

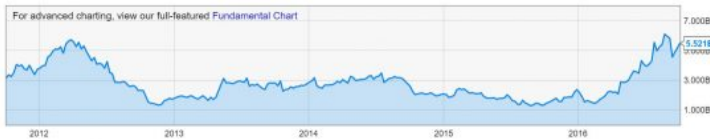
## AMD is back!



For years we haven't been complaining on this blog what AMD was lacking and what needed to be improved. And as you might have concluded from the title of this blogpost, there has been a lot of progress.

AMD is back! It will all come together in the beginning of 2017, but you'll see a lot of progress already the coming weeks and months.

AMD quietly recognised and solved various totally new problems in HPC, becoming the hidden innovator everybody needed. This blog is to give an overview of how AMD managed to come back and what it took to get to there. Their market cap supports it, as you can see.



AMD's market cap is back at 2012 levels ([source](#))

## Hardware

### HSA - mixing different flavours of silicon in a single chip, an AMD initiative



In 2006 AMD bought ATI and told the world they would integrate the GPU and CPU, under code-name "[Fusion](#)". Problem was that combining the two different worlds was a lot harder than anybody expected. From shared silicon to heating issues, the technology was certainly not ready back then.

AMD took lead to fix the problems that come in these very heterogeneous processors: [memory-sharing and task-dispatching](#). The HSA-capabilities can (soon) also be found on processors of ARM, Imagination Technologies, MediaTek, Qualcomm, Samsung and Texas Instruments. HSA-optimised software will therefore also result in performance improvements on non-AMD processors.

### HSA goes further than combining CPU and GPU

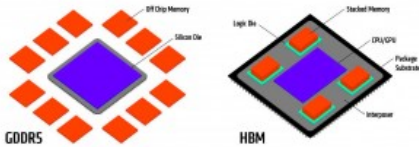
Nowadays X86 CPUs are full of specialised silicon like a h264-encoder/decoder. ARM processors are even more an exotic collection of various special-purpose silicon. With the increase of fabless, IP-selling design companies, the "exoticness" of

processors will only increase. Thanks to HSA it is possible to design special purpose silicon and have it integrate in processors of other vendors, as long as HSA design principles are used.

It comes to no surprise that HSA is getting a lot of recognition.

## HBM - reducing the memory bottleneck, an AMD invention

*HBM vs GDDR5:  
HBM shortens your information commute*



Isn't it interesting that the memories on all NVidia's GPU boards have been designed by AMD engineers? HBM is the next step up after GDDR. The bandwidth of HBM2 can reach up to 1 TB/s per GPU, using drastically less power.

The new memory is really different: it is asynchronous, smaller and faster. Just like it was a huge difference from OpenGL to Vulkan (evolved from Mantle, also an AMD invention). Read more on HBM [here](#).

## HSA-enabled high-performance APUs, an AMD product



See the specs of the upgraded PlayStation and Xbox and you'll see what the new APUs will deliver: around 5 TFLOPS by a single chip.

At the CPU side a lot has happened. You might have heard of Zen - the new CPU architecture that's coming soon. After over a decennium AMD finally leapfrogs Intel again - remember the Athlon, leapfrogged by Intel's Core architecture? What ofcourse also helps is no [illegal competition practises](#) by Intel for many years.

Upcoming APUs will be over 5 TFLOPS and thus directly competing with discrete GPUs.

## HSA-GPUs - the third generation AMD/ATI GPGPU



There was still a lot to do, to get from an programmable GPU to a real co-processor. This took three generations:

- The first GPUs capable of GPGPU used "VLIW", which were pretty hard to program.

The second generation used a scalar architecture, starting with the HD 7000 series and ending with Hawaii, called Graphics Core Next (GCN).

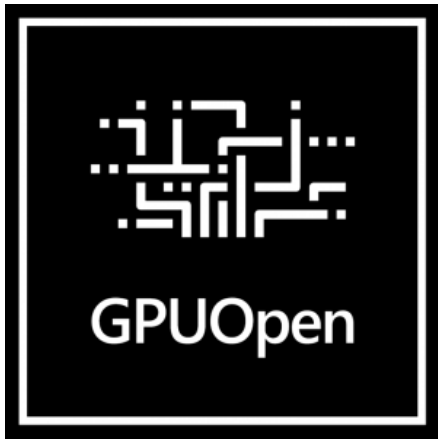
The third generation GPGPU, starting with Fiji, are HSA-capable GPUs with GDDR5 or HBM. This generation is now officially labelled "ROCm".

