## Flowpro Machines & AI Hyper-Parallel, Fault Tolerant, Lowest Power

Today's Implementations of Artificial Neural Network (ANN)



**C**PU, GPU, FPGA or ASIC implementations predominantly use Synchronous, <u>Clocked</u>, architecture resulting in higher power consumption.

**C**PUs and GPUs are large, <u>complex</u> hardware regardless of the size of the ANN.

Parallelism is being applied to ANN's to speed up learning and classification but as parallelism increases, complexity in implementing these ANN's also increases. Compromises are made and optimal parallelism, and hence optimal speed, is sometimes not achieved. Clock frequency can be a limiting factor in these <u>Synchronous</u> systems.

A significant amount of software written for these systems is <u>Serial</u> and not easily <u>Parallelized</u> without a considerable rewrite.

Flowpro Machine Feed Forward Neural Network



Flowpro Machines are

<u>Clock-less</u> and parallel resulting in lowest power consumption and fault tolerance.

Flowpro Machines produce hardware <u>sized</u> for the ANN and are particularly well suited for ANN applications at the 'Edge" and support both CNN's and SNN's (Spiking Neural Networks).

**P**arallelism is a foundational quality of Flowpro Machines. Flowpro Machines execute in parallel natively in substrate or are multitasked when executing on Turing Machines. Both are equivalent parallel systems within limits set by a parameter. Flowpro Machines run as fast as they can. They are bound by processor speed on Turing Machines and propagation delay when executing in substrate as Clock-less <u>Asynchronous</u> or Ordered Machines.

**C**ompanion patent US 9,003,383 <u>Analytic Engine to Parallelize</u> <u>Serial Code</u> imports other code into Flowpro Machine Objects for execution as parallel Flowpro Machines. It has been shown that converting real time industrial controls at GM Powertrain to Flowpro Machines improves speed.



Convolutional Neural Networks (CNN) Learned Features

## Convolutional Neural Networks (CNN) Learned Features

Classifier & Control

## Executing Millions of Parallel Asynchronous Flowpro Machines (FMs) per chip is achievable, perhaps fostering A New AI?

Today's Machine Learning and Deep Learning systems work well and continue to incrementally improve. Implementing the current learning algorithms with Flowpro Machines will greatly reduce the power requirements of these same algorithms, probably increase the speed of inference and perhaps incrementally improve recognition probabilities? A parallel approach to AI using the Flowpro Machines has yet to be scaled, but with Chips approaching 30 billion transistors, hyper-parallel AI is possible. US patent #10,181,003 provides a parallel hierarchical technology to construct and asynchronously execute Action, Test, And Task transistor circuits directly from asynchronous flowchart Models called Flowpro Machines.

Best Estimate of the Number of Transistors Needed to Implement a Flowpro Machine (FM) Parallel Deep Learning 4 Layer CNN Chip

Total **FMs Needed** = 4K TV Pixels / (5x5) Convolutions x (5 FM-MACs ) + (((3 Convolution Layers))) + Classify & Control FMs Total FMs Needed ~= ((8,294,400 / 25) x 5) + (((69,025))) + (est. 5,000) ~= **1,732,905 Flowpro Machines** 

Total **Transistors Needed** = (Total FMs Needed x ave. 100 Blocks per FM x ave. 12 Transistors per Block ) x Design Cushion ~= (1,732,905 x 100 x 12) x 1.5 = 2,079,486,000 x 1.5 = **3,119,229,000 Transistors** 

Classifier