

Neural Engineering 2007

A High-Yield Area-Power Efficient DWT Hardware for Implantable Neural Interface Applications

Awais M. Kamboh, Andrew Mason, Karim Oweiss
{Kambohaw, Mason, Koweiss} @msu.edu

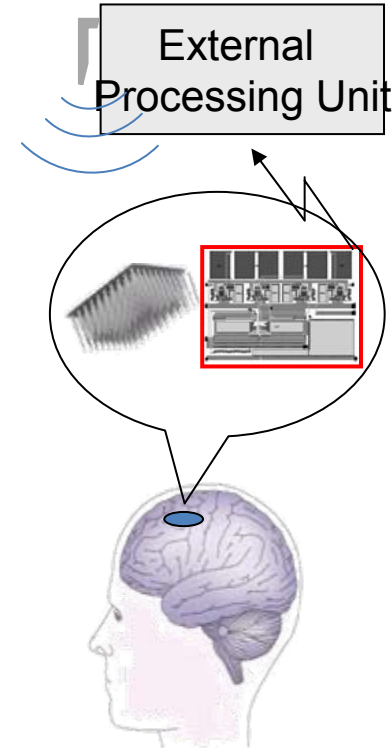
Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI, USA



May 4th, 2007

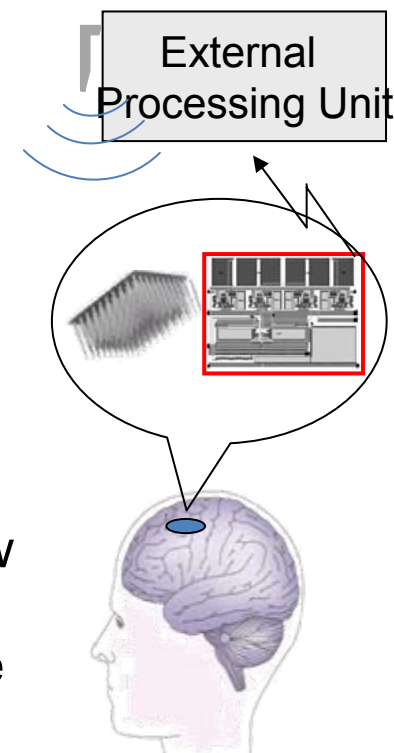
Motivation

- High density - hundreds of electrodes
- Wireless transmission requires large bandwidth - for example
 - a typical recording experiment
 - 100-electrode array,
 - sampled at 25 kHz per channel,
 - with 12-bit precision
 - yields an aggregate data rate of 30 Mbps
 - well beyond the reach of state-of-the-art telemetry links for biological applications.



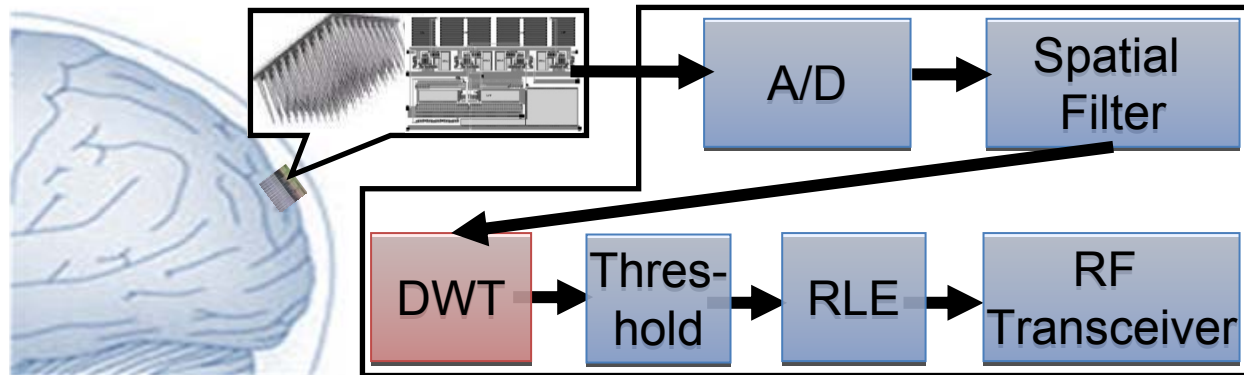
Problem Definition

- Challenge
 - Microelectrodes generate large amount of data
 - Wireless transmission - requires prohibitively large bandwidth
- Resource constraints
 - Area - physical implantation challenges
 - need to fit implanted circuitry within $\sim 1 \text{ cm}^2$ for the entire signal processing system.
 - Power
 - limited/transmitted power source
 - possible damage to neighboring tissues - very low power (no more than 8-10 mW) to prevent temperature rise above 1°C and avoid neural tissue damage [K. Oweiss, 2006].



