Toward General-Purpose Code Acceleration with Analog Computation

Amir Yazdanbakhsh[o]

Bradley Thwaites[o]

Hadi Esmaeilzadeh[o]

Luis Ceze[*]

Renee St. Amant[+]

Jongse Park[o]

Arjang Hassibi[+]

Doug Burger[ζ]

[o] Georgia Institute of Technology

[+] The University of Texas at Austin

[*] University of Washington

[ζ] Microsoft Research

Approximate Computing (Hardware)

Truffle [ASPLOS 2012]

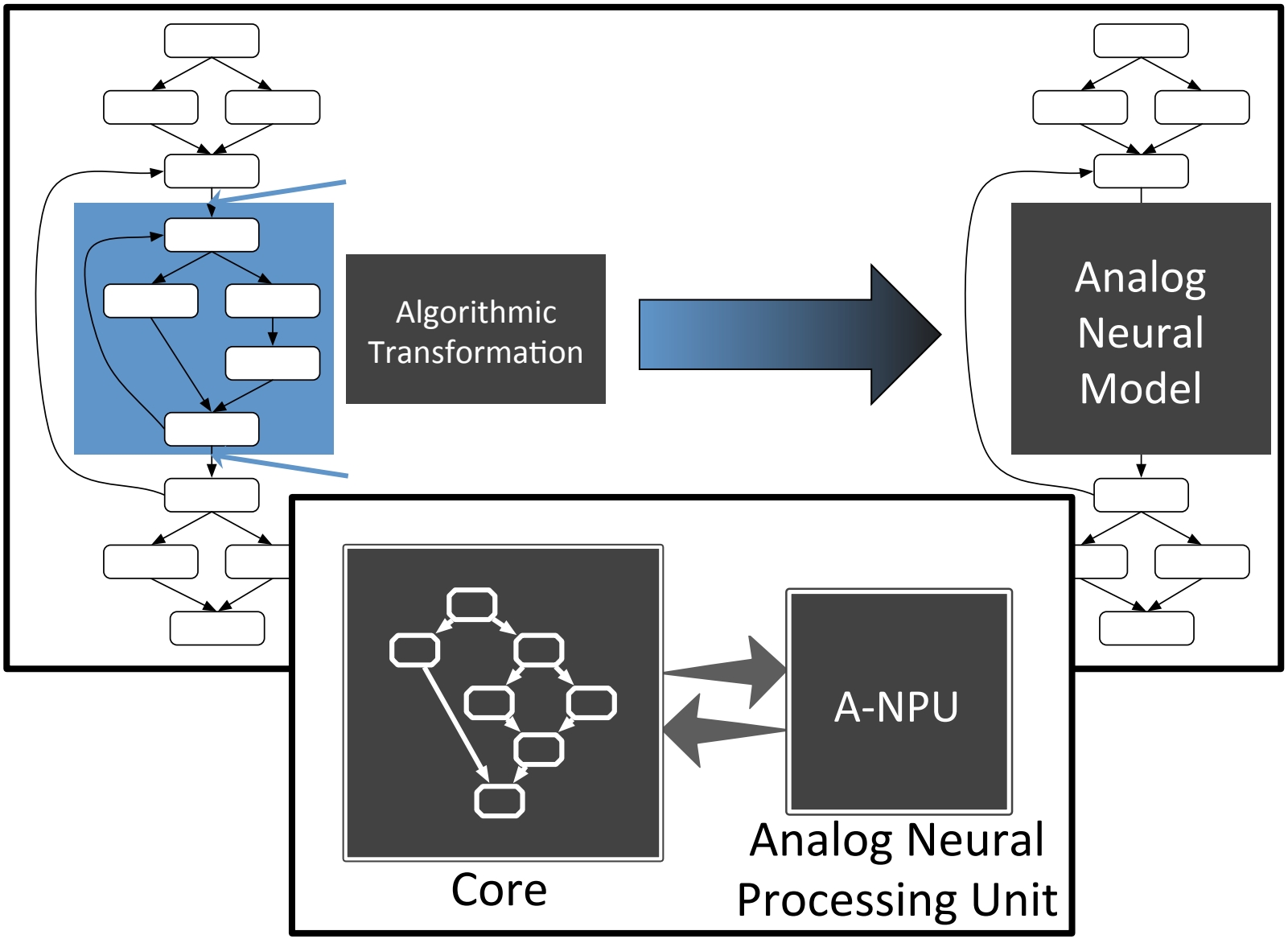
Probabilistic CMOS [Rice, Georgia Tech]

Stochastic processors [Illinois]

Flikker [ASPLOS 2011]

Execute approximable region(s) of the code on the 'Analog' circuit.

