LOW POWER STRATEGIES FOR ADAPTIVE FILTERING AND EQUALIZATION

A. O. Hero EECS, University of Michigan, Ann Arbor

3rd Year ARO review, July 1999

Accomplishments in years 1-3

Guiding principle: include power constraints up-front

1. **Register-length, power, and optimal bit allocation (Gupta)**
   1.1. Reduced register-length non-adaptive filtering
   1.2. Reduced register-length adaptive filtering
2. **Successive weight updating, power, and convergence (Godavarti)**
3. **Proximal point bundle methods for function optimization (Chretien)**

Impact on SP Applications

- matched filters and correlators
- channel equalization
- space-time processing
- adaptive anti-jam and noise cancelation
- adaptive multipath combining
- adaptive nulling and beamforming arrays
- adaptive source separation

I. Register-Length vs. Power

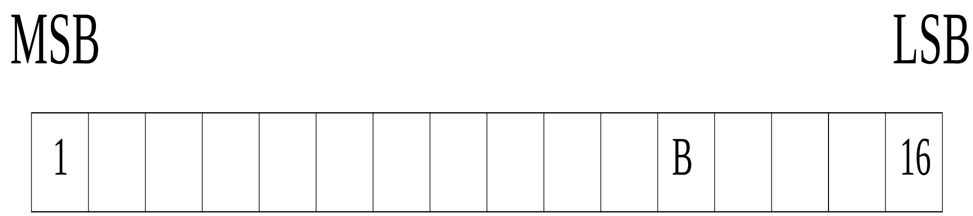


Figure 1: *Fixed point B-bit register*

Define:

- B – number of bits available
- $R(k)$ – process auto-correlation $E[d_l d_{l-k}^*]$
- η – power dissipated per bit transition

Power dissipation per unit time for B bit register with Gaussian data:

$$P_B \;\leq\; B\eta \cdot \left[1 - \frac{1}{2}\mathrm{erf}\left(\left[2^B\sqrt{2R(0) - 2R(1)}\right]^{-1}\right)\right]$$

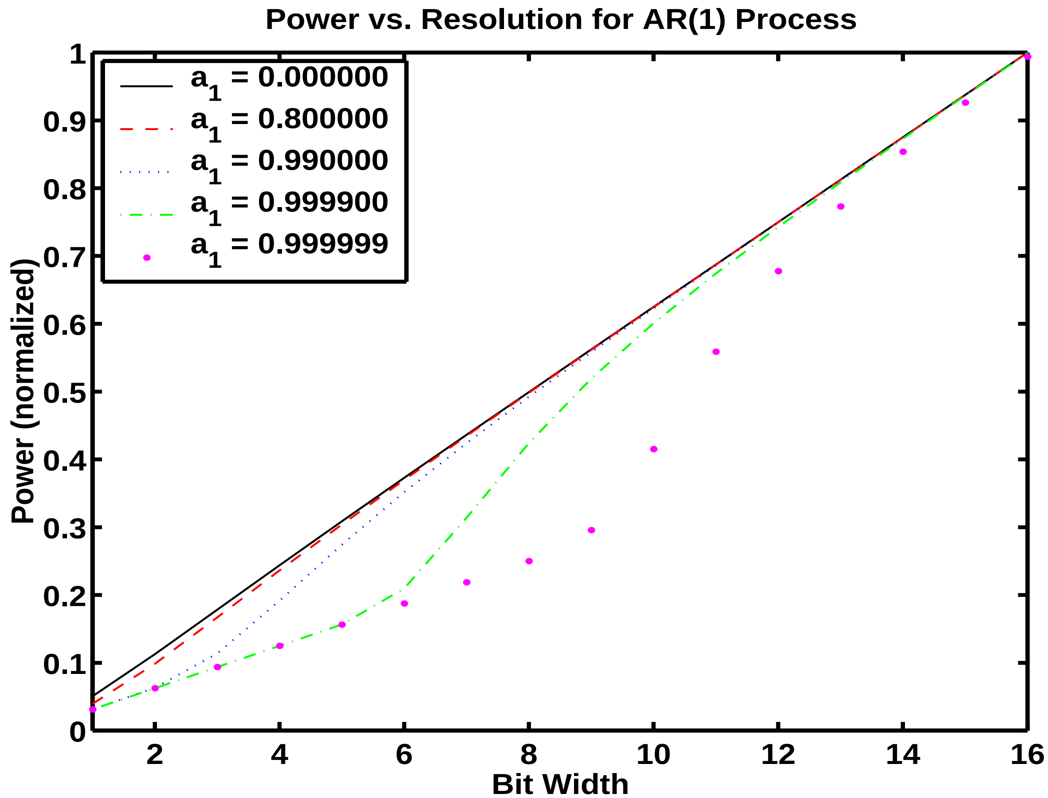


Figure 2: *Normalized power versus bit width b as a function of AR parameter a_1.*

For $|a_1| < 0.8$, power increases approximately linearly as a function of B.