Prior Work

- Solution – *in vivo* Data Compression

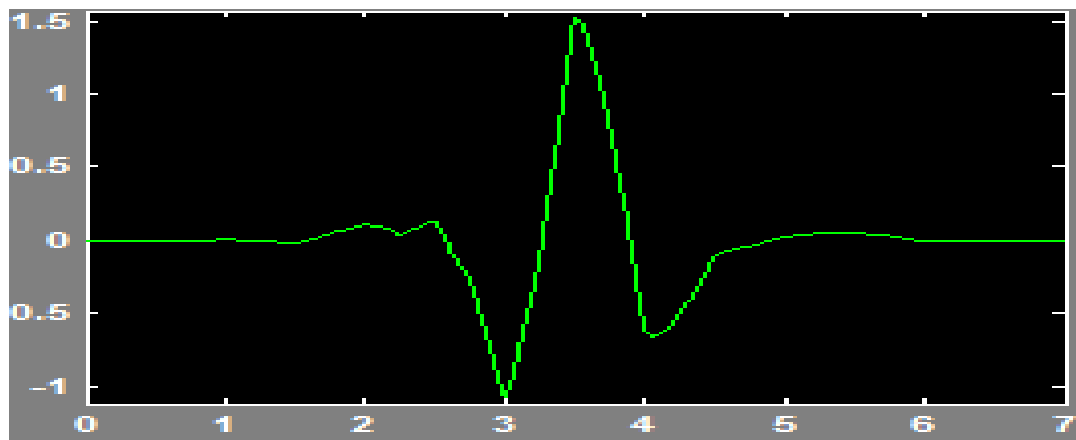


- Proposed *in vivo* data compression system

- Multi-channel electrodes
- Analog to digital converter (A/D)
- Spatial filter
- Discrete wavelet transform (DWT)
- Threshold
- Run-length encoder (RLE)

Prior Work (contd.)

- DWT - Time frequency domain processing
 - Application demands
 - multi-channel electrodes → multi-channel processing
 - multiple levels of decomposition → better compression



$$H(z) = -0.076 - 0.030z^{-1} + 0.498z^{-2} + 0.804z^{-3} + 0.298z^{-4} - 0.099z^{-5} - 0.013z^{-6} + 0.032z^{-7}$$
$$L(z) = -0.032 - 0.013z^{-1} + 0.099z^{-2} + 0.298z^{-3} - 0.804z^{-4} + 0.498z^{-5} + 0.030z^{-6} - 0.076z^{-7}$$

Prior Work (contd.)

- DWT - Time frequency domain processing
 - Application demands
 - multi-channel electrodes → multi-channel processing
 - multiple levels of decomposition → better compression
 - Implement with Symmlet 4 basis
 - Especially suited; maintains signal fidelity for neural signals
 - Two, eight-tap filters

$$H(z) = -0.076 - 0.030z^{-1} + 0.498z^{-2} + 0.804z^{-3} + 0.298z^{-4} - 0.099z^{-5} - 0.013z^{-6} + 0.032z^{-7}$$

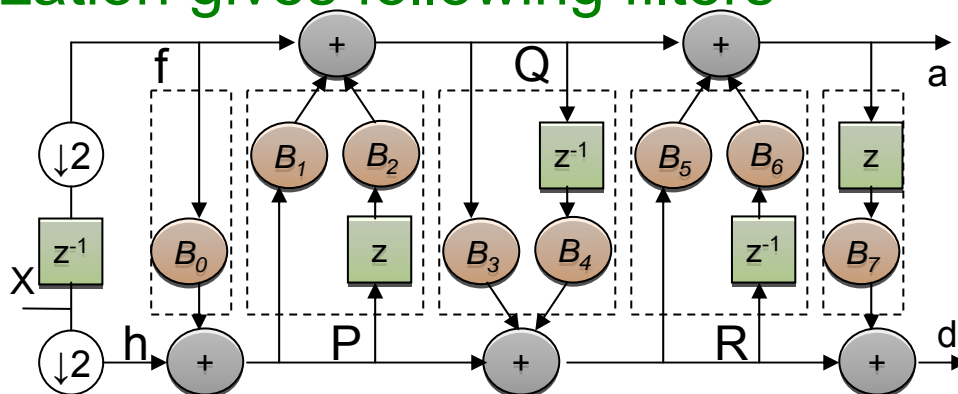
$$L(z) = -0.032 - 0.013z^{-1} + 0.099z^{-2} + 0.298z^{-3} - 0.804z^{-4} + 0.498z^{-5} + 0.030z^{-6} - 0.076z^{-7}$$

