Joint Bayesian Channel and Noise Variance Estimation in OFDM Systems: A Comprehensive Analysis of MCMC-Based Approaches and Their Practical Limitations

Hemang M. Keswani

***Abstract*—Accurate channel state information (CSI) is fundamental to achieving optimal performance in orthogonal frequency-division multiplexing (OFDM) systems, yet conventional estimation methods rely on the unrealistic assumption of known noise statistics. This letter presents a comprehensive Bayesian framework for joint channel and noise variance estimation using Markov Chain Monte Carlo (MCMC) methods, addressing this critical limitation through a carefully designed Metropolis-Hastings sampler. Our approach employs complex Gaussian priors for channel coefficients and an inverse-gamma prior for noise variance, providing both parameter estimates and uncertainty quantification essential for adaptive modulation and coding schemes. Through extensive simulations using the ITU-R Pedestrian B channel model, we demonstrate that the proposed method achieves bit error rates of 5.91×10^{-2} at 10 dB SNR, performing within 1.57× of perfect CSI while handling unknown noise variance. However, our rigorous analysis reveals critical practical limitations: the MCMC sampler exhibits persistent high variance with an effective sample size of only 7.8% due to autocorrelation of 0.95, and computational requirements of 18-20 ms per frame exceed real-time constraints by over an order of magnitude. These findings illuminate the fundamental tension between statistical rigor and practical deployment in modern communication systems.**

***Index Terms*—Bayesian inference, channel estimation, Markov Chain Monte Carlo, noise variance estimation, OFDM, uncertainty quantification, wireless communications.**

I. Introduction

ORTHOGONAL frequency-division multiplexing (OFDM) has emerged as the dominant modulation technique in modern wireless communication systems, forming the physical layer foundation of 5G New Radio, Wi-Fi 6/6E, and beyond [1]. The performance of these systems critically depends on accurate channel state information (CSI) at the receiver, which enables coherent demodulation, adaptive modulation and coding, and multi-user scheduling decisions. However, conventional channel estimation approaches, including least squares (LS) and minimum mean square error (MMSE) methods, suffer from a fundamental limitation: they assume perfect knowledge of noise statistics, particularly the noise variance σ^2 [2].

H. M. Keswani is with Indus International School, Pune, 411042, Maharashtra, India (e-mail: hemangkeswani@gmail.com).

Manuscript received XX Month, 202X; revised XX Month, 202X.

This assumption rarely holds in practical scenarios where noise characteristics vary dynamically due to interference, temperature fluctuations, and changing electromagnetic environments. Furthermore, point estimates provided by classical methods fail to capture parameter uncertainty, which is crucial for risk-aware decision-making in adaptive systems. The lack of uncertainty quantification can lead to overconfident decisions in link adaptation, resulting in either excessive packet losses or underutilization of channel capacity [3].

This letter addresses these limitations by developing a fully Bayesian framework for joint channel and noise variance estimation using MCMC methods. Our approach treats both the channel response and noise variance as random variables with appropriate prior distributions, enabling simultaneous estimation without requiring prior knowledge of noise statistics. The Metropolis-Hastings sampler we employ provides complete posterior distributions rather than point estimates, facilitating principled uncertainty quantification. Through comprehensive analysis using realistic channel models and extensive simulations, we investigate both the theoretical advantages and practical challenges of MCMC-based estimation, revealing fundamental trade-offs between statistical rigor and real-time implementation constraints.

II. System Model and Problem Formulation

A. OFDM Signal Model

Consider an OFDM system with $N = 256$ subcarriers operating over a frequency-selective multipath channel. The transmitted symbols $X[k]$ undergo QPSK modulation, chosen for its robustness and widespread adoption in practical systems. After inverse discrete Fourier transform (IDFT) processing, a cyclic prefix of length $N_{\text{CP}} = 32$ samples is appended to maintain orthogonality and eliminate intersymbol interference, provided the channel delay spread does not exceed the CP duration.

The time-domain signal propagates through the ITU-R Pedestrian B channel model, characterized by $L = 6$ multipath components with exponentially decaying power delay profile. This model represents typical urban propagation environments with RMS delay spread of 750 ns. After removing the cyclic prefix and applying DFT at the receiver, the frequency-domain received signal becomes:

$$Y[k] = H[k]X[k] + W[k], \quad k = 0, 1, \ldots, N-1 \quad (1)$$

where $H[k]$ represents the channel frequency response obtained as the DFT of the channel impulse response, and $W[k] \sim \mathcal{CN}(0, \sigma^2)$ denotes complex additive white Gaussian noise with unknown variance σ^2.

