ENERGY-EFFICIENT WIRELESS SENSORS: FEWER BITS, MOORE MEMS

FRED CHEN,
ANANTHA CHANDRAKASAN, VLADIMIR STOJANOVIC
Department of EECS, Massachusetts Institute of Technology, Cambridge, MA
August 29th, 2011
Technical Challenges: Node reliability and energy → cost

Global Sensory Network

Structural conditions (buildings, bridges, etc.)
Active nodes:
- Expensive to communicate: nJ/bit vs. pJ/bit
- Trade communication for computation → reduce the data

Idle nodes:
- Leakage: every circuit consumes a little... but for a long time
- Ideally we don’t consume anything → eliminate leakage

[Diagram of sensor field with active and idle sensor nodes, base station, and subsystems: Sensing Subsystem, ADC, Sensor Processing Subsystem, Energy Subsystem, Communication Subsystem, Digital Baseband, RAM, DSP, CPU, DC-DC Converter, MICS Radio, UWB Radio, state-of-the-art academic papers graph showing energy performance across different components: UWB Radio, MICS Radio, μP/DSP, ADC, and Amp]
**OPTIMIZE EFFICIENCY ACROSS DESIGN LAYERS**

**> 10X DATA COMPRESSION**

\[
\begin{bmatrix}
Y \\
\vdots \\
Y_M
\end{bmatrix} = \begin{bmatrix}
\Phi \\
\vdots \\
\Phi
\end{bmatrix} \begin{bmatrix}
F \\
\vdots \\
F_N
\end{bmatrix}
\]

Subject to \( y = \Phi \Psi \)

**Signal Recovery**

- Find sparse solution
- Reconstruct signal

**> 10X ENERGY EFFICIENCY**

- Analog
  - Digital
  - Target

**2500X MORE EFFICIENT W/ DIGITAL**

**ARCHITECTURE & CIRCUITS**

Sensor Node

- AFE
- CS Encoder

Base Station

- Mixer
- Integrator

Sensor

**ALGORITHMS**

- \( f[n] \)
- \( Q_{CS} \)

\( \Phi_{0}[n] \)

\( f/N \)

\( f_s \)

\( f_s \)

\( \Phi \)

\( Q_{M} \)

\( Q_{S} \)

\( Q_{CS} \)

\( Q_{CS} \)

\( f \)

\( \Psi \)

\( \Phi \)

\( \Psi \)

**DEVICES**

- CMOS (100F)
- CMOS (25F)
- Relay (Au)
- Relay (W, 25F)
- Relay (W, 100F)
- Relay (W, 100F, buff)

**CS Area:** 200\( \mu \)m \( \times \) 450\( \mu \)m

**ADC**

- 8-bit address

**MIT Thesis Defense**
Minimizing Active Sensor Power

Many sensor signals have a low information rate compared to their data rate

How best to leverage this characteristic?

Signal Type | Sampling Rate | Frequency of Events
--- | --- | ---
Extracellular APs | 30 kHz | 10 – 150 /s
EMG | 15 kHz | 0 – 10 /s
EKG | 250 Hz | 0 – 4 /s
EEG, LFP | 100 Hz | 0 – 1 /s
**DATA REDUCTION IN WIRELESS SENSORS**

- **What we want:**
  - **Compression:** minimize data, but retain the *information*
  - **Low cost:** must be less than transmission savings - leakage
  - **Generality:** we don’t want to customize every sensor design
Compressed sensing (CS)

- Similar compression performance as source encoders
- Simpler hardware
- Generically applicable to sparse signals
**Signal Sparsity**

- CS relies on the signal of interest, $f$, being **sparse** in *some* basis, $\Psi$.

\[ i.e. \ f = \Psi x = \sum_{i=1}^{L} x_i \varphi_i , \text{where most } x_i = 0 \]
**Signal Sparsity**

- **CS relies on the signal of interest,** $f$, **being sparse in some basis,** $\Psi$.

- **Sinusoid example:**

- Many sensor signals of interest are sparse

  E.g. EEG, EKG, etc...

\[ f = \Psi x = \sum_{i=1}^{L} x_i \phi_i , \text{where most } x_i = 0 \]
CS Framework for Wireless Sensors

- **Tx/Encoder:** energy constrained
  - $\Phi$ linearly maps $N$ input samples ($f$) into $M$ measures ($y$)
  - $\Phi$ can be ‘random’ (e.g. $\pm 1$) to capture the necessary info
Rx/Decoder: not as energy constrained
- Assumes that $f$ is sparsely represented by $\Psi$ (few non-zero $x$).
- Finds sparse solution (e.g. minimum $l_1$-norm)
HOW TO DESIGN THE CS SYSTEM?