- Convolution-based filtering
 - Computationally expensive
 - Large memory requirement
- Lifting implementation

Theory

- Lifting factorization
 - Reduces number of calculations
 - Factorizes the pair of filters into smaller filters
 - ‘symmlet 4’ factorization gives following filters

$$\begin{aligned}
 P_{-1} &= h + B_0 f \\
 Q_{-1} &= f + B_1 P_{-1} + B_2 P_0 \\
 R_0 &= P_0 + B_3 Q_0 + B_4 Q_{-1} \\
 a_0 &= Q_0 + B_5 R_0 + B_6 R_{-1} \\
 d_{-1} &= R_{-1} + B_7 a_0
 \end{aligned}$$



Coeff.	B₀	B₁	B₂	B₃
Value	0.391	-0.124	-0.339	-1.420
Coeff.	B₄	B₅	B₆	B₇
Value	0.162	0.431	0.146	-1.049

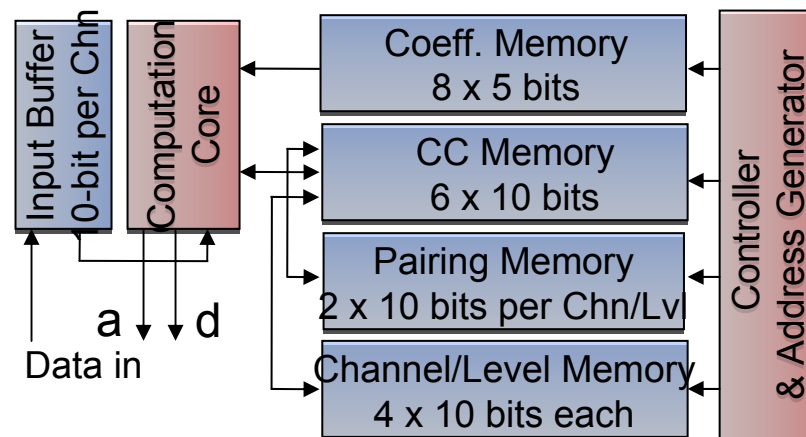
- Subscripts represent time-samples
- ‘a’ and ‘d’ are approximation and detail coefficients

Sequential Reuse of Hardware

- Sequential reuse of hardware
 - We use long computation intervals ($\sim 40\mu\text{s}$) to our benefit which allows us to reuse the same hardware
 - Sequential reuse provides area efficiency
 - Results in increased design complexity

- Architecture Blocks

- Computation core
- DWT controller
- Computation core memory
- Coefficient memory
- Channel & Level memory
- Pairing memory
- Input buffer



Computation Core

- One general equation
 - Two custom designed adders and two multipliers for area-power efficiency

$$P_{-1} = h + B_0 f$$

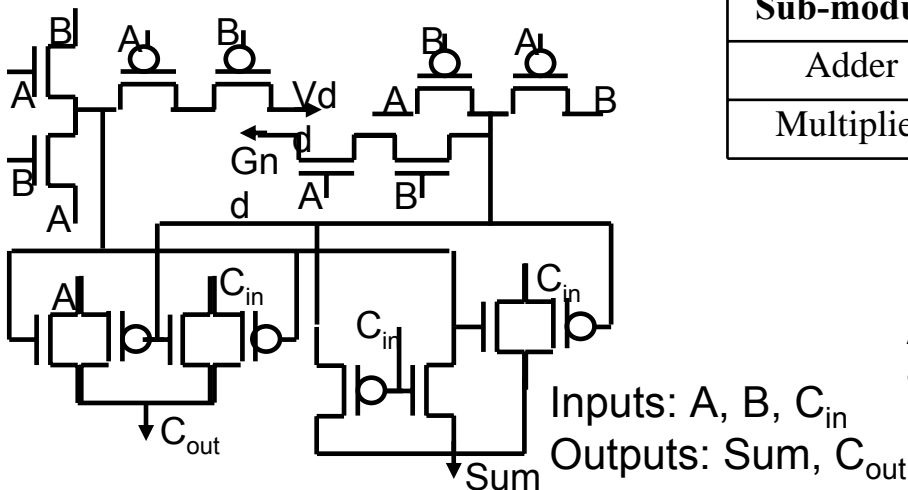
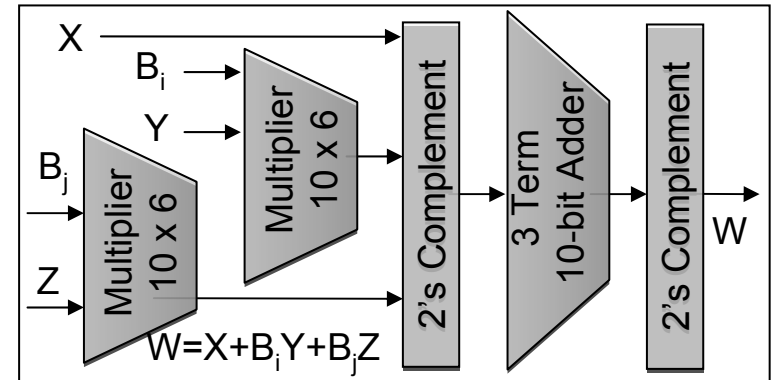
$$Q_{-1} = f + B_1 P_{-1} + B_2 P_0$$

