{\sigma_h^2}{\frac{1 + n^2 \frac{2}{\pi} \sigma_h^2}{\pi}}. \quad \text{(15)}
\]

The proof of Theorem 3 is given in Appendix B. The corollary below relates the MSE using adaptive dithered single-bit quantization to the MSE achievable with unquantized observations. Let \( \hat{h}_{\text{MMSE}}(Y^n) = E[H|Y^n] \), i.e., \( \hat{h}_{\text{MMSE}}(Y^n) \) is the MMSE estimate of \( H \) based on \( Y^n \); the MSE of \( \hat{h}_{\text{MMSE}}(Y^n) \) is given by [20, Example IV.B.2]

\[
E \left[ (H - \hat{h}_{\text{MMSE}}(Y^n))^2 \right] = \frac{\sigma_h^2}{1 + n^2 \frac{2}{\pi} \sigma_h^2}. \quad \text{(16)}
\]

**Corollary 1:** Suppose that \( |\hat{h}(z^n)| < \infty \) for any \( z^n \). The estimates \( \hat{h}(Z^n) \) and any dither signal \( D_i = \tau_i(Z^{i-1}) \), \( i = 1, 2, \ldots, n \), satisfy

\[
\lim_{n \to \infty} E \left[ (H - \hat{h}(Z^n))^2 \right] \geq \lim_{n \to \infty} \frac{1 + n^2 \frac{2}{\pi} \sigma_h^2}{1 + n^2 \frac{2}{\pi} \sigma_h^2} = \frac{\pi}{2}. \quad \text{(17)}
\]

Based on Corollary 1, any estimator using quantized observations asymptotically loses at least \( 10 \log_{10}(\pi/2) \approx 1.96 \) dB in MSE compared to the MMSE estimator employing unquantized observations.

An asymptotic loss of \( \pi/2 \) for single-bit adaptively dithered quantization compared to unquantized observations was also derived in [13, Section III-C] for the case of estimating a bounded non-random parameter using an unbiased estimator. Note, however, that the bound of [13] does not apply in our setting due to the random nature of \( H \). Moreover, for Theorem 3 to hold, the estimate \( \hat{h}(Z^n) \) need not be unbiased.

We remark that a factor of \( \pi/2 \) also arises as a multiplicative factor relating the low-SNR capacity of the binary-input AWGN channel with single-bit symmetric output quantization and with continuous output. The corresponding capacities are given by [21, (3.4.19)]

\[
C_{\text{AWGN}} \approx \frac{E_s}{N_0}, \quad E_s/N_0 \ll 1, \quad \text{(18)}
\]

for unquantized channel outputs, and by [21, (3.4.20)]

\[
C_Q \approx \frac{2E_s}{\pi N_0}, \quad E_s/N_0 \ll 1, \quad \text{(19)}
\]

for single-bit symmetric output quantization, where \( E_s/N_0 \) denotes the SNR. Hence, the use of hard decisions obtained from a symmetric quantizer causes a power loss of roughly 2 dB at low SNR. Recently, it was shown that this power loss can be removed if asymmetric signaling and asymmetric quantization is employed [22].

**B. Tightening the BCRLB for short observations**

The simulation results in Section VI suggest that the BCRLB of Theorem 3 is almost tight for large values of \( n \), whereas it is loose for small \( n \). We state an alternative version of the BCRLB in the next theorem.

**Theorem 4:** Suppose that \( |\hat{h}(z^n)| < \infty \) for any \( z^n \). Then for any dither signal \( D_i = \tau_i(Z^{i-1}) \), \( i = 1, 2, \ldots, n \), the MSE of \( \hat{h}(Z^n) \) is lower bounded by

\[
E \left[ (H - \hat{h}(Z^n))^2 \right] \geq \frac{\sigma_h^2}{\frac{2}{\pi} \sigma_h^2} + \gamma^2(\sigma_h^2, \sigma_h^2) \quad \text{if } n = 1
\]

\[
\geq \frac{\sigma_h^2}{1 + \frac{\sigma_h^2}{2\pi \sigma_h^2} \sqrt{2\pi}} + \gamma^2(\sigma_h^2, \sigma_h^2) \quad \text{if } n \geq 2,
\]

where

\[
\gamma(\sigma_h^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} hQ \left( \frac{h}{\sigma_h} \right) e^{-h^2/(2\sigma_h^2)} dh \quad \text{(21)}
\]
Theorem 4 provides a better lower bound on $\gamma$ in (21), (22), and (23) must be evaluated. Unfortunately, the integrals in Section VI show that especially at medium and high SNRs, Theorem 4 provides a better lower bound on the MSE than Theorem 3 for small $n$, whereas it is loose for large $n$ because the bound decreases exponentially in $n$.

To avoid numerical integration, we can bound $\gamma^2(\sigma_n^2, \sigma_h^2)$ and apply Lemma 1 from Appendix B to the integrands of $\tilde{\gamma}(\sigma_n^2, \sigma_h^2)$ and $\tilde{\gamma}(\sigma_h^2)$, leading to the following Theorem.

**Theorem 5:** Suppose that $|\hat{h}(z^n)| < \infty$ for any $z^n$. Then for any dither signal $D_i = \tau_i(Z_i-1)$, $i = 1, 2, \ldots, n$, the MSE of $\hat{h}(Z^n)$ is lower bounded by (24) at the top of the page.

A proof of Theorem 5 is given in Appendix D.

V. DESIGN OF ESTIMATORS AND DITHERING STRATEGIES

In this section, we design estimators and dither sequences $d_i = \tau_i(z_i-1)$. We first summarize the approach of [14], where both $d_i$ and the estimator for $H$ are chosen as the linear MMSE estimate of $H$ based on $z_i-1$. Next, we derive two other dither and estimation schemes that considerably outperform those of [14] at high SNR.

A. The linear MMSE estimate as estimator and dither signal

The linear MMSE estimator of $H$ given $z_{i-1} = [z_{i-1}, z_{i-2}, \ldots, z_1]^T$ is

$$\hat{h}_{\text{lin}}(z_{i-1}) = w_{i-1}^T(z_{i-1} - m_{i-1}),$$

where $m_{i-1} = E[Z_{i-1}]$ and $w_{i-1} = R_{i-1}^{-1}r_{i-1}$, with

$$R_{i-1} = E[Z_{i-1}Z_{i-1}^T] - E[Z_{i-1}]E[Z_{i-1}^T],$$

$$r_{i-1} = E[HZ_{i-1}].$$

The strategy of [14] is to use $\hat{h}_{\text{lin}}(z_{i-1})$ as the dither at time $i$, i.e., $d_i = \hat{h}_{\text{lin}}(z_{i-1})$. Since a closed-form solution is not available for $m_{i-1}$, $R_{i-1}$, and $r_{i-1}$ due to the non-linearity of the quantizer and the feedback of the dither signal, estimates of $m_{i-1}$, $R_{i-1}$, and $r_{i-1}$ are calculated by Monte-Carlo simulations over the statistics of $H$ and the noise [14], yielding an approximation of $\hat{h}_{\text{lin}}(z_{i-1})$.

