Neural Networks are "Hot Hot Hot!!!"

- Neural networks are layered structures consisting of neurons useful for:
  - Classification
  - Regression
  - Soft computing
  - They have broad use in high tech and big data, e.g., Google [1]
  - They also enable alternative computing methodologies
    - Automatic parallelization, e.g., ASC [2]
    - Approximate computing, e.g., NPU [3]

Our Vision of General-Purpose Neural Network Computing

1. Treat neural networks as a functional primitive backed by hardware acceleration [1]
2. Follow a transaction model for neural network computation
3. Exploit anticipated sharing of neural networks across applications
   - We think about neural acceleration like floating point acceleration
   - We break this into two distinct contributions:
     - X-FILES – Software/Hardware extensions for transaction management
     - DANA – One possible backend accelerator that interfaces with X-FILES [1]

Usage with a RISC-V Microprocessor

1. Grab a Rocket Chip RISC-V Microprocessor [1]
2. Build a RISC-V toolchain
3. Grab a copy of our X-FILES/DANA accelerator [2]
4. Build Rocket + X-FILES/DANA for FPGA
5. User processes can throw transactions at X-FILES

Open Source Availability and Acknowledgments

- On GitHub: github.com/bu-icsg/xfiles-dana
- This work was supported by the following:
  - A NASA Space Technology Research Fellowship
  - NSF CAREER awards

Table 1: Evaluated FANN-provided datasets

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>abelone</td>
<td>10 x 8 x 1</td>
<td></td>
</tr>
<tr>
<td>bank32f</td>
<td>32 x 16 x 1</td>
<td></td>
</tr>
<tr>
<td>bank32nh</td>
<td>32 x 16 x 1</td>
<td></td>
</tr>
<tr>
<td>building</td>
<td>14 x 8 x 3</td>
<td></td>
</tr>
<tr>
<td>kin32f</td>
<td>32 x 20 x 1</td>
<td></td>
</tr>
<tr>
<td>pumadyn_32f</td>
<td>32 x 16 x 8 x 4 x 1</td>
<td></td>
</tr>
<tr>
<td>robot</td>
<td>48 x 16 x 3</td>
<td></td>
</tr>
</tbody>
</table>

Classification
- diabetes 8 x 10 x 2
- gene 120 x 20 x 3
- soybean 82 x 12 x 19
- thyroid 21 x 10 x 3

Learning-on-chip using Fixed Point Arithmetic for Neural Network Accelerators

Schuyler Eldridge† (schuye@bu.edu) Marcia Sahaya Louis† Thomas Unger‡ Jonathan Appavoo‡ Ajay Joshi†
Boston University Departments of Electrical and Computer Engineering† and Computer Science‡