ACCURACY-ENERGY TRADE-OFF WITH DYNAMIC ADEQUATE OPERATORS

MPSoC 2017 | Anca Molnos | 06/07/2017

OVERVIEW

- **Context: adequate/approximate computing**
- **Hardware**
  - Design methodology for dynamic–accuracy operators
- **Software use-case**
  - Trading sharpness with energy consumption in an autofocus application
- **Conclusions**

Many emerging applications are error tolerant (or error resilient)

- **E.g., human perception, noisy input, recognition/classification algorithms**

- **Opportunity: increased performance or reduced energy consumption without significant impact in the quality of result**

  - Probabilistic computing, imprecise computing, approximate computing, …
  - **Adequate computing** = run-time accuracy-energy trade-off
    - Accommodate time-varying error tolerance
  - This idea may be applied at various levels: algorithm, data encoding, arithmetic precision, **operator implementation**, …

STATE-OF-THE-ART

1. **Approximate/inexact circuits:**

   - Mostly adders and multipliers
   - *Kyaw, Goh and Yeo, EDSSC'10, Huang, Lach and Robins, DAC'12, Farshchi, Saeed and Fakhraie, CADS'13, Jiang, Han and Lombardi, GLSVLSI'15, Bhardwaj, Mane and Henkel, Int'l Symp. on QED'15, A Lingamneni, et al. TECS'13, etc.*

2. **Approximate synthesis:**

   - Generalization of the previous techniques to any netlist
   - *Shin and Gupta, ATS'08, Venkataramani et al, DAC'12, Miao, Gerstlauer and Orshansky, ICCAD'13, etc.*

3. **Quality-configurable circuit architectures:**

   - Voltage over-scaling + error correction
   - Adders, multipliers, etc.
   - *De la Guya Solaz, Han, Conway, IEEE Trans. On CAS'11, Kahng and Kang, DAC'12, Ye et al, ICCAD'13, Liu, Han and Lombardi, DATE'14, etc.*
   - Voltage scalable meta-functions.
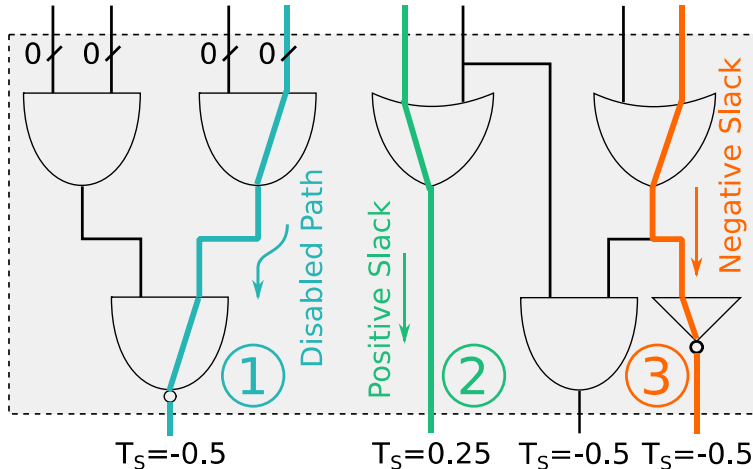   - *Mohapatra, Chippa, Raghunathan and Roy, DATE'11*

4. **Dynamic Voltage and Accuracy Scaling (DVAS):**

   - Use *technological knobs* only (no design modifications)
   - Reduce bit-width of input operands (LSB bits to 0) and voltage down-scaling
   - *Moons and Verhelst, ISLPED'15, etc.*

DYNAMIC ACCURACY OPERATORS

- **Limitation of DVAS:**
  - In standard ASIC implementation flow: most timing paths have a delay close to the critical one
  - Lower VDD → the entire operator is slowed-down.