Having received $z_{n}$, the approximation of $\hat{h}_{\text{lin}}(z_{n})$ is used as an estimator of $H$.

B. The MMSE estimate as estimator and dither signal

Given $m_{i-1}$, $R_{i-1}$, and $r_{i-1}$, the linear MMSE estimator $\hat{h}_{\text{lin}}(z_{i-1})$ can be computed efficiently. However, for a fixed (non-adaptive) dither sequence, the (non-linear) MMSE estimator $\hat{h}_{\text{MMSE}}(z_{i-1})$ is the optimal estimator of $H$, since we are using square-error cost in this paper. Since the overall system is non-linear and non-Gaussian due to the single-bit quantizer, it is intuitive that $\hat{h}_{\text{MMSE}}(z_{i-1})$ may considerably outperform the linear MMSE estimator $\hat{h}_{\text{lin}}(z_{i-1})$. The strategy of employing $\hat{h}_{\text{MMSE}}(z_{i-1})$ as an estimator of $H$ is combined with also using $\hat{h}_{\text{MMSE}}(z_{i-1})$ as the dither signal at the next time instance, i.e., by choosing $d_i = \hat{h}_{\text{MMSE}}(z_{i-1})$. We next discuss how to compute $\hat{h}_{\text{MMSE}}(z_{i-1})$ efficiently.

The integration over

$$h \cdot p_{H|Z_{i-1}}(h|z_{i-1}) = \int_{-\infty}^{\infty} h \cdot p_{H|Z_{i-1}}(h|z_{i-1}) \, dh$$

in (28) cannot be solved in closed form, since the term $P_{Z_{i-1}|H}(z_{i-1}|h)$ is a product of $i-1$ terms involving the $Q$-function. But an approximation $\hat{h}_{\text{MMSE}}(z_{i-1})$ of $\hat{h}_{\text{MMSE}}(z_{i-1})$ can be computed based on a recursively updated discrete approximation of $p_{H|Z_{i-1}}(h|z_{i-1})$ combined with interpolation. To that end, we form a discrete random variable $H$ by sampling from the distribution of $H$. The variable $H$ takes on values in $\mathcal{H}^{(0)} = \{-\Delta, -\Delta(1 - \frac{b}{\sqrt{\sigma_h^2}}), -\Delta(1 - \frac{2}{\sqrt{\sigma_h^2}}), -\Delta(1 - \frac{3}{\sqrt{\sigma_h^2}}), \ldots, \Delta\}$, where $|\mathcal{H}^{(0)}| = B$, so that $H$ has probability mass function $P_H(h)$, i.e., $\sum_{h \in \mathcal{H}^{(0)}} P_H(h) = 1$. The parameter $\Delta$ is chosen such that $P_{\{h \in \mathcal{H}^{(0)}\}} = 0.99995$; since $H \sim N(0, \sigma_h^2)$, we get $\Delta = \sqrt{2\sigma_h\text{erf}^{-1}(0.99995)} \approx 4.056\sigma_h$. Defining $\hat{P}^{(0)}(h|z_i^0) = P_H(h)$, $h \in \mathcal{H}^{(0)}$, we can recursively compute an approximation of $p_{H|Z_{i-1}}(h|z_{i-1})$ and of $\hat{h}_{\text{MMSE}}(z_i)$ by the following algorithm.

1) Update step: For $h \in \mathcal{H}^{(i-1)}$, compute

$$P^{(i)}(h, z_i|z_{i-1}) = P^{(i-1)}(h|z_{i-1}) \cdot \begin{cases} Q \left( \frac{h - d_i}{\sigma_h} \right) & \text{if } z_i = 0 \\ \left( 1 - Q \left( \frac{h - d_i}{\sigma_h} \right) \right) & \text{if } z_i = 1. \end{cases}$$
2) Interval expansion and interpolation step: Let $P_{\text{opt}}(i) = \max_{h \in \mathcal{H}(i)} P^{(i)}(h, z_i | z^{i-1})$, and let
\begin{align}
    h_{\text{min}}^{(i)} &= \arg\min_{h \in \mathcal{H}(i)} P^{(i)}(h, z_i | z^{i-1}) \\
    h_{\text{max}}^{(i)} &= \arg\max_{h \in \mathcal{H}(i)} P^{(i)}(h, z_i | z^{i-1}),
\end{align}

where both (31) and (32) are subject to $P^{(i)}(h, z_i | z^{i-1}) > \epsilon p_{\text{max}}, \epsilon > 0$. Defining $\eta^{(i)} = (h_{\text{max}}^{(i)} - h_{\text{min}}^{(i)})/(B - 1)$, the set $\mathcal{H}^{(i)}$ of size $B$ is given by
\begin{equation}
    \mathcal{H}^{(i)} = \{h_{\text{min}}^{(i)} + \eta^{(i)} n, \eta^{(i)} n + 2\eta^{(i)} n, \ldots, h_{\text{max}}^{(i)}\}. 
\end{equation}

Next, compute the term $P^{(i)}(h, z_i | z^{i-1})$, $h \in \mathcal{H}^{(i)}$, from $P^{(i)}(h, z_i | z^{i-1})$, $h \in \mathcal{H}(i)$, by interpolating $P^{(i)}(h, z_i | z^{i-1})$ at the $B$ points $h \in \mathcal{H}^{(i)}$; then we obtain an approximation $P^{(i)}(z_i | z^{i-1})$ of the conditional probability $P_{Z_i | Z^{i-1}}(z_i | z^{i-1})$ through
\begin{equation}
    P^{(i)}(z_i | z^{i-1}) = \sum_{h \in \mathcal{H}^{(i)}} \tilde{P}^{(i)}(h, z_i | z^{i-1})
\end{equation}

and finally,
\begin{equation}
    P^{(i)}(h | z^i) = \frac{\tilde{P}^{(i)}(h, z_i | z^{i-1})}{P^{(i)}(z_i | z^{i-1})},
\end{equation}

where the normalization with $P^{(i)}(z_i | z^{i-1})$ ensures that $\sum_{h \in \mathcal{H}^{(i)}} P^{(i)}(h | z^i) = 1$.

3) The estimate of $\hat{h}_{\text{MMSE}}(z^i)$ is given by
\begin{equation}
    \bar{h}_{\text{MMSE}}(z^i) = \sum_{h \in \mathcal{H}^{(i)}} h P^{(i)}(h | z^i).
\end{equation}

In a practical implementation (cf. Section VI), choosing $B$ on the order of 100 and $\epsilon$ around $10^{-5}$ yields excellent performance at reasonable computational complexity since only a small number of samples describing $p_{H | Z}(h | z^{i-1})$ needs to be updated at each time step.