Neural algorithmic transformation



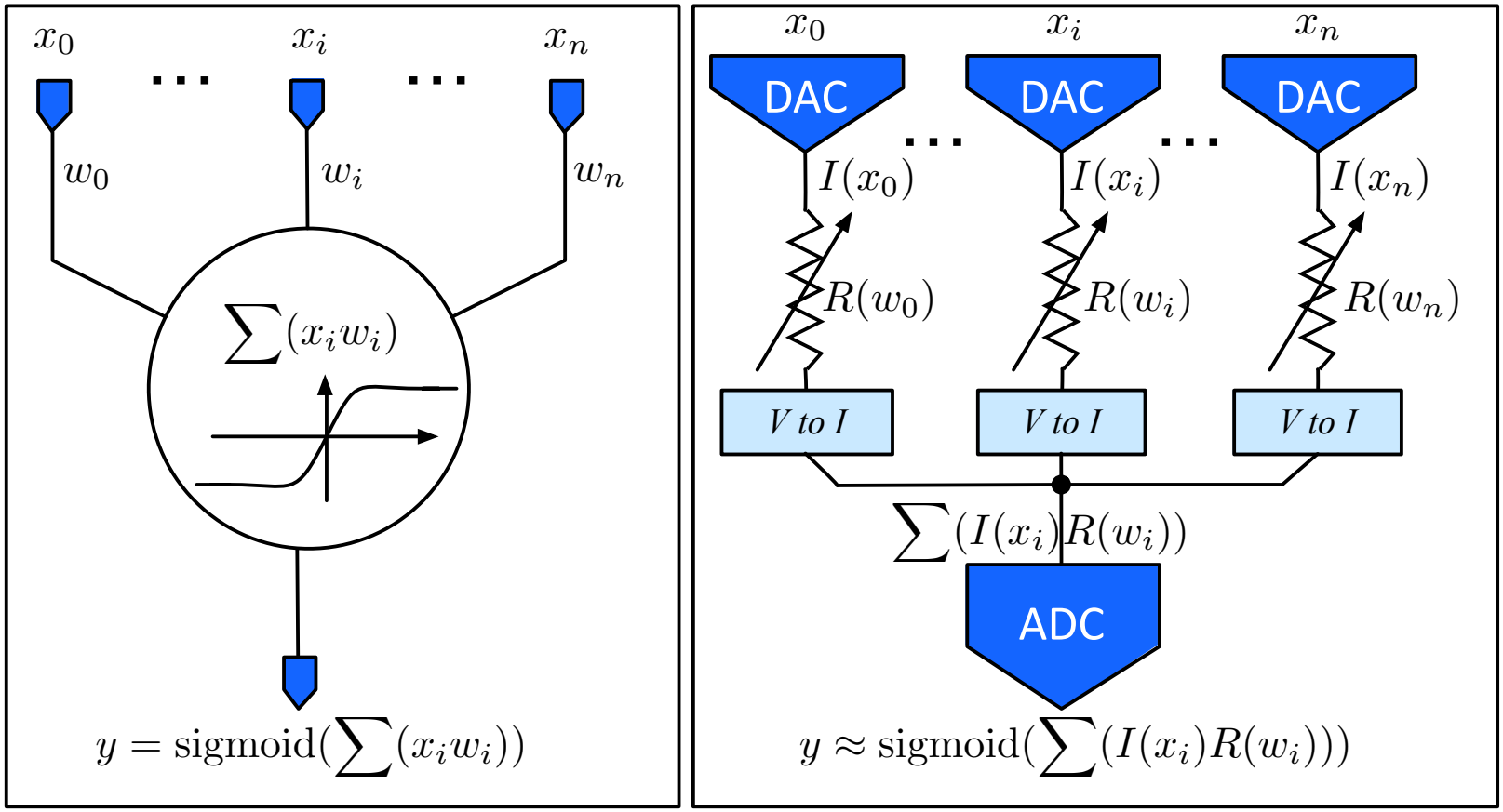
Programming Model

edgeDetection

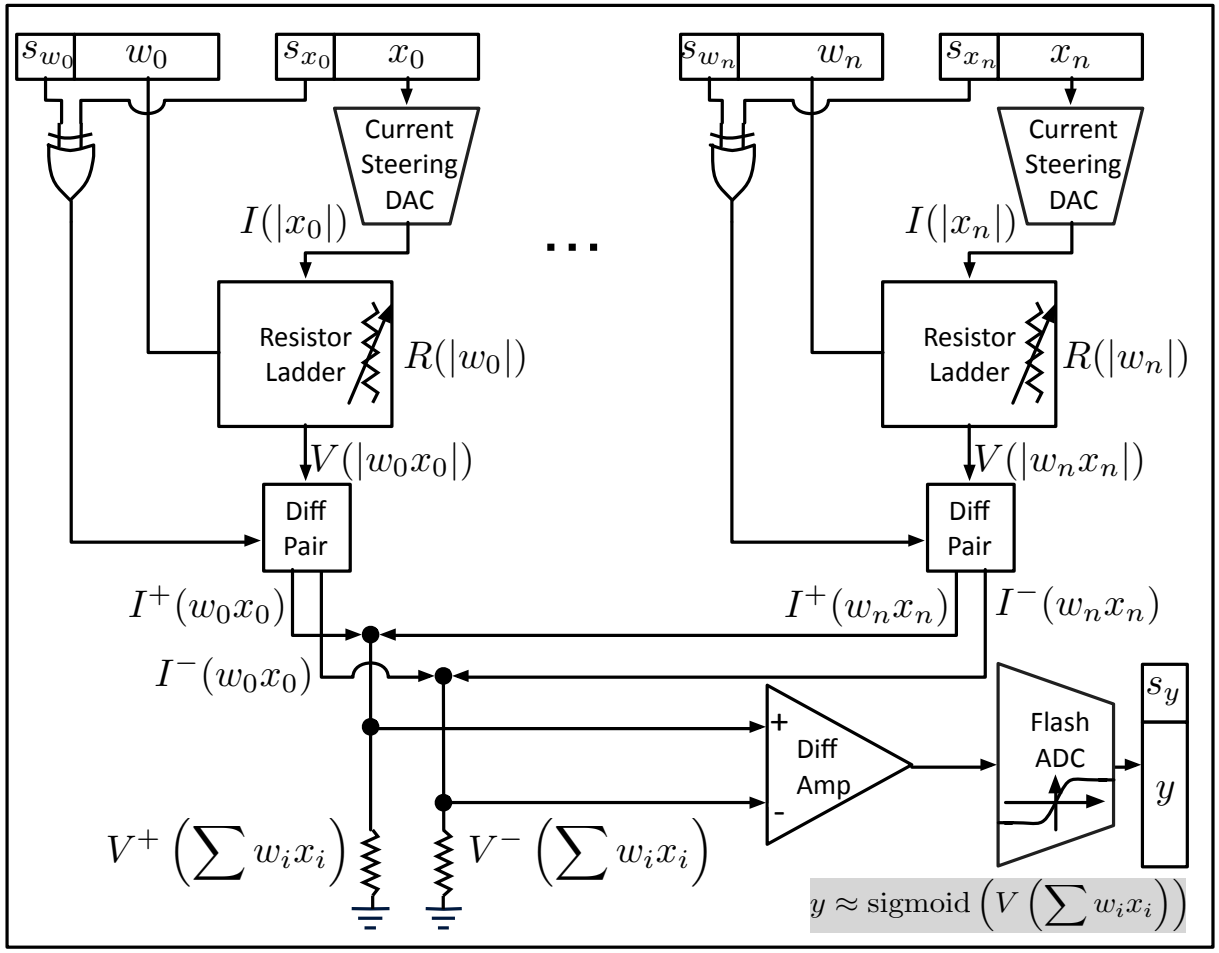
```
float grad [[candidate]] (float[3][3] p)
{
  …
}
```

```
void edgeDetection(
    Image &src, Image &dst) {

   grayscale(src);

   for (int y = …)
      for (int x = …) {
         dst[x][y] =
         grad(window(src, x, y));
      }
   }
```