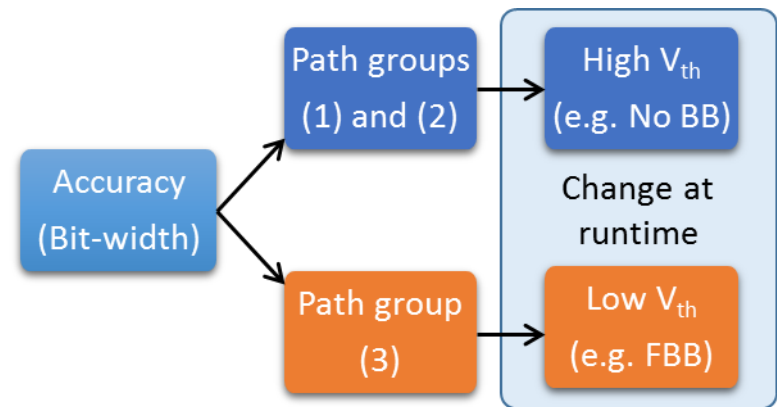- **One possibility: multi-VDD**
 - Requires level shifters
  - Excessive power overhead for an FU
- **Alternative:**
  - **Combine DVAS with FDSOI's back bias scaling**.
  - Fine-grain threshold voltage (Vth) tuning

$T_S=-0.5$ $T_S=0.25$ $T_S=-0.5$ $T_S=-0.5$

Need fine-grain power/delay tuning!

D. Pagliari et al. "A Methodology for the Design of Dynamic Accuracy Operators by Runtime Back Bias", DATE 2017

CIRCUIT PARTITIONING IN V_{BB} DOMAINS

- **Proposed partitioning:** regular tiling
  - Start from placed design
  - Divide the area in N identical V_{BB} domains
  - Assign each cell to the "closest" domain

Original placement

Insulation Trench

Placement with V_{BB} Domains
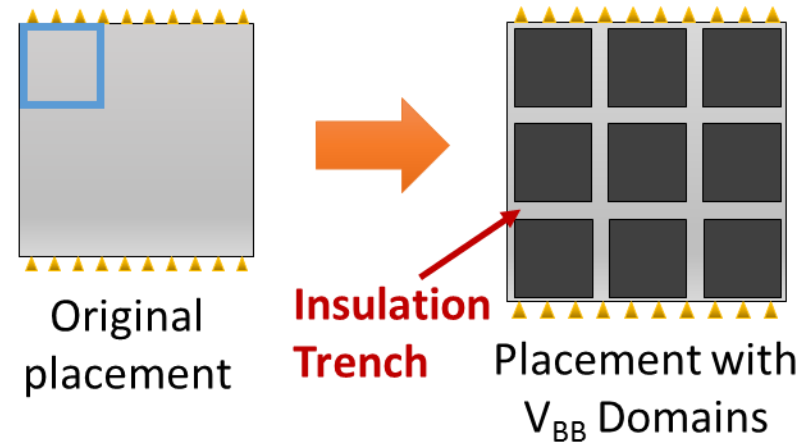
- **Pros:**
  - Regularity of design
  - Easy to incorporate in an EDA flow
  - Minimal displacement of cells w.r.t. a placement with standard constraints
    → Minimal timing, area and power overheads at maximum bit-width.