\section*{C. The dither signal that minimizes the MSE}

While the MMSE estimator $\hat{h}_{\text{MMSE}}(z^n)$ minimizes the MSE of estimating $H$ for a fixed dither sequence $d^n$, it is not clear if employing $d_i = \bar{h}_{\text{MMSE}}(z^{i-1})$ is an optimal dither strategy. Instead of dithering using $\bar{h}_{\text{MMSE}}(z^{i-1})$, the dither signal should be selected for optimal MSE performance at time $n$, i.e., given $z^{i-1}$, $i \leq n$, the MSE-optimal dither sequence $d_{\text{opt}, i}^n$ is formally given by
\begin{equation}
    d_{\text{opt}, i}^n = \arg\min_{d_i} E\left[(H - \hat{h}_{\text{MMSE}}(Z^n, d_i))^2 | Z^{i-1} = z^{i-1}\right].
\end{equation}

where we make the dependency of $\hat{h}_{\text{MMSE}}(z^n, d_i)$ on $d_i$ explicit. For complexity reasons, we will not attempt to solve Problem (37) for $n > i$; instead, we solve (37) for $n = i$ only, i.e., the dither signal $d_i$ is chosen based on $z^{i-1}$ such that the MSE at the next time instance, i.e., at time step $i$, is minimized. We refer to this dither as the “one-step look-ahead” (OSLA) dither signal.

Suppose that $Z^{i-1} = z^{i-1}$, so that the conditional distribution $p_{H | Z^{i-1}}(h | z^{i-1})$ is fixed. Then, the dither signal $d_i^*$ is given by
\begin{equation}
    d_i^* = \arg\min_{d_i} E\left[(H - \hat{h}_{\text{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right].
\end{equation}

Given $z^{i-1}$ and some $d_i$, we can view the conditional expectation $E[(H - \hat{h}_{\text{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}]$ as the MSE of estimating the random parameter $H$ with prior $p_{H | Z^{i-1}}(h | z^{i-1})$ using the estimator $\hat{h}_{\text{MMSE}}(z^i, d_i)$. Due to the properties of MMSE estimation, we have
\begin{equation}
    \begin{align}
        & E\left[(H - \hat{h}_{\text{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right] \\
        & = E[H^2 | Z^{i-1} = z^{i-1}] - E[\hat{h}_{\text{MMSE}}^2(Z^i, d_i) | Z^{i-1} = z^{i-1}],
    \end{align}
\end{equation}

and since $E[H^2 | Z^{i-1} = z^{i-1}]$ is fixed for a given $z^{i-1}$, the minimization in (38) is equivalent to
\begin{equation}
    d_i^* = \arg\max_{d_i} E\left[\hat{h}_{\text{MMSE}}^2(Z^i, d_i) | Z^{i-1} = z^{i-1}\right].
\end{equation}

Problem (40) is hard to solve since $\hat{h}_{\text{MMSE}}(z^i, d_i)$ cannot be found in closed form as a function of $z_i^i$ and $d_i$; we therefore work with a discrete approximation $\bar{h}_{\text{MMSE}}(z^i, d_i)$ of $\hat{h}_{\text{MMSE}}(z^i, d_i)$, similar to Section V-B. As before, let $P^{(i-1)}(h | z^{i-1}), h \in \mathcal{H}(i-1)$ be the approximation of the conditional density $p_{H | Z^{i-1}}(h | z^{i-1})$ obtained from the received sequence $z^{i-1}$. Based on $P^{(i-1)}(h | z^{i-1})$ and
\begin{equation}
    \begin{align}
        \tilde{P}^{(i)}(z_i | z^{i-1}) &= \sum_{h \in \mathcal{H}(i-1)} P_{Z_i | H Z^{i-1}}(z_i | h, z^{i-1}) P^{(i-1)}(h | z^{i-1}) \\
        &= \sum_{h \in \mathcal{H}(i-1)} h P_{Z_i | H Z^{i-1}}(z_i | h, z^{i-1}) P^{(i-1)}(h | z^{i-1})
    \end{align}
\end{equation}

we can compute the estimate $\bar{h}_{\text{MMSE}}(z^i, d_i)$ (omitting the interval expansion and interpolation step) of $\hat{h}_{\text{MMSE}}(z^i, d_i)$, given by
\begin{equation}
    \bar{h}_{\text{MMSE}}(z^i, d_i) = \sum_{h \in \mathcal{H}(i-1)} h P_{Z_i | H Z^{i-1}}(z_i | h, z^{i-1}) P^{(i-1)}(h | z^{i-1})
\end{equation}

where the right-hand side depends on $d_i$ through
\begin{equation}
    P_{Z_i | H Z^{i-1}}(z_i | h, z^{i-1}) = \delta(z_i) Q \left(\frac{h - d_i}{\sigma_n}\right) + \delta(z_i - 1) \left(1 - Q \left(\frac{h - d_i}{\sigma_n}\right)\right) 
\end{equation}

Consequently, the cost function approximating the one in (40) is
\begin{equation}
    V(d_i | z^{i-1}) = \sum_{z_i} P^{(i)}(z_i | z^{i-1}) \bar{h}_{\text{MMSE}}^2(z^i, d_i),
\end{equation}

and the approximate solution to Problem (40) is
\begin{equation}
    d_i^* = \arg\max_{d_i} V(d_i | z^{i-1}).
\end{equation}

In all the simulations that we performed, we observed that $V(d_i | z^{i-1})$ is a strictly quasiconcave function of $d_i$, for any $z^{i-1}$. However, we were unable to formally verify this observation. Nevertheless, we solve Problem (45) with a gradient
descent algorithm [23, Chapter 9.3] combined with backtracking line search [23, Chapter 9.2], observing excellent convergence behavior. For such a gradient descent method, we need the first derivative of $V(d_i, z^{-1})$ with respect to $d_i$, which is

$$\frac{\partial V(d_i, z^{-1})}{\partial d_i} = \sum_{z_i} \hat{h}_{\text{MMSE}}(z^i, d_i) \left[ 2 \hat{P}^{(i)}(z_i | z^{-1}) \frac{\partial \hat{h}_{\text{MMSE}}(z^i, d_i)}{\partial d_i} + \hat{h}_{\text{MMSE}}(z^i, d_i) \frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} \right].$$

(46)

Given

$$\frac{\partial P_{Z_i|HZ^{-1}}(z_i | h, z^{-1})}{\partial d_i} = \frac{1}{\sqrt{2\pi \sigma_n}} e^{-(h-d_i)^2/(2\sigma_n^2)} (\delta(z_i) - \delta(z_i - 1)), \quad (47)$$

it is straightforward to compute

$$\frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} = \sum_{h \in \mathcal{H}^{(i-1)}} P^{(i-1)}(h | z^{-1}) \frac{\partial P_{Z_i|HZ^{-1}}(z_i | h, z^{-1})}{\partial d_i}$$

(48)

and

$$\frac{\partial \hat{h}_{\text{MMSE}}(z^i, d_i)}{\partial d_i} = \frac{1}{P^{(i)}(z_i | z^{-1})} \sum_{h \in \mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{-1}}(z_i | h, z^{-1})}{\partial d_i} P^{(i-1)}(h | z^{-1})$$

$$- \frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} \sum_{h \in \mathcal{H}^{(i-1)}} h P_{Z_i|HZ^{-1}}(z_i | h, z^{-1}) P^{(i-1)}(h | z^{-1})$$