Full Resolution FIR Filter

$$\hat{Y}_k = \underline{W}^H \underline{X}_k$$

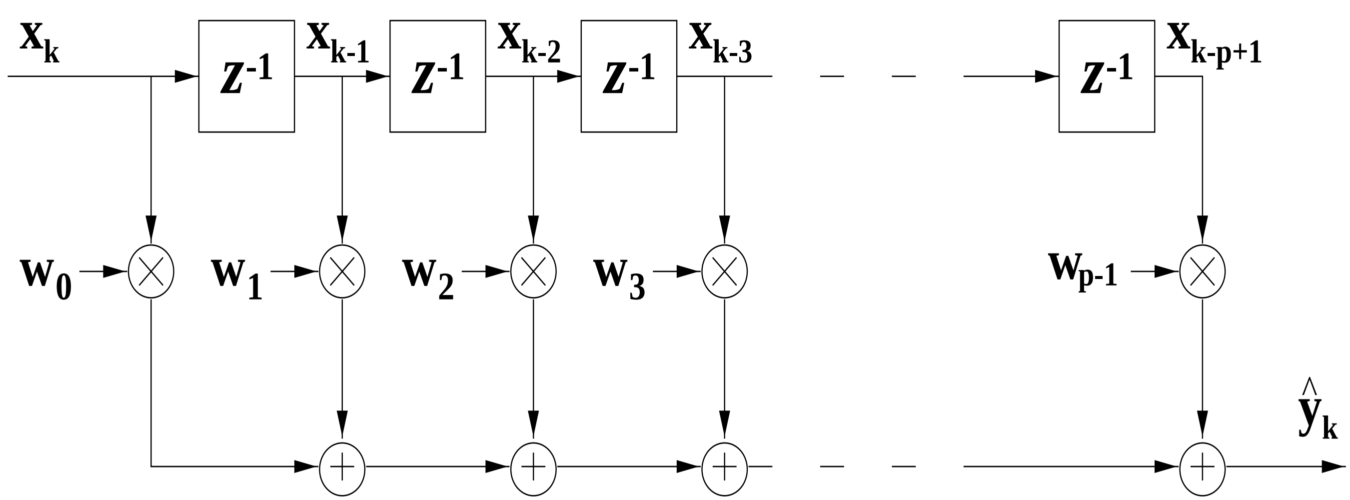


Figure 3: *Infinite precision FIR filter implemented as tapped delay line*

Reduced Resolution FIR Filter

$$Q_d(\hat{Y}_k) = Q_d\left(Q_c(\underline{W})^H Q_d(\underline{X}_k)\right)$$

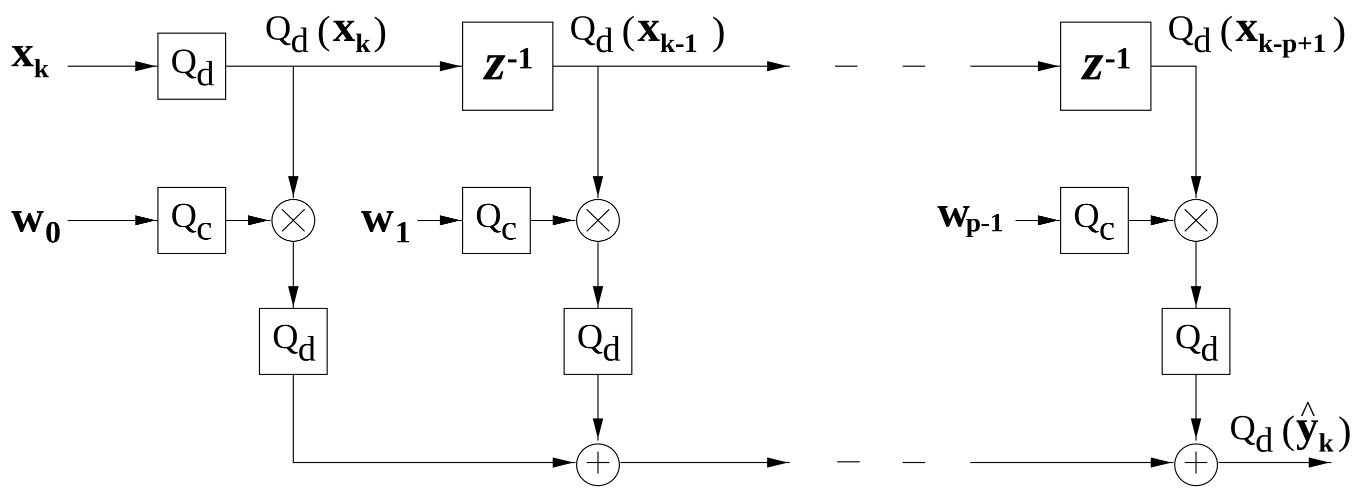


Figure 4: *Finite precision FIR filter implemented as tapped delay line*

Full Resolution FIR Adaptive Filter

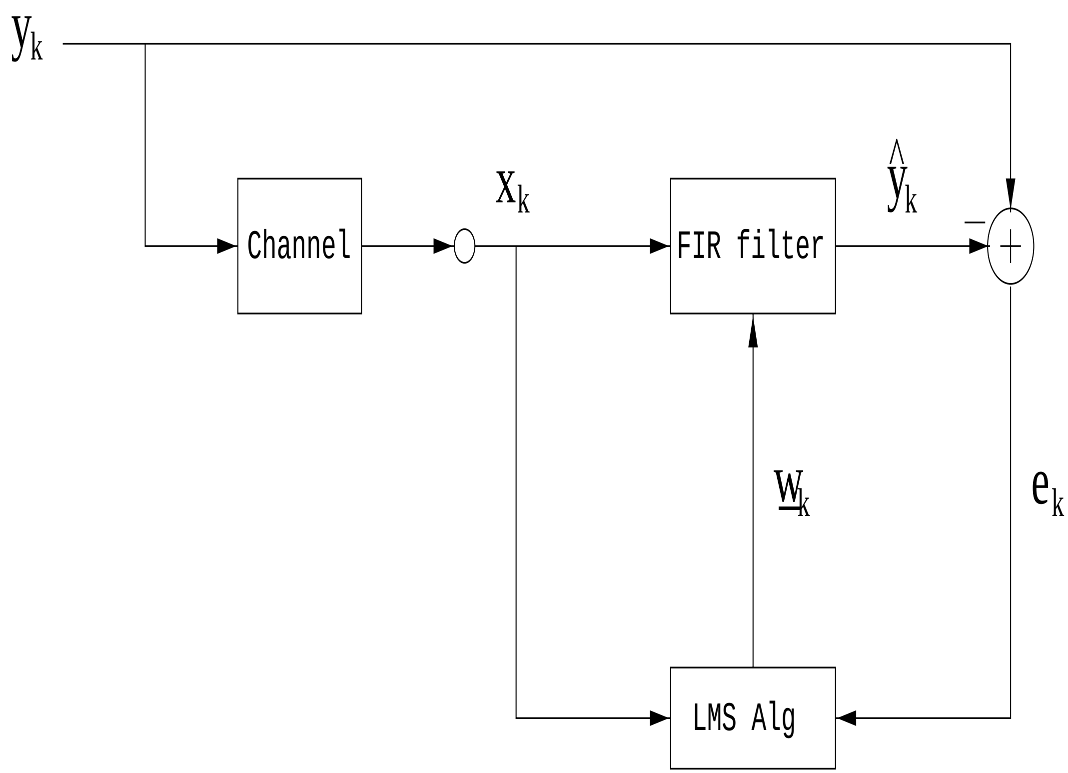


Figure 5: *Adaptive channel equalizer using LMS with training sequence y_k.*

Reduced Resolution FIR Adaptive Filter

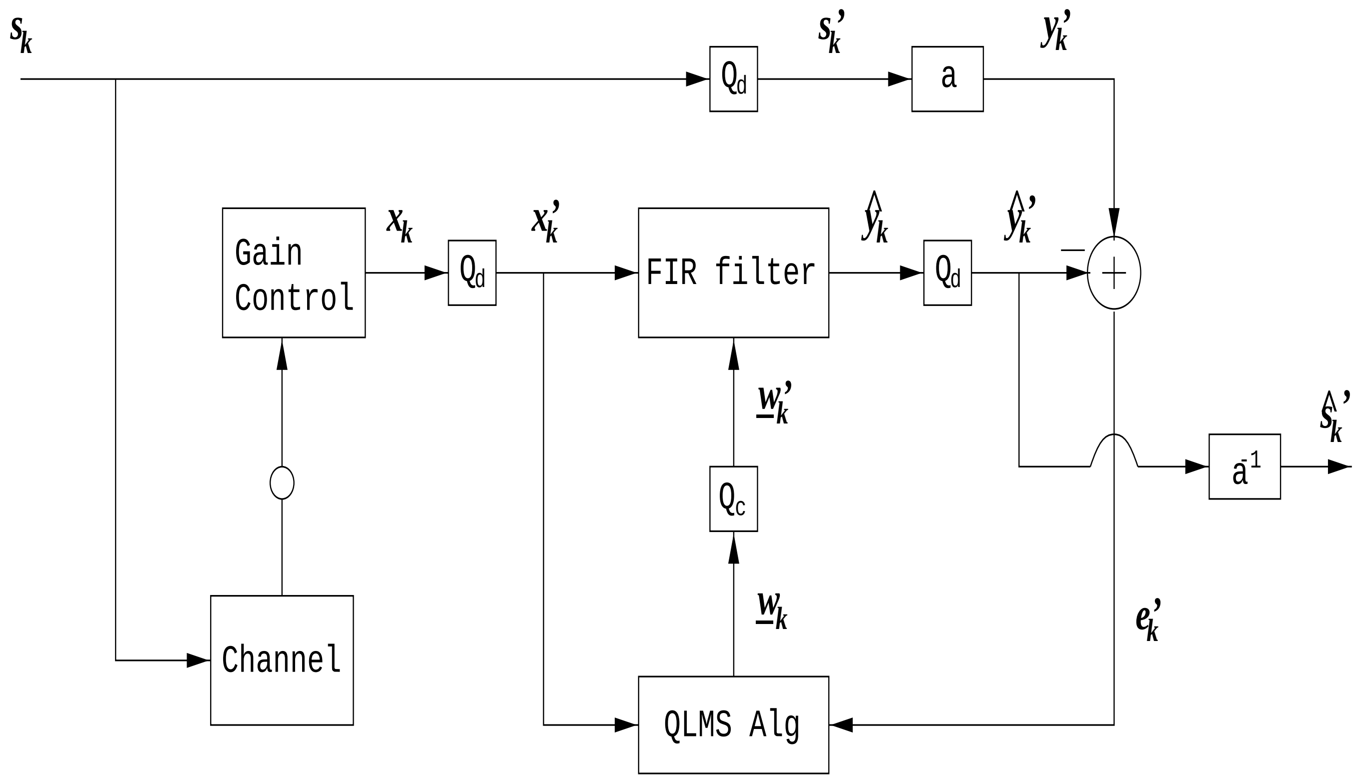


Figure 6: *Adaptive channel equalizer using LMS with training sequence s_k. Q_d and Q_c are uniform scalar quantizers using $B_d + 1$ and $B_c + 1$ bits, respectively. Scaling factor a is used to prevent overflow.*