- **Cons:**
  - For a given accuracy, if a single domain contains cells that require different Vth values, all cells must receive the lowest Vth
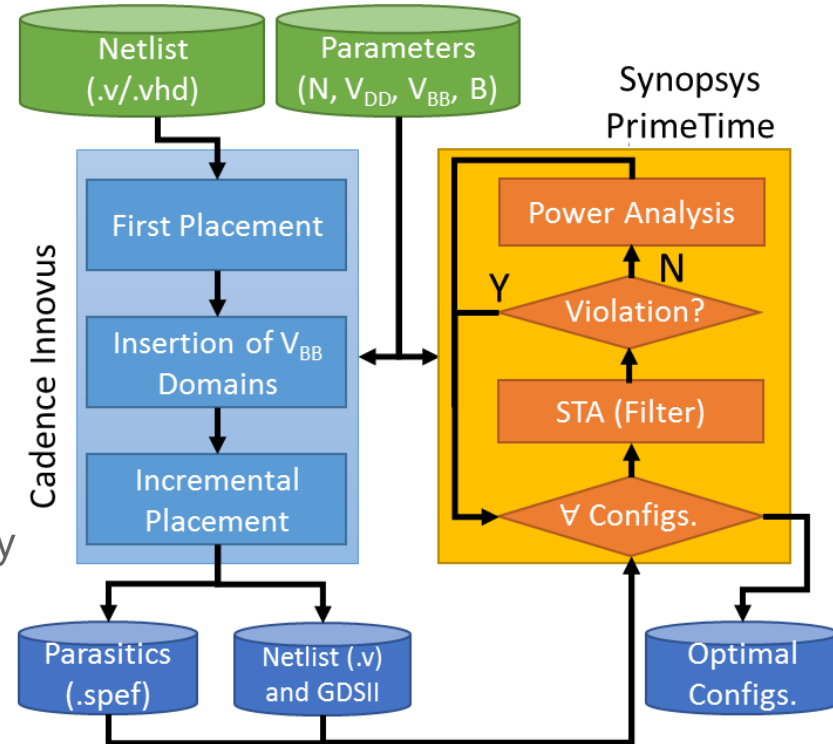 - Leakage overhead!

- ## Implementation Phase
  - Partition circuit in VBB domains using regular tiling.
  - Incremental placement:
    - Insert well-taps
   - Fix possible constraints violations in nominal operating conditions due to cell displacement.

- ## Analysis&Optimization Phase
  - Exhaustive exploration of all possible configurations of Accuracy, V_{BB}, and possibly global V_{DD}
  - Static Timing Analysis (STA) to prune unfeasible configurations (timing violations)
  - Power analysis on feasible configurations
 - Many configurations (thousands), but fast analysis. Feasible for < 10-15 V_{BB} domains

- **Implemented several operators circuits (16-bit fixed-point)**
  - Booth multiplier, FFT butterfly unit, 30-tap FIR filter
  - Technology: 28nm UTBB FDSOI from ST Microelectonics
- **Operating conditions:**
  - V_{DD} = [0.6V, 0.7V,…1.0V]
  - Forward BB: V_{BB} = ±1.1V (N-Well/P-Well)
  - Nominal condition (for 1st placement): 1.0V + FBB in all domains.