(49)

$$= \frac{1}{P^{(i)}(z_i | z^{-1})} \sum_{h \in \mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{-1}}(z_i | h, z^{-1})}{\partial d_i} P^{(i-1)}(h | z^{-1})$$

$$- \frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} \hat{h}_{\text{MMSE}}(z^i, d_i) \frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} \hat{h}_{\text{MMSE}}(z^i, d_i)$$

Inserting (48) and (49) into (46), we have

$$\frac{\partial V(d_i, z^{-1})}{\partial d_i} = \sum_{z_i} \hat{h}_{\text{MMSE}}(z^i, d_i) \left[ 2 \sum_{h \in \mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{-1}}(z_i | h, z^{-1})}{\partial d_i} P^{(i-1)}(h | z^{-1}) \right] + \hat{h}_{\text{MMSE}}(z^i, d_i) \frac{\partial \hat{P}^{(i)}(z_i | z^{-1})}{\partial d_i} \hat{h}_{\text{MMSE}}(z^i, d_i)$$

(50)

The approximation $\hat{h}_{\text{MMSE}}(z^{-1})$ of the estimate $\tilde{h}_{\text{MMSE}}(z^{-1})$ (cf. Section V-B) is chosen as a starting point for the gradient descent method solving Problem (45).

VI. SIMULATION RESULTS

Simulation results for various SNRs are shown in Figs. 3 to 7, where $\sigma_n^2 = 1$ without loss of generality. In addition to the BCRLBs of Theorems 3 and 4 we also show the MSE of the MMSE estimator $\tilde{h}_{\text{MMSE}}(Y^n)$ (cf. (16)) using unquantized observations $Y^n$. The BCRLB of Theorem 5 is not shown since it almost coincides with the BCRLB of Theorem 4.

The performance of the three dither schemes presented in Section V is determined by means of simulation for the linear/non-linear MMSE and the OSLA dither schemes.

At an SNR of 0 dB, the performance of all dither strategies is almost indistinguishable, and very close to the lower bound provided by the BCRLB. The non-linear MMSE and OSLA feedback schemes continue to perform close to the BCRLB for higher SNRs, while the linear feedback strategy exhibits a considerable performance gap if the SNR is 20 dB or higher. While the performance difference between the non-linear MMSE and OSLA dither is not large, it is most pronounced at very high SNR. While the tightened bound of Theorem 4 still has a gap of approximately 10 dB in MSE for small $n$ and high SNR, the slope of the bound is a good match to the slope of the MSE curves obtained from simulation, for small $n$.

VII. CONCLUSIONS

We studied the parameter estimation problem using a single-bit dithered quantizer. By bounding the BCRLB for that problem, we derived lower bounds on the MSE which hold
Fig. 5. MSE for an SNR of 20 dB.

Fig. 6. MSE for an SNR of 30 dB.

Fig. 7. MSE for an SNR of 40 dB.

for all dither strategies. We showed that the performance of single-bit dithered parameter estimation cannot approach the performance of estimation using unquantized observations; in particular, the estimator based on continuous observations outperforms the estimator based on the quantized observations asymptotically by at least 1.96 dB. We also designed dither sequences that are computed by strictly causal processing of the quantizer output sequence. By means of simulations, we showed that the derived bounds on the MSE can be closely approached.

Appendix A
Proof of Theorem 2

Multiplying both sides of (10) with $p_\theta(\theta)$ and differentiating gives

$$\frac{\partial p_\theta(\theta)\tilde{B}(\theta)}{\partial \theta} = -\sum_{z^n} P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)$$

$$+ \sum_{z^n} \left[\hat{\theta}(z^n) - \theta\right] \frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)}{\partial \theta}.$$  (51)

Integrating both sides with respect to $\theta$ yields

$$p_\theta(\theta)\tilde{B}(\theta)\bigg|_{-\infty}^{\infty} = -1 + \int_{-\infty}^{\infty} \sum_{z^n} \left[\hat{\theta}(z^n) - \theta\right] \frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)}{\partial \theta} d\theta,$$  (52)

and the assumptions in (11) and (12) ensure that

$$p_\theta(\theta)\tilde{B}(\theta)\bigg|_{-\infty}^{\infty} = 0,$$  (53)

so that we have

$$\sum_{z^n} \int_{-\infty}^{\infty} \left[\hat{\theta}(z^n) - \theta\right] \frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)}{\partial \theta} d\theta = 1.$$  (54)

Next, observe that

$$\frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)}{\partial \theta} = \frac{\partial \ln \left(P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right)}{\partial \theta} P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta).$$  (55)

Substituting (55) into (54) and rewriting, we have

$$\sum_{z^n} \int_{-\infty}^{\infty} \left[\frac{\partial \ln \left(P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right)}{\partial \theta} P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right]$$

$$\left[\left[\hat{\theta}(z^n) - \theta\right] \sqrt{P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)}\right] d\theta = 1.$$  (56)

and, by applying the Schwarz inequality to the integral in the summation, we have

$$\sum_{z^n} \left(\int_{-\infty}^{\infty} \left[\frac{\partial \ln \left(P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right)}{\partial \theta} P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right] d\theta\right)^2$$

$$\times \left(\int_{-\infty}^{\infty} \left[\hat{\theta}(z^n) - \theta\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\theta(\theta)\right) d\theta \geq 1.$$  (57)
Next, apply the Cauchy inequality to the summation in (57) to obtain
\[
\left( \sum_{z^n=-\infty}^{\infty} \left[ \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \right)^2 \geq \frac{1}{\sum_{z^n=-\infty}^{\infty} \left[ \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta}, \tag{58}
\]

