Approximating Computation and Data for Energy Efficiency

Daniele
Jahier Pagliari

EDA Group
Politecnico di Torino
Torino, Italy

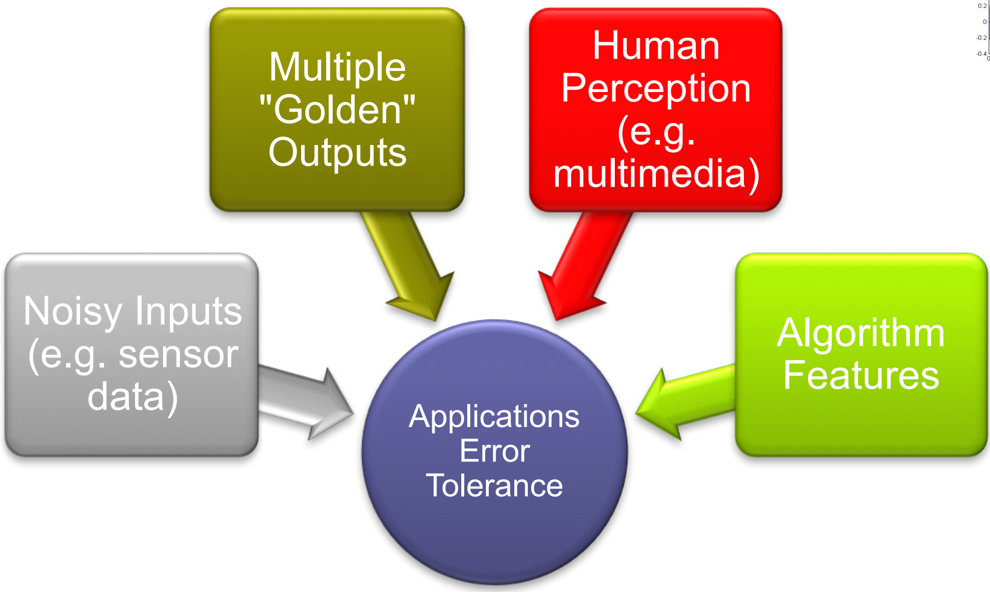
1st IWES
September 20th , 2016, Pisa, Italy

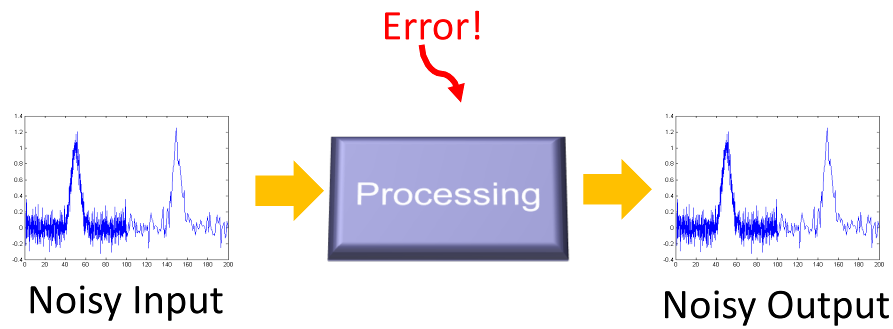
Outline

- Error Tolerance and Approximate Computing
- Our View
- AC in Processing
- AC in Interconnects
- AC in Actuators (OLED Displays)
- Conclusions

Error Tolerance

- Many emerging applications are ***error tolerant*** (or ***resilient***)





Original Image

Errors in 3 LSBs

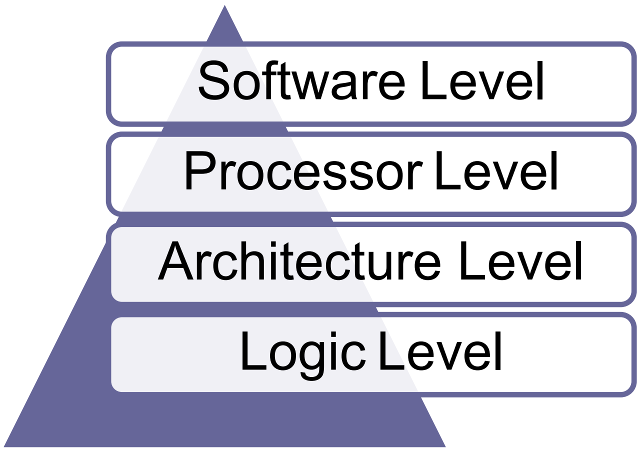
Approximate Computing

ESs Design Challenges:

Performance Demands
- "Smart" Systems
- Internet of Things

Energy Budget
- Battery-operated
- Energy-autonomous

Abstraction Levels:



Approximate Computing (AC):

Tradeoff *energy consumption* and *output quality* leveraging applications error tolerance.

Classic AC:

- Design-time approximations (fixed error).

