MATENVMED

Δράση 3.2

ΥΛΟΠΟΙΗΣΗ ΣΕ FPGAs ΚΑΙ ΠΟΛΥΠΥΡΗΝΑ ΣΥΣΤΗΜΑΤΑ (MULTICORES / MANYCORES)

Νίκος Μπέλλας

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ

Πανεπιστήμιο Θεσσαλίας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Deliverable 3.2

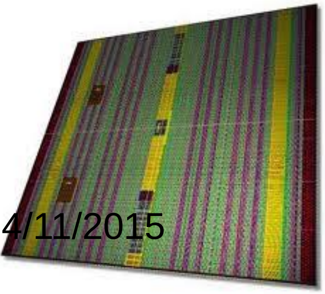
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

- Select and study computationally demanding kernels of WP2 that can potentially be accelerated when mapped to a massively parallel platform (GPUs, FPGAs).
- Design and implement appropriate tools or use third-party tools to help with mapping these kernels to such platforms (compilers, CAD tools)
- Map the kernels to FPGAs and GPUs
- Performance analysis and comparison wrt. to CPU implementation

What is an FPGA?

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

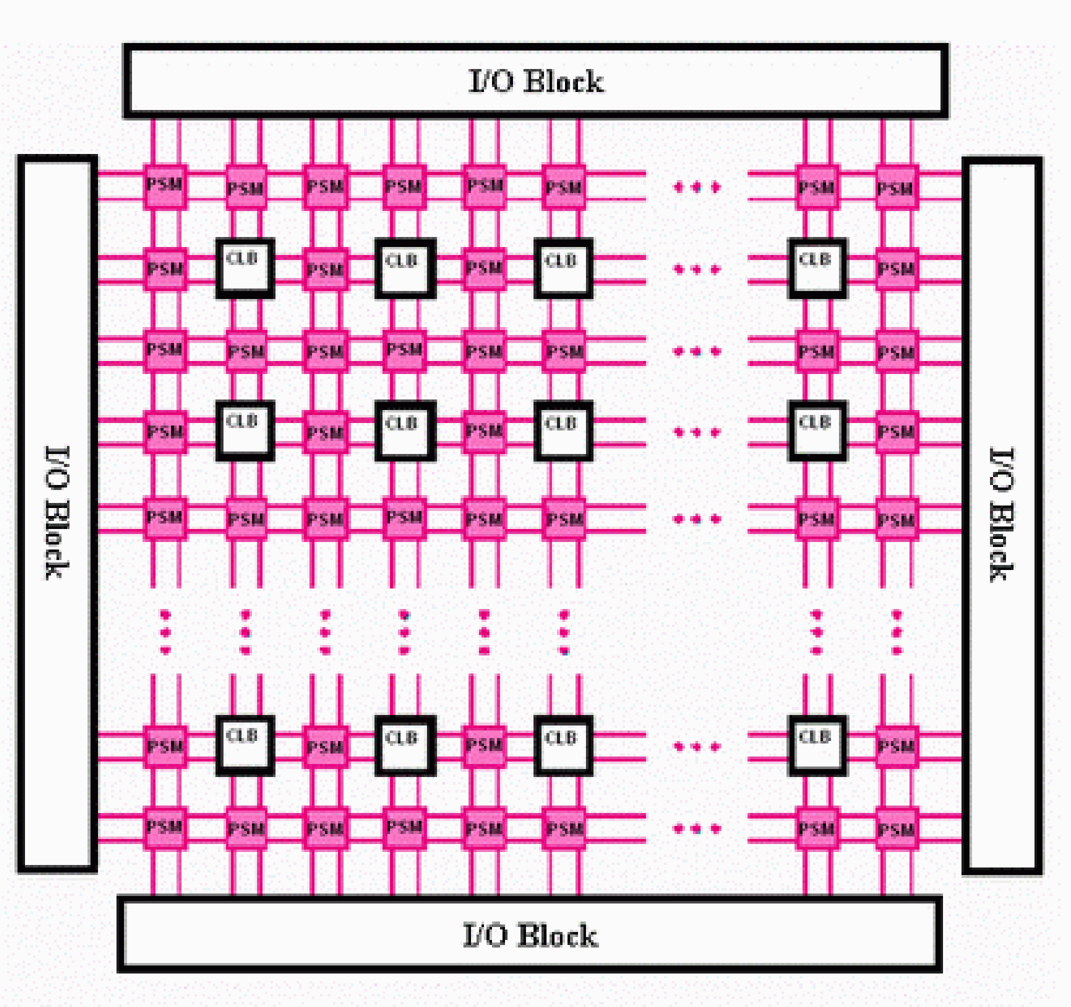
- *Field Programmable Gate Array (FPGA)* is the best known example of Reconfigurable Logic
- Hardware can be modified post chip fabrication
- Tailor the Hardware to the application
  - Fixed logic processors (CPUs/GPUs) only modify their software (via programming)
- FPGAs can offer superior performance, performance/power, or performance/cost compared to CPUs and GPUs.