or, equivalently,
\[
1 \leq \sum_{z^n=-\infty}^{\infty} \int \left[ \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \times \sum_{z^n=-\infty}^{\infty} \int \left[ \hat{\theta}(z^n) - \theta \right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \tag{59}
\]
\[
= E \left[ \left( \frac{\partial \ln(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta))}{\partial \Theta} \right)^2 \right] E \left[ (\hat{\Theta} - \Theta(Z^n))^2 \right]. \tag{60}
\]
Rearranging the above inequality yields the BCRLB in terms of the squared first derivative of \( \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)) \). To derive the BCRLB in terms of the second derivative of \( \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)) \), note that
\[
\sum_{z^n} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) = p_\Theta(\theta). \tag{61}
\]
Next, differentiation of (61) twice on both sides with respect to \( \theta \) yields
\[
\sum_{z^n} \frac{\partial}{\partial \theta} \left[ \frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial \theta} \right] = \frac{\partial^2 p_\Theta(\theta)}{\partial \theta^2}. \tag{62}
\]
Now, substitution of (55) into the left-hand side of (62) yields
\[
\sum_{z^n} \frac{\partial}{\partial \theta} \left[ \frac{\partial P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial \theta} \right] = \sum_{z^n} \frac{\partial}{\partial \theta} \left[ \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) \tag{63}
\]
\[
= \sum_{z^n} \left[ \frac{\partial^2 \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta^2} + \left( \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right)^2 \right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) \tag{64}
\]
\[
= \frac{\partial^2 p_\Theta(\theta)}{\partial \theta^2}, \tag{65}
\]
and by integrating with respect to \( \theta \), we have
\[
\sum_{z^n=\infty}^{\infty} \int \left[ \frac{\partial^2 \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta^2} \right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \tag{66}
\]
\[
= \int_{-\infty}^{\infty} \frac{\partial^2 p_\Theta(\theta)}{\partial \theta^2} d\theta = \frac{\partial p_\Theta(\theta)}{\partial \theta} \bigg|_{-\infty}^{\infty} = 0, \tag{67}
\]
where the last equality holds due to Condition 2 of Theorem 2. Therefore, we have
\[
E \left[ \left( \frac{\partial \ln(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta))}{\partial \Theta} \right)^2 \right] = E \left[ \frac{\partial^2 \ln(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta))}{\partial \Theta^2} \right]. \tag{68}
\]
Plugging (68) into (60), we have
\[
1 \leq E \left[ \frac{\partial^2 \ln(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta))}{\partial \Theta^2} \right] E \left[ (\hat{\Theta} - \Theta(Z^n))^2 \right], \tag{69}
\]
which is Theorem 2.
Finally, we comment on the tightness of Theorem 2. Equality in the application of the Schwarz inequality in (57) holds if and only if
\[
\frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} = \left[ \hat{\theta}(z^n) - \theta \right] c(z^n), \tag{70}
\]
for all \( z^n \) and \( \theta \), where \( c(z^n) \) is a function of \( z^n \). Furthermore, equality in (59) holds if and only if there is a constant \( \hat{c} \in \mathbb{R} \) such that
\[
E \left[ \left( \int_{-\infty}^{\infty} \left[ \frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} \right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \right)^{\frac{1}{2}} \right] = \hat{c} \left( \int_{-\infty}^{\infty} \left[ \hat{\theta}(z^n) - \theta \right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) d\theta \right)^{\frac{1}{2}}, \tag{71}
\]
for all \( z^n \). Combining (70) and (71), we see that equality holds in Theorem 2 if and only if
\[
c(z^n) = \hat{c} \quad \forall z^n, \tag{72}
\]
\[
\frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta} = \left[ \hat{\theta}(z^n) - \theta \right] \hat{c}. \tag{73}
\]
Differentiating (73) with respect to \( \theta \) gives the condition
\[
\frac{\partial^2 \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial \theta^2} = -\hat{c}. \tag{74}
\]

\section*{APPENDIX B}
\textbf{PROOF OF THEOREM 3}

\subsection*{A. Conditions for the applicability of Theorem 2}
We first show that the conditions for the applicability of Theorem 2 are satisfied. To show that Condition 1) is satisfied, let \( h(z^n) \) be any estimator with \( |\hat{h}(z^n)| < \infty \) for all \( z^n \), and let
\[
\bar{B}(h) = \sum_{z^n} \left[ \hat{h}(z^n) - h \right] P_{Z^n|H}(z^n|h), \tag{75}
\]
so that
\[
\lim_{h \to \pm \infty} \bar{B}(h)p_H(h) = \sum_{z^n} \hat{h}(z^n) \lim_{h \to \pm \infty} P_{Z^n|H}(z^n|h)p_H(h)
\]
Since $0 \leq P_{Z^n|H}(z^n|h) \leq 1$ for any $z^n$ and $h$, and since $H \sim \mathcal{N}(0, \sigma_H^2)$, we have
\begin{align*}
lim_{h \to \pm \infty} P_{Z^n|H}(z^n|h)p_H(h) = 0, \quad \forall z^n, \quad (77)\\nlim_{h \to \pm \infty} h P_{Z^n|H}(z^n|h)p_H(h) = 0, \quad \forall z^n, \quad (78)
\end{align*}
and therefore
\begin{equation}
\lim_{h \to \pm \infty} \bar{B}(h)p_H(h) = 0. \quad (79)
\end{equation}
To check Condition 2), we compute
\begin{equation}
\frac{\partial p_H(h)}{\partial h} = -\frac{h}{\sigma_H^2} \frac{1}{\sqrt{2\pi\sigma_H}} e^{-h^2/(2\sigma_H^2)}, \quad (80)
\end{equation}
so that clearly
\begin{equation}
\lim_{h \to \pm \infty} \frac{\partial p_H(h)}{\partial h} = 0. \quad (81)
\end{equation}

**B. The second derivative of $\ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right)$**

Let $p_N(\alpha)$ denote the distribution of a zero-mean Gaussian random variable with variance $\sigma_n^2$, i.e.,
\begin{equation}
p_N(\alpha) \triangleq \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\alpha^2/(2\sigma_n^2)}, \quad (82)
\end{equation}
and let $Q(x)$ be the Q-function, i.e.,
\begin{equation}
Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-u^2/2} du, \quad (83)
\end{equation}
whose derivative is
\begin{equation}
\frac{\partial Q(x)}{\partial x} = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (84)
\end{equation}

With
\begin{align*}
P_{Z^n|H}(z^n|h) &= \prod_{i=1}^{n} P_{Z_i|H^{z_i-1}}(z_i|h, z^{i-1}) \\
&= n \prod_{i=1}^{n} \left[ \delta(z_i)Q\left(\frac{h - \tau_i(z^{i-1})}{\sigma_n}\right) \right. \\
&\left. \quad + \delta(z_i - 1)\left(1 - Q\left(\frac{h - \tau_i(z^{i-1})}{\sigma_n}\right)\right) \right], \quad (85)
\end{align*}
we have
\begin{align*}
\ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right) \\
= \ln p_H(h) + \sum_{i=1}^{n} \ln P_{Z_i|H^{z_i-1}}(z_i|h, z^{i-1}) \\
= -\ln(\sqrt{2\pi\sigma_H}) - \frac{h^2}{2\sigma_H^2} \\
&+ \sum_{i=1}^{n} \ln \left( \delta(z_i)Q\left(\frac{h - d_i}{\sigma_n}\right) + \delta(z_i - 1)\left(1 - Q\left(\frac{h - d_i}{\sigma_n}\right)\right) \right), \quad (87)
\end{align*}
where we write $d_i = \tau_i(z^{i-1})$ for brevity, and keep in mind that $d_i$ is a function of $z^{i-1}$. We have
\begin{equation}
\frac{\partial}{\partial h} \ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right) \\
= -\frac{h}{\sigma_H^2} + \sum_{i=1}^{n} p_N(h-d_i) \left[ \frac{\delta(z_i - 1)}{1 - Q\left(\frac{h - d_i}{\sigma_n}\right)} - \frac{\delta(z_i)}{Q\left(\frac{h - d_i}{\sigma_n}\right)} \right], \quad (88)
\end{equation}
and
\begin{align*}
\frac{\partial^2}{\partial h^2} \ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right) \\
= \frac{1}{\sigma_H^4} + \sum_{i=1}^{n} p_N(h-d_i) \left[ \frac{\delta(z_i)\left(\frac{h - d_i}{\sigma_n}\right)}{Q\left(\frac{h - d_i}{\sigma_n}\right)^2} - \frac{p_N(h - d_i)}{Q\left(\frac{h - d_i}{\sigma_n}\right)^2} \right] \\
&\quad - \delta(z_i - 1)\left(1 - Q\left(\frac{h - d_i}{\sigma_n}\right)\right) + p_N(h - d_i) \left(1 - Q\left(\frac{h - d_i}{\sigma_n}\right)^2\right) \right], \quad (90)
\end{align*}
Since (90) is not a constant, we conclude based on (74) that there exists no estimator of $H$ that achieves the BCRB
\begin{equation}
\left\{ -E \left[ \frac{\partial^2}{\partial H^2} \ln \left( P_{Z^n|H}(Z^n|h)p_H(H) \right) \right] \right\}^{-1} \quad (91)
\end{equation}
with equality.