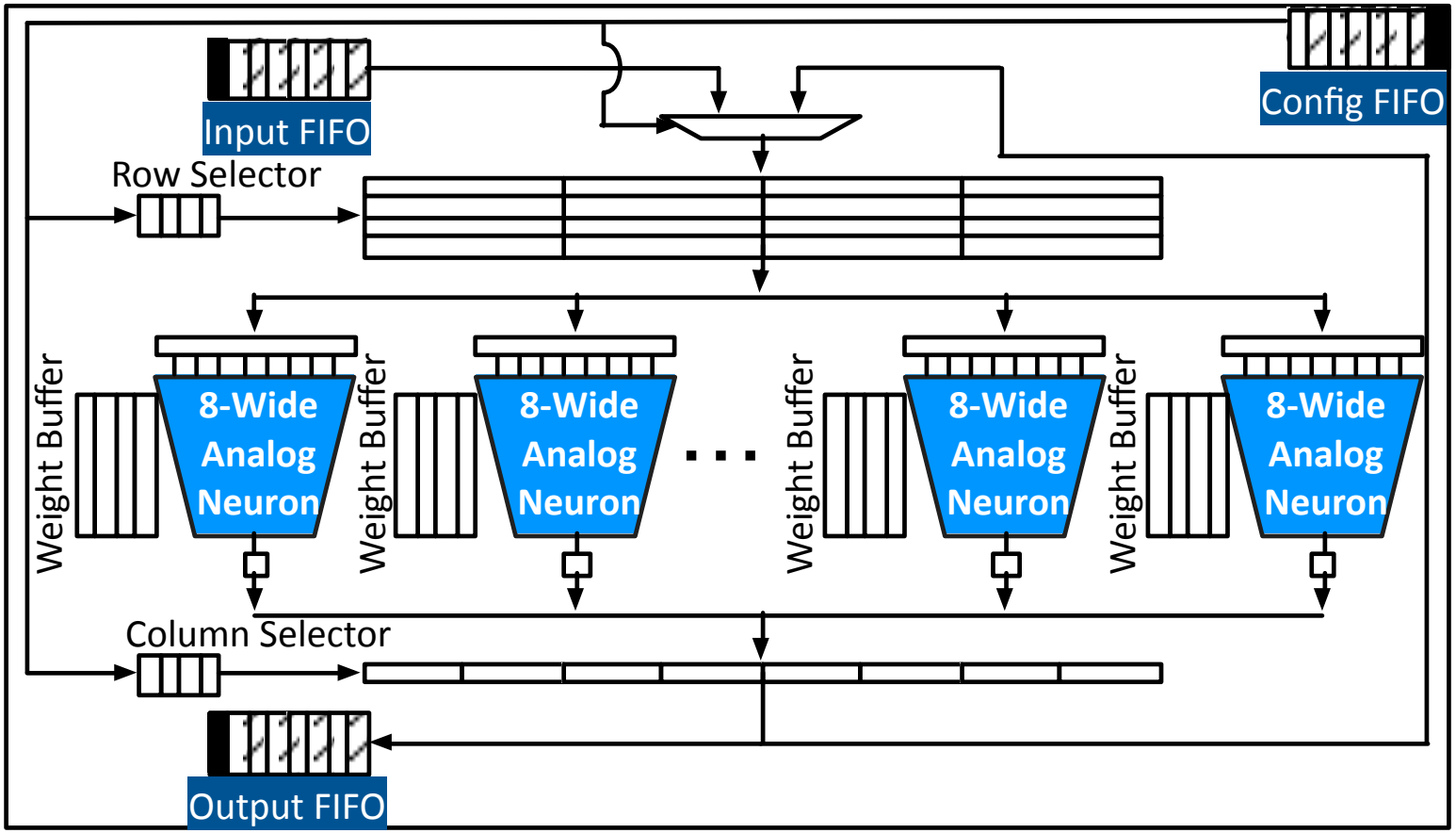
Neuron and its Conceptual Analog Circuit