Quantized Algorithm: Mathematical Model

- $Q_d()$ quantizes to $B_d + 1$ data bits
- $Q_c()$ quantizes to $B_c + 1$ coefficient bits

Quantized FIR Digital Filter

$$Q_d(\hat{y}_k) = Q_d\left(Q_d(\underline{X}_k)^H Q_c(\underline{W}_k)\right) = \underline{X}_k^H \underline{W}_k + q_k^{\hat{y}}$$

Quantized LMS Algorithm

$$\underline{W}_{k+1} = \underline{W}_k + Q_c\left(\mu\, Q_d(\underline{X}_k) e_k\right) = \underline{W}_k + \underline{q}_k^w$$

$$e_k = Q_d(y_k) - Q_d(\hat{y}_k) = y_k - \hat{y}_k + q_k^e$$

Define:

- total bit width: $B_T = B_d + B_c$
- data bit allocation factor: $\rho = B_d / B_T$

Under white q_k assumption

$$\mathrm{MSE}_{\mathrm{excess}} =: \xi_{\mathrm{q}} \;=\; \alpha_c \, 2^{-2(1-\rho)B_T} + \alpha_d \, 2^{-2\rho B_T}$$

where

$$\alpha_c = \frac{p\sigma_x^2}{6}, \quad \alpha_d = \frac{\|\underline{W}\|^2 + p}{6}, \;\text{(for Quantized FIR filter)}$$

$$\alpha_c = \frac{p}{12\mu a^2}, \; \alpha_d = \frac{\|\underline{w}^o\|^2 + p}{6a^2}, \;\text{(for Quantized LMS)}$$

Filter Power Dissipation

Define

- η_t = power per table-lookup per bit
- η_g = power per logic gate
- p = vector length (# of filter taps)

total power/iteration of complex FIR filter

Table lookup mult.

$$P_T = [(32p - 12)B_d + 16pB_c - 8p - 4]\,\eta_g + [8pB_d + 4pB_c]\,\eta_t,$$

Partial product mult.

$$P_T = [28pB_dB_c + (52p - 12)B_d + 28pB_c + 36p - 4]\eta_g$$

total power/iteration of complex LMS filter:

$$P_T = [24p(3B_d + B_c - 2) + 32p]\,\eta_g + 24pB_d\eta_t, \quad \text{(Table lookup mult.)}$$

$$P_T = [56pB_d^2 + 138pB_d + 24pB_c + 72p]\eta_g, \quad \text{(Partial product mult.)}$$

LMS Power Dissipation

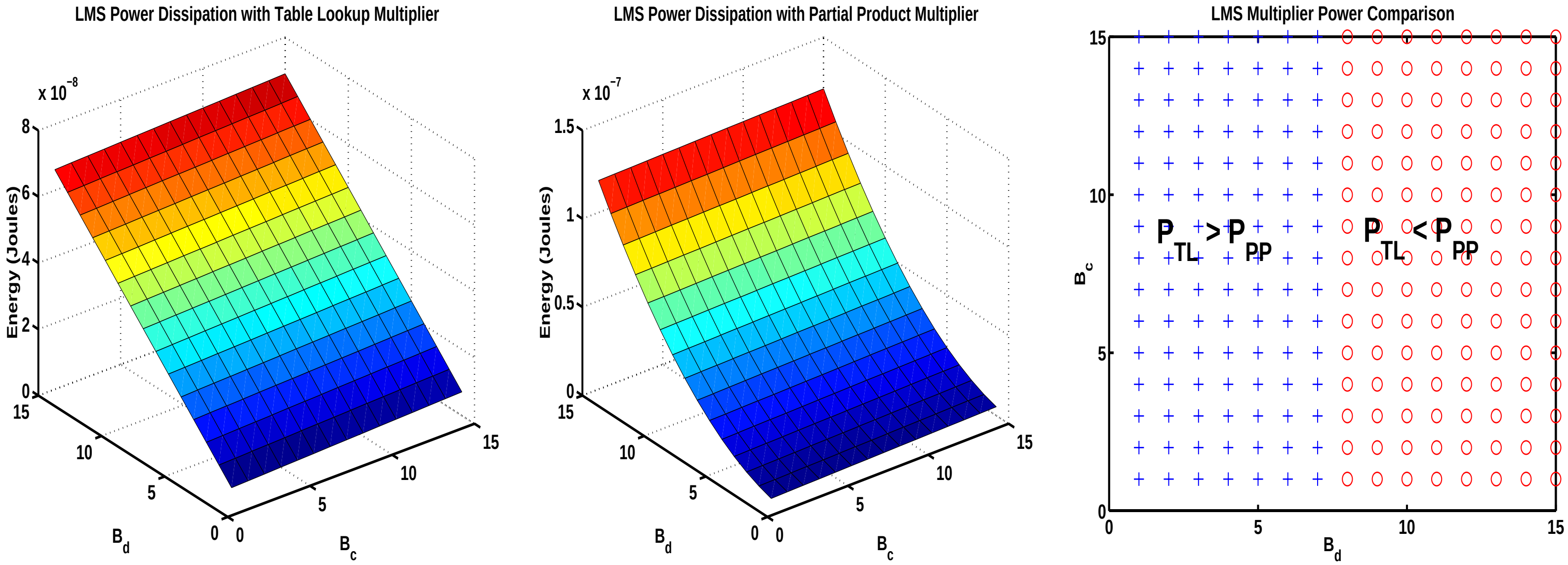


Figure 7: *LMS Power Dissipation vs. B_d and B_c with table lookup and partial product accumulation multipliers*

FIR Filter Power Dissipation

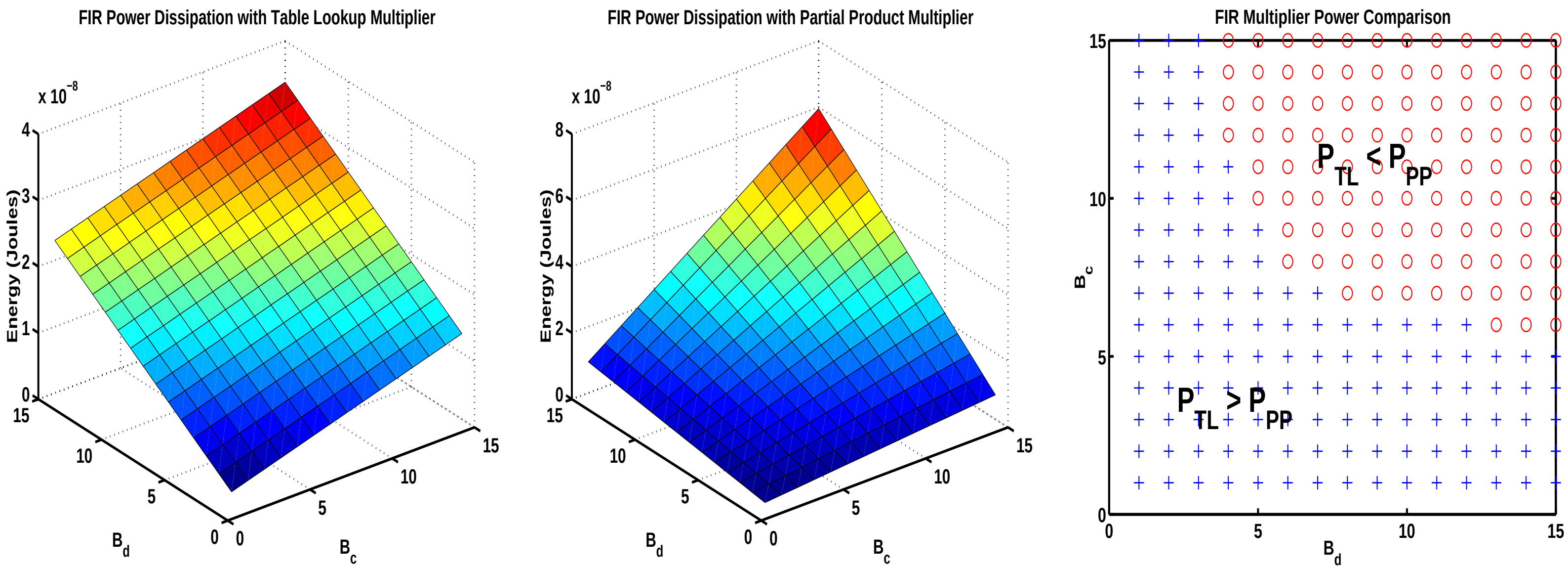
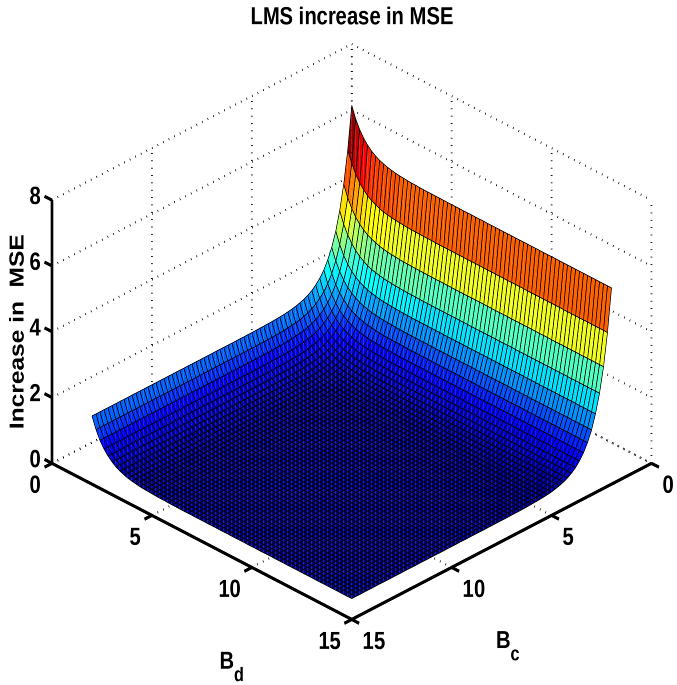