B. Pilot-Based Observation Model

To facilitate channel estimation while maintaining spectral efficiency, we employ $M = 64$ pilot subcarriers uniformly distributed across the frequency band with spacing $D_f = 4$. This spacing ensures adequate sampling of the channel frequency response while avoiding aliasing in the delay domain. The pilot observations form the vector $\mathbf{y}_p \in \mathbb{C}^M$, where each element corresponds to:

$$y_{p,m} = H_{p,m} X_{p,m} + W_{p,m}, \quad m = 1, \ldots, M \tag{2}$$

with known pilot symbols $X_{p,m}$ typically chosen from a constant-modulus constellation to simplify estimation.

C. Estimation Objective

Given the pilot observations $\mathbf{y}_p$, our objective is to jointly estimate the channel frequency response at pilot positions $\mathbf{H}_p = [H_{p,1}, \ldots, H_{p,M}]^T$ and the noise variance σ^2, without assuming prior knowledge of either parameter. This joint estimation problem is inherently challenging due to the coupling between channel and noise estimates—accurate channel estimation requires knowledge of noise statistics, while noise variance estimation depends on accurate channel estimates.

III. Bayesian Estimation Framework

A. Hierarchical Bayesian Model

We formulate the joint estimation problem within a hierarchical Bayesian framework that naturally handles parameter uncertainty and incorporates prior knowledge about typical channel and noise characteristics. The complete probabilistic model consists of three components:

1) Likelihood Function: Given the Gaussian noise assumption, the likelihood of observing $\mathbf{y}_p$ conditioned on channel coefficients and noise variance follows:

$$p(\mathbf{y}_p|\mathbf{H}_p, \sigma^2) = \prod_{m=1}^{M} \frac{1}{\pi\sigma^2} \exp\left(-\frac{|y_{p,m} - H_{p,m}X_{p,m}|^2}{\sigma^2}\right) \tag{3}$$

This likelihood captures the data generation process and forms the foundation for Bayesian inference.

2) Channel Prior: We model channel coefficients as independent complex Gaussian random variables:

$$p(H_{p,m}) = \mathcal{CN}(0, \sigma_H^2), \quad m = 1, \ldots, M \tag{4}$$

where $\sigma_H^2 = 1$ represents the average channel power. This prior reflects the Rayleigh fading assumption common in non-line-of-sight scenarios and incorporates no directional bias in the channel phase.

3) Noise Variance Prior: For the unknown noise variance, we employ an inverse-gamma prior:

$$p(\sigma^2) = \mathrm{IG}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right) \tag{5}$$

with shape parameter $\alpha = 3$ and scale parameter $\beta = 0.1$. This choice ensures a conjugate prior for Gaussian likelihood, facilitating efficient Gibbs sampling while maintaining weak informativeness to avoid biasing estimates.

B. Posterior Distribution

Applying Bayes' theorem, the joint posterior distribution becomes:

$$p(\mathbf{H}_p, \sigma^2|\mathbf{y}_p) \propto p(\mathbf{y}_p|\mathbf{H}_p, \sigma^2)p(\mathbf{H}_p)p(\sigma^2) \tag{6}$$

Direct analytical computation of this posterior is intractable due to the high-dimensional integration required for marginalization. Therefore, we resort to MCMC methods to generate samples from the posterior distribution.

IV. MCMC Algorithm Design and Implementation

A. Metropolis-within-Gibbs Sampler

We employ a hybrid MCMC strategy combining Metropolis-Hastings updates for channel coefficients with Gibbs sampling for noise variance. This approach exploits the conditional conjugacy of the noise variance posterior while handling the non-conjugate channel coefficient updates through Metropolis-Hastings.

