"Project Denver"
Processor to Usher in a New Era of Computing

Bill Dally

January 5, 2011

Project Denver Announced

On January 5, 2011, NVIDIA announced that it is developing high-performance ARM-based CPUs designed to power future products ranging from personal computers to servers and supercomputers.

Known under the internal codename “Project Denver,” this initiative features an NVIDIA CPU running the ARM instruction set, which will be fully integrated on the same chip as the NVIDIA GPU.

This initiative is extremely important for NVIDIA and the computing industry for several reasons.
Heterogeneous Computing Platform

NVIDIA’s Project Denver will usher in a new era for computing by extending the performance range of the ARM instruction-set architecture, enabling the ARM architecture to cover a larger portion of the computing space.

Coupled with an NVIDIA GPU, it will provide the heterogeneous computing platform of the future by combining a standard architecture with awesome performance and energy efficiency.
Project Denver: High-end ARM CPU

ARM is already the standard architecture for mobile devices.

Project Denver extends the range of ARM systems upward to PCs, data center servers, and supercomputers.

ARM’s modern architecture, open business model, and vibrant eco-system have led to its pervasiveness in cell phones, tablets, and other embedded devices.

Denver is the catalyst that will enable these same factors to propel ARM to become pervasive in higher-end systems.
Denver frees PCs, workstations and servers from the hegemony and inefficiency of the x86 architecture.

For several years, makers of high-end computing platforms have had no choice about instruction-set architecture.

The only option was the x86 instruction set with variable-length instructions, a small register set, and other features that interfered with modern compiler optimizations, required a larger area for instruction decoding, and substantially reduced energy efficiency.
Choice

- Denver provides a choice.

- System builders can now choose a high-performance processor based on a RISC instruction set with modern features such as fixed-width instructions, predication, and a large general register file.

- These features enable advanced compiler techniques and simplify implementation, ultimately leading to higher performance and a more energy-efficient processor.
Windows on ARM

Microsoft’s announcement that it is bringing Windows to ultra-low power processors like ARM-based CPUs provides the final ingredient needed to enable ARM-based PCs based on Denver.

Along with software stacks based on Android, Symbian, and iOS, Windows for ultra-low power processors demonstrates the huge momentum behind low-power solutions that will ultimately propel the ARM architecture to dominance.
CPU+GPU

- An ARM processor coupled with an NVIDIA GPU represents the computing platform of the future.

- A high-performance CPU with a standard instruction set will run the serial parts of applications and provide compatibility while a highly-parallel, highly-efficient GPU will run the parallel portions of programs.
The result is that future systems – from the thinnest laptops to the biggest data centers, and everything in between — will deliver an outstanding combination of performance and power efficiency.

Their processors will provide the best of both worlds, while enabling increased battery life for mobile solutions.

We’re really excited to help engineer smarter brains for the next major era in computing.
FPGAs in a Denver World

Mike Butts

FPGA 2011 Pre-Conference Workshop 2/27/11: The Role of FPGAs in a Converged Future with Heterogeneous Programmable Processors
Scalable Silicon = GPUs and FPGAs

- Only massively parallel silicon scales with Moore’s Law
- GPUs and FPGAs have complimentary strengths

**GPUs**
- Processor parallelism
- Software programming
- Max TFLOPS, TF/W, TF/$
- Max memory bandwidth
- Max PCIe bandwidth

**FPGAs**
- Boolean parallelism
- Hardware programming
- Non-arithmetic datatypes
- Irregular fine-grain control
- 1000 programmable pins
PCI Express is the system connection for GPUs and CPUs
- PCIe Gen 3 has latency optimizations for accelerators

FPGA as sensor or comm front-end: 1000 programmable pins
- Medical ultrasound: 64-256 12-16-bit ADCs @ 50-100 MHz: 400 Gb/s
- FPGA integer beamforming front-end feeds GPU compute software

FPGA as accelerator: App-specific datatypes, fine irregular control
- Get good at low latency with PCIe Gen3