Starting with Radeon R9 Fury, Radeon R9 Nano and FirePro S9300x2, AMD's GPU architecture was evolved to have compute-performance, power-efficiency and HSA-capabilities in one package.

The upcoming Polaris brings down the costs - the RX 460 will cost around ?100 and the RX 490 ?300, and are to be launched in two weeks. The main competitor's GPUs cost double(!) the price.

## Software

### Open Source - enabling more possibilities



One thing that holds back innovation in broad, complex areas is closed source software. If the keeper of the software disagrees with directions, this will delay progress. Also a bug in a driver can be very costly, as only the keeper knows how to best write around it. For example doing something very different than deep learning on NVidia GPUs or VR on AMD GPUs.

AMD boldly decided that everything should be open source, as far patents allowed. This includes the (Linux) drivers, as described further below. This could mean that the hardware can follow the drivers instead of the other way around.

[Check out this list on GPUOpen.com](#) to find out what is open sourced for HPC. The list is pretty long

### New drivers built from the ground up



AMD has had a bad name if it comes to drivers. Drastic measures were needed and the complete driver-stack was built up from

scratch: Crimson.

The new driver is built on top of HSA, and thus needs at least Fiji. Because of that, the support for older hardware has been reduced to the most critical updates: "AMD Radeon R5 235X, Radeon R5 235, Radeon R5 230, Radeon? R5 220, Radeon HD 8470, Radeon HD 8350, Radeon HD 8000 (D/G variants), Radeon HD 7000 Series (HD 7600 and below), Radeon HD 6000 Series, and Radeon HD 5000 Series Graphics products have been moved to a legacy support model and no additional driver releases are planned". This will come as bad news for owners of that hardware. I understood this was necessary, to be able to fully focus on third-generation GPUs. I hope that HSAIL (3rd gen GPGPUs) can be converted to AMDIL (2nd gen GPGPUs) in the future, but that seems to be quite a task.

The good news is that new driver is much more stable and very performant on 3rd gen GPUs.

### Open source Linux driver

As said above, the Linux drivers are open source now. [You can find the Linux kernel driver on Github](#). For HPC (mostly Linux) this is very important, as bug fixing is under full control of the HPC software builders (like StreamHPC).

[Follow Phoronix for updated benchmarks of the new Linux driver](#).

## ROCm - easier programming of "exotic" hardware



There are two parts important here: HCC and HIP.

HCC is a C++ compiler inspired by C++AMP and C++14. It offers the following modes (taken from the GPUOpen-website):

- C++ AMP: Microsoft C++ AMP is a C++ accelerator API with support for GPU offload. This mode is compatible with version 1.2 of the C++ AMP specification.
- C++ Parallel STL: HCC provides an initial implementation of the parallel algorithms described in the ISO C++ Extensions for Parallelism, which enables parallel acceleration for certain STL algorithms.
- OpenMP: HCC supports OpenMP 3.1 on CPU. The support for OpenMP 4.x accelerator offloading is currently in development.

The HIP-tool converts CUDA code to HIP (with some restrictions), HIP to HSA (via HCC compiler) and HIP to NVidia PTX (via NVCC compiler). This way it's possible to indirectly run your CUDA-code on AMD FirePro (3rd gen GPGPUs).

I will write more on ROCm very soon, so leave it with this for now.

## AMD's return is important for the industry

Not an AMD fan? No worries, because it is also good for you.

Both Intel and NVidia have been asking for more and more money for their hardware over the years. NVidia now charges about [\\$10,000](#) for a GPU, and also Intel has slowly increased their CPU-prices. We can simply blame AMD for not democratising the industry for years.

Innovation also has been slower-paced due to lack of AMD's competitiveness - even though you might disagree here, you will agree that with AMD's return, it will bring about an increase in innovation.