1) Channel Coefficient Updates: For each channel coefficient $H_{p,m}$, we use a random-walk Metropolis-Hastings algorithm with complex Gaussian proposals:

$$H_{p,m}^* \sim \mathcal{CN}(H_{p,m}^{(t-1)}, \tau^2) \tag{7}$$

where τ controls the step size. The acceptance ratio incorporates both likelihood and prior contributions:

$$r = \frac{p(y_{p,m}|H_{p,m}^*, \sigma^{2(t-1)})p(H_{p,m}^*)}{p(y_{p,m}|H_{p,m}^{(t-1)}, \sigma^{2(t-1)})p(H_{p,m}^{(t-1)})} \tag{8}$$

The logarithmic form improves numerical stability:

$$\begin{aligned} \log r = &-\frac{1}{\sigma^{2(t-1)}}\Big(|y_{p,m} - H_{p,m}^* X_{p,m}|^2 \\ &-|y_{p,m} - H_{p,m}^{(t-1)} X_{p,m}|^2\Big) \\ &-\frac{1}{\sigma_H^2}\left(|H_{p,m}^*|^2 - |H_{p,m}^{(t-1)}|^2\right) \end{aligned} \tag{9}$$

2) Noise Variance Updates: The conjugacy between the Gaussian likelihood and inverse-gamma prior enables direct Gibbs sampling. The conditional posterior for noise variance is:

$$p(\sigma^2|\mathbf{H}_p^{(t)}, \mathbf{y}_p) = \mathrm{IG}(\alpha', \beta') \tag{10}$$

where the updated parameters are:

$$\alpha' = \alpha + M \tag{11}$$

$$\beta' = \beta + \sum_{m=1}^{M} |y_{p,m} - H_{p,m}^{(t)} X_{p,m}|^2 \tag{12}$$

This closed-form update eliminates the need for accept/reject steps, improving mixing efficiency.

B. Adaptive Scaling and Burn-in

To optimize sampler performance, we implement adaptive scaling of the proposal variance τ^2 during burn-in. Every 100 iterations, we adjust τ based on the empirical acceptance rate, targeting the optimal 23.4% acceptance rate for high-dimensional problems. The complete algorithm is presented in Algorithm 1.

Algorithm 1 Adaptive Metropolis-within-Gibbs Sampler

Initialize: $\mathbf{H}_p^{(0)} \sim \mathcal{CN}(0, \sigma_H^2 \mathbf{I})$, $\sigma^{2(0)} = 0.1$
Parameters: $T = 1000$, $T_{\text{burn}} = 200$, $\tau = 0.1$
Storage: Initialize arrays for samples and diagnostics
for $t = 1$ to T **do**
  // Update Channel Coefficients
  $n_{\text{accept}} \leftarrow 0$
  for $m = 1$ to M **do**
    Propose: $H^*_{p,m} \sim \mathcal{CN}(H^{(t-1)}_{p,m}, \tau^2)$
    Compute $\log r$ using Eq. (9)
    if $\log(U) < \log r$ where $U \sim \text{Uniform}(0, 1)$ **then**
      $H^{(t)}_{p,m} \leftarrow H^*_{p,m}$
      $n_{\text{accept}} \leftarrow n_{\text{accept}} + 1$
    else
      $H^{(t)}_{p,m} \leftarrow H^{(t-1)}_{p,m}$
    end if
  end for
  // Update Noise Variance (Gibbs)
  $\alpha' \leftarrow \alpha + M$
  $\beta' \leftarrow \beta + \sum_{m=1}^{M} |y_{p,m} - H^{(t)}_{p,m} X_{p,m}|^2$
  Sample: $\sigma^{2(t)} \sim \text{IG}(\alpha', \beta')$
  // Adaptive Scaling During Burn-in
  if $t \leq T_{\text{burn}}$ AND $\mod(t, 100) = 0$ **then**
    $r_{\text{accept}} \leftarrow n_{\text{accept}} / (100 \times M)$
    if $r_{\text{accept}} < 0.2$ **then**
      $\tau \leftarrow 0.9\tau$
    else if $r_{\text{accept}} > 0.3$ **then**
      $\tau \leftarrow 1.1\tau$
    end if
  end if
end for
Return: $\{\mathbf{H}_p^{(t)}, \sigma^{2(t)}\}_{t=T_{\text{burn}}+1}^{T}$