FPGA architecture

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

- A generic island-style FPGA fabric
- Configurable Logic Blocks **(CLB)** and Programmable Switch Matrices **(PSM)**
- **Bitstream** configures functionality of each CLB and interconnection between logic blocks



FPGA discussion

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

- Advantages
  - Hardware tailored to application: potential for (near) optimal performance for a given application
  - Various forms of parallelisms can be exploited

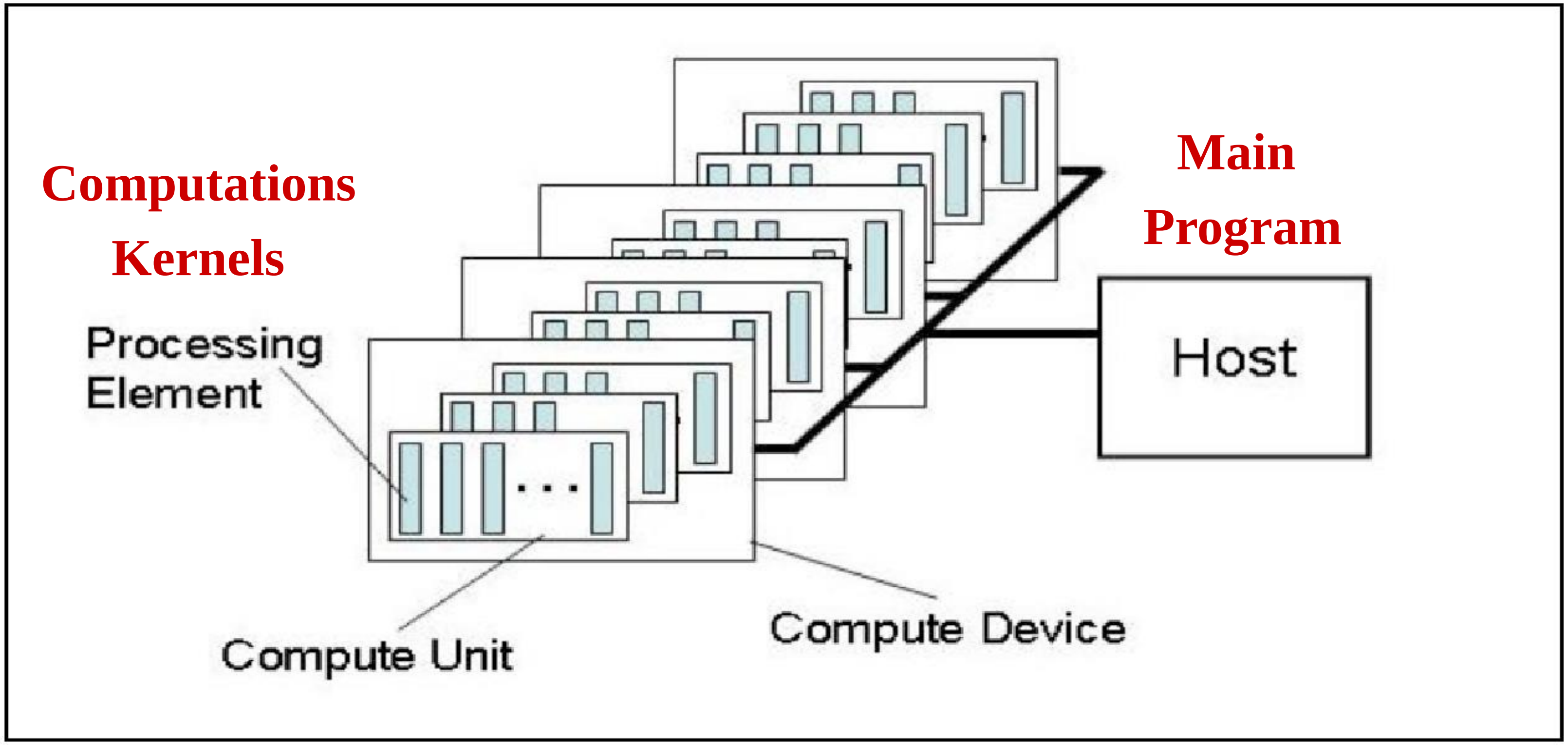
- Disadvantages
  - Programmable mainly at the hardware level using Hardware Description Languages (BUT, this can change)
  - Lower clock frequency (< 300 MHz) compared to CPUs (~ 3GHz) and GPUs (~1.5 GHz)

OpenCL for Heterogeneous Systems

- OpenCL (Open Computing Language) : A unified programming model aims at letting a programmer write a portable program once and deploy it on any heterogeneous system with CPUs and GPUs.

- Became an important industry standard after release due to substantial industry support.