**C. Evaluation of the BCRB**

Defining
\begin{equation}
G(z^n, h) \triangleq \frac{\partial^2}{\partial h^2} \ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right) \quad (92)
\end{equation}
and
\begin{align*}
g(z^i, h) &\triangleq p_N(h - d_i) \left[ \frac{\delta(z_i)\left(\frac{h - d_i}{\sigma_n}\right)}{Q\left(\frac{h - d_i}{\sigma_n}\right)^2} - \frac{p_N(h - d_i)}{Q\left(\frac{h - d_i}{\sigma_n}\right)^2} \right] \\
&\quad - \delta(z_i - 1)\left(1 - Q\left(\frac{h - d_i}{\sigma_n}\right)\right) + p_N(h - d_i) \left(1 - Q\left(\frac{h - d_i}{\sigma_n}\right)^2\right) \right], \quad (93)
\end{align*}
we have
\begin{align*}
G(z^n, h) &= \frac{1}{\sigma_H^4} + \sum_{i=1}^{n} g(z^i, h), \quad (94)
\end{align*}
so that
\begin{align*}
-E[G(Z^n, H)] \\
&= \int_{-\infty}^{\infty} \sum_{i=1}^{n} \left( \frac{1}{\sigma_H^4} - \sum_{i=1}^{n} g(z^i, h) \right) P_{Z^n|H}(z^n|h)p_H(h) dh \\
&= \frac{1}{\sigma_H^4} \sum_{i=1}^{n} \int_{-\infty}^{\infty} g(z^i, h)P_{Z^n|H}(z^n|h)p_H(h) dh \\
&= \frac{1}{\sigma_H^4} \sum_{i=1}^{n} \sum_{z^i} \int_{-\infty}^{\infty} g(z^i, h)P_{Z^n|H}(z^n|h)p_H(h) dh \\
&= \frac{1}{\sigma_H^4} \sum_{i=1}^{n} \sum_{z^i} \sum_{z_{i+1}} \cdots \sum_{z_{i+n}} P_{Z_{i+1}^{n+1}|H}(z_{i+1}^{n+1}|h, z^i) \right] dh \quad (97)
\end{align*}
where we write $d_i = \tau_i(z^{i-1})$ for brevity, and keep in mind that $d_i$ is a function of $z^{i-1}$. We have
\begin{equation}
\frac{\partial}{\partial h} \ln \left( P_{Z^n|H}(z^n|h)p_H(h) \right) \\
= -\frac{h}{\sigma_H^2} + \sum_{i=1}^{n} p_N(h-d_i) \left[ \frac{\delta(z_i - 1)}{1 - Q\left(\frac{h - d_i}{\sigma_n}\right)} - \frac{\delta(z_i)}{Q\left(\frac{h - d_i}{\sigma_n}\right)} \right], \quad (89)
\end{equation}
Lemma 1: Let

\[ \lambda(x) = \frac{e^{-x^2}}{Q(x)[1 - Q(x)]} \]

with \( x \in \mathbb{R} \). We have \( \lambda(x) \leq 4e^{-1 \cdot 2 / 4 \pi} x^2 \approx 4e^{-0.36342} x^2 \).

Proof: First, we express \( \lambda(x) \) in terms of the error function

\[ \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} \, du, \]

yielding

\[ \lambda(x) = \frac{4e^{-x^2}}{1 - \text{erf} \left( \frac{x}{\sqrt{2}} \right)^2}. \]

To bound the square of the error function in (105), we employ an upper bound due to Williams [24] and Pólya [25], which was complemented with a lower bound by Chu [26]; the bound is

\[ \text{erf}(x) \leq \sqrt{1 - e^{-4 \pi^2 x^2}}, \quad x \geq 0. \]

Consequently, since \( 1 - \text{erf} \left( \frac{x}{\sqrt{2}} \right)^2 \) is symmetric around the origin, i.e., \( 1 - \text{erf} \left( -\frac{x}{\sqrt{2}} \right)^2 = 1 - \text{erf} \left( \frac{x}{\sqrt{2}} \right)^2 \), we have

\[ 1 - \text{erf} \left( \frac{x}{\sqrt{2}} \right)^2 \geq e^{-2x^2 / 4 \pi}, \quad x \in \mathbb{R}, \]

and therefore

\[ \lambda(x) \leq \frac{4e^{-x^2}}{e^{-2x^2 / 4 \pi}} = 4e^{-(1 - \frac{1}{4}) x^2}, \quad x \in \mathbb{R}. \]

To find an upper bound on \( -E[G(Z^n, H)] \), we applyLemma 1 to obtain, for any \( d_i \) and \( \forall h \),

\[ \frac{e^{-(h - d_i)^2 / \sigma_n^2}}{Q\left( \frac{h - d_i}{\sigma_n} \right)} \left[ 1 - Q\left( \frac{h - d_i}{\sigma_n} \right) \right] \leq 4e^{-(1 - \frac{1}{4}) \left( \frac{h - d_i}{\sigma_n} \right)^2} \leq 4. \]

Inserting (109) into (102) yields

\[ -E[G(Z^n, H)] \leq \frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \int_{-\infty}^{\infty} \sum_{z_i} \int_{-\infty}^{\infty} \sum_{z_i} P_{Z_{i-1}|H}(z_{i-1}|h) \frac{4}{2 \pi \sigma_n^2} p_H(h) \, dh. \]

Therefore, we have

\[ E \left[ (H - \hat{h}(Z^n))^2 \right] \geq \frac{1}{\sigma_h^2 + n \frac{\sigma^2}{\pi \sigma_n^2}} = \frac{\sigma_h^2}{1 + n \frac{\sigma^2}{\pi \sigma_n^2}}, \]

which completes the proof of Theorem 3. \( \blacksquare \)

Appendix C

Proof of Theorem 4

To prove Theorem 4 for \( n = 1 \), we compute \( E[(H - \hat{h}_{\text{MMSE}}(Z_1))^2] \) exactly without using the BCRLB. Due to the properties of MMSE estimation, we have

\[ E \left[ (H - \hat{h}_{\text{MMSE}}(Z_1))^2 \right] = E[H^2] - E \left[ \hat{h}_{\text{MMSE}}^2(Z_1) \right] \]

\[ = \sigma_h^2 - E \left[ \hat{h}_{\text{MMSE}}^2(Z_1) \right]. \]