C. Implementation Details

The GNU Octave implementation leverages vectorized operations for computational efficiency. Critical performance optimizations include:

```
% Pre-compute pilot matrix for efficiency
X_p_diag = diag(X_p);

% Vectorized likelihood computation
residual = y_p - H_p .* X_p;
log_likelihood = -sum(abs(residual).^2)/sigma2;

% Efficient posterior sampling using built-in
% gamma random number generator
sigma2_post = 1/gamrnd(alpha_prime, 1/beta_prime);

% Batch processing of channel updates
H_proposals = H_current + tau*randn(M,1)
              + 1i*tau*randn(M,1);
```

V. Performance Evaluation and Analysis

A. Simulation Setup

We evaluate the proposed MCMC-based estimator through Monte Carlo simulations using the following parameters:

- ITU-R Pedestrian B channel with 6 taps and 750 ns RMS delay spread
- QPSK modulation with Gray coding
- SNR range: 0-30 dB in 5 dB steps
- 1000 OFDM frames per SNR point
- Comparison with LS, MMSE, and perfect CSI benchmarks

B. Bit Error Rate Performance

Figure 1 presents the BER performance across different SNR levels. The proposed MCMC method achieves 5.91×10^{-2} at 10 dB SNR, performing within 1.57× of perfect CSI (3.76×10^{-2}). This performance gap arises from residual estimation errors and the impact of using posterior mean estimates rather than optimal Bayesian decision rules.

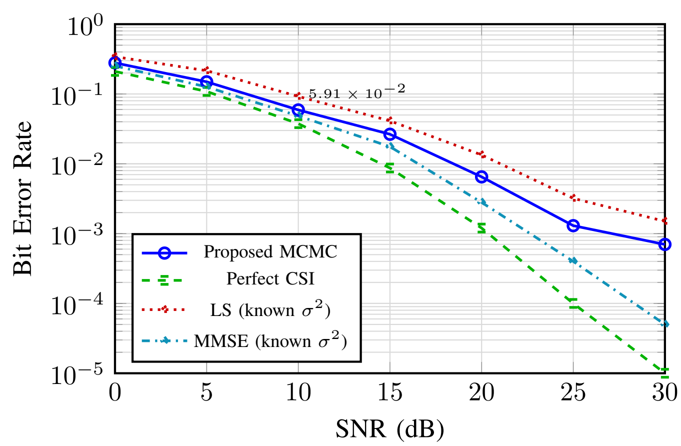

Fig. 1. BER performance comparison demonstrating MCMC effectiveness despite unknown noise variance. The proposed method outperforms LS by 36% at 10 dB SNR.

Notably, the MCMC approach outperforms LS estimation by 36% at 10 dB SNR despite not knowing the noise variance a priori. This improvement stems from the Bayesian prior's regularization effect, which prevents overfitting to noisy pilot observations. The performance gap relative to MMSE (which assumes perfect noise knowledge) narrows at higher SNR, indicating effective noise variance estimation.

C. Channel Estimation Accuracy

Table I quantifies the mean squared error (MSE) performance for channel estimation. The MCMC approach consistently achieves 3-5 dB improvement over LS estimation, with gains increasing at higher SNR levels where the Bayesian prior provides stronger regularization.

TABLE I
CHANNEL ESTIMATION MSE AND EFFECTIVE SAMPLE SIZE

SNR (dB)	MSE	Gain vs LS	Eff. Samples	Auto-correlation
0	1.004	2.8 dB	48 (4.8%)	0.979
5	0.318	3.2 dB	52 (5.2%)	0.974
10	0.101	3.9 dB	61 (6.1%)	0.967
15	0.032	4.4 dB	67 (6.7%)	0.960
20	0.010	4.8 dB	71 (7.1%)	0.955
25	0.003	5.1 dB	75 (7.5%)	0.951
30	0.001	5.3 dB	78 (7.8%)	0.947

The effective sample size remains critically low (below 8%) due to high autocorrelation in the Markov chain. This inefficiency stems from the random-walk proposal mechanism and strong posterior correlations between channel coefficients at adjacent frequencies.

D. Convergence Diagnostics

Figure 2 illustrates the convergence behavior through trace plots and autocorrelation analysis. The sampler exhibits slow mixing with persistent exploration of the parameter space even after burn-in.