$$R_0 = P_0 + B_3 Q_0 + B_4 Q_{-1}$$

$$a_0 = Q_0 + B_5 R_0 + B_6 R_{-1}$$

$$d_{-1} = R_{-1} + B_7 a_0$$

$$W = X + B_i Y + B_j Z$$

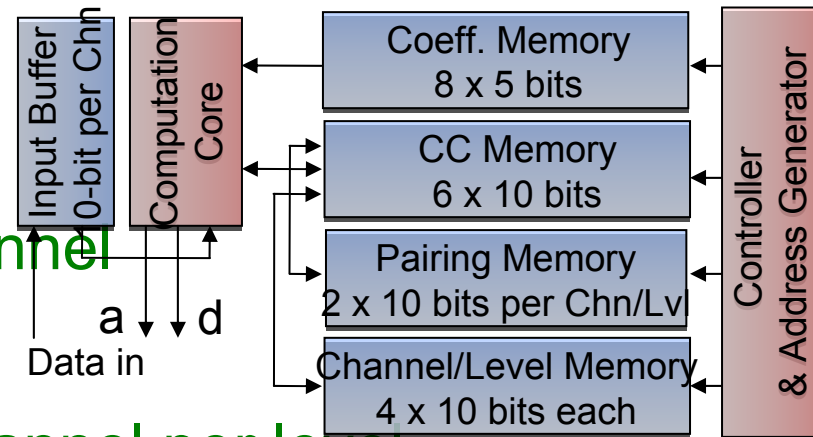


Sub-module	# of TX	Area(μm ²)	Power(μW)	Delay (ns)
Adder	160	418	3.834	0.404
Multiplier	733	2858	42.76	4.12

A. Shams, T. K. Darwish, and M. A. Bayoumi, "Performance analysis of low-power 1-bit CMOS full adder cells," *IEEE Trans. VLSI systems*, vol. 10, no. 1, pp. 20-19, Feb 2002.

Memory Modules

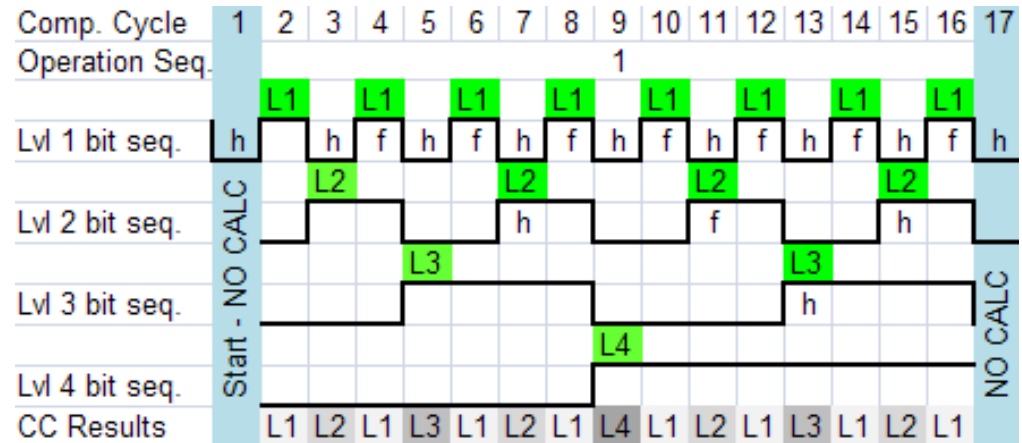
- Computational Core memory
 - 10 bit Flip Flop, 6 values
- Coefficient Memory
 - 6 bit ROM, 8 values
- Input Buffer
 - 10 bit SRAM, 1 value per channel
- Pairing Memory
 - 10 bit SRAM, 2 values per channel per level
- Channel/Level Memory
 - 10 bit SRAM, 4 values per channel per level