Figure 8: *FIR Filter Power Dissipation vs. B_d and B_c with table lookup and partial product accumulation multipliers*

Increase in MSE due to quantization



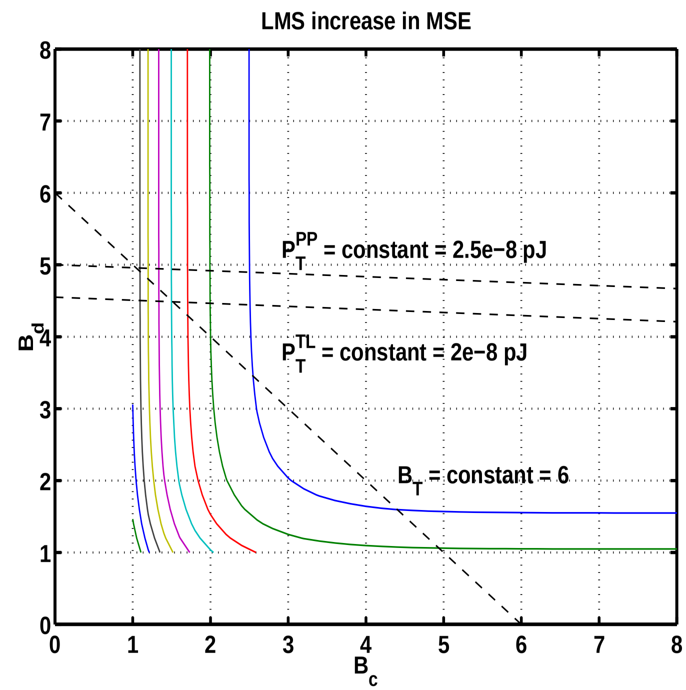


Figure 9: *Excess MSE as a function of B_d and B_c for single pole IIR channel.*

Optimal bit allocation strategy for fixed P_T

Relation between total bit allocation and power (table lookup)

$$B_T = \frac{P_T + 16p\eta_g}{p[\rho(48\eta_g + 24\eta_t) + 24\eta_g]}$$

Optimal bit allocation factor ρ:

$$\min_{\rho} \xi_q \;=\; \alpha_c\, 2^{-2(1-\rho^{**})B_T} + \alpha_d\, 2^{-2\rho^{**} B_T}$$

where

$$\rho^{**} \;=\; \frac{\log_2\left[\frac{24\eta_g\alpha_d}{(72\eta_g+24\eta_t)\alpha_c}\right]\frac{24\eta_g p}{P_T+16\eta_g p} + 2}{-\log_2\left[\frac{24\eta_g\alpha_d}{(72\eta_g+24\eta_t)\alpha_c}\right]\frac{(48\eta_g+24\eta_t)p}{P_T+16\eta_g p} + 4}$$

MSE Performance vs. P_T

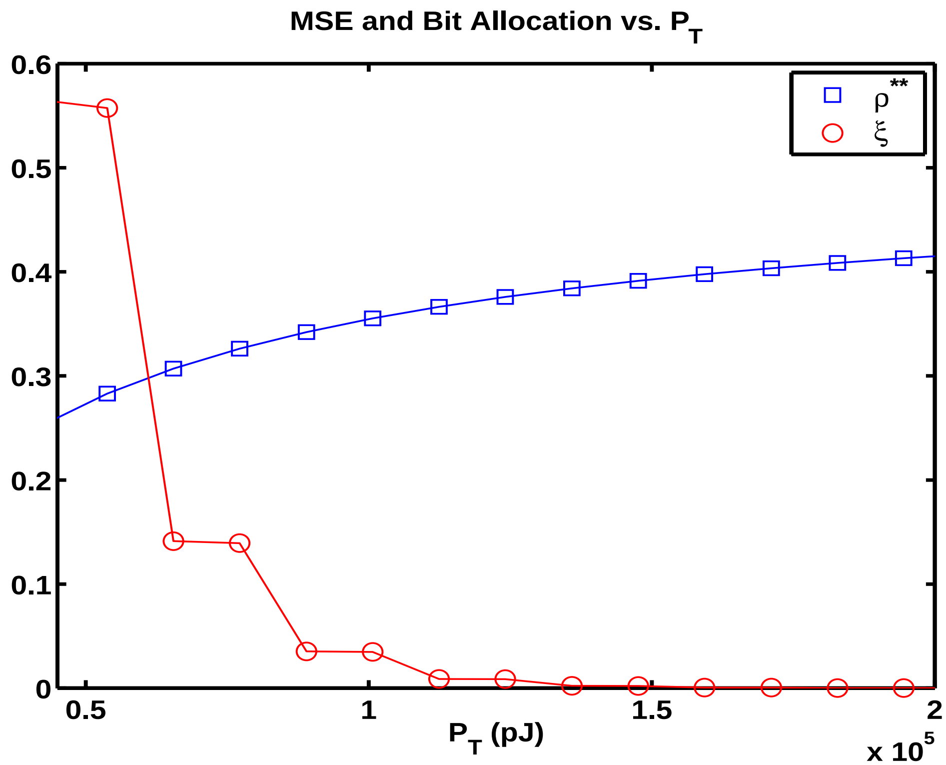


Figure 10: *Optimal data bit allocation factor under P_T constraint and MSE as a function of P_T.*