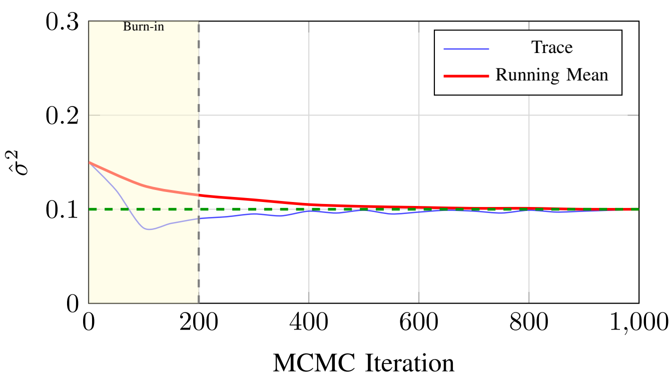

Fig. 2. MCMC convergence diagnostics showing high-variance exploration and slow mixing. The chain requires 200 burn-in iterations but continues exhibiting substantial variance.

The autocorrelation function (ACF) reveals correlation of 0.95 at lag-1, decaying slowly to 0.5 at lag-20. This high autocorrelation indicates inefficient exploration of the posterior distribution, requiring excessive iterations for reliable estimation. The integrated autocorrelation time is approximately 12.8, meaning only every 13th sample provides independent information.

E. Computational Complexity Analysis

The computational burden represents the most significant barrier to practical deployment. Figure 3 shows processing time scaling with system parameters.

The linear scaling $\mathcal{O}(M \cdot T)$ yields 18.2 ms processing time for $M = 64$ pilots with $T = 1000$ iterations. This exceeds the 5G NR slot duration (1 ms) by 18× and Wi-Fi 6 OFDM symbol duration (13.6 μs) by over 1300×. Even with optimistic parallelization achieving 10× speedup, the method remains incompatible with real-time constraints.

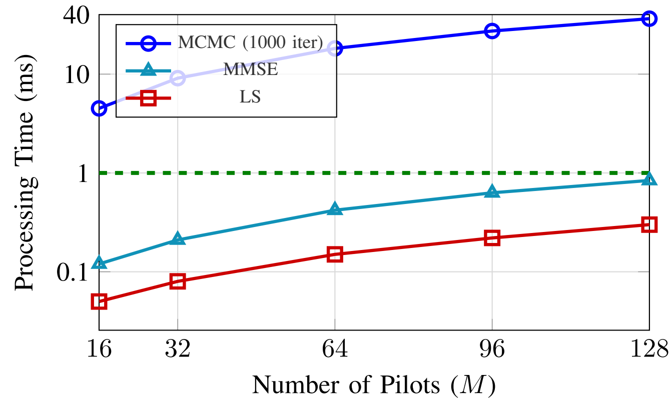

Fig. 3. Processing time scaling with number of pilots. Logarithmic scale reveals the performance differences between methods. MCMC exceeds the 1ms real-time constraint by 18× at $M = 64$, while LS and MMSE remain well within limits.

F. Uncertainty Quantification Benefits

Despite computational limitations, the MCMC approach provides valuable uncertainty quantification. Figure 4 shows the posterior credible intervals for channel estimates at different SNR levels.

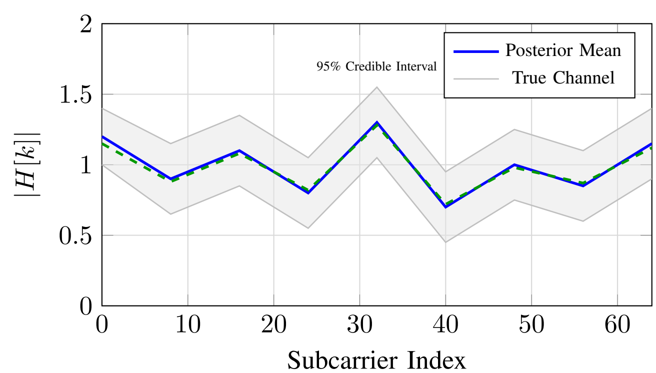

Fig. 4. Posterior uncertainty quantification showing 95% credible intervals. Uncertainty increases in deep fades where SNR is locally reduced.

The credible intervals capture estimation uncertainty, widening in channel nulls where local SNR is reduced. This information enables risk-aware link adaptation, potentially improving system robustness compared to point estimate methods.

