Elastic Fidelity: Trading-off Computational Accuracy for Energy Reduction

Georgios Tziantzioulis, Ali Murat Gok, S M Faisal, Ke Liu, Sourya Roy, Tyler Clemons, Nikos Hardavellas, Seda Ogrenci Memik, Srinivasan Parthasarathy

**Trend: Increasing Power Dissipation**
- Power dissipation becoming an unmanageable problem
- Computing’s energy consumption (2010): 408 TWh
- Google datacenters’ energy needs: 1/3 nuclear plant
- Datacenters carbon footprint: ~Czech Republic

**Trend: Scaling Degrades Reliability**
- Aggressive technology scaling degrades reliability
- Fabrication variation causes error surge
- Future computers will be built by inherently unreliable components

**Thinking Beyond**
- Question: What if we relax the guarantees on reliability and intentionally allow components of the processor to fail sometimes with tolerable error rates?
  - We can gain power savings at quadratic rate with voltage reduction.
  - We have to accommodate these timing errors in the architecture and software layers.

**Elastic Fidelity Computing**
- Different code and data segments of an application exhibit variable sensitivity to errors.
- Language constructs declare reliability requirements of each code/data segment.
- Hardware steers computation to components and adjusts their voltage to reach the reliability target.
- A program may still appear to execute correctly if it returns acceptable results from the user’s perspective, even if there are inaccuracies in the computation.

**Simple Example**
- imprecise[25%] int a[n];
  - int b[n];
- Execution units (e.g., ALUs)
- Fidelity requirement translated into voltage level
- Execute error-tolerant computation on low-voltage ALU
- Store error-tolerant data in low-voltage storage

**Methodology**
- Software wrappers inject errors in computations at run time to simulate elastic-fidelity ALUs.
- Software wrappers inject errors in loaded data at run time to simulate elastic-fidelity storage.
- Calculate the output quality by comparing the results of error-free and error-injected runs.

**Summary**
- Computing is unsustainable (energy, environment)
- Computing devices will sustain massive errors
- Elastic Fidelity: exploit the inherent error-tolerance of applications to lower the energy consumption, and withstand the massive errors of future computing.
  - Allow some data to be imprecise
  - Programming language constructs and ISA extensions pass the fidelity requirements from application to hardware
  - Hardware models adjust the voltage to maintain fidelity guarantees
  - The execution system steers imprecise computations to components with low voltage
- Error-tolerant data are stored at low-voltage storage

**Power Consumption (JPEG-D)**

**Power & Energy Savings (JPEG-D)**
- Processor limit study @ 50% Vdd (Gem5/McPAT):
  - 14% dynamic, 27% static power savings
- Cache limit study @ 70% Vdd (CACTI):
  - 23% dynamic energy savings, 14% static power savings; even higher by foregoing unnecessary ECC

**McCormick**
Northwestern Engineering