Booth multiplier
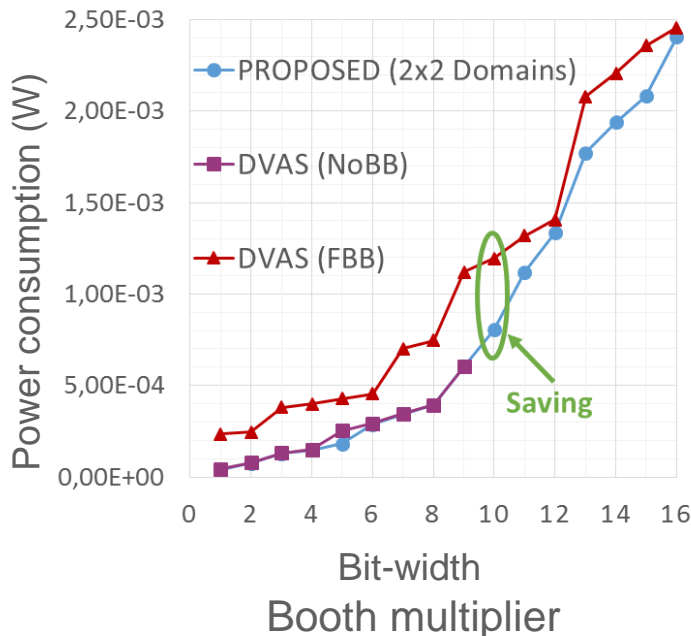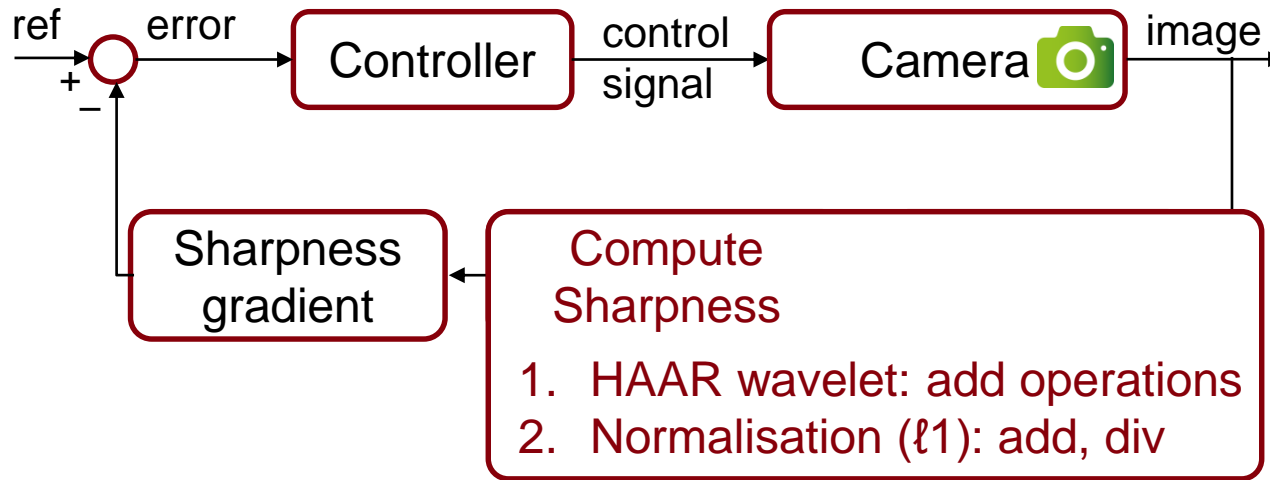
32.7% Saving w.r.t. DVAS @ 10-bit accuracy
Area overhead: 15%

FIR filter
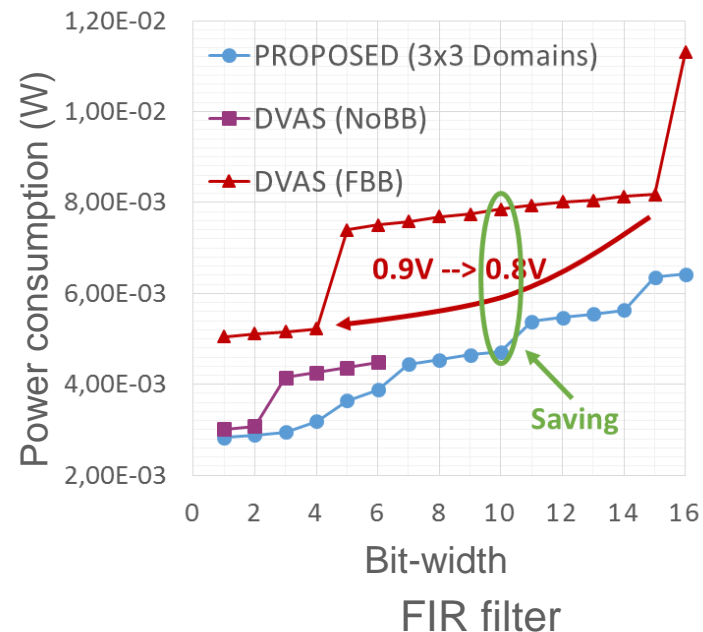
39.9% Saving w.r.t. DVAS @ 10-bit accuracy
Area overhead: 16%

OVERVIEW

- **Context: adequate/approximate computing**
- **Hardware**
  - Design methodology for dynamic–accuracy operators
- **Software use-case**
  - Trading sharpness with energy consumption in an autofocus application
- **Conclusions**

LENS AUTO-FOCUS USE CASE

- **Auto-focus control loop on a region of interest (ROI)**
  - adaptive, integral controller
  - sharpness → modified Haar-wavelet and normalisation
    - only the horizontal and vertical high-pass components of the transform are taken into the consideration
    - simple computation (many addition operations and few division operation)
 - *Zarudniev et al. NEWCAS'15*

ref → error → **Controller** → control signal → **Camera** 📷 → image

Sharpness gradient ← **Compute Sharpness**
1. HAAR wavelet: add operations
2. Normalisation (ℓ1): add, div

N. Gonthier et. al "Trading sharpness with energy consumption in a lens autofocus application", in Workshop on Approximate Computing, ESWEEK, 2016.