Recent Trend:

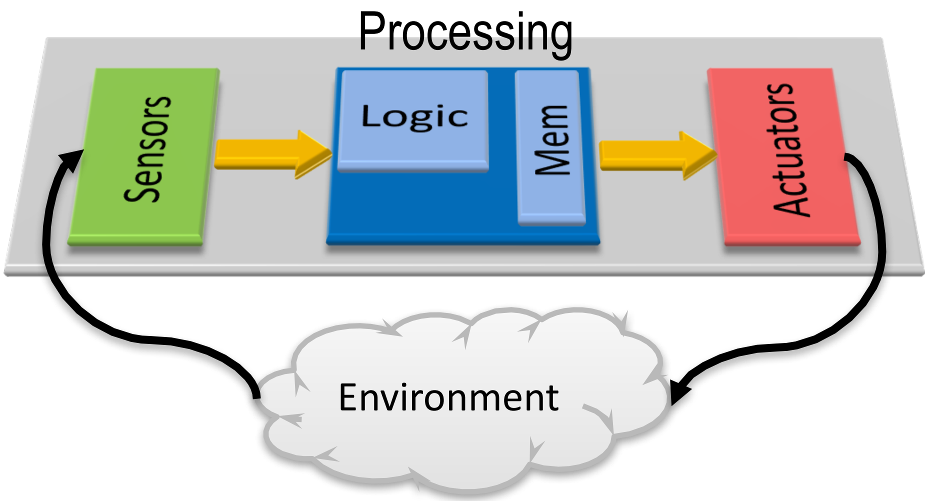
- Runtime-reconfigurable error.

Issues:

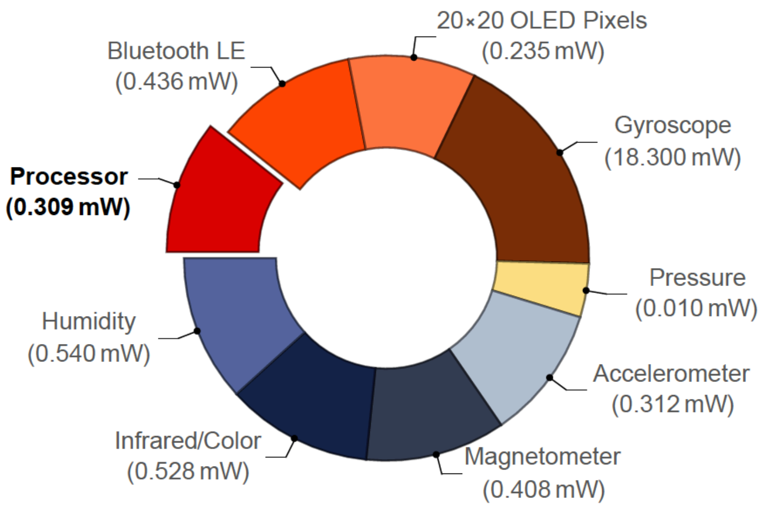
- What about **system-level**?.
- What about **automation?**

Motivation

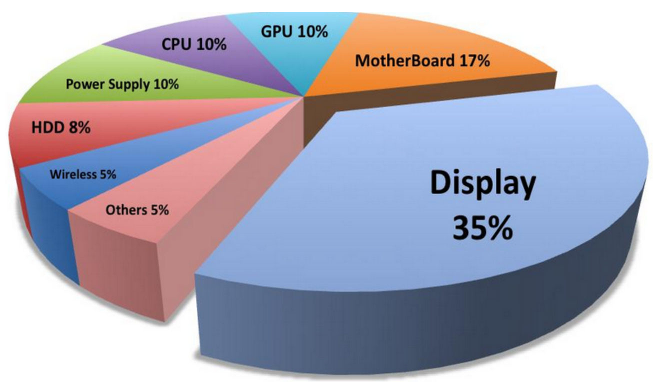
- **Embedded System:**



- **ES Energy Breakdown:**



Stanley-Marbell et al. DAC'16



EETimes

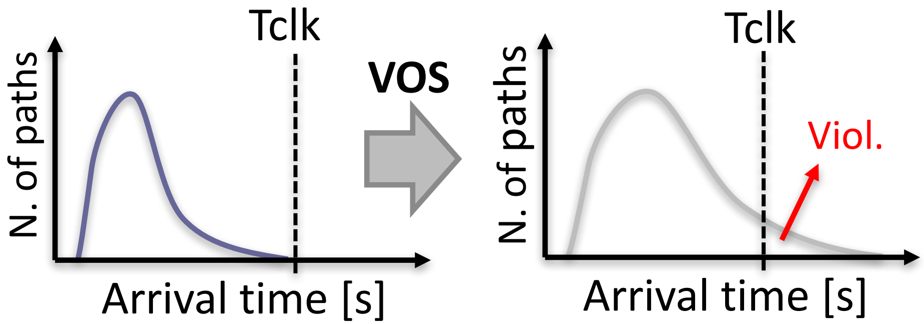
In Summary:

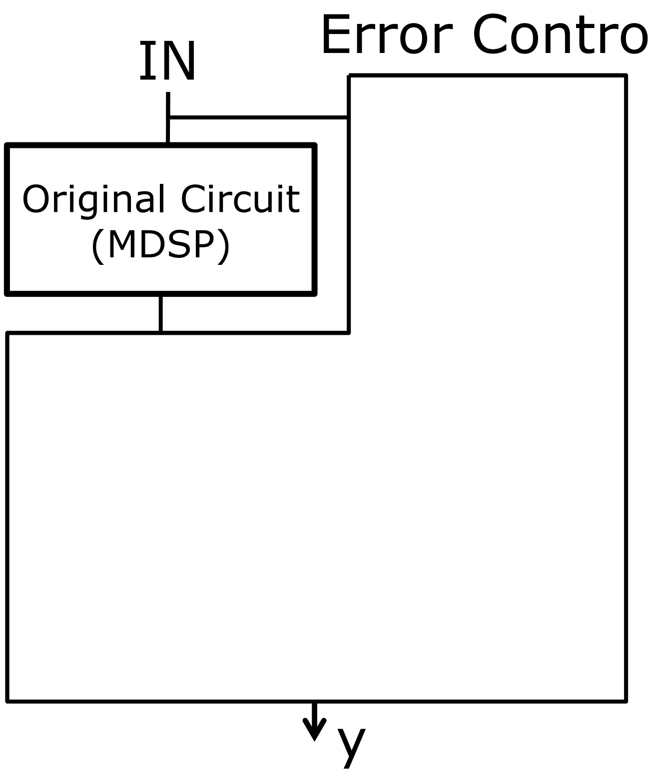
- **Sensors, actuators** and **interconnects** are relevant contributors to consumption.
- The breakdown is strongly **system-dependent**
- **Approach: approximation as a system-level design knob!**

AC in Processing: RPR

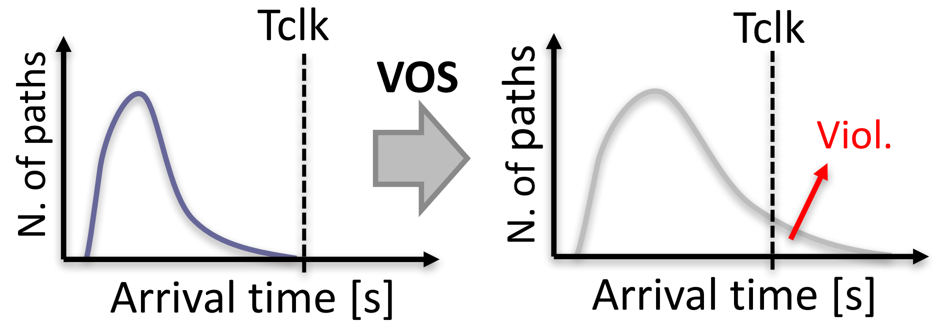
Idea: [Shanbag et al. ISLPED'99]

- **Voltage Over-Scaling** (VOS) on the original circuit (MDSP)
- **Error Control Block** (EC-Block) to mitigate the effect of timing errors.





AC in Processing: RPR



Idea: [Shanbag et al. ISLPED'99]

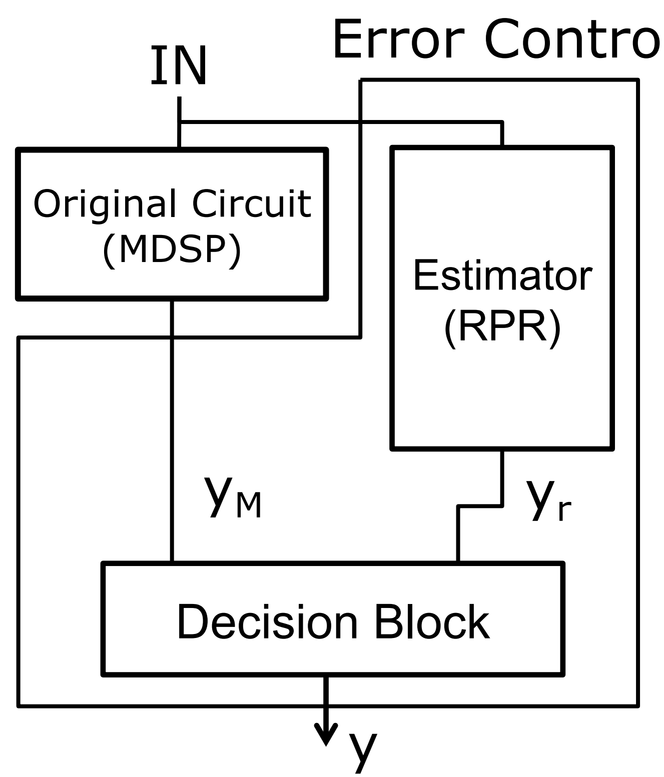
- **Voltage Over-Scaling** (VOS) on the original circuit (MDSP)
- **Error Control Block** (EC-Block) to mitigate the effect of timing errors.

EC Block Structure:

- **Estimator** of the error-free output
- **Decision block** to select between MDSP and Estimator outputs

Estimator Implementation:

- **Reduced Precision Replica (RPR)**



AC in Processing: Our Contribution

Classic RPR has limitations:

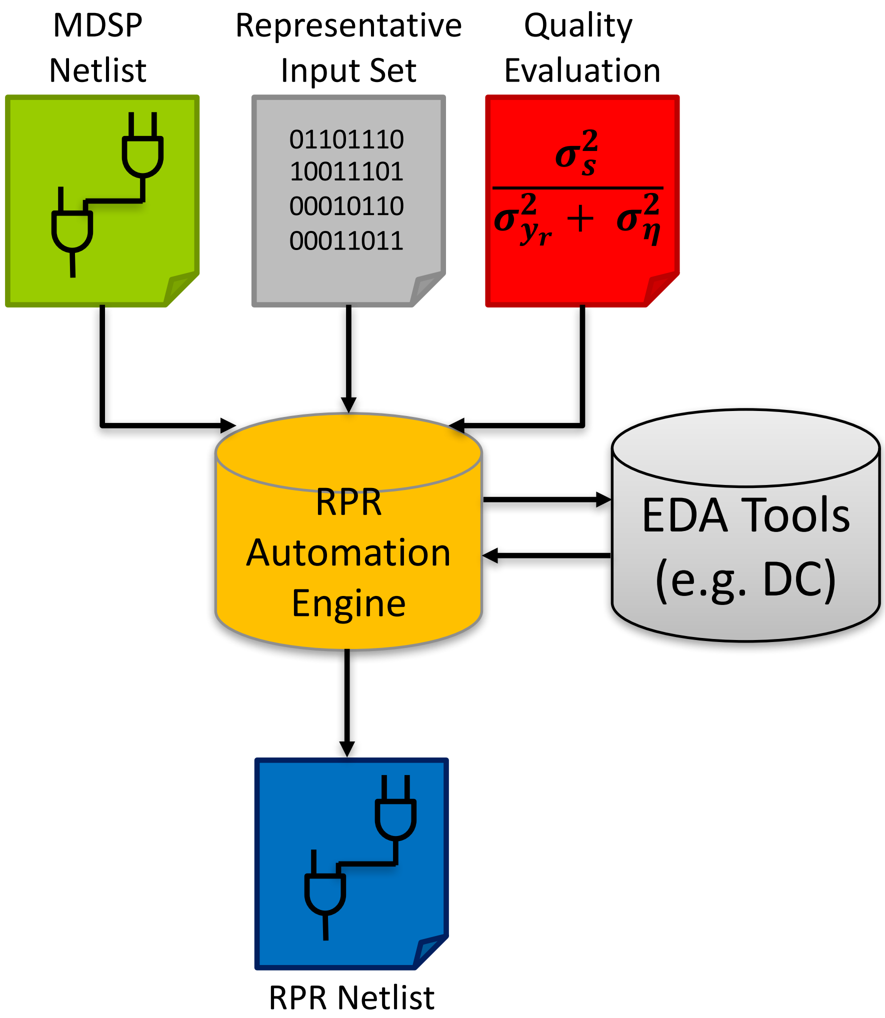
- Replica design and error estimation require knowledge of functionality (**design specific**)
- Uses **simplified and unrealistic assumptions** (e.g. on input statistics)

Proposed Framework:

- Automatically add RPR to existing gate-level netlist of a datapath circuit.

Features:

- **Functionality-agnostic.**
- **Simulation-based.**
- Integrated with **state-of-the-art tools** for synthesis and simulation.



AC in Processing: Results

Setup:

- 45nm library from STM.
- Opencores designs, realistic quality constraints

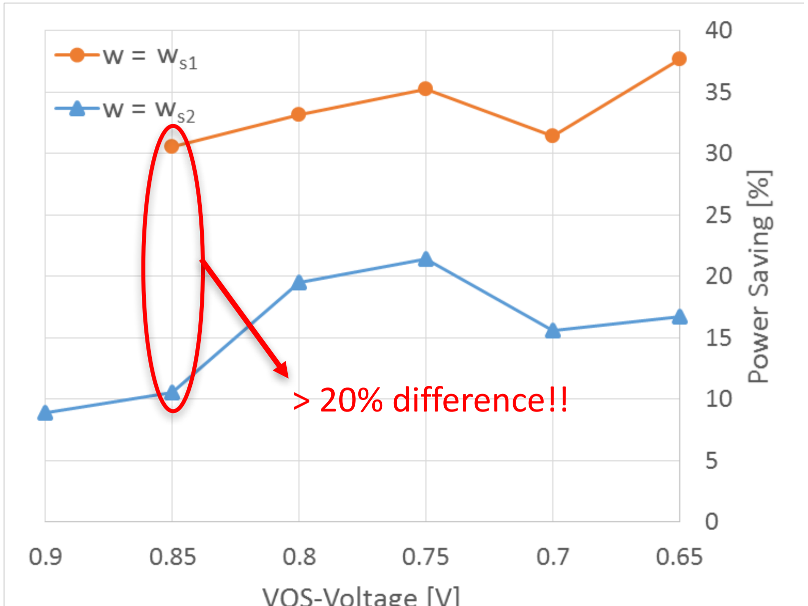
Generality:

- Successfully applied RPR to **previously untested designs** (CORDIC, SRU).
- **Comparable savings w.r.t. ad-hoc approach** on FIR and FFT.

Benchmark	Tot. Power Saving [%]	Area Ovr. [%]
FIR Filter	**44.96**	82.39
FFT Butterfly	**49.66**	133.20
RM-CORDIC	**42.05**	127.64
SRU	**47.91**	143.32

Benefits of simulation-based approach:

- Different input stimuli cause different error rate on the MDSP, at the same V_{VOS}.
- Consequently, a larger/smaller replica can be used to obtain the same quality.
- **Strong impact of inputs on the obtainable power savings.**



RPR power saving vs voltage for a FIR filter, with different input stimuli (same quality constraint).

