Traveling the Wild Frontiers of Ultra-Low Voltage Design

Jan M. Rabaey
Director Gigascale Silicon Research Center
Co-Director Berkeley Wireless Research Center
University of California at Berkeley
Why Ultra-Low Voltage?

- Power and Energy Limiting Integration and Scaling
- Exploring the Bounds and Frontiers of Computation
- The Brave New World of Ubiquitous Electronics

Meso-scale low-cost wireless transceivers for ubiquitous wireless data acquisition that
- are fully integrated
  - Size smaller than 1 cm³
- are dirt cheap (“the Dutch treat”)
  - At or below 1$
- minimize power/energy dissipation
  - Limiting power dissipation to 100 μW enables energy scavenging
- and form self-configuring, robust, ad-hoc networks containing 100’s to 1000’s of nodes
Why Worry about Ultra-Low Voltage?

- Maximum integration density ultimately limited by energy dissipated per unit volume.
- Technology scaling leads to linear increase in energy density (for same switching activity and voltage)
- Only options:
  - Reduce computational density – lowering frequency and/or activity
  - Reduce supply voltage
Power density: $p$ [W/cm$^2$]

Design rule [µm]

Power and Energy Limiting Integration

The Picture of Old

$\rho = \rho_{\text{DYNAMIC}} + \rho_{\text{LEAK}}$

Constant V scaling

$\rightarrow \rho_{\text{DYNAMIC}} \propto \kappa^3$

V scaled as $\kappa^{-1}$

$I_{DS} \propto (V_{GS} - V_{TH})^{1.3}$

$\rightarrow \rho_{\text{DYNAMIC}} \propto \kappa^{0.7}$

(Sakurai, 2003)
Power and Energy Limiting Integration
The Roadmap Perspective

Active power density: $k^{1.7}$

Leakage power density: $k^{3.4}$

Compute density: $k^3$

2003 ITRS – Low operating power scenario
The Reality May Be Worse!

1st Mega Trend: Slowing $V_{CC}$ / Growing Power

Voltage scaling is slowing / stopping ~1.0V

Scott Thompson, TI Fellows meeting 2004.
Better Option: Slow down or reverse compute density increase

– Use slack to control power (that is, voltage) and leakage
There is Room: Minimum Operational Voltage of Inverter

- Swanson, Meindl (April 1972)
- Further extended in Meindl (Oct 2000)

Limitation: gain at midpoint > -1

\[ V_{dd}(\text{min}) = 2\left(\frac{kT}{q}\right)[1 + \frac{C_{fs}}{(C_{ox} + C_d)}] \ln(2 + \frac{C_d}{C_{ox}}) \]

Cfs: fast surface state capacitance
Cox: gate capacitance
Cd: diffusion capacitance

For ideal MOSFET (60 mV/decade slope):

\[ V_{dd}(\text{min}) \cong 2 \left(\ln 2\right)\frac{kT}{q} = 1.38\frac{kT}{q} = 0.036 \text{ V at } 300^\circ \text{K}. \]
Gain is the Limiting Factor

Voltages normalized to $U_T = kT/q$

From E. Vittoz, Ch. 16, Low Power Electronics, Ed. C. Piguet, 2005.
Confirmed for Current Technologies

Min $V_{dd}$ (inverter)

Degradation due to asymmetry

Min $V_{dd}$ (NOR)

For $n = 1.6$, $V_{dd_{min}} = 1.9 \frac{kT}{q} = 48 \text{ mV}$

90 nm CMOS (simulation – nominal process parameters)

Source: M. Stan, L. Alarcon
Minimum Energy per Operation

• Predicted by von Neumann: $kT \ln(2)$

• Based on previous result – moving one electron over $V_{dd min}$:
  – $E_{min} = QV_{DD}/2 = q \frac{2(\ln2)kT}{2q} = kT \ln(2)$
  – Would be approximately three times larger for CMOS inverter with PMOS twice the size of NMOS
  – At room temperature (300K): $E_{min} = 0.29 \times 10^{-20} \text{ J}$

• Minimum sized CMOS inverter at 90 nm operating at 1V
  – $E = CV_{dd}^2 = 0.8 \times 10^{-15} \text{ J}$, or 5 orders of magnitude larger!

How Close Can We Get?
Option 1: Subthreshold Operation

Making Leakage Work You!

Example: Energy-Aware FFT

- Energy aware FFT architecture scales gracefully from 128 to 1024 point lengths and supports 8b and 16b precision.

