Retiming and Dual-supply Voltage Based Energy Optimization for DSP Applications

Teng Xu, and Miodrag Potkonjak

University of California, Los Angeles

Background

- **DSP Applications**

Video Compression

Audio Signal Processing

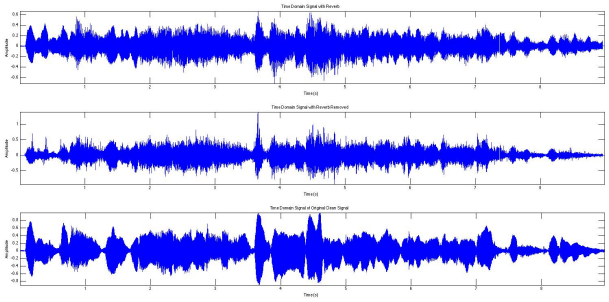
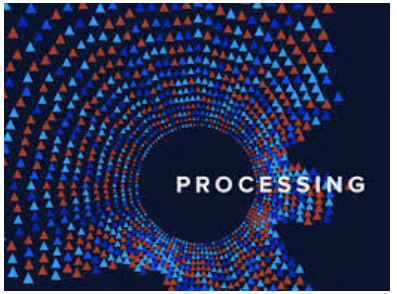
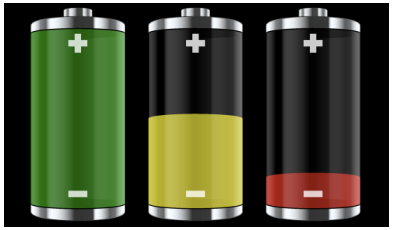


Image Processing



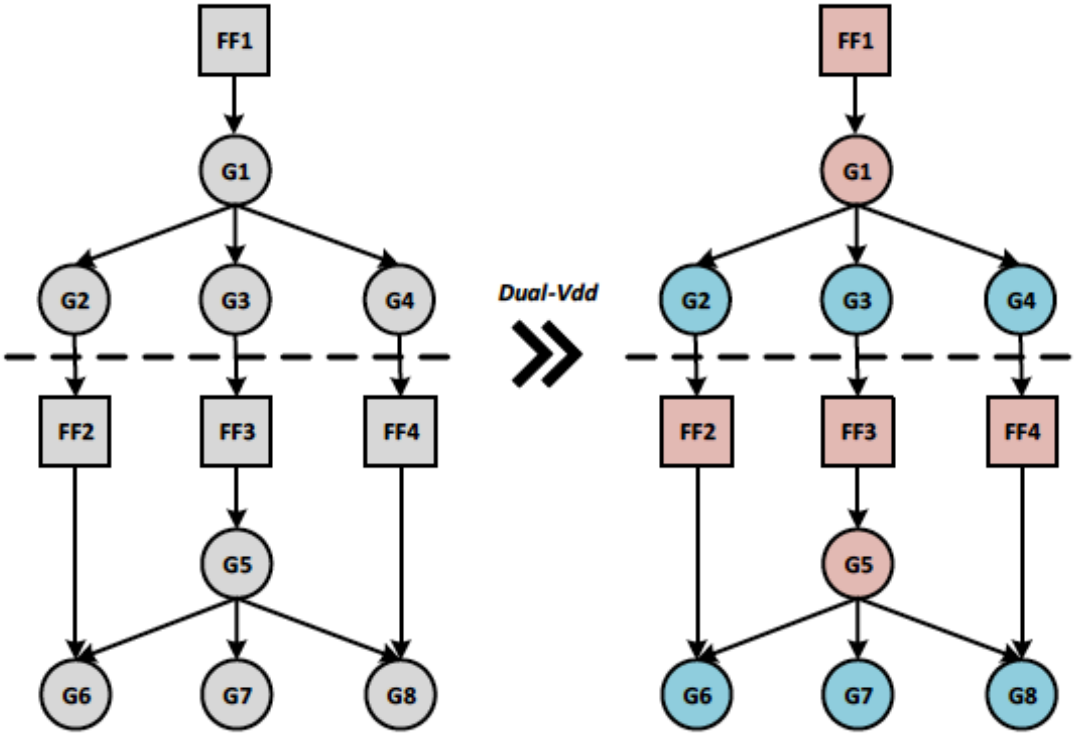
- **Hardware Implemented DSP Applications**
  - ASIC
  - FPGA



- **Constraints**
  - Cost, Size, Latency, Energy
  - Among all constraints, energy becomes especially important