AC in Interconnects: Motivation

Serial buses:

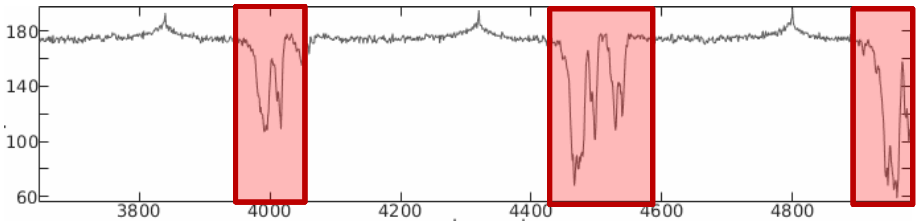
- De-facto standard for interfacing **sensors, actuators and I/O controllers**
- Higher frequencies, no jitter issue, reduced crosstalk
- Lower costs (less pins and easier wiring)
- **SPI, I2C, MIPI**, etc.

Motivation:

- PCB traces have large capacitive loads that have not scaled as transistors!
- **Transmission of one 12 bit sample ≈ execution of 1 instruction!**[1][2]
- **Tens of serial connections in a system!**

Error Tolerant Bus Traces:

- Sensor ICs/multimedia actuators (audio DAC, displays)
- Long **"idle"** (roughly constant) phases.
- Short **"bursty"** (fast and large variations) phases.
- **Example:** *Lena* image (red channel)



(Most) information conveyed by the bursty phases!

[1] P. Stanley-Marbell and M. Rinard. Value-deviation-bounded serial data encoding for energy-efficient approximate communication. 2015
[2] N. Ickes, et al. . A 10-pJ/instruction, 4-MIPS micropower DSP for sensor applications. 2008.

AC in Interconnects: ST0/ADE

Two Encodings with common Principles:

- **Exploit idle phases for power saving!**
- **Avoid redundancy** (introduces large overheads in serial buses)
- **Allow runtime-reconfiguration of accepted error.**
- **Simple implementation** (CODEC HW overheads must not offset gains).
- **Serial T0 (ST0):**
  - Selectively transmit the correct datum or a special 0-Transitions pattern (interpreted as "*repeat previous datum*").
  - $|b(t) - b(t')| > Th$ → **Send correct data**
  - $|b(t) - b(t')| \leq Th$ → **Send 0-T pattern.**
- **Approximate Differential Encoding (ADE):**
  - Based on bitwise **Differential Encoding (DE)**: $B(t) = b(t) \oplus b(t-1)$
  - Enhanced with **LSB-saturation** to reduce transitions also during bursty phases

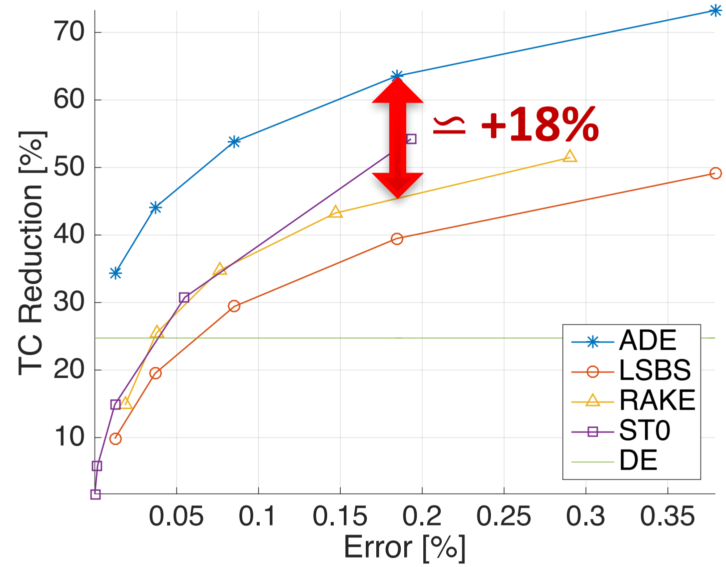
AC in Interconnects: Results

Comparison:

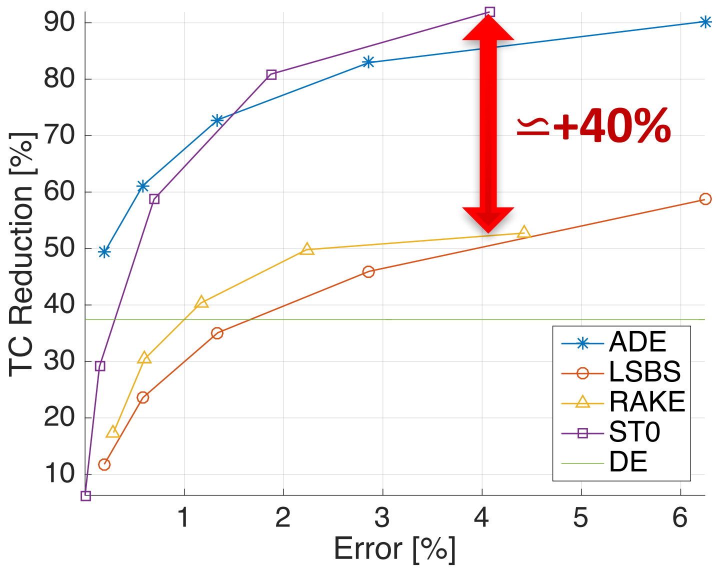
- **Rake** [Stanley-Marbell, DAC'16]
- **LSBS** and **Accurate DE**

Results:

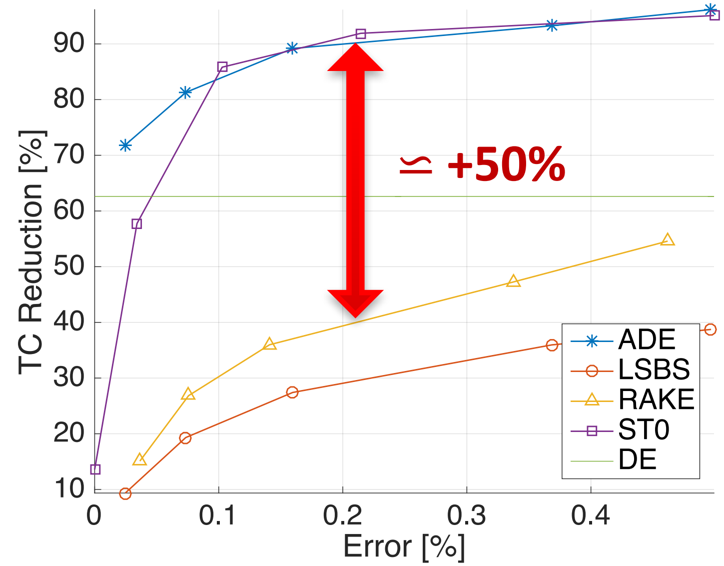
- ADE and ST0 are **both superior to state-of-the-art**
- ST0 better for "**strong burstiness**", ADE superior for **more random data**.



Accelerometer [PSR Database]



Images [Kodak Database]



ECG [Physionet Database]

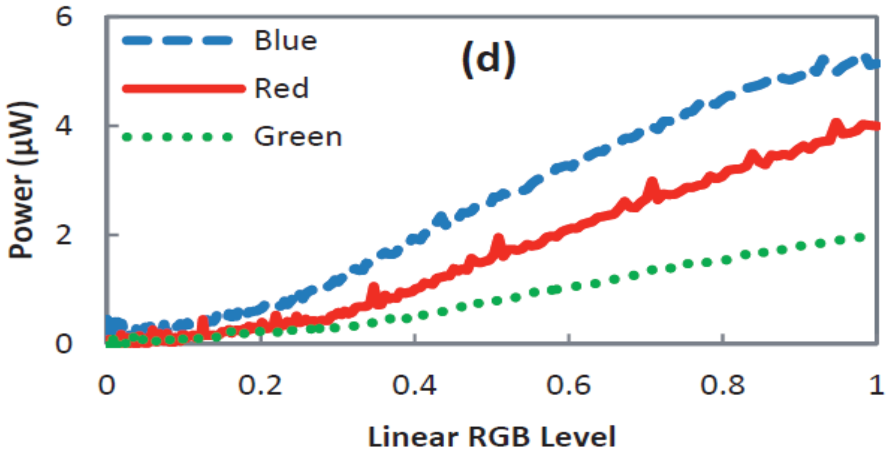
AC in Actuators: OLED Displays

OLED Displays:

- Brighter and better viewing angles w.r.t. LCDs
- Thinner and/or flexible panels

OLEDs are *emissive*:

- **Power strongly depends on pixels luminance and (secondarily) color**



- Power optimization can be achieved with an **image transformation!** (≠ LCD)

Motivation:

- Transformations for **general images** must preserve contrast while reducing power consumption.
- Existing solutions are **computationally intensive**.
  - **Power overhead?**
  - **Realtime applicability?**

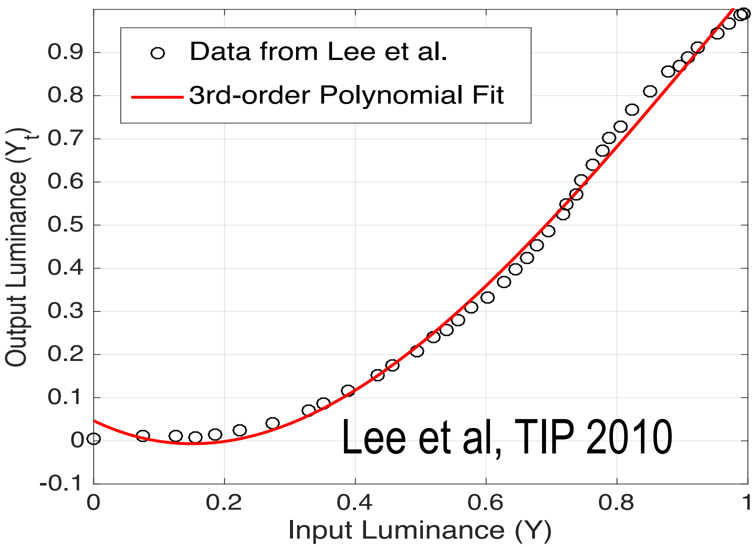
Claim:

- Similar transformations can be obtained by much simpler (**approximate**) computations.