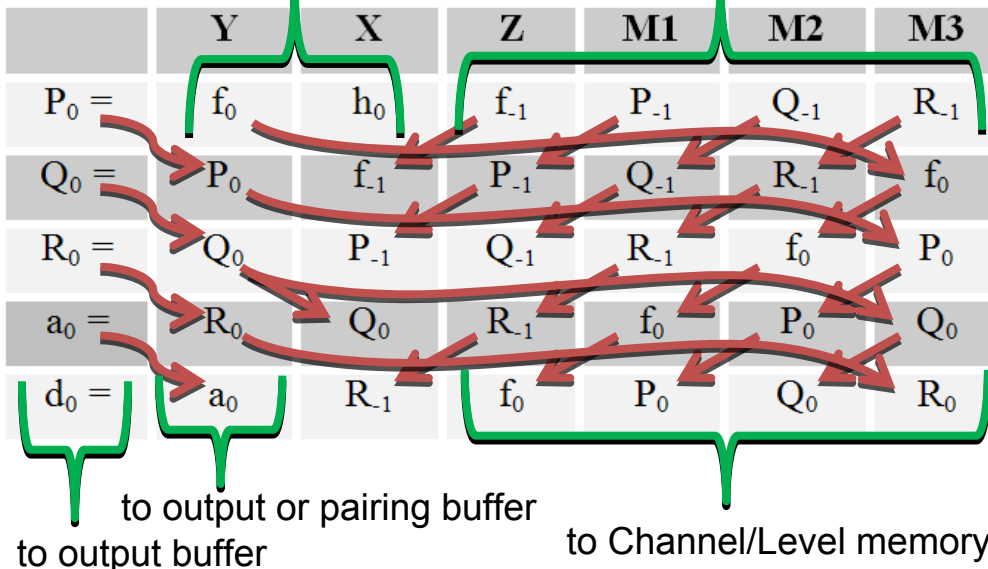
Controller Design

- Right

- CC use for calculation of interleaved results at higher levels of DWT



from input or pairing buffer from Channel/Level memory



- Left

- Complex data movement in the CC memory registers and resulting accesses to other memory modules

Analysis & Results

- Transistor counts and area of hardware modules
- Scalable for different channels and levels
- 0.18 μm process

Module	No. of Tx	Area (μm^2)
Controller	987	3834
Comp Core	1854	13828
Comp Core Memory	2800	1775
Coefficient Memory	122	1114
Pairing Memory per channel/level	$120 * C * (L - 1)$	Varies
Input Buffer per channel	$60 * C$	Varies
Per Channel or Level Memory	$240 * C * L$	Varies
Complete DWT system: 32 Chn, 4 Lvl	52078	163377

*Counts exclude transistors for address decoding circuitry

•C = Number of Channels

•L = Number of Levels

•Transistor count and area for hardware modules

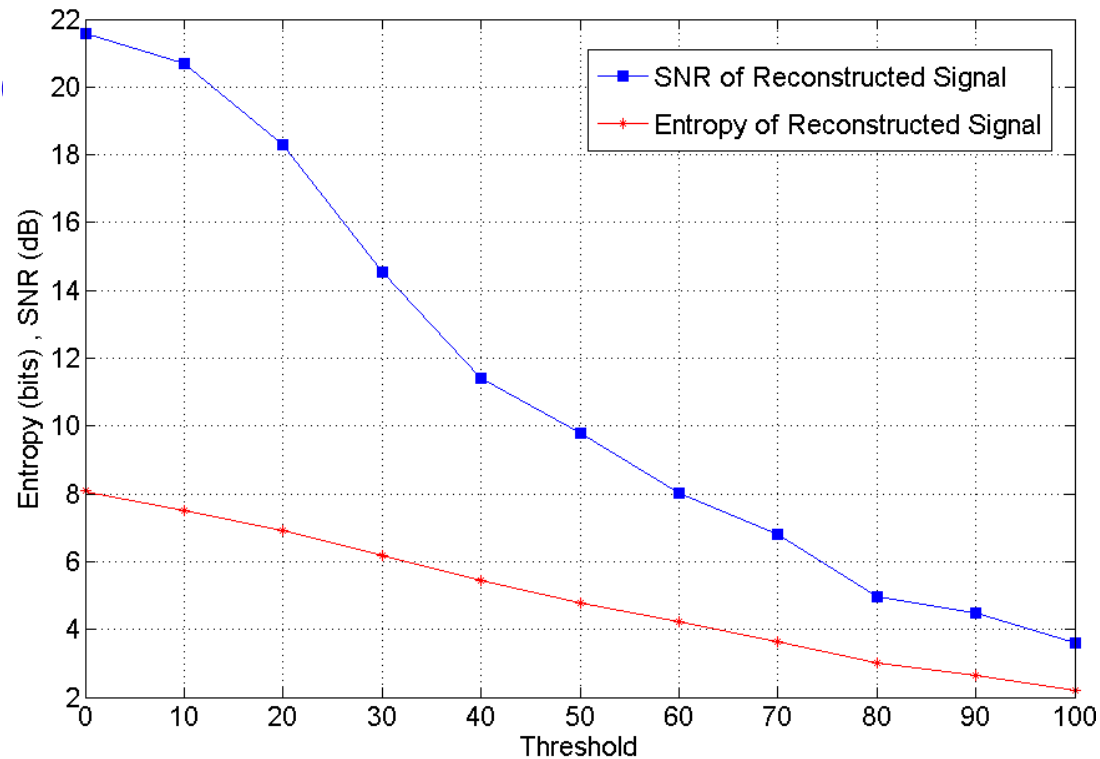
Analysis & Results (contd.)