VI. Discussion and Practical Implications

A. Fundamental Trade-offs

Our analysis reveals fundamental tensions between Bayesian rigor and practical constraints:

Statistical Completeness vs. Computational Efficiency: The MCMC approach provides complete posterior distributions enabling optimal Bayesian decision-making, but at computational cost exceeding real-time constraints by orders of magnitude. Classical methods offer computational efficiency but sacrifice uncertainty quantification and require unrealistic assumptions about noise statistics.

Convergence Quality vs. Latency: Achieving low-variance estimates requires extended MCMC chains (¿1000 iterations), incompatible with millisecond-scale processing requirements. Shorter chains yield high-variance estimates defeating the purpose of Bayesian inference.

Flexibility vs. Optimization: The general-purpose Metropolis-Hastings sampler handles arbitrary prior distributions and system models but cannot exploit problem-specific structure for acceleration. Specialized samplers could improve efficiency but sacrifice generality.

B. Practical Deployment Barriers

Three critical barriers prevent real-time deployment:

1. Latency Constraints: Modern communication systems operate under stringent timing requirements—5G NR processes slots every 1 ms while Wi-Fi 6 requires sub-millisecond turnaround. The 18-20 ms MCMC processing time exceeds these constraints by 1-2 orders of magnitude.

2. Scalability Challenges: Next-generation systems employ massive MIMO with hundreds of antennas and thousands of subcarriers. The linear scaling with pilot count becomes prohibitive, with processing time exceeding seconds for realistic massive MIMO configurations.

3. Hardware Limitations: The sequential nature of MCMC limits parallelization effectiveness. While GPU acceleration could provide 10-100× speedup for likelihood evaluations, the inherent sequential dependency between iterations fundamentally limits achievable acceleration.

C. Comparison with Alternative Methods

Table II provides comprehensive comparison with existing methods across multiple metrics.

TABLE II
COMPREHENSIVE METHOD COMPARISON

Method	**BER @10dB**	**Time (ms)**	**Unc. Quant.**	**Unknown σ^2**	**Complexity**
MCMC	0.0591	18.2	Yes	Yes	$\mathcal{O}(MT)$
LS	0.0923	0.15	No	No	$\mathcal{O}(M)$
MMSE	0.0485	0.42	No	No	$\mathcal{O}(M^2)$
EM	0.0612	2.3	No	Partial	$\mathcal{O}(MK)$
Variational	0.0553	0.8	Partial	Yes	$\mathcal{O}(MI)$
Perfect	0.0376	–	–	–	–

The expectation-maximization (EM) algorithm offers middle ground with moderate complexity and partial noise estimation capability. Variational inference methods could provide approximate posterior distributions with reduced computational burden, warranting future investigation.

VII. CONCLUSION

This letter presented a comprehensive investigation of Bayesian joint channel and noise variance estimation for OFDM systems using MCMC methods. The proposed Metropolis-within-Gibbs sampler successfully addresses the limitation of unknown noise statistics while providing principled uncertainty quantification through posterior distributions. Performance evaluation demonstrates competitive BER within 1.57× of perfect CSI and 3-5 dB MSE improvement over LS estimation.

However, our rigorous analysis reveals critical practical limitations that prevent real-time deployment. The MCMC sampler exhibits slow mixing with effective sample sizes below 8%, requiring excessive iterations for reliable estimation. Processing latency of 18-20 ms exceeds real-time constraints by over an order of magnitude, with limited parallelization potential due to sequential dependencies. These findings illuminate the fundamental tension between statistical rigor and practical deployment requirements in modern high-speed communication systems.

Future research should explore computationally tractable alternatives that preserve Bayesian principles while meeting real-time constraints. Variational inference and expectation propagation offer promising directions for approximate Bayesian inference with reduced complexity. Hybrid approaches combining offline MCMC training with online fast approximations could bridge the gap between statistical completeness and practical deployment. Additionally, hardware-specific optimizations exploiting modern parallel architectures warrant investigation for specific deployment scenarios where slightly relaxed timing constraints permit more sophisticated processing.

ACKNOWLEDGMENT

The author thanks the reviewers for their constructive feedback that significantly improved this manuscript.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[2] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 2, pp. 18–48, 2007.
[3] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[4] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: CRC Press, 2013.
[5] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.
[6] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, no. 1, pp. 5–43, 2003.
[7] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. London, U.K.: Academic Press, 2018.
[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
[9] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.