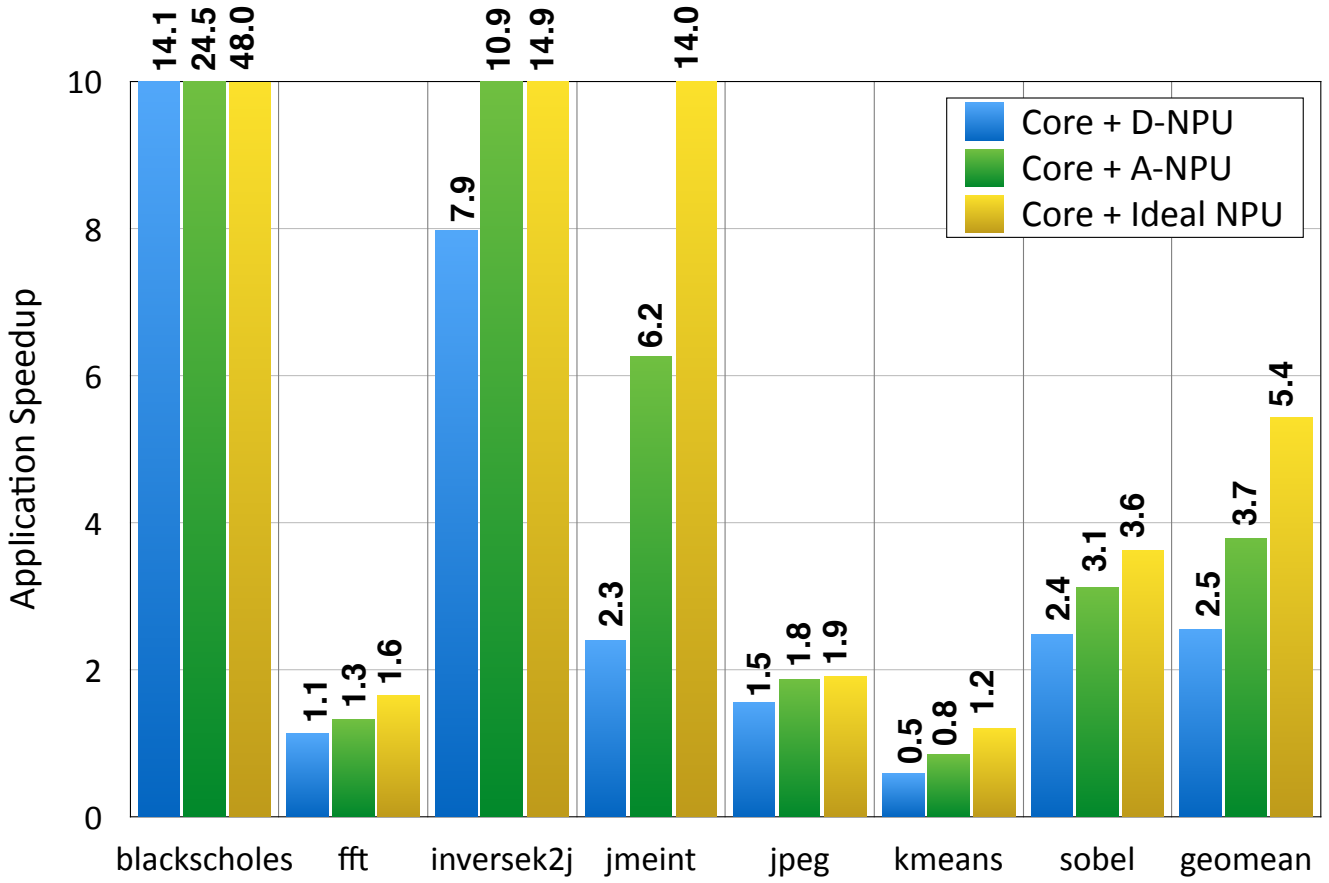
A Single Analog Neuron



Mixed-signal neural accelerator (A-NPU)