- Threshold
- Signal to Noise Ratio
 - Units: dB

$$SNR = 10 \log_{10} \frac{\text{signal_energy}}{\text{noise_energy}}$$

- Entropy
 - Units: Bits/symbol

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2 p(x_i)$$



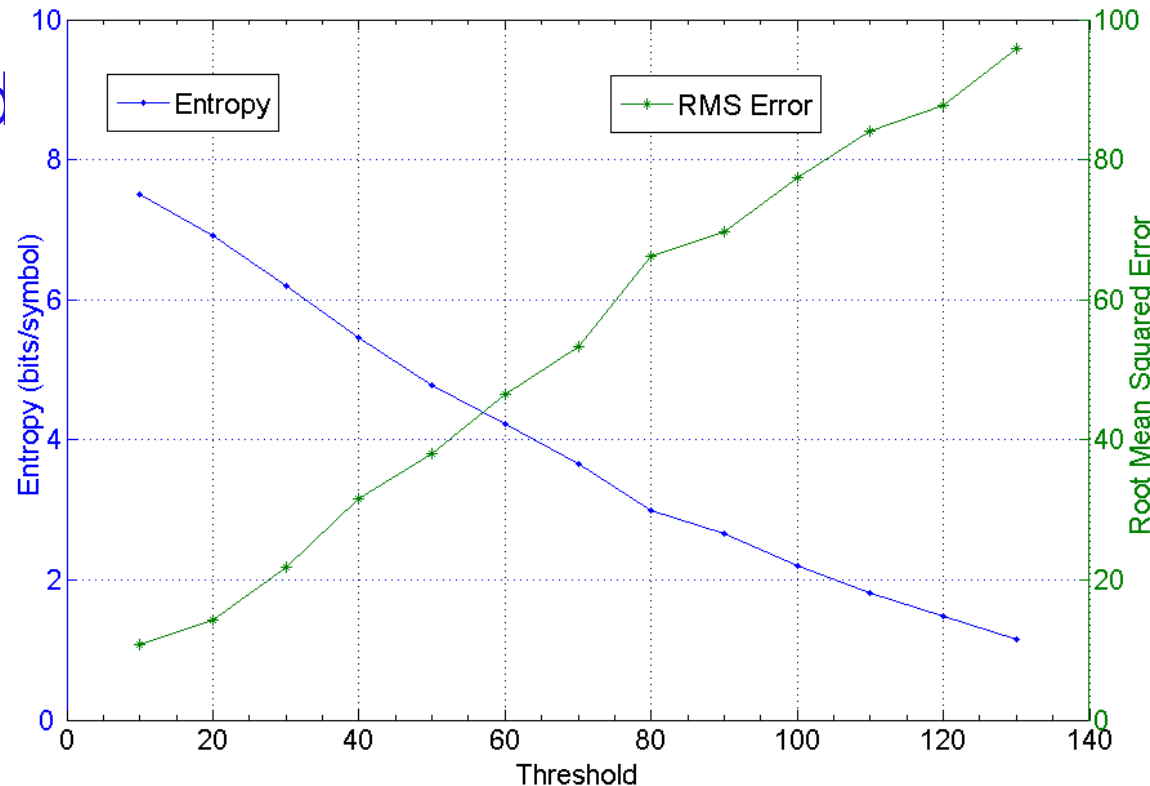
Analysis & Results (contd.)