Performance under different bit allocations

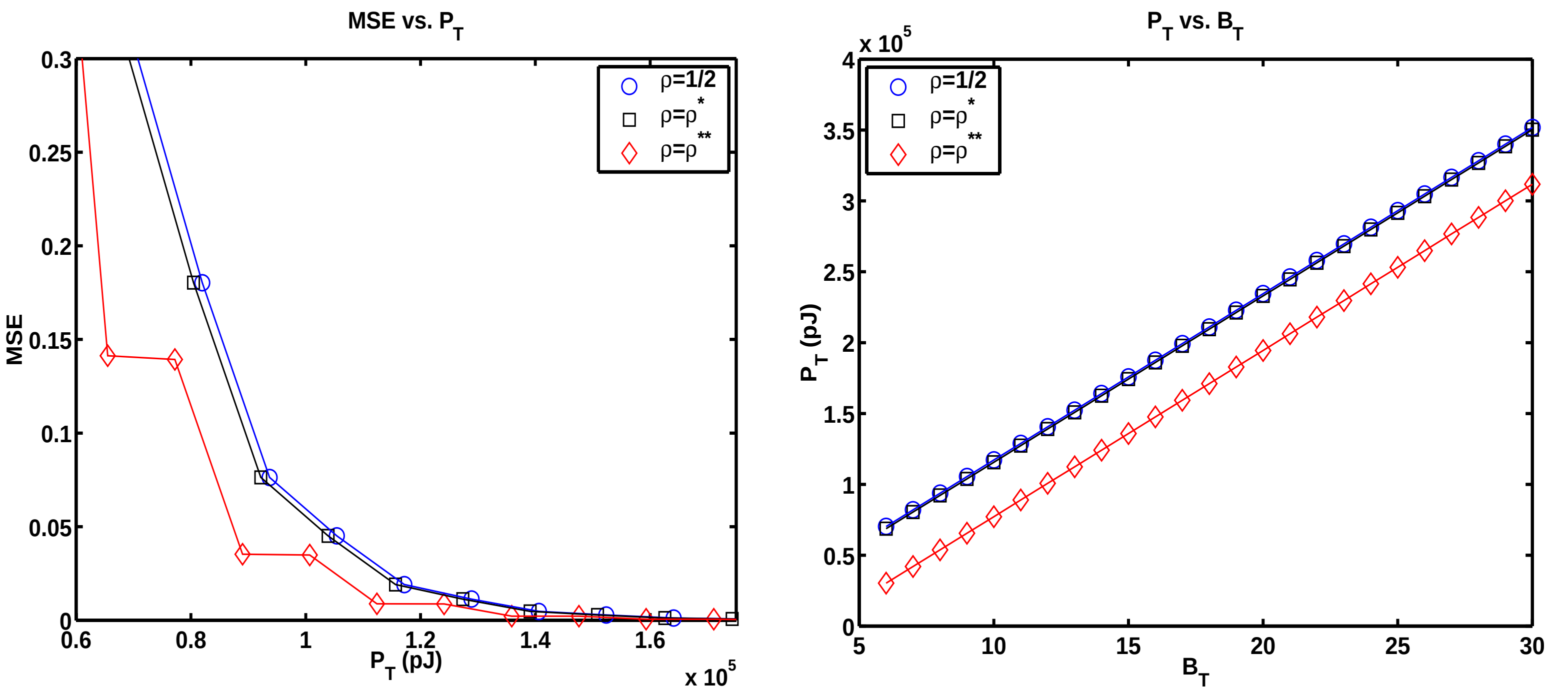


Figure 11: *MSE as a function of P_T for various bit allocation factors.*

Experimental Validation

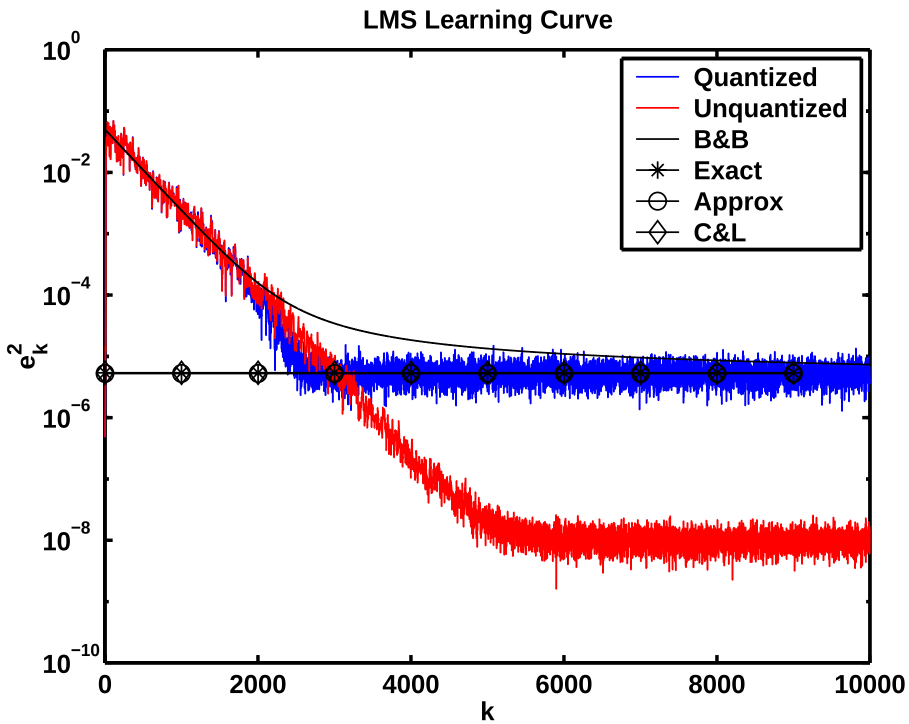
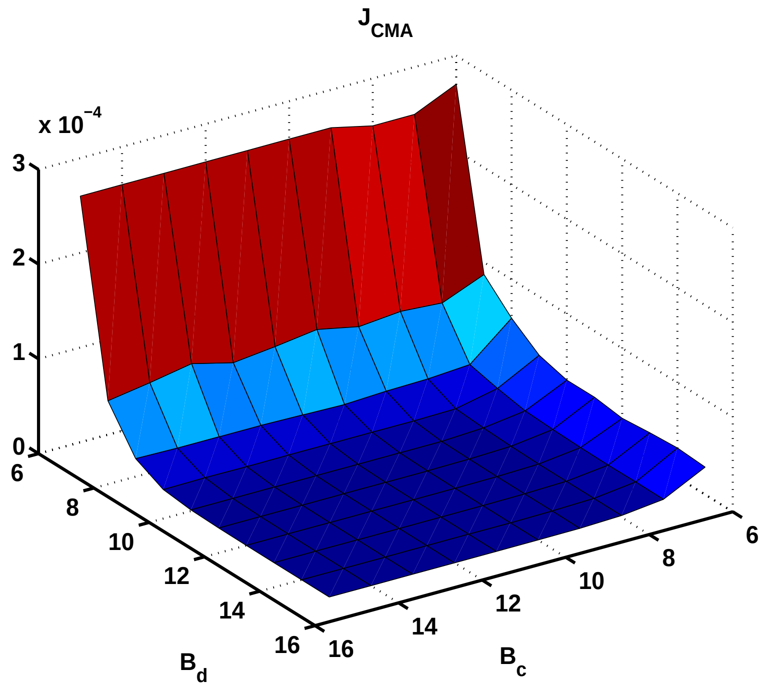


Figure 12: *Quantized LMS (channel identification) learning curve. Complex White Gaussian training sequence y_k with additive noise, Training sequence passed through 31-tap FIR channel. Parameters are: $\sigma_y^2 = 0.1$, $N_0 = 10^{-8}$, $B_c = B_d = 12$, $\mu = 1/32$, $p = 31$, $\xi_{min} = 10^{-8}$.*

Experimental Results for Blind Equalization (CMA)