OpenCL Platform Model



One host and one or more Compute Devices (CD)

Each CD consists of one or more Compute Units (CU)

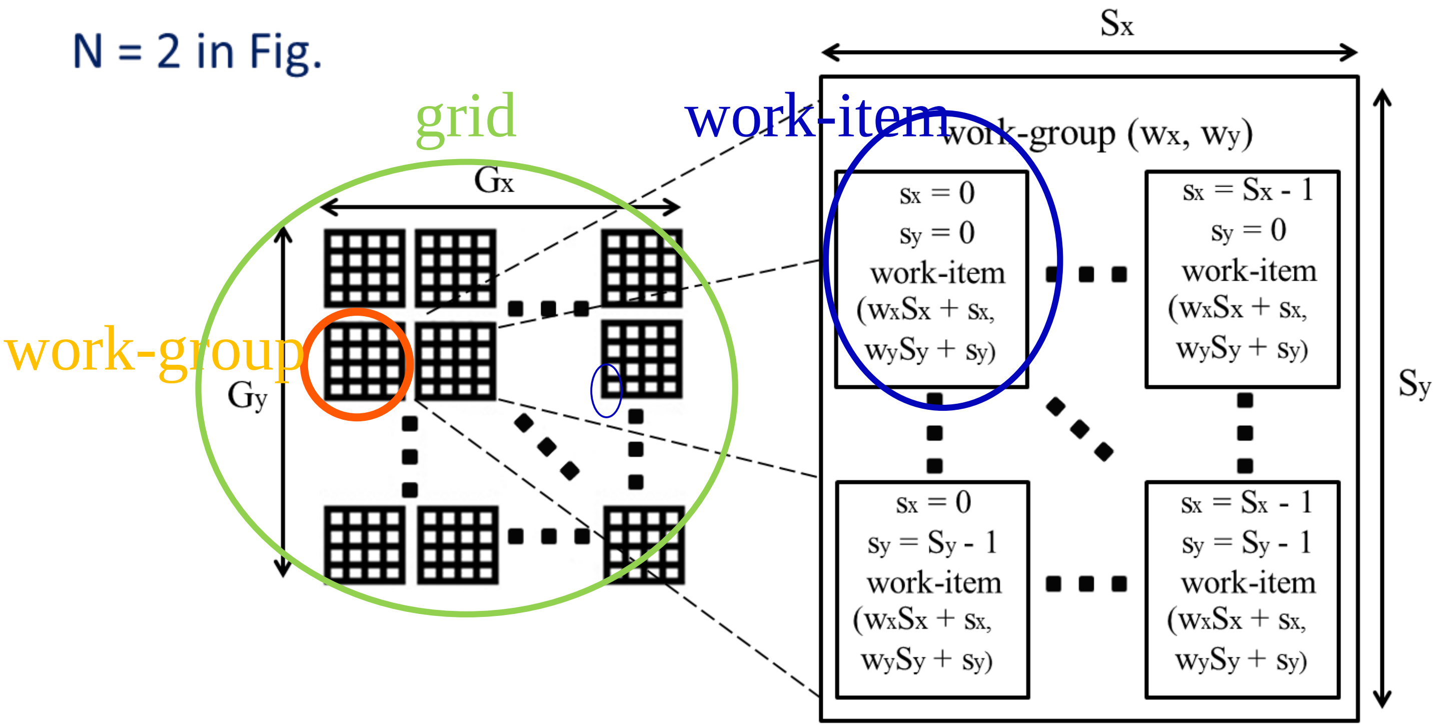
Each CU is further divided into one or more Processing Elements (PE)

OpenCL Kernel Execution Geometry

- OpenCL defines a geometric partitioning of grid of computations
- Grid consists of N dimensional space of work-groups
- Each work-group consists of N dimensional space of work-items.

$1 \leq N \leq 3$

N = 2 in Fig.



OpenCL Simple Example

- OpenCL kernel describes the computation of a work-item
  - Finest parallelism granularity
- e.g. add two integer vectors (N=1)

Run-time call

Used to differentiate execution

for each work-item

```
void add(int* a,
        int* b,
        int* c) {
for (int idx=0; idx<sizeof(a); idx++)
   c[idx] = a[idx] + b[idx];
}
```