Study: impact of adequate addition operations

EXPERIMENTAL RESULTS

- **Baseline: precise additions, energy level = 1**
- **39 images, ROI with 256x256 pixels → 17K addition operations**
  - 16K in Haar transform
  - 1K in norm ℓ1
- **2 cases**
  - all operations (Haar wavelet and normalization) are approximate
  - additions in the Haar wavelet are approximate and the normalization is precise
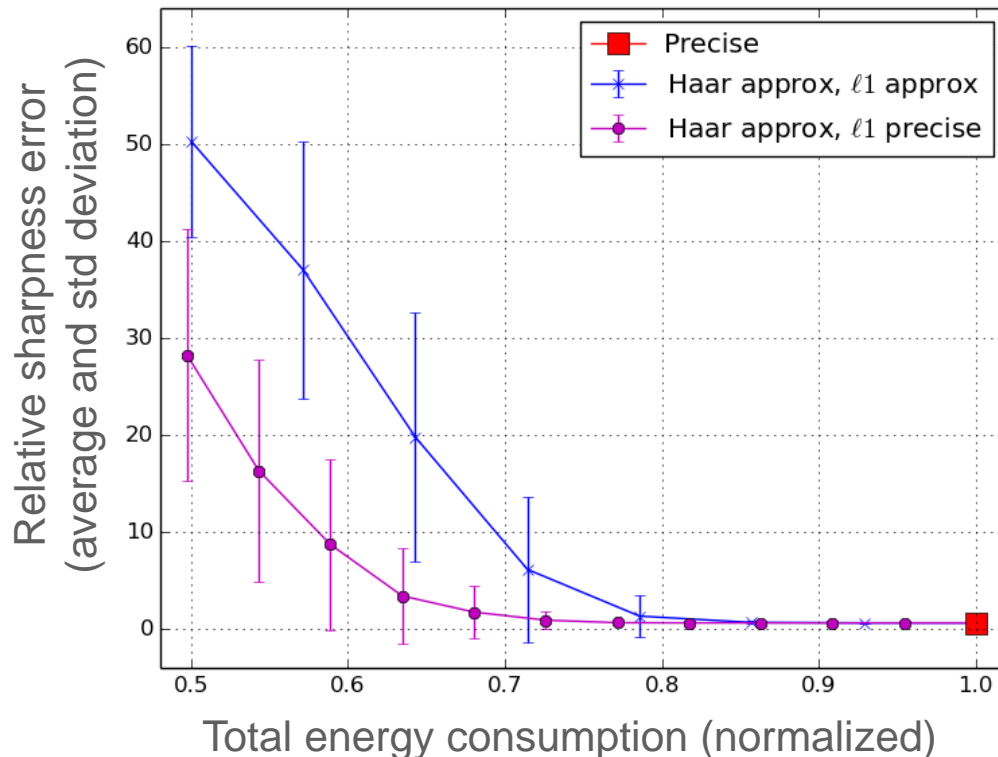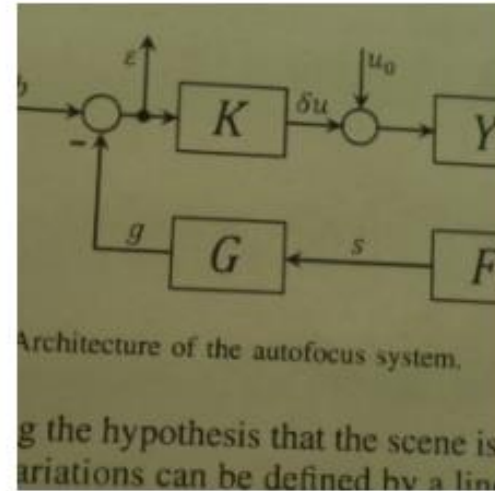