- **CS Parameters:**
  - **N** – size of compression block
  - **Q_{CS}** – resolution of the input
  - **M** – number of measurements
  - **B** – measurement resolution
**How to Design the CS System?**

- **CS Parameters:**
  - N – size of compression block
  - \( Q_{CS} \) – resolution of the input
  - M – number of measurements
  - B – measurement resolution

- **Objective:**
  - *Minimize bits over the channel

- **Constraint:**
  - We’d like \( f_{cs} \) as good as \( f_Q \)
  - How close (or far) are we?

\[
PRD = 100 \sqrt{\frac{\sum_{n=1}^{N} |\hat{f}[n] - f[n]|^2}{\sum_{n=1}^{N} |f[n]|^2}}
\]
- Reconstruction is probabilistic: always some probability of getting a “bad” reconstruction
- “Bad” = quantitatively large, perceptually small
**Average vs. Net Reconstruction Error**

- **$\text{PRD}_Q$**
  - Distortion due to just quantization

- **$\text{PRD}_{\text{net}}$**
  - Total error

- **$\text{PRD}_{\text{avg}}$**
  - Time-averaged error per block (average case)
Reconstruction Error – $M$ and $Q_{CS}$

- When $M$ sets net: trade-off between resolution (of $f_{CS}$) and compression

For $M=50$ and $Q=12$:

- PRD$_Q$: 1000 hits
- PRD$_{avg}$: 800 hits
- PRD$_{net}$: 600 hits

For $M=50$ and $Q=10$:

- PRD$_Q$: 800 hits
- PRD$_{avg}$: 600 hits
- PRD$_{net}$: 400 hits

For $M=50$ and $Q=8$:

- PRD$_Q$: 700 hits
- PRD$_{avg}$: 500 hits
- PRD$_{net}$: 300 hits

For $M=100$ and $Q=12$:

- PRD$_Q$: 1200 hits
- PRD$_{avg}$: 1000 hits
- PRD$_{net}$: 800 hits

For $M=100$ and $Q=10$:

- PRD$_Q$: 900 hits
- PRD$_{avg}$: 700 hits
- PRD$_{net}$: 500 hits

For $M=100$ and $Q=8$:

- PRD$_Q$: 800 hits
- PRD$_{avg}$: 600 hits
- PRD$_{net}$: 400 hits

PRD (dB) range:

- $-15$ to $15$
Quantization and Measurement Space

- Recovered signal error: a combination of reconstruction error due to $M$ and quantization error due to $Q_{CS}$

\[ n(Q_{CS}) \quad n(M) \]

\[ f \quad f_{CS} \]

\[ \text{PRD}_{\text{avg}} \]

\[ \text{PRD}_{\text{net}} \]

- Measurement limited
- Quantization limited

MIT Thesis Defense
How many bits needed to represent the signal with an error $\leq Q$-bit quantizer?

- Net case degrades for $Q > 8$ bits
- Average case only limited by $Q_{CS}$, tracks entropy
Huffman & LZW suffer from error propagation
What is the cost of transmission errors in CS?
CS Over a Wireless Channel

- Huffman & LZW suffer from error propagation
- What is the cost of transmission errors in CS?

\[ n_f = N_{in} \]

\[ f \rightarrow + \rightarrow \text{ADC} \rightarrow Q_{CS} \rightarrow \text{CS} \rightarrow B \rightarrow + \rightarrow \text{Reconstruct} \rightarrow f_{CS} \]

\[ n_{ch} = N_0 \]

\[ \text{PRD}_{avg}, M=50 \]

\[ \text{PRD}_{net}, M=50 \]

Noise floor set by \( M \)
Transmission errors affect CS measurements similarly to raw samples.

- ~20X lower energy for the same recovered signal error.
**Noisy Inputs**

- Input signal noise is another performance limiter
- CS filters some uncorrelated noise: can relax amp noise

\[ n_j = N_{in} \]

\[ f \rightarrow \text{ADC} \rightarrow Q_{CS} \rightarrow \text{CS} \rightarrow B \rightarrow \text{Reconstruct} \rightarrow f_{CS} \]

\[ f_{CS} = n_f, n(Q_{CS}), n_{ch}, n(M) \]

---

**Graphs**

- Minimum Energy/sample \( (N_0) \)
- Input Noise Level
- PRD \( (\text{dB}) \)
- Quantization only

- 6 dB PRD = 16X noise power

---

8/29/2011

MIT Thesis Defense
When we do get errors, sometimes it is BAD, sometimes it is OK

Can we prevent just the BAD cases?
(Unequal) Error Correction

- Try BCH block codes to protect the message, starting with the MSBs (PRD_{avg} shown)
  - ~4X lower energy compared to unprotected CS
  - Diminishing returns for more protection

![Diagram showing MSB and LSB with BCH block codes]

![Graph showing minimum energy/sample (N_0) vs. PRD (dB) for Uncoded CS, BCH-5, BCH-10, BCH-15]
**Error Detection**