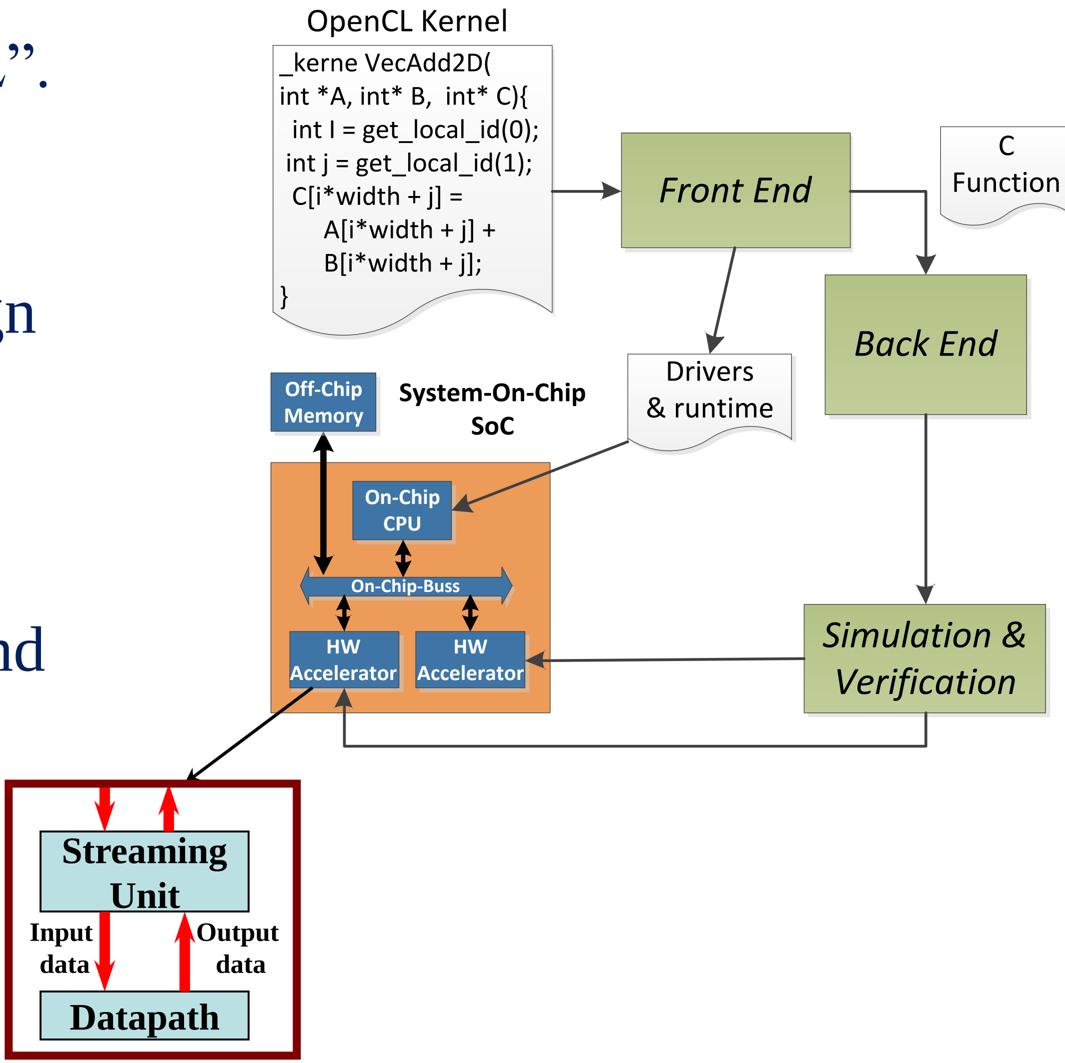
C code

```
__kernel void vadd(
         __global int* a,
         __global int* b,
         __global int* c) {
     int idx= get_global_id(0);
     c[idx] = a[idx] + b[idx];
}
```

OpenCL kernel code

Silicon OpenCL

- Silicon-OpenCL “SOpenCL”.
- A tool flow to convert an **unmodified** OpenCL application into a SoC design with HW/SW components.
- A template-based hardware accelerator generation.
- Decouple data movement and computations.



OpenCL Kernel

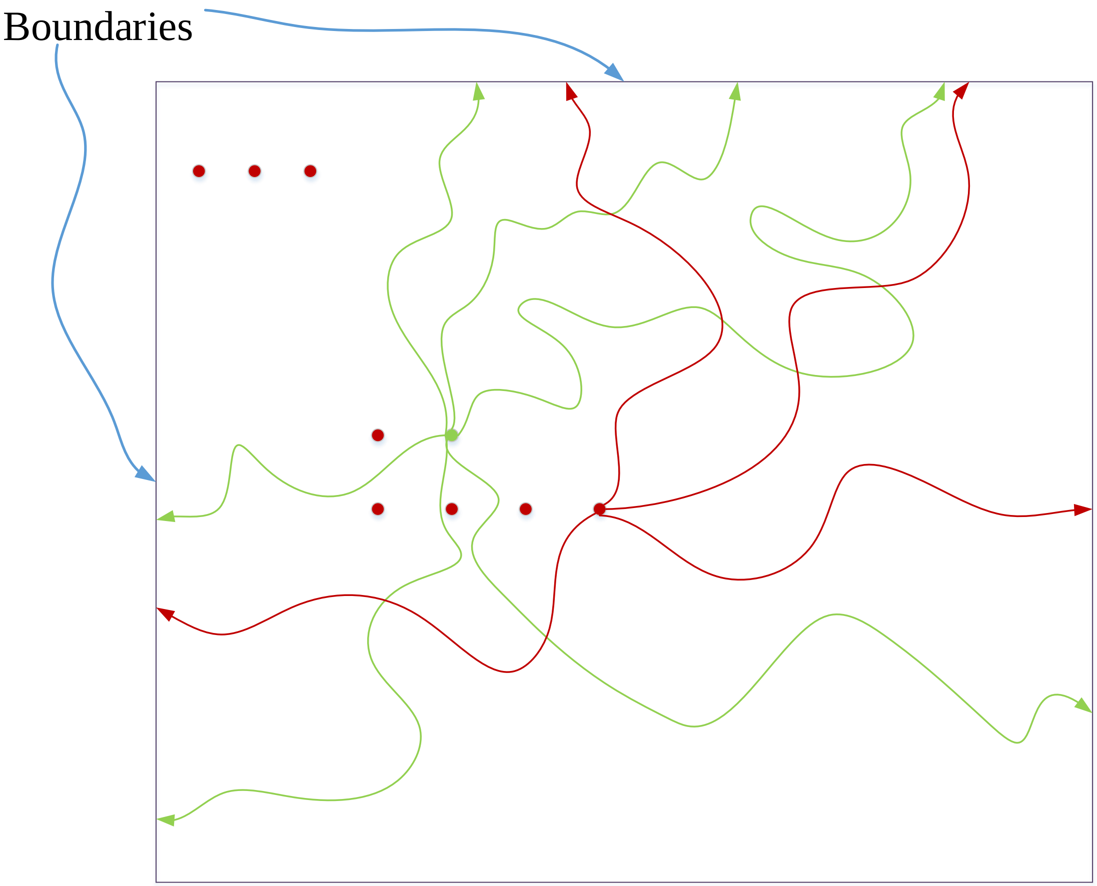
```
_kerne VecAdd2D(
int *A, int* B,  int* C){
 int I = get_local_id(0);
int j = get_local_id(1);
 C[i*width + j] =
     A[i*width + j] +
     B[i*width + j];
}
```