Since \( H \) is a zero-mean random variable, the initial quantizer threshold is \( d_1 = \tau_1(z^0) = 0 \), so that \( P_{Z_1}(1) = P_{Z_1}(0) = 1/2 \), and the MMSE estimator has the symmetry property

\[ \hat{h}_{\text{MMSE}}(z_1) = \int h \left( 1 - Q \left( \frac{h}{\sigma_n} \right) \right) p_H(h) \, dh \]

\[ = \frac{1}{P_{Z_1}(1)} \int_{-\infty}^{\infty} \int_{-\infty}^{0} h p_H(h) \, dh \]

\[ = - \frac{1}{P_{Z_1}(0)} \int_{-\infty}^{\infty} \int_{-\infty}^{0} h Q \left( \frac{h}{\sigma_n} \right) p_H(h) \, dh \]

\[ = - \hat{h}_{\text{MMSE}}(z_1) = 0. \]
Consequently, the expectation of $\hat{h}_{\text{MMSE}}^2(Z_1)$ is given by
\begin{equation}
E \left[ \hat{h}_{\text{MMSE}}^2(Z_1) \right] = \sum_{z_1} P_{Z_1}(z_1) \hat{h}_{\text{MMSE}}^2(z_1) 
= \hat{h}_{\text{MMSE}}^2(z_1 = 0) 
= \frac{2}{\pi \sigma_h^2} \left( \int_{-\infty}^{\infty} h Q \left( \frac{h}{\sigma_h} \right) e^{-h^2/(2\sigma_h^2)} dh \right)^2. 
\end{equation}

Defining
\begin{equation}
\gamma(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} h Q \left( \frac{h}{\sigma_h} \right) e^{-h^2/(2\sigma_h^2)} dh, 
\end{equation}
we have
\begin{equation}
E \left[ (H - \hat{h}_{\text{MMSE}}(Z_1))^2 \right] = \sigma_h^2 - \frac{2}{\pi \sigma_h^2} \gamma(\sigma_n^2, \sigma_h^2). 
\end{equation}

For $n \geq 2$, the key idea to tighten the bound is to not employ Lemma 1 to bound (102) but to use that $P_{Z_{i-1} \mid H}(z_i^{-1} \mid h) \leq 1$ for all $z_i^{-1}$ and $h$. First consider the term for $i = 1$ in the summation in (102); since $z_1^{-1}$ is the empty sequence for $i = 1$, that term is
\begin{equation}
\int_{-\infty}^{\infty} \frac{e^{-(h-d_1)^2/\sigma_n^2}}{2\pi \sigma_n^2 Q \left( \frac{h-d_1}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_1}{\sigma_n} \right) \right]} p_H(h) dh = \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_h^2}}{2\pi \sigma_h^2 Q \left( \frac{h}{\sigma_h} \right) \left[ 1 - Q \left( \frac{h}{\sigma_h} \right) \right]} p_H(h) dh, 
\end{equation}
where the equality follows since $d_1 = \tau_1(z_i^0) = 0$. Since there seems to be no closed-form solution available for the integral in (124), we compute
\begin{equation}
\tilde{\gamma}(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_h^2} e^{-h^2/(2\sigma_n^2)}}{Q \left( \frac{h}{\sigma_h} \right) \left[ 1 - Q \left( \frac{h}{\sigma_h} \right) \right]} dh 
\end{equation}
by numerical integration so that (124) becomes
\begin{equation}
\tilde{\gamma}(\sigma_n^2, \sigma_h^2) = \frac{\gamma(\sigma_n^2, \sigma_h^2)}{2\pi \sigma_n^2 \sqrt{2\pi} \sigma_h}. 
\end{equation}

Next, consider the summation in (102) for $i \geq 2$ and note that $P_{Z_{i-1} \mid H}(z_i^{-1} \mid h) \leq 1$ for all $z_i^{-1}$ and $h$. Therefore, we have
\begin{equation}
\sum_{i=2}^{n} \sum_{z_{i-1}^{-1}} P_{Z_{i-1} \mid H}(z_i^{-1} \mid h) \frac{p_H(h) e^{-(h-d_i)^2/\sigma_n^2}}{2\pi \sigma_n^2 Q \left( \frac{h-d_i}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_i}{\sigma_n} \right) \right]} dh 
\leq \sum_{i=2}^{n} \sum_{z_{i-1}^{-1}} \int_{-\infty}^{\infty} \frac{p_H(h) e^{-(h-d_i)^2/\sigma_n^2}}{2\pi \sigma_n^2 Q \left( \frac{h-d_i}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_i}{\sigma_n} \right) \right]} dh. 
\end{equation}
Since $p_H(h) \leq 1/(\sqrt{2\pi} \sigma_h)$ for all $h$, we can upper bound the integral in (127) by
\begin{equation}
\int_{-\infty}^{\infty} \frac{e^{-(h-d_i)^2/\sigma_n^2} p_H(h)}{2\pi \sigma_n^2 Q \left( \frac{h-d_i}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_i}{\sigma_n} \right) \right]} dh \leq \int_{-\infty}^{\infty} \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi \sigma_n^2 Q \left( \frac{h-d_i}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_i}{\sigma_n} \right) \right]} \frac{1}{\sqrt{2\pi} \sigma_h} dh
\end{equation}
for any $d_i = \tau_i(z_i^{-1})$. Unfortunately, a closed-form solution for the integral in (129) does not seem to be available. Therefore, we compute
\begin{equation}
\tilde{\gamma}(\sigma_n^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_h^2}}{Q \left( \frac{h}{\sigma_h} \right) \left[ 1 - Q \left( \frac{h}{\sigma_h} \right) \right]} dh
\end{equation}
by numerical integration. Consequently, we have
\begin{equation}
\sum_{i=2}^{n} \sum_{z_{i-1}^{-1}} P_{Z_{i-1} \mid H}(z_i^{-1} \mid h) \frac{p_H(h) e^{-(h-d_i)^2/\sigma_n^2}}{2\pi \sigma_n^2 Q \left( \frac{h-d_i}{\sigma_n} \right) \left[ 1 - Q \left( \frac{h-d_i}{\sigma_n} \right) \right]} dh 
\leq \sum_{i=2}^{n} \sum_{z_{i-1}^{-1}} \frac{\tilde{\gamma}(\sigma_n^2)}{2\pi \sigma_h^2 \sqrt{2\pi} \sigma_h} \left( \gamma(\sigma_n^2, \sigma_h^2) + (2^n - 2) \tilde{\gamma}(\sigma_n^2) \right), 
\end{equation}
and for $n \geq 2$ the lower bound on the MSE becomes
\begin{equation}
E \left[ (H - \hat{h}(Z^n))^2 \right] \geq \frac{\left( \frac{1}{\sigma_h^2} + \frac{1}{2\pi \sigma_n^2 \sqrt{2\pi} \sigma_h} \left( \gamma(\sigma_n^2, \sigma_h^2) + (2^n - 2) \tilde{\gamma}(\sigma_n^2) \right) \right)^{-1}}{1 + \frac{\sigma_h^2}{2\pi \sigma_n^2 \sqrt{2\pi} \sigma_h} \left( \gamma(\sigma_n^2, \sigma_h^2) + (2^n - 2) \tilde{\gamma}(\sigma_n^2) \right)} 
\end{equation}