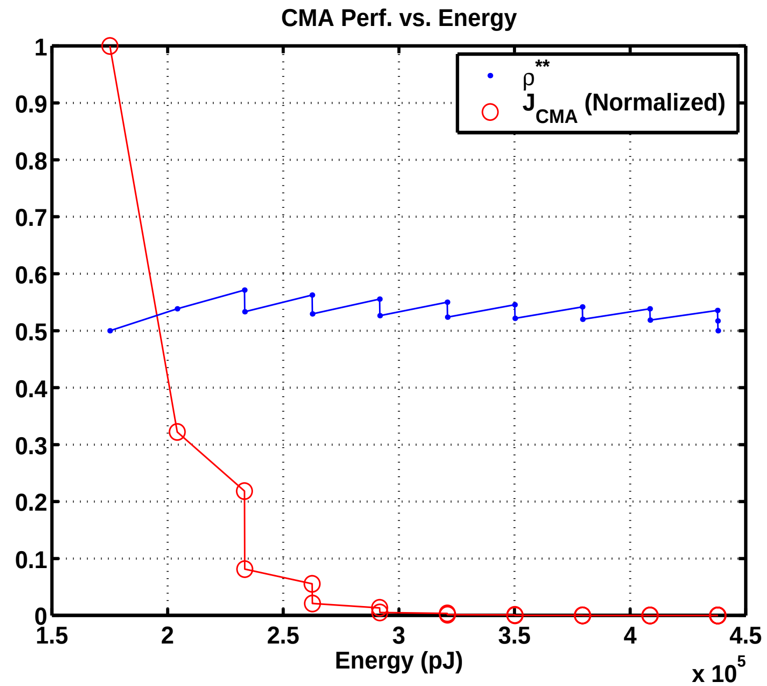


Figure 13: *CMA Objective function, J_{CMA}, vs B_d, B_c and power-optimal bit allocation*

Main conclusions for reduced resolution strategies

- Significant B_T and P_T reductions are possible for many DSP applications
- Analysis yields MSE-optimal LMS bit allocation strategies for fixed $B_T = B_d + B_c$ and P_T.
  - $B_c = B_d$ is MSE-optimal for high power
  - $B_c > B_d$ is MSE-optimal for low power
- For FIR matched filter $B_c = B_d$ is nearly optimal for B_T and P_T
- simulations have borne out theoretical results for medium to high B_T regimes

Future Work on Low-Power Adaptive Filtering

- Increase accuracy of MSE approximations
  - Nonlinear quantizer models
  - Non-white noise models
- Extend analysis to filters with different resolutions for each coefficient:

$$Q_c(\underline{w}_k) = \left[Q_c^0(w_k^0), Q_c^1(w_k^1), \ldots, Q_c^{p-1}(w_k^{p-1})\right]$$

- Extend to Probability of Error determination for typical communications settings
- Extend to Blind Equalization (CMA)

II. Partial Update LMS and Power

Partial Update LMS: only p_o of p coefficients updated/iteration

Advantages:

- Computational savings
- Memory savings
- Power savings

$$P_T^{PU-LMS} = P_T^{LMS}\,\frac{p_o}{p} + \epsilon$$

Requirement: condition on gain μ to guarantee convergence

Sequential Partial Update LMS Algorithm

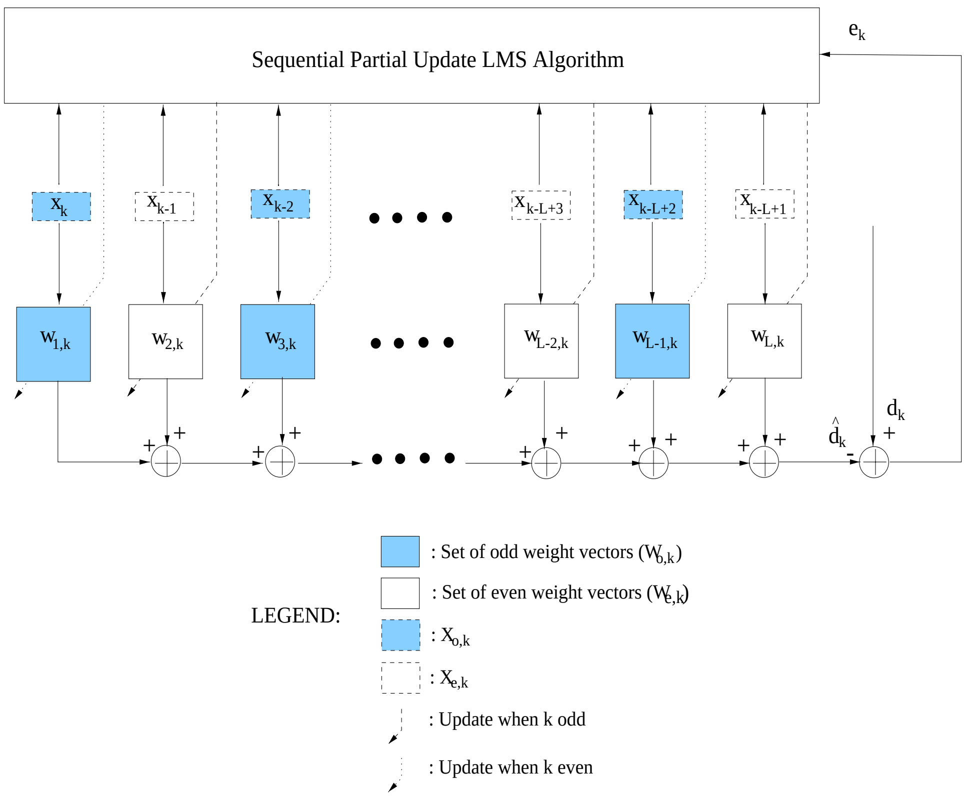


Figure 14: *Block diagram of the Sequential Partial Update LMS algorithm*

Comparison of Weight Trajectories

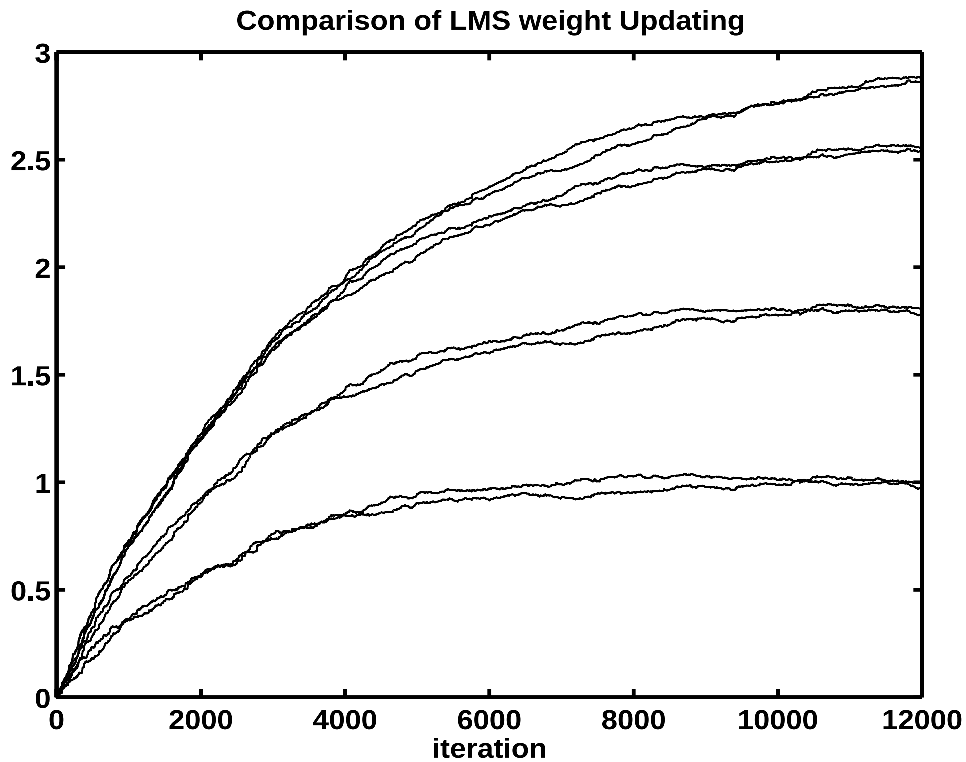


Figure 15: *Weight Update Trajectories for* $\mu = 0.2$ *and* $\mu = 0.4$

Description: Partial Update LMS algorithm

- Update Equations:

$$\begin{bmatrix} W_{e,k+1} \\ W_{o,k+1} \end{bmatrix} = \begin{cases} \begin{bmatrix} W_{e,k} \\ W_{o,k} \end{bmatrix} + \begin{bmatrix} \mu e_k^* X_{e,k} \\ 0 \end{bmatrix} & \text{for odd } k \\[2ex] \begin{bmatrix} W_{e,k} \\ W_{o,k} \end{bmatrix} + \begin{bmatrix} 0 \\ \mu e_k^* X_{o,k} \end{bmatrix} & \text{for even } k \end{cases}$$