Architectural Template

OpenCL Implementation

- Our plan was to use the same code base (e.g. OpenCL) to explore different architectures
  - OpenCL used for multicore CPU, GPU, FPGA (SOpenCL)
- Fast exploration based on area, performance and power requirements

Monte Carlo random walks



All random walks from
all points are independent.

Double precision arithmetic
with non-constant number
of iterations

Compute-bound algorithm

```
$do\{$
$rad = rand\_uniform(seed) * d;$
$\}while(\frac{4 * rad}{(d * d)} * log(\frac{d}{rad}) < \frac{4 * rand\_uniform()}{e * d});$
```

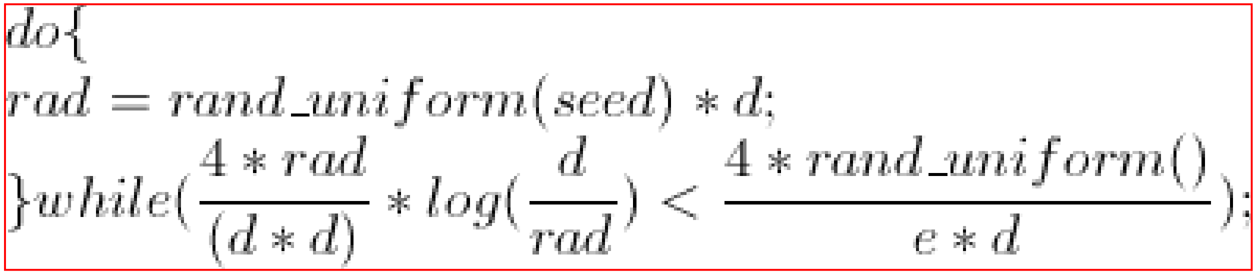
CPU - GPU implementation

- Intel Core i7-4820K CPU clocked at 3.70GHz
  - 8 threads
  - Each thread performs a random walk
- GeForce GTX 680 GPU (Kepler architecture)
  - 3 Teraflops peak performance, 288 Watts power dissipation (TDP). 1536 cores. 3.5 billion transistors.
  - N*768 threads (N is the number of points in the grid)
  - Each thread performs a random walk

FPGA implementation (I)

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

- Good match for FPGA technology:
  - MC algorithm massively parallel
  - Independent multiple path traversal from multiple points
- Poor match for FPGA technology:
  - Double precision (DP) Trigonometric, Log, Additions, Multiplications functions for each walk
  - DP arithmetic takes up a lot of area and is slow



```
$do\{$
$rad = rand\_uniform(seed) * d;$
$\}while(\frac{4 * rad}{(d * d)} * log(\frac{d}{rad}) < \frac{4 * rand\_uniform()}{e * d});$
```

FPGA implementation (II)