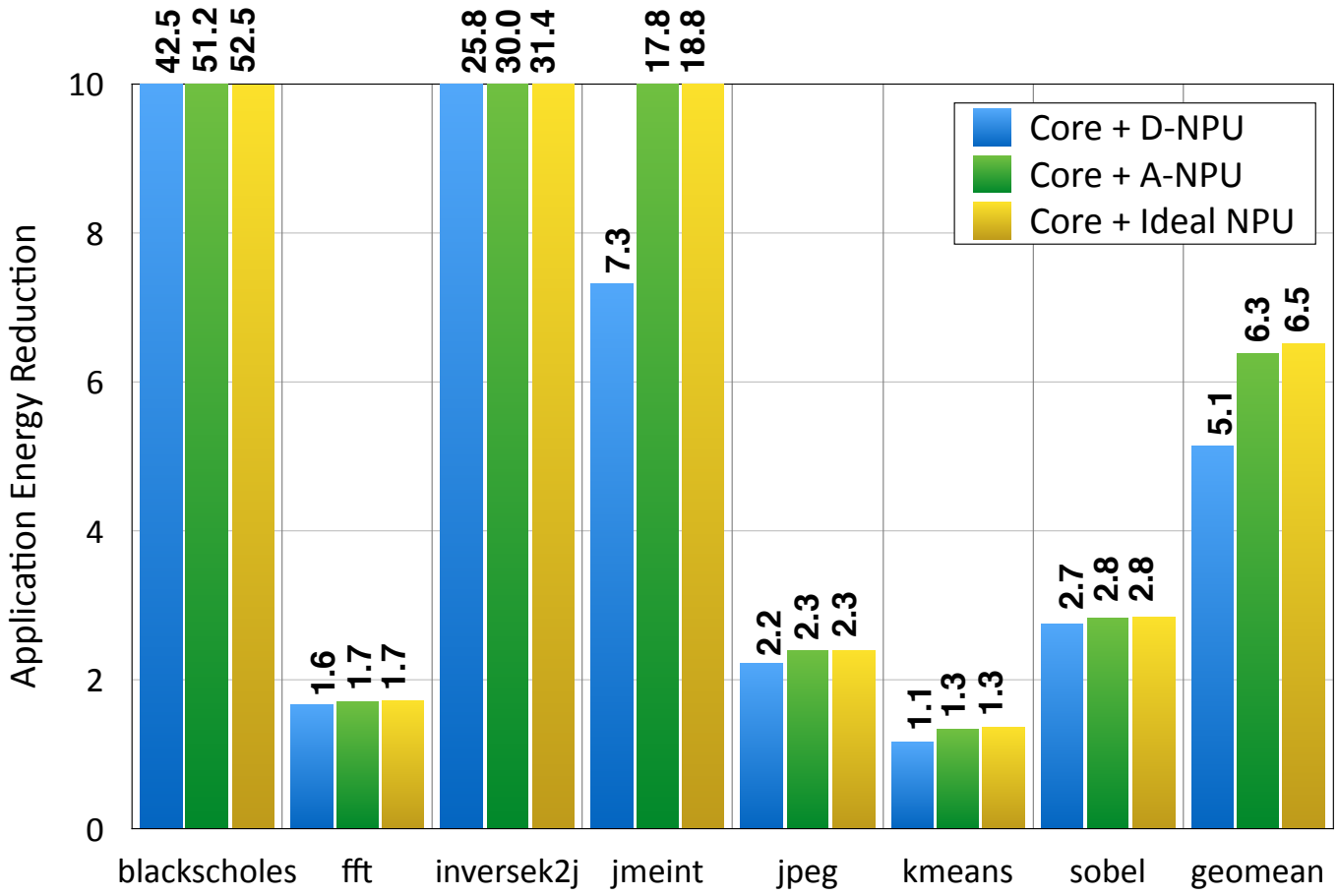
Speedup with A-NPU



~1.5x average speedup over D-NPU

Up to 24.5x speedup over all-CPU execution

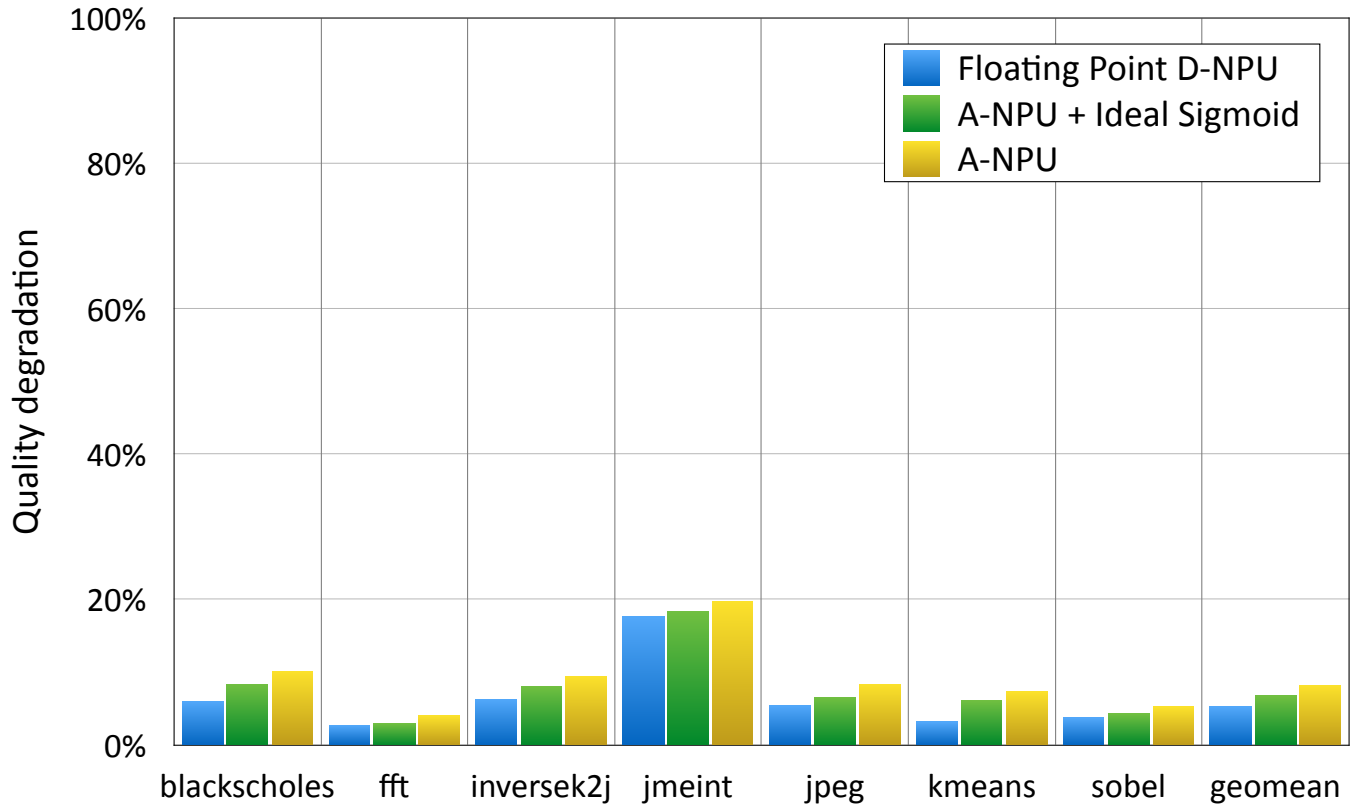
Energy saving with A-NPU



6.3x average energy reduction

Very close to the ideal NPU

Application quality degradation



Quality loss in all but one application is less than 10%

Customized quality metric for each application