AC in Actuators: OLED Displays

3rd Order Polynomial Fit:

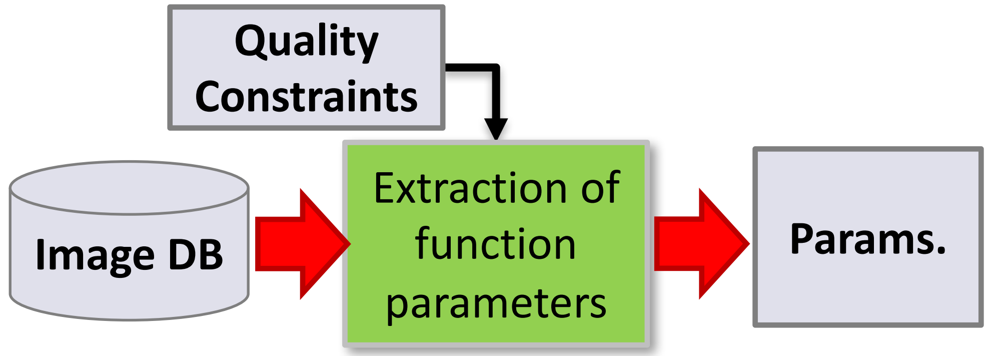
- Transform images according to a 3rd order **polynomial function of the input luminance** (YUV space)



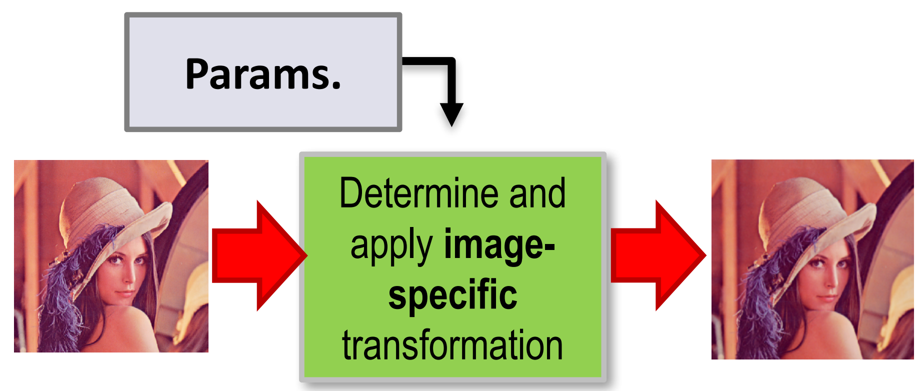
- Polynomial evaluation vs. histogram processing, etc.
- **Simpler and fewer operations** (ADD, MULT)

Approach:

1. **Offline Training Phase** (Computationally Intensive):



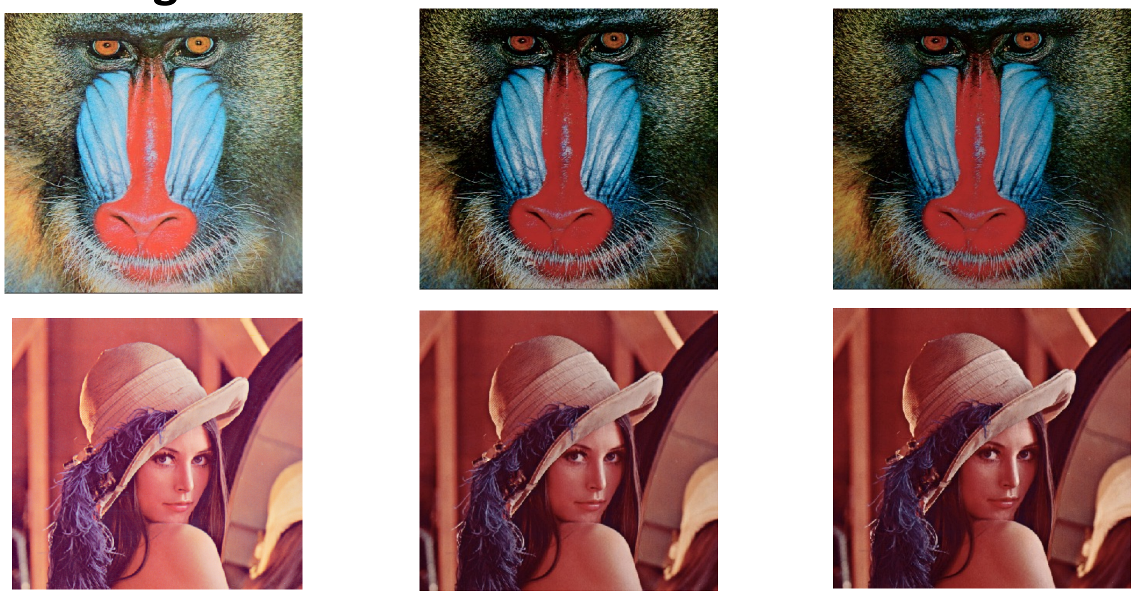
2. **Online Transformation** (Linear Complexity):



AC in Actuators: Results

- **Comparable quality at iso-savings w.r.t. state-of-the-art**
  - Visually and quantitatively:

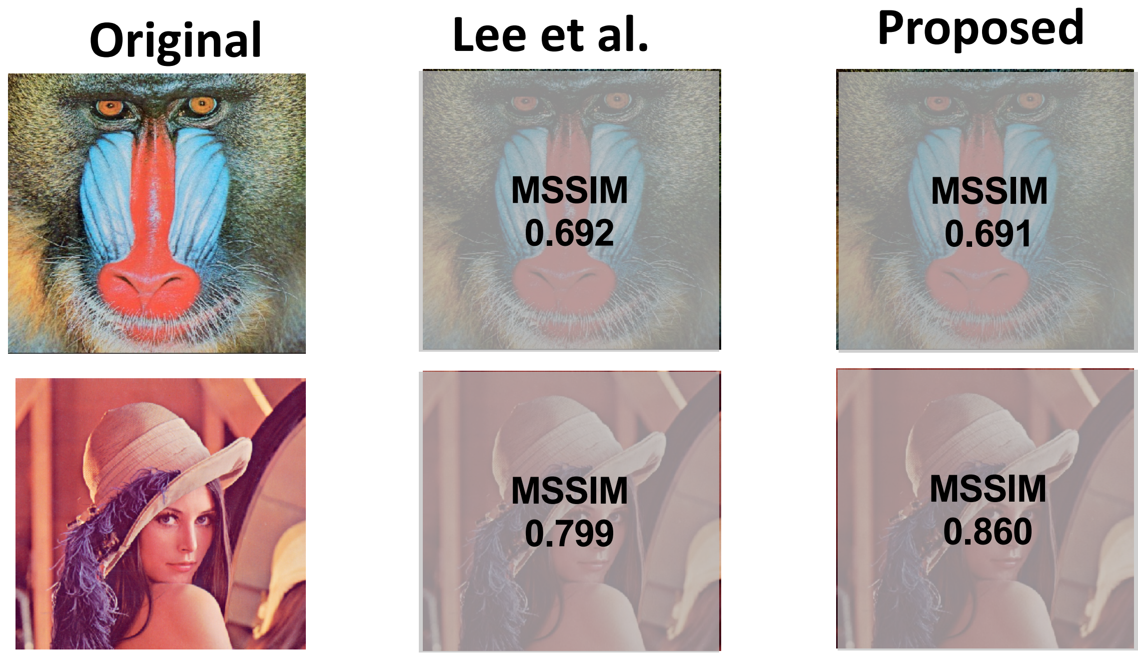
Original | **Lee et al.** | **Proposed**



- **Much lower complexity!** (SW or HW)
  - **10x faster** than Lee et al.
  - **Minimal power overhead** for HW implementation.

AC in Actuators: Results

- **Comparable quality at iso-savings w.r.t. state-of-the-art**
  - Visually and quantitatively:

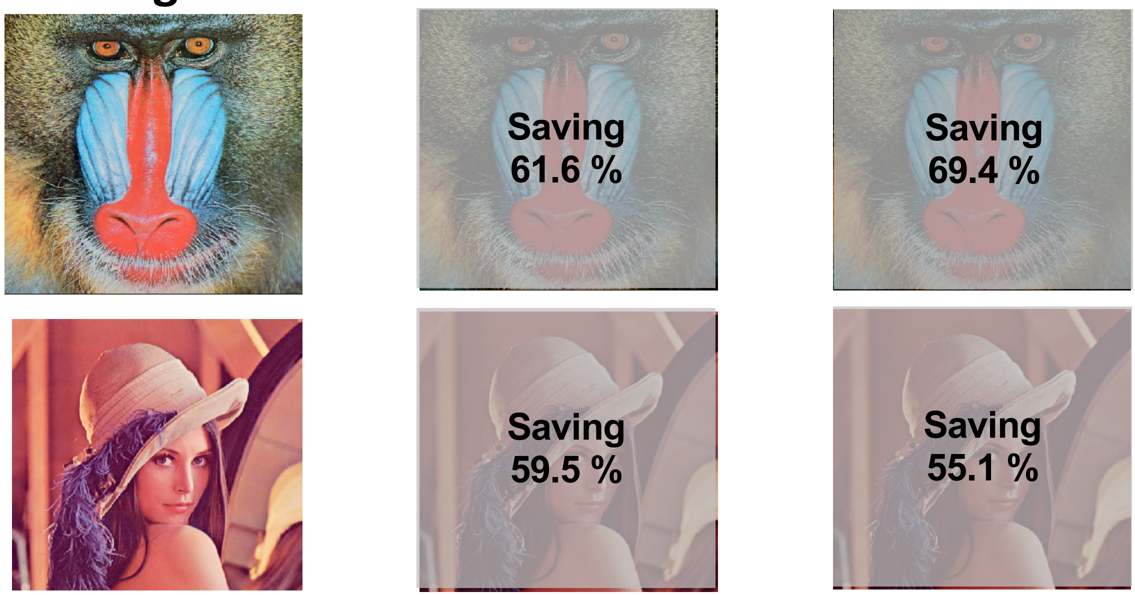


- **Much lower complexity!** (SW or HW)
  - **10x faster** than Lee et al.
  - **Minimal power overhead** for HW implementation.

AC in Actuators: Results

- **Comparable quality at iso-savings w.r.t. state-of-the-art**
  - Visually and quantitatively:

Original **Lee et al.** **Proposed**



- **Much lower complexity!** (SW or HW)
  - **10x faster** than Lee et al.
  - **Minimal power overhead** for HW implementation.

Conclusions

- Exploring the **energy versus quality** tradeoff can be interesting at **system level:**
  - The **computation part** is not always the one to blame.
- **Automation** aspects are key to the widespread diffusion of these design techniques.
- **Open Issues/Future Work:**
  - AC in **memories**?
  - AC in **sensors** (and/or, ADCs)?
  - How to **combine** AC techniques in different parts of the system to maximize **total power savings**?
    - *(e.g., encoding + RPR)*

THANK YOU !