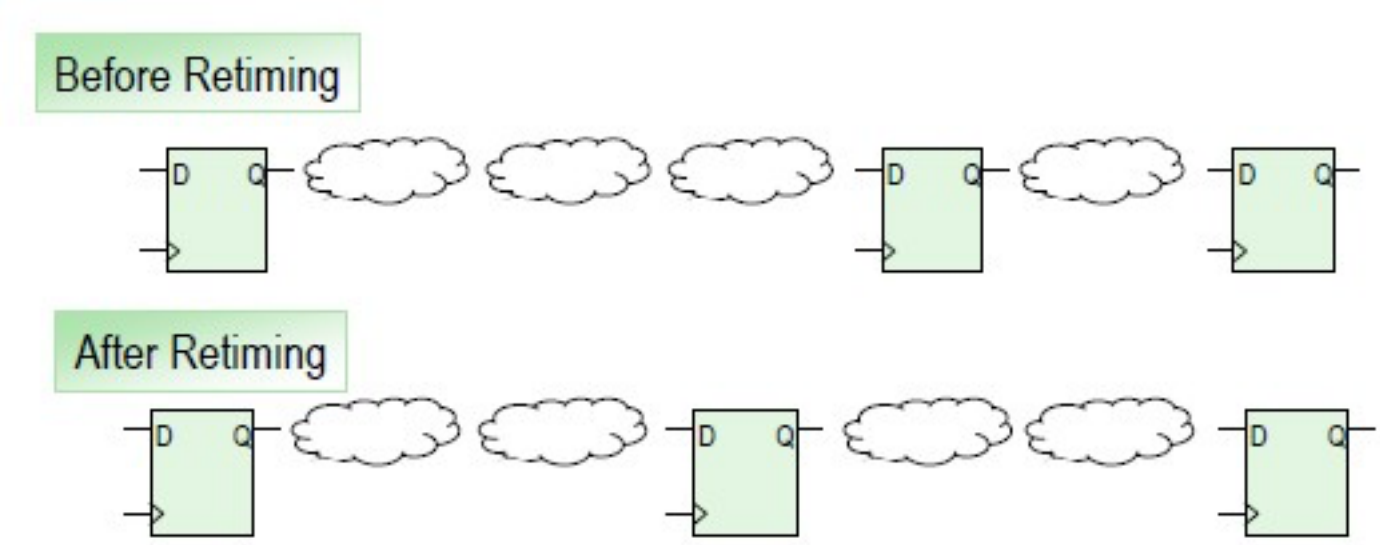
Circuit Energy Optimization

- **Dual supply voltages (dual-vdd)**
  - Assign two supply voltages to a circuit
  - Gates with high voltage drive gates with low voltage
  - Flip-flops are set to high voltage