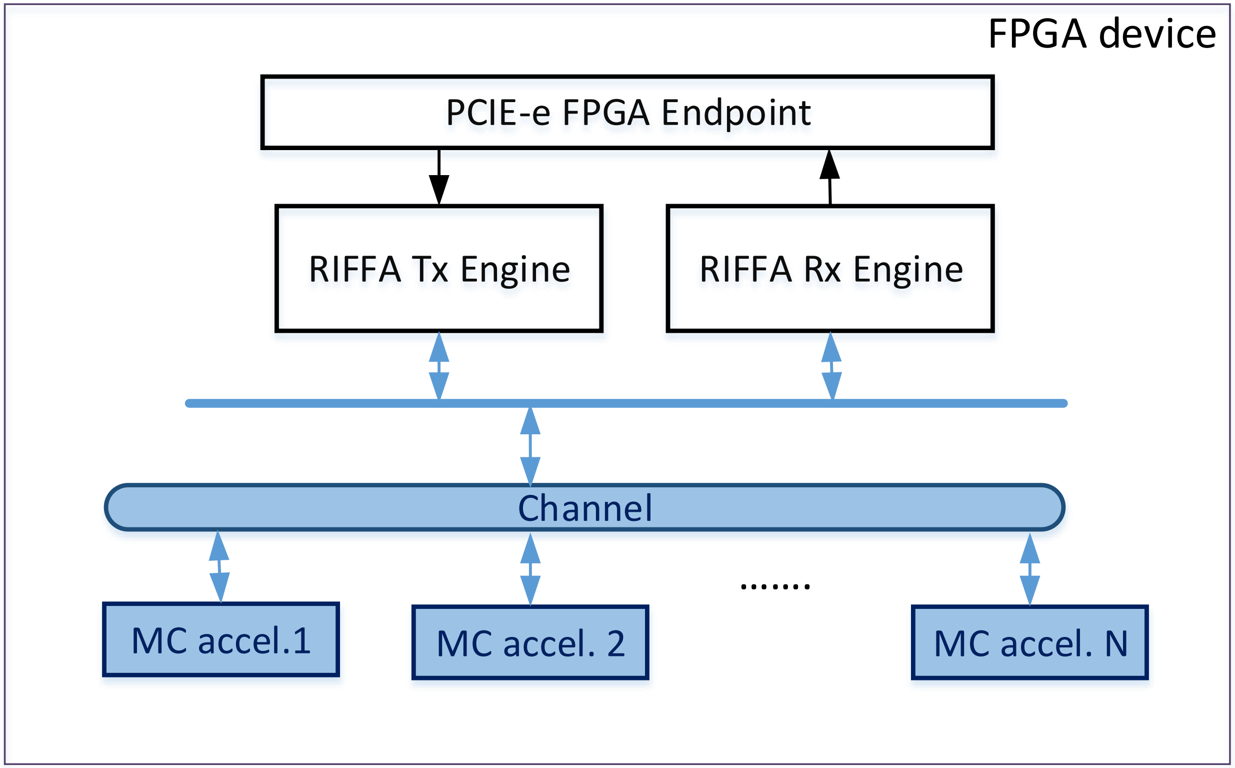
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

- Use SOpenCL and Xilinx Vivado HLS (High Level Synthesis) tools to automatically generate MC hardware accelerators
- Multiple accelerators with different performance vs. area characteristics can be generated very fast
- Best approach is when one accelerator traverses all walks from as many points as possible
- Target clock 250 MHz

FPGA architecture

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Ευρωπαϊκή Ένωση Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΣΠΑ 2007-2013 πρόγραμμα για την ανάπτυξη ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Connected to a Linux box