- Normally, error detection \( \rightarrow \) re-transmit packet, does not lower energy
- Take advantage of the naturally redundancy in CS
  - Detect errors per measurement (CRC)
  - Throw away bad measurements
- Similar gains as error correction, but simpler hardware

![Graph showing energy consumption vs. PRD (dB)](image-url)
Implementing CS in Hardware

- Connect system performance (M, Q_CS, etc..) to hardware cost
  - Cost (active + leakage) must be less than transmission savings
- CS Encoder: a matrix multiply $\Phi f$
- Choose $\Phi$ to be Bernoulli ‘random’ matrix (i.e. ±1 entries)
  - Simplifies multiply operation
**CS Encoder Architectures: Analog**

Analog(-to-Information)

- Multiply and integrate in analog domain
- Quantize $M$ measurements ($B$-bits) with ADC
- $M$, slower ADCs, but also more full-rate amplifiers & mixers
- Random integration reduces effective headroom (tighter noise constraint at amplifier)
CS Encoder Architectures: Digital

- Quantize inputs first
- Multiply and accumulate in the digital domain
- $M$ accumulators, 1 full-rate ADC, but only 1 full-rate amplifier
- Unlike analog integrator, headroom can be extended
Analog power largely dominated by amplifiers
- Same spec. as digital but headroom gets reduced by integration

Digital power dominated by leakage (low freq.) and ADC (high freq.)
Compression vs. Gain vs. Resolution:
- $Q_{CS}=8$, $M=50$, $BW_f=200$ Hz, 90 nm CMOS

Higher resolution, gain, compression favors digital
Vice versa for analog $\rightarrow$ noise constrained
Matrix Generation

N > 500, M=50 \rightarrow How do we generate the matrix entries?

\[ y_1[k] = \int_{n}^{n+N} \frac{f_s}{N} f[n] \Phi_1[n] \, dn \]

\[ y_2[k] = \int_{n}^{n+N} \frac{f_s}{N} f[n] \Phi_2[n] \, dn \]

\[ \vdots \]

\[ y_{50}[k] = \int_{n}^{n+N} \frac{f_s}{N} f[n] \Phi_{50}[n] \, dn \]
Matrix Generation – 2 Mixed PRBS

- N > 500, M=50 → How do we generate the matrix entries?
- State of 1st PRBS x Output of 2nd PRBS: 65 flip-flops
  - Memory: >25 kb
  - Independent PRBS per measurement: 750 flip-flops
Matrix Generation – 1 PRBS?

- N > 500, M=50 → How do we generate the matrix entries?

- Shifted versions of a single PRBS are also uncorrelated, would that work?
I MPACT O F M ATRIX C Hosen

- Single PRBS works… if input is not oversampled
- Mixed PRBS reconstructs accurately in both cases
**Testchip Results: EEG Signal**

- 90 nm CMOS testchip\(^1\)
- A continuous EEG acquisition example
  - N = 1000, M=50
  - ADC: 8-bit, CS: 16-bit
  - 10X compression

Practical Application: EKG

- Original signal from the MIT-BIH arrhythmia database
- Good results at ~10X compression
Performance is leakage limited

- Measured 1.9 µW @ 0.6V for f < 20 kHz
- Model predicted ~0.6 µW with no clocks, buffers or control
OVERCOMING LEAKAGE LIMITATIONS

- Consider MEM relays
  - Zero leakage
  - Sharp sub-threshold slope
- Could enable E/op to scale
- How to design CS encoder circuits using relays?
STRUCTURE AND OPERATION OF MEM SWITCH

OFF Switch:

\[ |V_{gb}| < V_{po} \] (pull-out voltage)

ON Switch:

\[ |V_{gb}| > V_{pi} \] (pull-in voltage)

MEM Switches as Logic Elements

- Switches are ambipolar: absolute gate to body voltage determines actuation
- Conductance independent of drain/source voltage: non-inverting logic possible
CMOS: 30 transistors

Relays: 12 relays

- **CMOS:** delay set by electrical time constant
  - Quadratic delay penalty for stacking devices (pass transistor logic)
  - Buffer & distribute logical/electrical effort over many stages

- **Relays:** delay dominated by mechanical movement
  - Want all relays to switch simultaneously
  - Implement logic as a single complex gate
Delay Comparison vs. CMOS

- Single mechanical delay vs. several electrical gate delays
- For reasonable loads, relay delay unaffected by fan-out/fan-in (no need to size or buffer)

Area Comparison vs. CMOS

- Larger individual devices
- Fewer devices needed to implement the same logic function

B-bit Relay-based Adder

- 12 relays for full adder (vs. 24 for CMOS)
- Cascade FA cells to create larger adder
- Stack of B relays—still 1 mechanical delay

