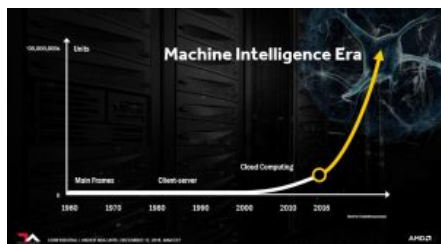


AMD gets into Machine Intelligence with "MI" range of hardware and software



Always good to have a share out of that curve.

In June we wrote on "[AMD is back!](#)", where this is one of the blog posts with more details in a specific direction. This post is about AMD specifically targeting machine learning with the MI (= Machine Intelligence) range of hardware and software.

With all the news around AMD's new processors Ryzen (CPU) and VEGA (GPU), it became apparent that AMD wants a good share of the Deep Learning market.

And they seem to succeed. Here is the current status.

Hardware: 25 TFLOPS @ 16-bit

Recently released have been the "Radeon Instinct" series, which purely focus on compute. How the new naming of AMD is organised will be discussed in a separate blog post.

For fast deep learning you need two things: extremely fast memory and lots of FLOPS at 16-bit. AMD happens to have developed HBM2, the world's fastest memory and now available to everybody. So AMD only needed to beat the NVIDIA P100 on FLOPS, and they did: the AMD "MI25" is expected to deliver around 25 TFLOPS for 16-bit operations. If you want to know more, lots of new links show up daily on [Google](#).

This means that AMD is beating NVIDIA's top-range GPUs again. Add NVlink-competitor [CCIX](#) and it's clear that AMD is a strong competitor again, as they used to. The only problem is that much of the software is written in CUDA...

Software: porting from CUDA

AMD's Greg Stoner, Director of Radeon Open Compute, opened up today on the current state of their software (typos fixed):

If you guys saw the Radeon Instinct launch you will find we finally announced our big push into Deep Learning. Here is good article <http://www.anandtech.com/show/10905/amd-announces-radeon-instinct-deep-learning-2017>

We will be delivery HIP version of Caffe, Tensorflow, Torch7, MxNet, Theano, CNTK, Chainer, all supporting our new MIOpen - our new Deep Learning solver.

Since the everyone is interested in Tensorflow

- HipEigen 35/43 GPU tests pass today.

Stream Executor integrated with hipFFT, hipRNG, hipDNN (MIOpen) + other interefaces
https://github.com/tensorflow/tensorflow/tree/master/tensorflow/stream_executor/cuda

All kernels in tensorflow/tensorflow/core/kernels/ have been ported

Note this will run on AMD and NVIDIA hardware

[\(source\)](#)

The status of Eigen is "35 out of 43", which is a rather vague description but an indication nevertheless. Eigen is a very important part of TensorFlow. A good promise that the code will be ready when the new VEGA hardware is launched.

Also interesting the the mention of MIOpen. It has been discussed on [TechReport](#):

This library offers a range of functions pre-optimized for execution on Radeon Instinct cards, like convolution, pooling, activation, normalization, and tensor operations. AMD says that convolution operations performed with MIOpen are nearly three times faster than those performed using the widely-used "general matrix multiplication" (GEMM) function from the standard [Basic Linear Algebra Subprograms specification](#). That speed-up is important because convolution operations [make up the majority of program run time for a convolutional neural network](#), according to Google TensorFlow team member Pete Warden.

The reason why they can deliver so many software-ports in such limited time with a small team, is because of [HIP](#). This makes it possible to port CUDA code to HIP, which runs on both AMD and NVIDIA.

We personally also had good experience with porting code to HIP. If you need CUDA code to be ported to AMD, know we tend to make the code faster and solve previously undiscovered bugs during the porting process.