**FFT Energy-Performance Curves**

\[ E_{\text{Switching}} = a \cdot C_L \cdot V_{DD}^2 \]

\[ E_{\text{Leakage}} = I_s \cdot V_{DD} \cdot 10^{\frac{V_{th}}{S}} \cdot T \]

**Minimum Energy Point @ VDD = 0.35V and VT = 0.475V**

(estimated from switching and leakage models for a 0.18\(\mu\)m process)

Courtesy: A. Wang, A. Chandrakasan, MIT
SubThreshold FFT

Process Details
- 0.18\(\mu\)m CMOS process
- 6 layer metal
- 628k transistors

Operational down to 180 mV (fclock = 64 Hz)
The FFT operates between $V_{DD}=180mV-900mV$ and clock frequency of 164Hz-6MHz.

The minimum energy dissipated is 155nJ/FFT at 350 mV for a 1024-point 16b FFT. The clock frequency is 10kHz and the FFT processor dissipates 0.6$\mu$W.
• Explores Minimum-Energy Processor
  - subthreshold operation
  - 3pJ per instruction at 350mV operation
  - 10X less energy than previously reported
  - 41 year operation on 1g Li-ion battery

• Research Areas
  - Processor architectural trade-offs
  - Low voltage memory design
  - Process variation tolerance

 Courtesy: D. Blaauw, T. Austin, UMICH
Is Sub-threshold The Way to Go?

- Achieves lowest possible energy dissipation
- But … at a dramatic cost in performance

130 nm CMOS
Option 2: Managing Leakage while Reducing Thresholds

Stacked transistors enable aggressive threshold scaling

- Ion/Ioff increases with increasing stack height (leakage suppression)
- More robust to correlated (tune or adapt) and random variations (self-cancel)
- Decreased short channel effect

![Graphs showing current vs. voltage for different stack heights and without stack effects.](Image)
Impact of Stacking Devices

Stack-depth 2

$V_{DD}$

$V_T$

1 ns

1 fJ
Impact of Stacking Devices

Stack-depth 4

$V_{DD}$

$V_T$

1 ns

1 fJ
Impact of Stacking Devices

Stack-depth 6

\( V_{DD} \)

\( V_T \)

1 ns

1 fJ
Impact of Stacking Devices

Stack-depth 8

$V_{DD}$

$V_T$

1 ns

1 fJ
Optimal EDP, Energy, Delay vs. Stack

- Minimum EDP
- Delay for Minimum EDP
- Energy for Minimum EDP
Complex Gates
Reducing thresholds while containing leakage

The return of PLAs?
- Regular
- Tunable
- NAND/NAND configuration

Or pass-transistor logic
- Current-steering
- Regular
- Balanced delay
- Programmable
Some ULV Challenges:
(1) Excessive Timing Variance

- Timing variance increases dramatically with $V_{dd}$ reduction
- Design for large yield means huge overhead at low voltages:
  - Worst case design at 300mV means over 200% overkill
Managing Systematic Variations Through Self-Adaptation

Test Module

Test inputs and responses

Module

Energy-performance trade-off

Adaptive Tuning
Worst Case, w/o Vth tuning
Nominal, w/ Vth tuning

Path Delay (ps)

10x

Move test onto the chip
Dynamically adjust supply and threshold design parameters to center the design!

Courtesy: K. Cao, Arizona
Some ULV Challenges:

(2) The Memory Data-Retention Voltage (DRV)

When $V_{dd}$ scales down to DRV, the Voltage Transfer Curves (VTC) of the internal inverters degrade to such a level that Static Noise Margin (SNM) of the SRAM cell reduces to zero.

**DRV Condition:**

$$\frac{\partial V_1}{\partial V_2} \bigg|_{\text{Left inverter}} = \frac{\partial V_1}{\partial V_2} \bigg|_{\text{Right inverter}}, \text{ when } V_{DD} = \text{DRV}$$

Source: Huifang Qin, QEDTI 2004
The Impact of Process Variations

DRV Spatial Distribution
(256*128 Cells)

Histogram of 32K SRAM cells
Reducing the DRV

Histogram of 32K SRAM cells

Solution II: Error-tolerant SRAM design (with Redundancy and ECC)

Option: ULV SRAM circuit optimization

Combination of sizing and aggressive ECC reduces DRV to 150 mV
How about Mixed-Signal?

- Reduced headroom challenges traditional mixed-signal design
- Process variation makes design centering tough
- Does further scaling help?

Courtesy: R. Rutenbar, CMU
The Lure of the Sub-Threshold Region

- Greater transconductance \( (g_m) \) for a given bias current
- Lower \( f_t \), but CMOS scaling helps
Baseband Processor

Technology: 0.13um CMOS
Power Supply: < 1 V
Chip Area: 2.1mm x 2.1mm (pad limited)
Total Power (Analog+Digital): 200 μW
Synchronization Header: Amplitude estimation: 10 bits
Timing estimation: 25 bits (worst case)

Courtesy: Yan-Mei Li, UCB, CICC 2005
Example: Energy-Efficient Data Conversion

Low-Voltage Low-Power Successive-Approximation A/D

Simplest architecture wins!