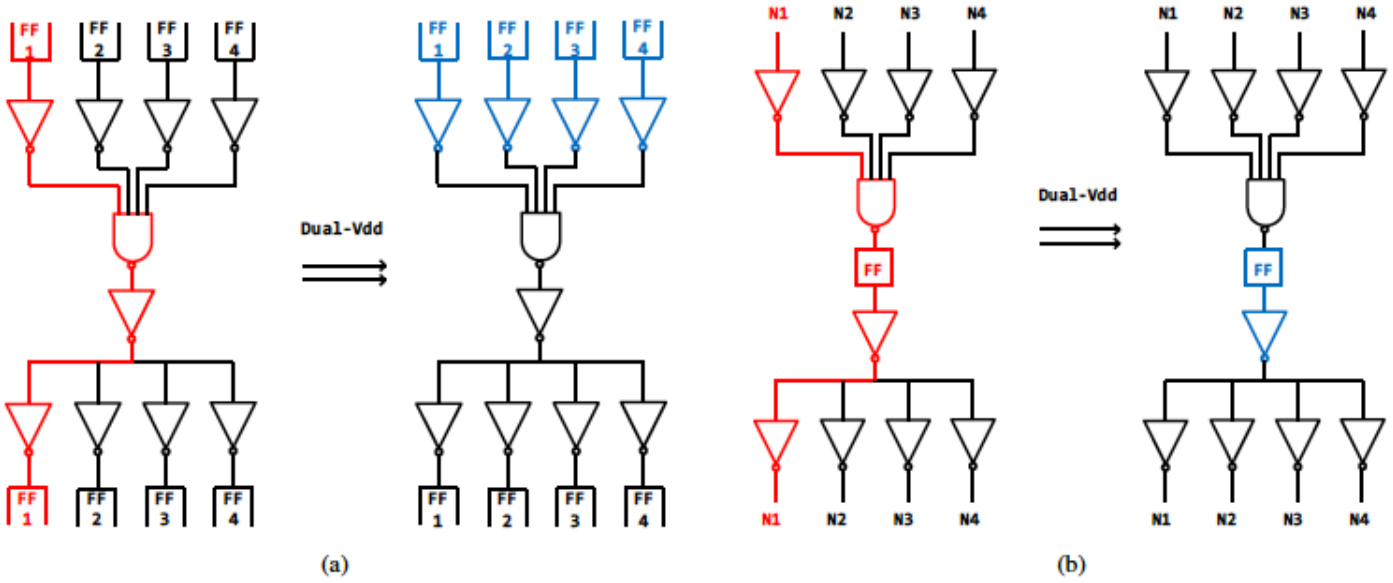
Circuit Energy Optimization

- **Retiming**
  - **Change circuit structure, e.g., position of flip-flops**
  - **Circuit functionality is not altered**
  - **Reduce circuit delay**



Objective

- **Retime circuit to facilitate dual-vdd optimization**
  - **Conventional retiming is not designed for dual-vdd**
  - **Minimize the number of flip-flops: fewer gates on high voltage**
  - **Not compromise the delay of circuit**



Circuit Modeling

- Delay and power models from Markovic et al. [1]
- Delay Model

$$D = \frac{k_{tp} \cdot C_L \cdot V_{dd}}{2 \cdot n \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot (\frac{kT}{q})^2} \cdot \frac{k_{fit}}{(\ln(e^{\frac{(1+\sigma)V_{dd}-V_{th}}{2 \cdot n \ cdot(kT/q)}}))^2}$$

$$C_L = C_{ox} \cdot L \cdot (\gamma \cdot W + W_{fanout})$$

- Power Model
  - Leakage Power
  - Switching Power

$$P_{leakage} = 2 \cdot n \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot (\frac{kT}{q})^2 \cdot V_{dd} \cdot e^{\frac{\sigma \cdot V_{dd}-V_{th}}{n \cdot (kT/q)}}$$

$$P_{switching} = \alpha \cdot C_L \cdot V_{dd}^2 \cdot f$$

[1] D. Markoví c, C. C. Wang, L. P. Alarcon, T.-T. Liu, and J. M. Rabaey, "Ultralow-power design in near-threshold region," Proceedings of the IEEE, vol. 98, no. 2, pp. 237–252, 2010.

Retime For Minimal Flip-flops

- **Apply minimal-cut on circuit**
- **Assign flip-flops on the minimal cut**
- **Only assign flip-flops that will not compromise circuit delay**

Algorithm 1 Retiming for minimum flip-flops (RTMF)

Input: C_0 - original circuit.
Input: CP_0 - critical path on C_0.
Input: FF_0 - flip-flops on C_0.
FF_{fix} is a vector that contains all the flip-flops that are fixed.

$FF_{fix} = \emptyset$
do
  for all ff_i in FF_0
    if ff_i is in CP_0
      FF_{fix}.append(ff_i)
    end if
  end for
  $C_{pre} = C_0$
  $(C_0, CP_0, FF_0) = Mincut((C_0, CP_0, FF_0 - FF_{fix}))$
while $CP_0 != CP_{pre}$
Output: C_0

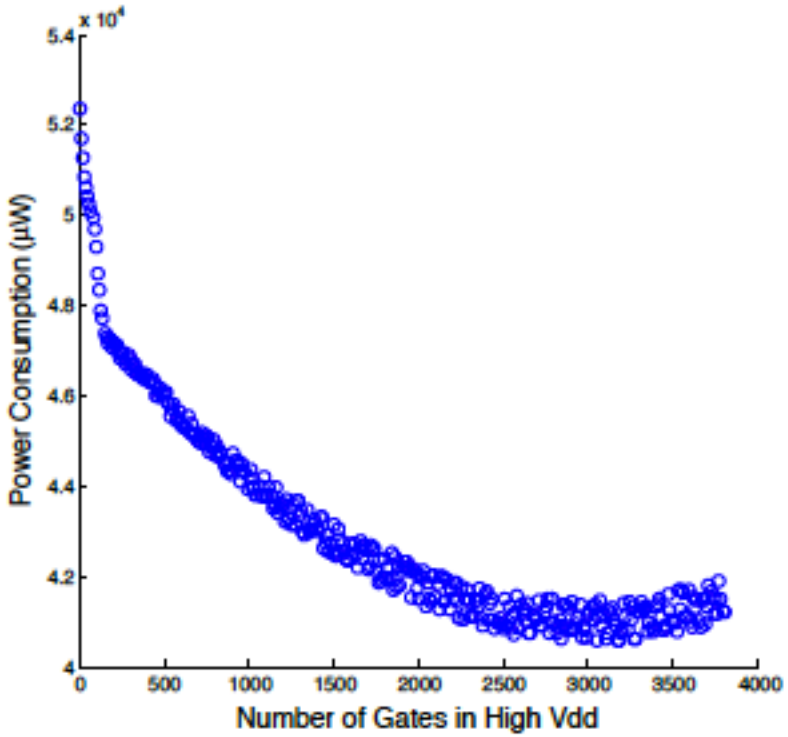
Dual Voltage Optimization

- **Two key issues**
  - **What voltages should be used?**
  - **Which part of the circuit should be assigned to high/low voltage?**