- Update Equations for expected weight error vector

$$E\begin{bmatrix} V_{e,k+1} \\ V_{o,k+1} \end{bmatrix} = \begin{cases} \begin{bmatrix} I - \mu R_{k,e} & -\mu R_{k,eo} \\ 0 & I \end{bmatrix} E\begin{bmatrix} V_{e,k} \\ V_{o,k} \end{bmatrix} & \text{for odd } k \\[2ex] \begin{bmatrix} I & 0 \\ -\mu R_{k,oe} & I - \mu R_{k,o} \end{bmatrix} E\begin{bmatrix} V_{e,k} \\ V_{o,k} \end{bmatrix} & \text{for even } k \end{cases}$$

Example

- 2-tap adaptive filter
- Model

$$\begin{aligned} d_k &= W^*_{1,opt} s_k + W^*_{2,opt} s_{k-1} + n_k \\ x_k &= s_k + v_k \end{aligned}$$

where $W_{1,opt} = 0.5$, $W_{2,opt} = 0.4$, n_k is white Gaussian with variance, 0.01 and v_k is white Gaussian with variance, 0.01.

- $\{s_k\}$: cyclo-stationary with period 2 having Autocorrelation matrices

$$\begin{aligned} R_1 &= \begin{bmatrix} 5.1354 & -0.5733 - 0.6381i \\ -0.5733 + 0.6381i & 3.8022 \end{bmatrix} \\ R_2 &= \begin{bmatrix} 3.8022 & 1.3533 + 0.3280i \\ 1.3533 - 0.3280i & 5.1354 \end{bmatrix} \end{aligned}$$

Example contd.

- Regular LMS condition gives $\mu = 0.33$
- Sufficient condition derived here gives $\mu = 0.0254$
- Eigenvalues of the update equation for $\mu = 0.33$ have magnitudes 1.0481 and 0.4605

Large Step Size Weight Trajectories

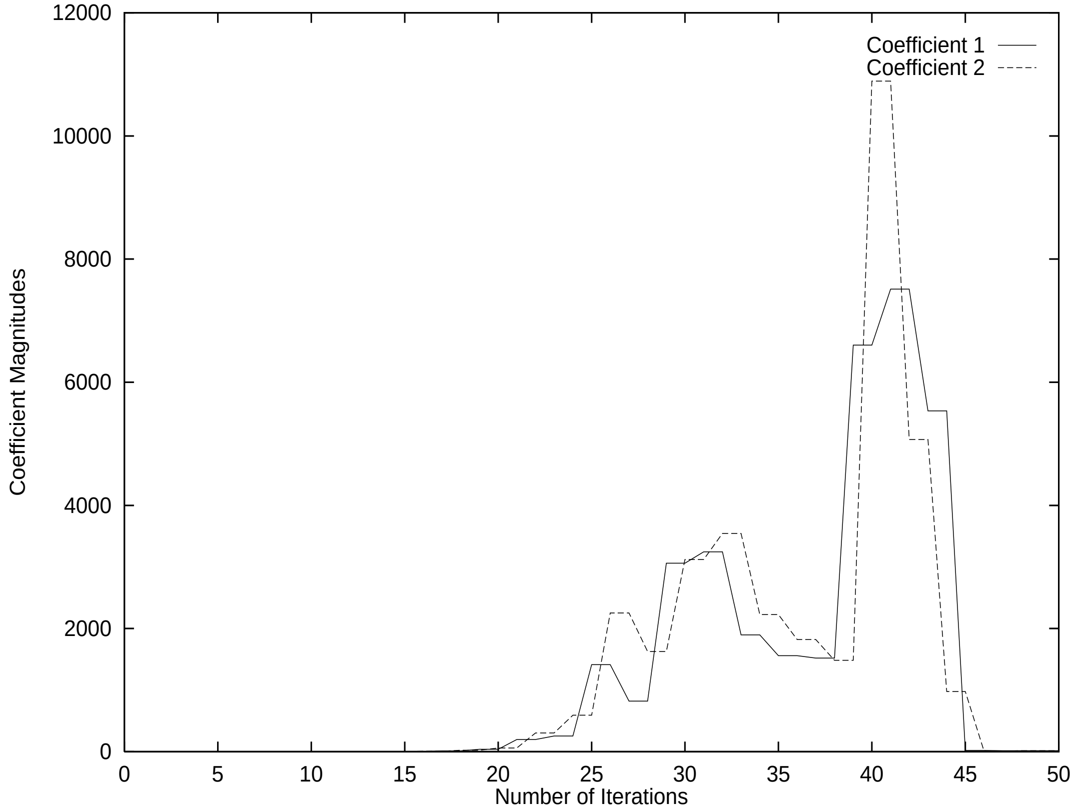


Figure 16: *Trajectory of* $w_{1,k}$ *and* $w_{2,k}$ *for* $\mu = 0.33$

Small Step Size Weight Trajectories

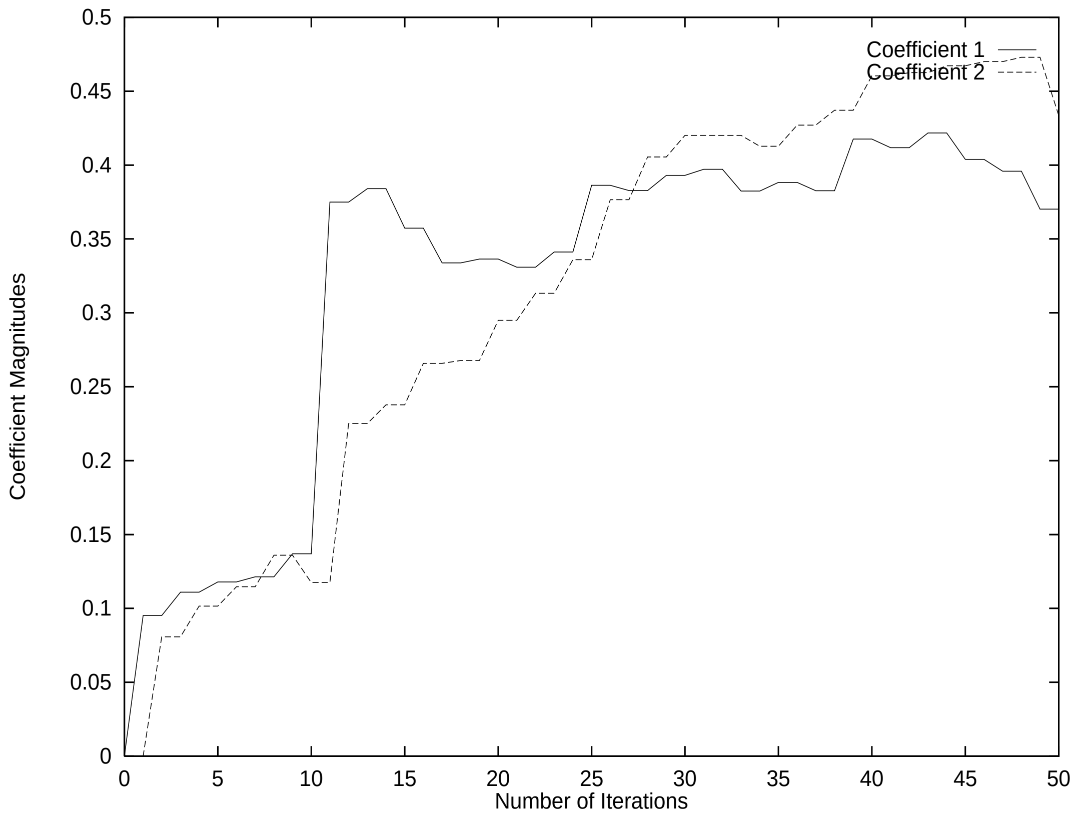


Figure 17: *Trajectory of* $w_{1,k}$ *and* $w_{2,k}$ *for* $\mu = 0.0254$

Conclusion and Future Work

- Conclusions:
  - Partial Update LMS algorithm can attain significant power savings w/o appreciable loss
  - Standard LMS condition for selecting μ doesn't guarantee convergence of the Partial Update LMS algorithm
  - Sufficient conditions for selecting μ ensuring convergence in mean were derived
- For future work:
  - Extension of current work to update of arbitrary subsets of filter weights
  - Derivation of theoretical results for mean square error convergence