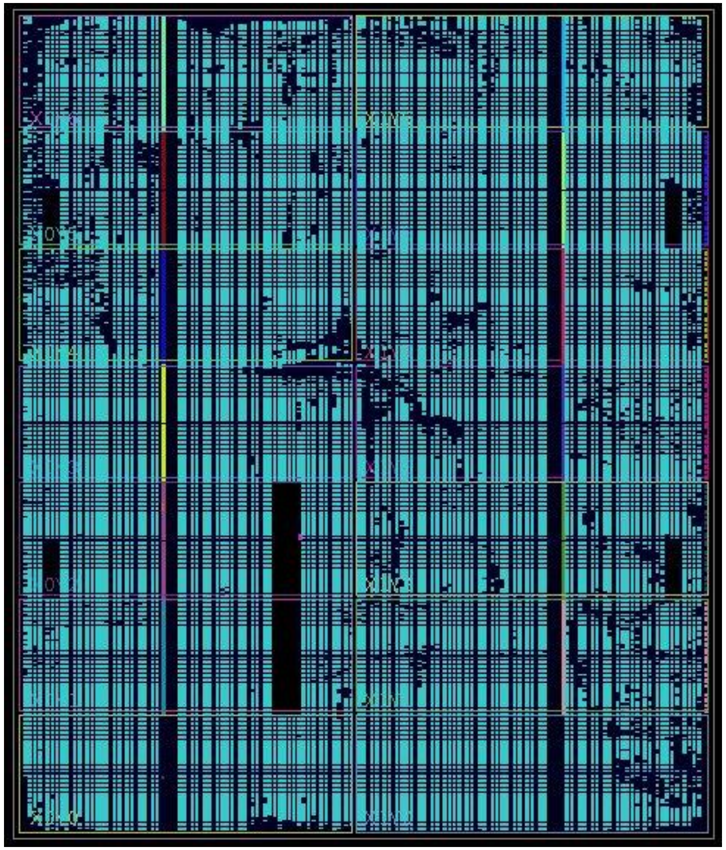


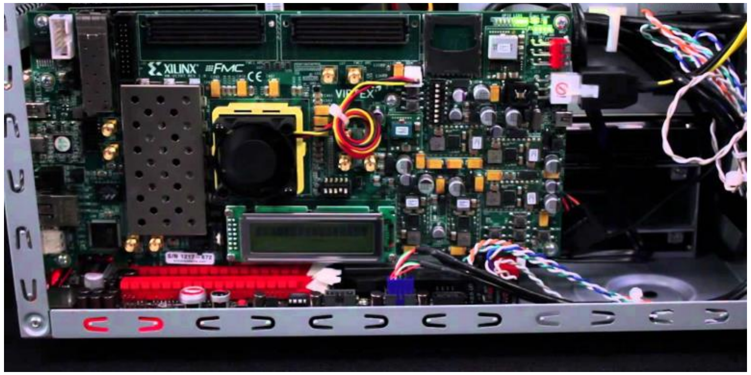
Parameterizable Architecture in Number of Accelerators
Automatically generated after tools in place
(SOpenCL and Vivado HLS)

Virtex-7 FPGA implementation

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

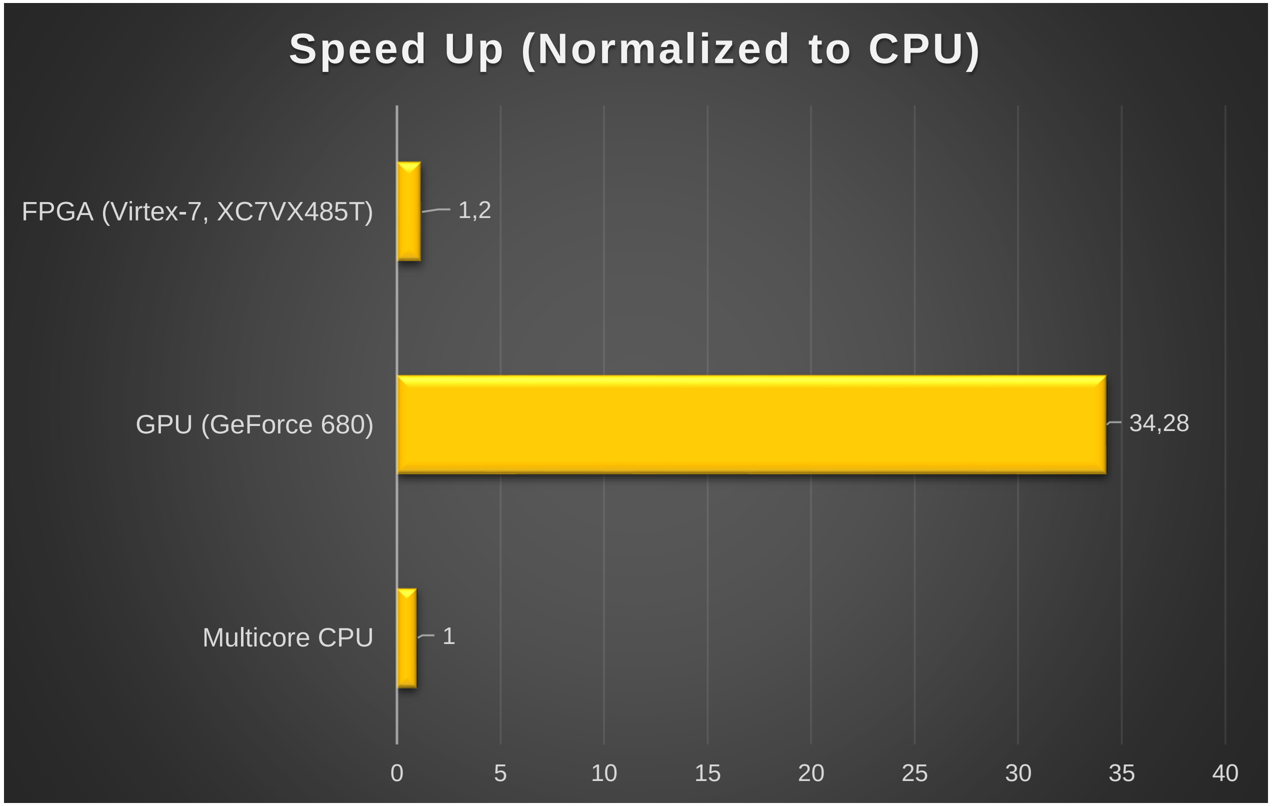
Desktop-based
High Performance Accelerator
5 MC accelerators shown





