Toward General-Purpose Code Acceleration with Analog Computation

Amir Yazdanbakhsh°
Bradley Thwaites°
Hadi Esmaeilzadeh°
Luis Ceze*

Renee St. Amant+
Jongse Park°
Arjang Hassibi+
Doug Burger†

° Georgia Institute of Technology
+ The University of Texas at Austin
* University of Washington
† 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

Algorithmic Transformation

Core

A-NPU

Analog Neural Processing Unit

Analog Neural Model
Programming Model

```c
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

\[
\begin{align*}
&\text{Current Steering DAC} \\
&\text{Resistor Ladder} \\
&\text{Diff Pair} \\
\end{align*}
\]

\[
\begin{align*}
I^+(w_0x_0) &\quad \ldots \\
I^-(w_0x_0) &\quad \ldots \\
V^+(\sum w_i x_i) &\quad \ldots \\
V^- (\sum w_i x_i) &\quad \ldots \\
\end{align*}
\]

\[
\begin{align*}
&\text{Flash ADC} \\
&y \approx \text{sigmoid} \left( V \left( \sum w_i x_i \right) \right) \\
&y \\
\end{align*}
\]
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
Quality loss in all but one application is less than 10%
Customized quality metric for each application