- Threshold
- Root Mean Squared Error

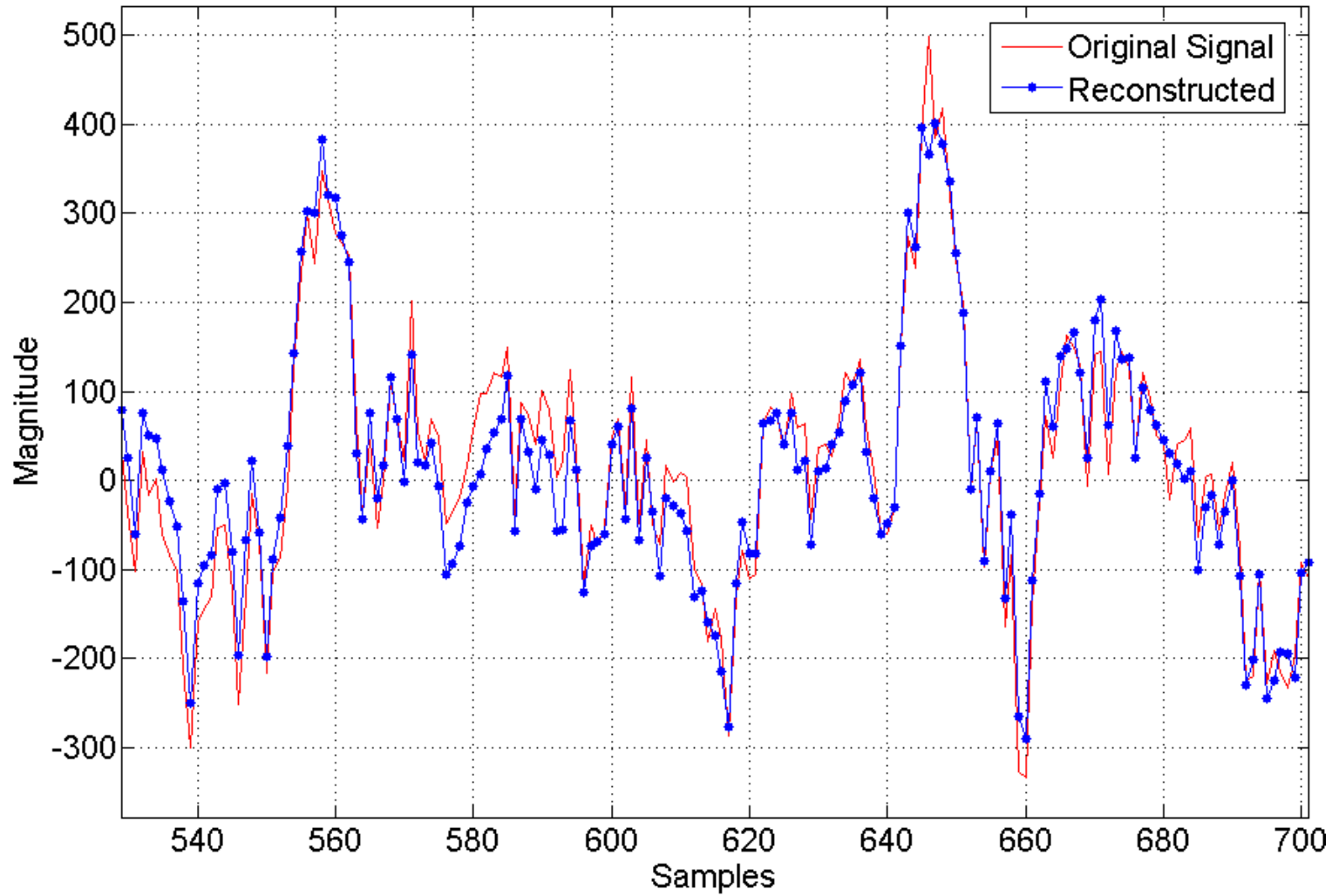
$$RMS_error = \sqrt{\frac{1}{N} \sum_n^N (x_n - \hat{x}_n)^2}$$