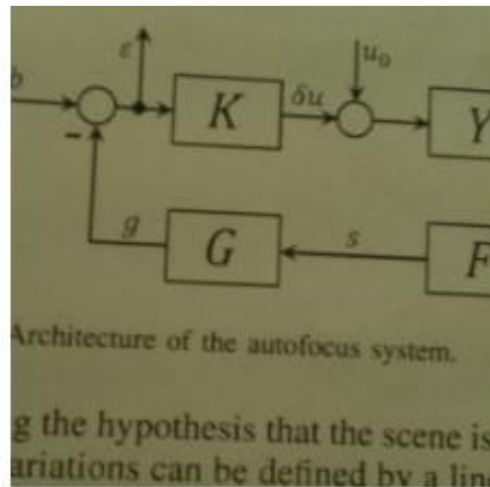30% reduction on energy → sharpness degradation of only 2%.
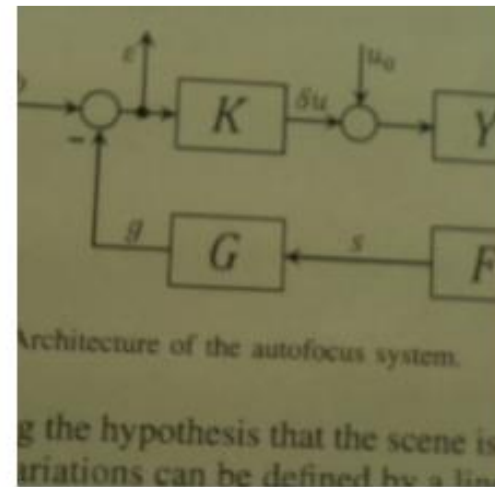
Maximum sharpness
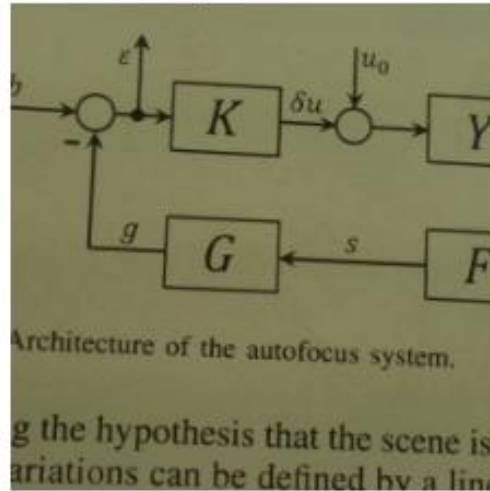
Relative sharpness error 10%

Relative sharpness error 25%

Relative sharpness error 50%

CONCLUSIONS

- **Summary and perspectives:**
  - Back-Bias is an effective knob for fine-grain delay/power tuning in quality-configurable functional units.
    - First ever application of Back-Biasing to quality-configurable systems (to our knowledge).
    - Easy to integrate with EDA flows
    - Many avenues for optimization: alternative partitioning techniques (irregular tiling), avoid exhaustive analysis of all V_{DD}, V_{BB}, and accuracy combinations, …
    - Compare with emerging multiple-precision fixed and floating point approaches
    - Approximate/adequate memory and storage
  - Energy gain potential for a small loss in final image sharpness in auto-focus use-case.
    - More applications to be investigated

- **Challenges at application level**
  - Analysis/programming/compiler support for adequate operations
    - Methods to find sensitive application parts
    - Analysis of error propagation for large applications
  - Algorithmic changes for better error-resilience

ACKNOWLEDGEMENTS

- **Yves Durand, CEA-Leti, France**
- **Edith Beigne, CEA-Leti, France**
- **Daniele Jahier Pagliari, Politecnico di Torino, Italy**
- **Massimo Poncino, Politecnico di Torino, Italy**

Thank you!

INSTITUT CARNOT
CEA LETI