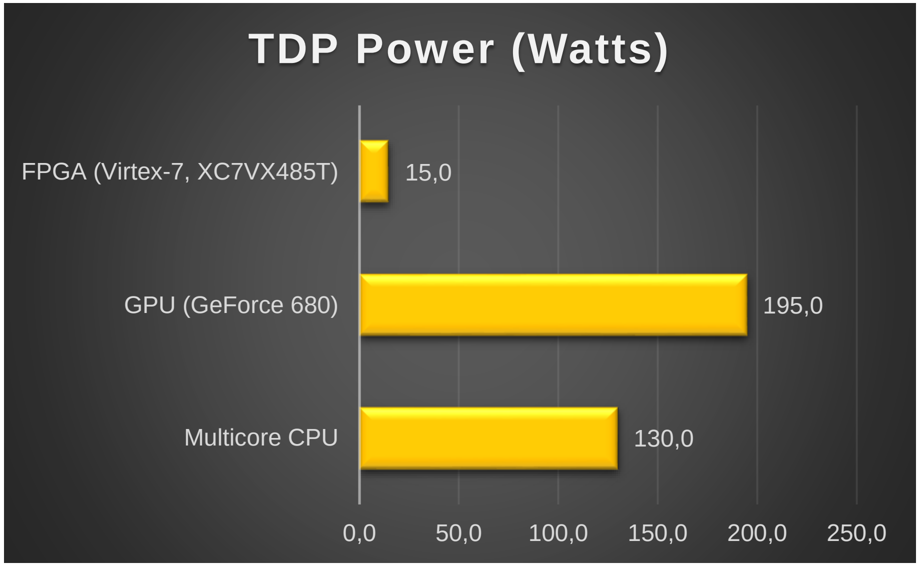
Performance

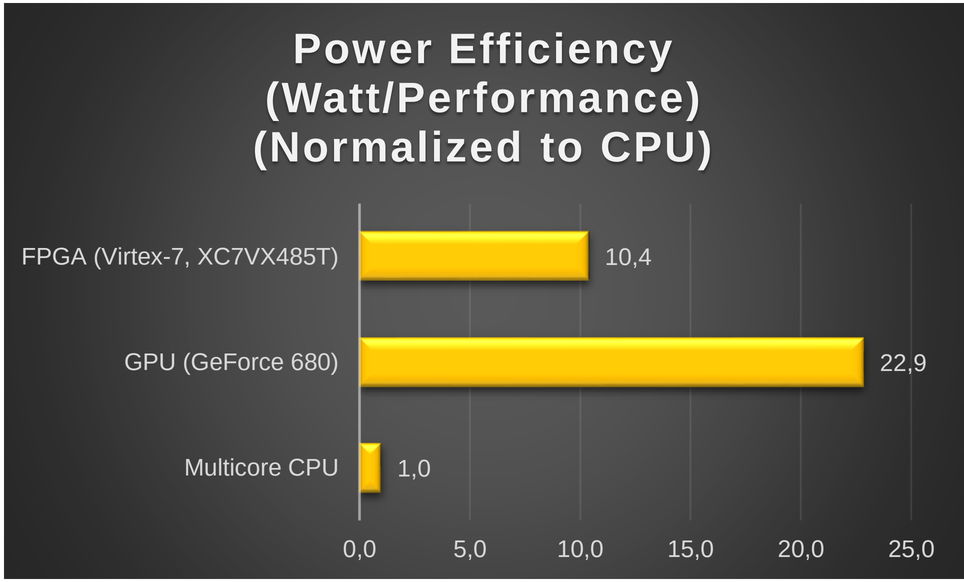
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



Power and Power Efficiency

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





MATENVMED-sponsored Publications (Δράση 3.2)

- Muhsen Owaida, Christos D. Antonopoulos, Nikolaos Bellas. A Grammar Induction Method for Clustering of Operations in Complex FPGA Designs. *IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM)*, May 11-13, 2014, Boston, MA.
- K. Krommydas, W.C. Feng, M. Owaida, C.D. Antonopoulos, and N. Bellas. On the Portability of the OpenCL Dwarfs on Fixed and Reconfigurable Parallel Platforms. *IEEE International Conference on Application-Specific Systems, Architectures and Processors (ASAP).* June 18-20, 2014, Zurich, Switzerland**. (Nominated for Best Paper Award)**
- Christos Antonopoulos. SOpenCL: An Infrastructure for Transparently Integrating FPGAs in Heterogeneous, Accelerator-Based Systems. *6th Conference on Numerical Analysis (NumAn)*. September 2-5, 2014. Chania.
- Panagiotis Skribonis, George Zindros, Ioannis Parnassos, Muhsen Owaida, Nikolaos Bellas, Paolo Ienne. Exploring Automatically Generated Platforms in High Performance FPGAs. *Parallel FPGA (ParaFPGA)*. September 1-4, 2015. Edinburgh, UK
- K. Krommydas, W.C. Feng, C.D. Antonopoulos, and N. Bellas. On the Characterization of OpenCL Dwarfs on Fixed and Reconfigurable Platforms. *Journal of Signal Processing Systems (JSSP)*. US Springer. October 2015.

Conclusion

- Graphics Processor Units (GPUs) are ideally positioned to run massively parallel applications
- FPGAs the same, but Double Precision Arithmetic is not their strong point
- Better Hardware Libraries for DP arithmetic are needed.
- Manycore parallel architectures mature to be used in framework-based computing