Energy vs. Delay: MEM Relays vs. CMOS

Energy/op vs. Delay/op across $V_{dd}$

- 30X capacitance gain
- Lower device $C_g$, $C_d$
- Fewer devices
- 2.4X $V_{dd}$ gain
- No leakage

- Compare vs. 32-bit Sklansky 90 nm CMOS adder\(^\text{[1]}\)
- For similar area: > 10x lower energy/op\(^\text{[2]}\)


\(^{[2]}\) F. Chen, et al. ICCAD '08
MEMORY ELEMENTS

- Same latch structure as in CMOS
  - Take advantage of non-inverting logic + equal pull-up/dn
- Same structure can be re-wired to create a 4 relay SRAM
MEM RELAY FLASH ADC

- Relay used as a dynamic comparator
- Sees same advantages as adder from smaller $C_g$'s
- Energy dominated by resistor string
  - Can eliminate resistor string by sizing devices for different $V_{pi}$

MEM Relay Demonstration Testchip

Oscillator & I/O

Logic

Memory & Timing

CLICKR1: 1 μm lithographic process

1F. Chen et. al., “Demonstration of Integrated MEM Switch Circuits for VLSI Applications,” ISSCC 2010. (Jack Raper Award)
**THINGS WE LEARNED – LAYOUT MATTERS!**

- Unbalanced current flow: flexures burn up!
- Parasitic $C_{gd}$, $C_{gs}$: can affect $V_{pi}$ (DIBL-like effect)

---

**MIT Thesis Defense**

8/29/2011

---

(a) Original Layout

(b) Improved Layout

$V_{DS} = 1V$

$V_{DS} = 2V$

$V_B = 0V$

$V_B = 0V$
MEM RELAY EVOLUTION

- Reduce parasitic capacitances:
  - $C_{gd}$, $C_{gs}$, $C_{gc}$, $C_{cb}$
- Added functionality (2nd set of drain/source terminals)
Adder: $\frac{1}{2}$ the number of devices in the PGK
ADC: 2\textsuperscript{nd} sampling phase for free
CS Encoder: \(~400\, \text{nW} \text{ vs. } 1.9\, \mu \text{W}\) in 90 nm CMOS (\(~5X\) gain)

- 10-bit accumulator, 20 kHz, 2V, (with output flops)

- ADC: on par with \textit{state-of-the-art} CMOS

- 5-bits, 2V, 20 kS/s, \(~45\, \text{nW}\) → FOM \(~70\, \text{fJ/conv. step}\)
CONCLUSIONS

- Compressed Sensing as a source encoder
  - ~Equivalent compression as LZW/Huffman
  - Fidelity to ~8-bits for compression of ~10-20X (trade-off!)
  - Robust over the channel, + can add low cost error detection

- CS Implementation Cost
  - Digital >> Analog for typical sensor specifications
  - Matrix generation is key: mixed PRBS is cheap and dynamic
  - Energy performance limited by leakage

- MEM Relays
  - Potential energy efficiency gains > 10X vs. CMOS
  - Optimal circuit design can expand application range to 10-100 MHz range
  - Relay circuits not limited to logic, but span the VLSI space
**Future Work**

- **CS Challenges**
  - Efficient reconstruction hardware
  - Signal basis selection

- **MEM Relay Challenges**
  - Contact engineering – non-oxidizing, durable
  - Device scaling
    - Residual stress, endurance
    - Releasing small gap distances (10-20 nm)
  - Reliability

- **CS Opportunities**
  - Wireless security: inherent layer of security in compression
  - Collaborative sensing (each node only takes a few measures)

- **MEMS Opportunities**
  - Could change fabless MEMS model from devices/components to systems
ACKNOWLEDGEMENTS

- Advisors: Vladimir Stojanović, Anantha Chandrakasan
- Committee: Vivek Goyal
- Collaborators
  - Compressed Sensing
    - Fabian Lim, Omid Abari
  - MEMS Circuits
    - MIT: Hossein Fariborzi
    - Berkeley: Elad Alon, Matthew Spencer, Abhinav Gupta
    - UCLA: Dejan Markovic, Chengcheng Wang, Kevin Dwan
  - MEMS Devices
    - Berkeley: Tsu-Jae King Liu, Hei Kam, Rhesa Nathanael, Jaeseok Jeon, Vincent Pott
  - Other:
    - Prof. Jing Kong/group, Prof. Carl Thompson/group, Hong Ma
- Sponsors
  - DARPA NEMS program, FCRP (C2S2, MSD), MIT CICS, Berkeley Wireless Research Center, NSF, Intel Graduate Fellowship, IBM TAPO
ACKNOWLEDGEMENTS

- Admins: Margaret Flaherty, Nora Luongo
- EECS, MTL, RLE Staff, Faculty & Students
- ISG and Ananthagroup members past & present (you know who you are)
- My wife Jes and her family
- My Mom