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fs</td>
<td>800KS/s</td>
</tr>
<tr>
<td>Resolution</td>
<td>6 bits</td>
</tr>
<tr>
<td>$V_{dd}$</td>
<td>.5V</td>
</tr>
<tr>
<td>$P_d$</td>
<td>$\sim$2 uW</td>
</tr>
</tbody>
</table>
Absolutely!

- Aggressive use of passives
- Unorthodox architectures to create gain (receiver) or increase efficiency (transmitter) at low voltage/current levels
- Stacking of components often helps (current re-use)
- Efficient oscillators are essential!
Extensive Use of (Innovative) Passives

- High Q-factor
- Small form factor
- MEMS/CMOS co-design
- Integration into IC process

Ruby et. al. (Ultrasonics Symposium 2001)

Carpentier et. al. (ISSCC 2005)
Exploring the Limits: (Almost) Passive 1.9GHz Receiver

- \( P_{RX} = 200\text{nW} \)
- \( BW_{-3\text{dB}} = 4\text{MHz} \)
- Sensitivity = -38dBm (12dB SNR)
- \(|S_{11}| = -9.3\text{dB} \)

Courtesy: B. Otis, N. Pletcher
Providing Gain: The Return of Super-Regenerative

- Super-regenerative receiver creates gain at low-current level
- Sub-threshold operation
- No external components (inductors, crystals, capacitors)
- 0.13µm CMOS

Total Rx: 380µW
SuperRegenerative: Gain at Low Current

\[ f_q = 100\text{kHz} \]

\[ -70\text{dBm} \]
SuperRegenerative: Gain at Low Current

Detector oscillator transient:
- OOK modulation
- -80dBm, 5kbps
Low-Voltage / Energy Oscillators

Low power design techniques
- Complementary Gm stages to reduce $I_{\text{bias}}$
- Large $R_b$ to reduce FBAR loading
- Sub-threshold MOSFET maximizes $g_m/I_d$
- Optimal choice of $C_1$ and $C_2$

Y.H. Chee, CICC 2005
Low-Energy FBAR Oscillator - Measurements

FBAR oscillator phase noise at 90\(\mu\)W power consumption

Instrument’s noise floor

FBAR oscillator voltage swing
(~140mV 0-pk @ 90 \(\mu\)W)
Dealing with Variations

- Calibrate LC oscillator with high accuracy reference (FBAR)
- Convert control voltage to digital signal and control oscillator frequency digitally
- Turn off FBAR oscillator and control loop after calibration

To calibrate over 200MHz span better than 500kHz accuracy, 400 steps (9 bits) required

High-accuracy (500ppm) FBAR oscillator (300μW)

Low-accuracy LC oscillator (<100μW)

\[ C_{\text{LSB}} = 1 \, fF \]
Digitally Tuned VCO

One bondwire and one integrated version implemented for comparison

**Bondwire oscillator performance**

- Supply voltage: 0.5V
- Power consumption: 100μW
- Nominal frequency: 1.9GHz
- Tuning Range: 150MHz
- Resolution: ~400kHz (9 bits)
- Phase noise: -115dBc/Hz @ 1MHz offset

Courtesy N. Pletcher, UCB, ESSCIRC 2005
Low Accuracy LC Oscillator - Results

0.5V supply

Minimum startup conditions:
- \( V_{dd} = 0.3V \)
- \( I_{bias} = 140\mu A \)
- \( (42\mu W) \)

Output Swing vs Bias Current

Differential \( V_{out} \) (mV, p-p)

Core bias current (\( \mu A \))

Tuning range (10 bit code)
Innovative Architectures: Injection Locked Transmitter

- Use LC power oscillator instead of a power amplifier
  - Self-drive reduces driver power
- Capacitive bank to tune oscillation within locked range
- Reference oscillator to lock the power oscillator to an accurate carrier frequency

Y.H. Chee et al, CICC 2005
TX Performance

- TX consumes an average power of 1.5mW while delivering 1mW OOK signal (32% efficiency).
- Degradation of TX efficiency due to driver stage (FBAR oscillator) is only 1%.

ST 0.13μm CMOS
Perspectives

There is plenty of room at the bottom!

- Further scaling of energy/operation (or current per function) is essential for scaling to produce its maximum impact
- Current digital gates 5 orders of magnitude from minimum
- Exciting opportunities offered by new paradigms in computing
- Innovations at circuit, architecture and system level are essential
- Ample opportunity still to tame some wild horses

The art of ingenuity

H. De Man
ISSCC 05