Algorithm 2 Dual-vdd Optimization (DV)

Input: C - original circuit.
Input: Vdd_0 - original supply voltage on C_0.
Vec_{high} is a vector that contains the gates with high vdd.
Vec_{low} is a vector that contains the gates with low vdd.

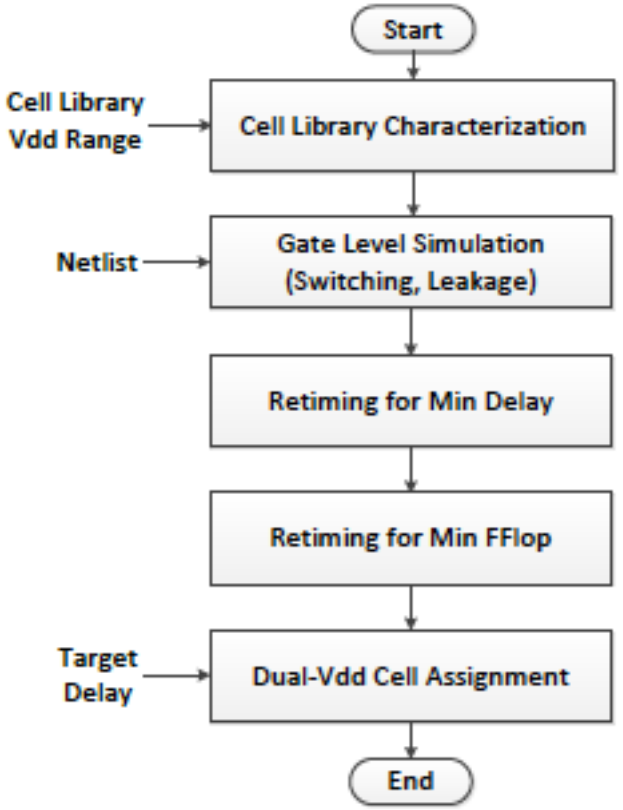
Vec_{high}= $\emptyset$.
Vec_{low}= all gates in C_0.
$Vdd_{high} = Vdd_{low} = Vdd_0$
do
  $CP_0 = CriticalPath(C, Vdd_{low}, Vdd_{high}$
    $Vec_{low}, Vec_{high})$
  $POW_0 = Power(C, Vdd_{low}, Vdd_{high}, Vec_{low}, Vec_{high})$
  for all $Gate_i$ in CP_0
    Vec_{high}.append($Gate_i$ with smallest slack)
    Vec_{low}.erase($Gate_i$ with smallest slack)
  end for
  for all possible $Vdd_{newlow} \leq Vdd_0$
    binary search $Vdd_{newhigh}$
    until $CP_0 = CriticalPath(C, Vdd_{newlow},$
      $Vdd_{newhigh}, Vec_{low}, Vec_{high})$
    $POW_1 = Power(C, Vdd_{lownew},$
      $Vdd_{highnew}, Vec_{low}, Vec_{high})$
    if $POW_1 < POW_0$
      $Vdd_{low} = Vdd_{newlow}$
      $Vdd_{high} = Vdd_{newhigh}$
      $POW_0 = POW_1$
    end if
  end for
while Vec_{low}!= $\emptyset$
Output: C



Results

- **Experimental Set up**
  - ISPD2012 cell library
  - Initial supply voltage: 0.7V
  - Initial threshold voltage: 0.3V
  - Energy reduction: 28.18% to 39.27%

- **Experimental Results**



Circuit	Target Delay (ns)	Initial Energy (mJ)	Supply Voltage(s) (V)			Energy Savings (% compared to initial)	
			Scaled (after RTMF)	(V_{min}, V_{max})		RTMF	RTMF+DV
FFT-64	47.35	38.75	0.57V	0.53V	0.64V	34.57 %	45.08%
FFT-128	47.35	153.15	0.55V	0.52V	0.61V	39.27 %	56.06%
DCT-8x8	59.83	117.3	0.56V	0.52V	0.63V	35.38 %	65.1%
DCT-16x16	55.39	2988.18	0.59V	0.56V	0.68V	28.18 %	39.58%
Average	-	-	-	-	-	34.35 %	51.46 %