\section*{Appendix D}
\section*{Proof of Theorem 5}
We begin the proof for $n = 1$, and an upper bound for $\gamma^2(\sigma_n^2, \sigma_h^2)$, which is
\begin{equation}
\gamma^2(\sigma_n^2, \sigma_h^2) = \left( \int_{-\infty}^{\infty} h Q \left( \frac{h}{\sigma_n} \right) e^{-h^2/(2\sigma_n^2)} dh \right)^2 
= \left( \frac{1}{2} \int_{-\infty}^{\infty} h \left( 1 - \text{erf} \left( \frac{h}{\sqrt{2\sigma_n}} \right) \right) e^{-h^2/(2\sigma_n^2)} dh \right)^2 
= \left( \int_{0}^{\infty} \text{erf} \left( \frac{h}{\sqrt{2\sigma_n}} \right) e^{-h^2/(2\sigma_n^2)} dh \right)^2 
\end{equation}
\[
\begin{align*}
\gamma^2(\sigma_n^2, \sigma_h^2) & \leq \int_0^\infty h e^{-h^2/(2\sigma_n^2)} \int_0^\infty h e^{-h^2/(2\sigma_h^2)} dh \\
& = \sigma_h^2 \int_0^\infty h e^{-h^2/(2\sigma_h^2)} dh,
\end{align*}
\]
where the last equality follows by \( \int_0^\infty x e^{-ax^2} dx = 1/(2a), \) \( a > 0. \) Applying the bound in (106) to (142) yields
\[
\gamma^2(\sigma_n^2, \sigma_h^2) \leq \frac{2}{\pi} \left( \frac{\pi \sigma_n^2 \sigma_h^2}{\pi \sigma_n^2 + 4\sigma_h^2} \right).
\]

Inserting (145) into (123), we have the bound
\[
E \left[ (H - h_{\text{MMSE}}(Z_1))^2 \right] \geq \frac{\sigma_n^2 - 2}{2} \left( \frac{\pi \sigma_n^2 \sigma_h^2}{\pi \sigma_n^2 + 4\sigma_h^2} \right)
\]
\[
= \left( 1 - \frac{2}{\pi} \right) \frac{\sigma_n^2}{\pi \sigma_n^2 + 4\sigma_h^2}.
\]

Next, consider the case \( n \geq 2 \) and apply Lemma 1 to obtain
\[
\gamma(\sigma_n^2, \sigma_h^2) \leq \int_{-\infty}^\infty \frac{e^{-h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)}}{Q \left( \frac{h}{\sigma_n} \right)} dh
\]
\[
\leq 4 \int_{-\infty}^\infty e^{-(1-2/\pi)h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)} dh
\]
\[
= \int_{-\infty}^\infty \frac{2\pi \sigma_n^2 \sigma_h^2}{\pi \sigma_n^2 + 4\sigma_h^2} \frac{1}{\pi \sigma_n^2 + \sigma_h^2},
\]
where we used \( \int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a}, a > 0. \) Likewise, we obtain the bound
\[
\bar{\gamma}(\sigma_n^2) \leq \int_{-\infty}^\infty \frac{e^{-h^2/\sigma_n^2}}{Q \left( \frac{h}{\sigma_n} \right)} dh
\]
\[
\leq 4 \int_{-\infty}^\infty e^{-(1-2/\pi)h^2/\sigma_n^2} dh
\]
\[
= 4 \int \frac{\pi \sigma_n^2}{(1-2/\pi)}.
\]

Inserting (150) and (153) into the bound of Theorem 4 yields the bound of Theorem 5.
Andrew C. Singer received the S.B., S.M., and Ph.D. degrees, all in electrical engineering and computer science, from the Massachusetts Institute of Technology. Since 1998, he has been on the faculty of the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, where he is currently a Professor in the ECE department and the Coordinated Science Laboratory. Prior to joining the faculty at Illinois, he was a Postdoctoral Research Affiliate at MIT and a Research Scientist at Sanders, A Lockheed Martin Company in Manchester, New Hampshire. At Sanders, he designed algorithms, architectures and systems for a variety of communications and surveillance applications. His research interests include signal processing and communication systems. He 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, in 2001 he received the Xerox Faculty Research Award, and in 2002 was named a Willett Faculty Scholar. Dr. Singer has served as an Associate Editor for the IEEE Transactions on Signal Processing and as a guest editor a number of special issues of the IEEE Transactions on Information Theory, IEEE Transactions on Communications, and IEEE Journal of Selected Topics in Signal Processing. He is a member of the MIT Educational Council and of Eta Kappa Nu and Tau Beta Pi. He is a Fellow of the IEEE.

In 2005, Prof. Singer was appointed as the Director of the Technology Entrepreneur Center (TEC) in the College of Engineering where he oversees educational and extra-curricular activities on entrepreneurship and innovation activities for the college. He co-founded Intersymbol Communications, a fabless semiconductor IC company, based in Champaign Illinois. A developer of signal processing enhanced chips for ultra-high speed optical communications systems, Intersymbol was acquired by Finisar Corporation in 2007. Prof. Singer serves on the board of directors of a number of technology companies and as an expert witness for electronics, communications, and circuit-related technologies.

Georg Zeitler received the M.S. degree in electrical engineering and computer science from the University of Illinois at Urbana-Champaign in 2007. Since 2007, he has been a research and teaching assistant at Technische Universität München, Germany, working towards his Dr.-Ing. degree. His research interests include the effects and optimization of low-precision quantization in communication systems. He received a best paper award at the IEEE International Conference on Communications (ICC) in 2009.

Gerhard Kramer is Alexander von Humboldt Professor and Head of the Institute for Communications Engineering at the Technische Universität München (TUM). He received the Dr. sc. techn. degree from the ETH Zürich in 1998. From 1998 to 2000, he was with Endora Tech AG, Basel, Switzerland, as a communications engineering consultant. From 2000 to 2008 he was with the Math Center, Bell Labs, Murray Hill, NJ, as a Member of Technical Staff. He joined the University of Southern California as Professor in 2009 and TUM in 2010.

Gerhard Kramer was awarded the Vodafone Innovations Prize in 2011, an Alexander von Humboldt Professorship in 2010, the IEEE Communications Society Stephen O. Rice Prize paper award in 2005, a Bell Labs President’s Gold Award in 2003, and an ETH Medal in 1998. He is an IEEE Fellow. He is serving as 1st Vice President of the IEEE Information Theory Society in 2012 and has been on the Society’s Board of Governors since 2009. He was Associate Editor, Guest Editor, and Publications Editor for the IEEE Transactions on Information Theory from 2004-2008. He was a member of the Emerging Technologies Committee of the IEEE Communications Society from 2009-2011.