- Entropy
 - Units: Bits/symbol

$$H(X) = -\sum_{i=1}^N p(x_i) \log_2 p(x_i)$$

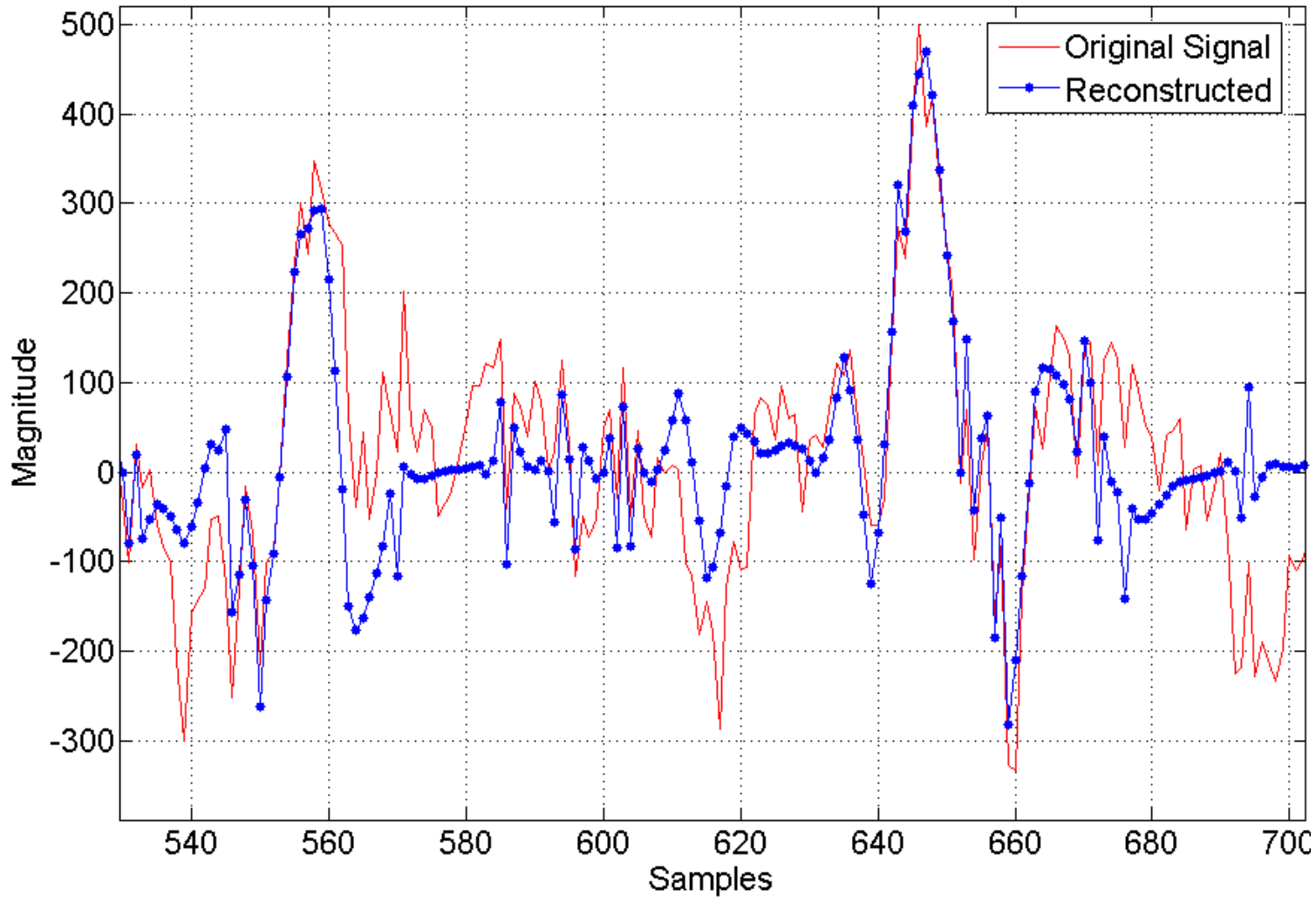


Analysis & Results (contd.)



- Threshold = 50, RMS Error = 37.95, Entropy = 4.78 bits/sym
- Sampling frequency 25kHz, Data width = 10 bits, Coefficient width = 6¹⁵ bits

Analysis & Results (contd.)



- Threshold = 120, RMS Error = 87.69, Entropy = 1.48 bits/sym.
- Sampling frequency 25kHz, Data width = 10 bits, Coefficient width = 6¹⁶ bits

Conclusions

- System designed to pseudo-simultaneously perform multi level DWT for multiple channels of incoming neural signals.
- Enables very high data compression rates while maintaining high signal fidelity.
- Implementation in 0.18 μm CMOS requires only 0.16mm² of area to process 32 channels of data at 4 levels of DWT in real time.
- Highly suitable for implantable high-density microelectrode array devices because of the small size and low power consumption of the system make it

Acknowledgement

- NIH grant # NS047516-01A2
- National University of Sciences and Technology, Islamabad, Pakistan.
- Advanced Microsystems and Circuits Research Group at Michigan State University.

Questions & Comments



Notes

- We used the 'symmlet4' wavelet basis, which was factorized into lifting steps, based on previous analysis [4]. Since this factorization is not unique, we selected the one with the scaling factors K and $K-1$ closest to one, so that the details at various levels have the same natural dynamic range. This resulted in 5 lifting steps and one scaling step, which was discarded in the integer case. The scaling factors were 1.57 and 0.63 and had to be included for computing the denoising threshold at each level of decomposition. For this orthogonal wavelet ($h = 8; g = 8$), the number of flops for the standard algorithm is 34, while the lifting algorithm only requires 20 flops
- The *Symlet* wavelet basis family was found to be the most effective in terms of energy compactness in the fewest number of coefficients, while enjoying a simple hardware implementation.