Proximal Point Methods

Proximal Point Algorithm (PPA) for optimizing function $J(\theta)$: (Rockafellar:SIAM76)

$$\theta^{k+1} = \text{argmax}_\theta \{J(\theta) - \lambda_k \|\theta - \theta^k\|^2\}, \qquad k = 1, 2, \ldots$$

PPA with Kullback Penalty (Chretien&Hero:SIAM99)

$$\theta^{k+1} = \text{argmax}_\theta \left\{J(\theta) - \lambda_k K(\theta,\ \theta^k)\right\}$$

- $K(\theta,\ \theta^k) = \int g(y;\theta) \ln \frac{g(y;\theta)}{g(y;\theta^k)}\, dy$, $g(y;\theta) \geq 0$, $\int g(y;\theta) dy = 1$
- $\{\lambda_k\} > 0$ sequence of relaxation parameters

$$\lambda_k > 0, \qquad \text{and} \qquad \lambda_k \to 0$$

Advantages:

1. Superlinear convergence rates for smooth $J(\theta)$
2. Can be applied to non-differentiable J, e.g. l_1 CMA (Chretien&Hero:SIAM99).
3. Obtain EM-ML algorithm for:

   $$J(\theta) = \ln f(Y;\theta),\ K(\theta,\ \theta^k) = E[\ln f(X;\theta)|Y;\theta^k) - \ln f(Y;\theta), \lambda^k = 1$$
4. Obtain new class of accelerated EM algorithms for $\lambda^k \neq 1$ (Chretien&Hero:ISIT98).
5. Successive iterates $\{\theta^k\}$ produce increasing $\{J(\theta^k)\}$.
6. Under local quadratic approximation to $\ln f(Y;\theta)$ Kullback-PPA becomes hybrid EM/Newton algorithm
7. Kullback-PPA generalizes to coordinatewise optimization: hybrid SAGE/Newton

Example: Maximum likelihood sequence estimation

$$y_k = \sum_{i=1}^{k} a_{k-i}\theta_i + n_k \quad k = 1, \ldots, n$$

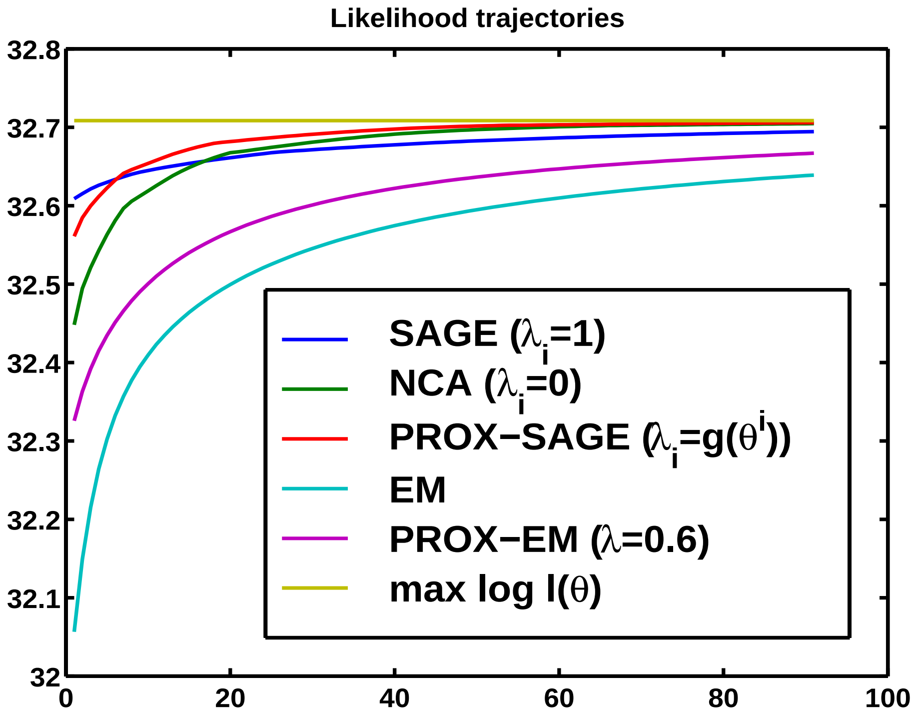


Figure 18: *Likelihood trajectory comparisons for ML sequence estimation*

Publications citing ARO-MURI

1. **“Power-performance tradeoffs and optimal bit allocation in reduced resolution adaptive filtering,” R. Gupta and A. O. Hero, submitted to IEEE Trans. on Sig. Proc. Dec. 1999.**
2. **“Generalized proximal point algorithms and bundle implementations,” S. Chretien and A. Hero, accepted with revisions, SIAM Journ. on Optimization, May 1999.**
3. **“Maximum Likelihood Digital Receiver Using Coordinate Ascent and the Discrete Wavelet Transform,” I. Sharfer and A. O. Hero, IEEE Transactions on Signal Processing, Vol. 47, No. 3, pp. 813-825, Mar. 1999.**
4. **“Stability bounds on the step size for the partial update LMS algorithm,” M. Godavarti and A. O. Hero, Proc. of 1998 Int. Conf. on Acoust., Speech, and Sig. Proc., Phoenix, March 1999.**
5. **“Theoretical analysis of power-performance tradeoffs in reduced resolution adaptive filtering,” R. Gupta and A. O. Hero, submitted to Proc. of 1998 Int. Conf. on Acoust., Speech, and Sig. Proc., Phoenix, March 1999.**
6. “Turbo codes for fading and burst channels,” J. H. Kang, W.E. Stark and A.O. Hero, Proceedings GLOBECOM, Sydney, Australia Oct. 1998.
7. “Power vs. Performance Tradeoffs for Reduced Resolution Adaptive Equalizers,” A.O. Hero and R. Gupta, IEEE Conf. on Military Communications (MILCOM), Oct. 1998.
8. ”Optimal Bit Allocation for the Quantized LMS Adaptive Algorithm,” R. Gupta and A. O. Hero, Proc. of the 1998 IEEE Workshop on Statistical Signal and Array Processing , Sept. 1998.
9. ”Data-recursive algorithms for blind channel identification in oversampled communication systems”, D. Goeckel, A. O. Hero, and W. E. Stark, IEEE Transactions on Signal Processing, Vol. 46, No. 8, pp. 2217-2220, Aug. 1998.
10. ”Modulation discrimination in digital communications using higher order moments”,” A. O. Hero and H. Hadinejad-Marham, Proc. of 1998 Int. Conf. on Acoust., Speech, and Sig. Proc., Seattle, May 1998.

Technology Transfer Activities

1. **ISIS GDR on Telecommunications, Telecom Paris**, "Bit allocation strategies for reduced-power mobile communications," June 1999. Contacts: Eric Moulines, Pierre Duhamel, Sylvie Mayrargue.
2. **Mathematics of Communications Research, Bell Laboratories, Lucent Technologies, Murray Hill**, "Power reduction design strategies for lower power adaptive filtering," May 1999. Contacts: Jim Mazo, Rajiv Laroia, Vivek Goyal, Tom Marzetta.
3. **Hughes Network Systems, Germantown, Maryland**, "Sucessive updating and bit allocation design for adaptive equalization," May 1999. Contacts: Basel Beidas, Richard Klumer, Neal Becker,Bill Kirchner, Roger Hammons,
4. **21st Biennial Communications Symposium, Kingston ON**, "EM algorithm and beyond," (Plenary presentation), June 1998). Contacts: Steve Blostein, Jerry Hayes.