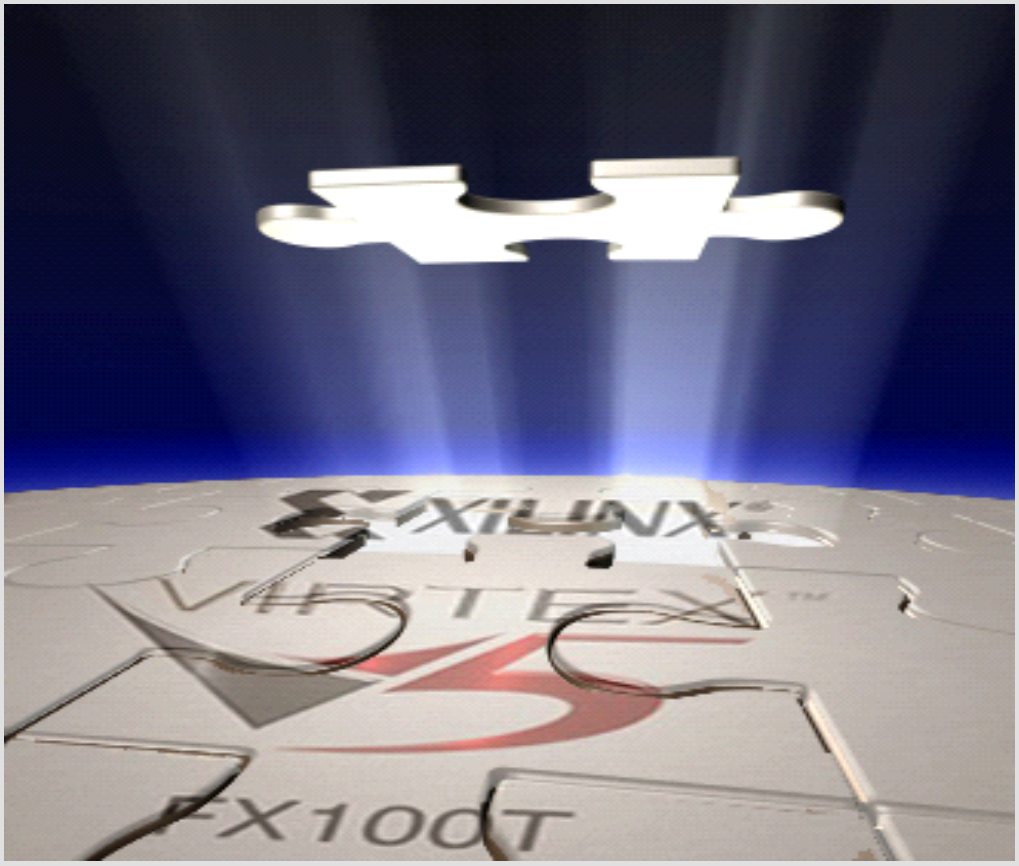
Towards Exascale Computing: The ECOSCALE Approach



Dirk Koch, The University of Manchester, UK (dirk.koch@manchester.ac.uk)

Motivation: let's build a 1,000,000,000,000,000,000 FLOPS Computer

(Exascale computing:
10^{18} FLOPS = one quintillion or a billion billion floating-point calculations per sec.)

1,000,000,000,000,000,000 FLOPS

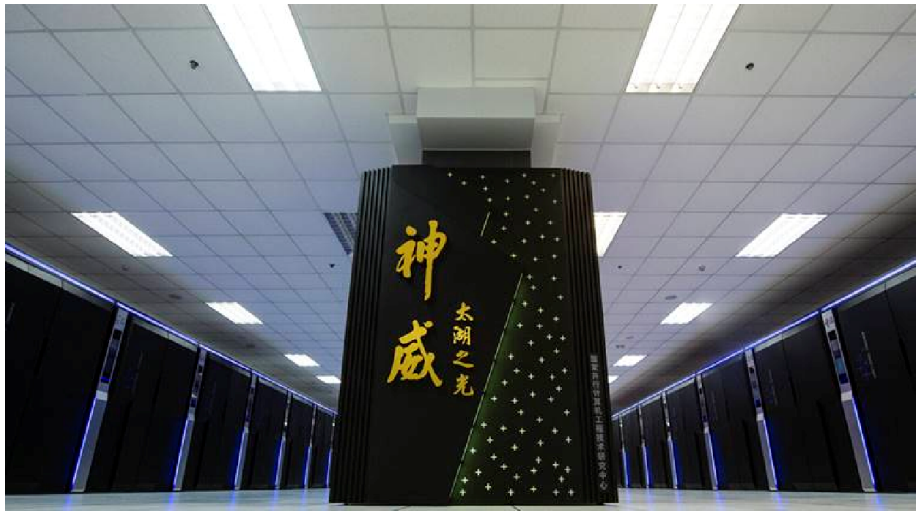
- 10,000,000,000,000,000,00 FLOPS

  1975: MOS 6502 (Commodore 64, BBC Micro)

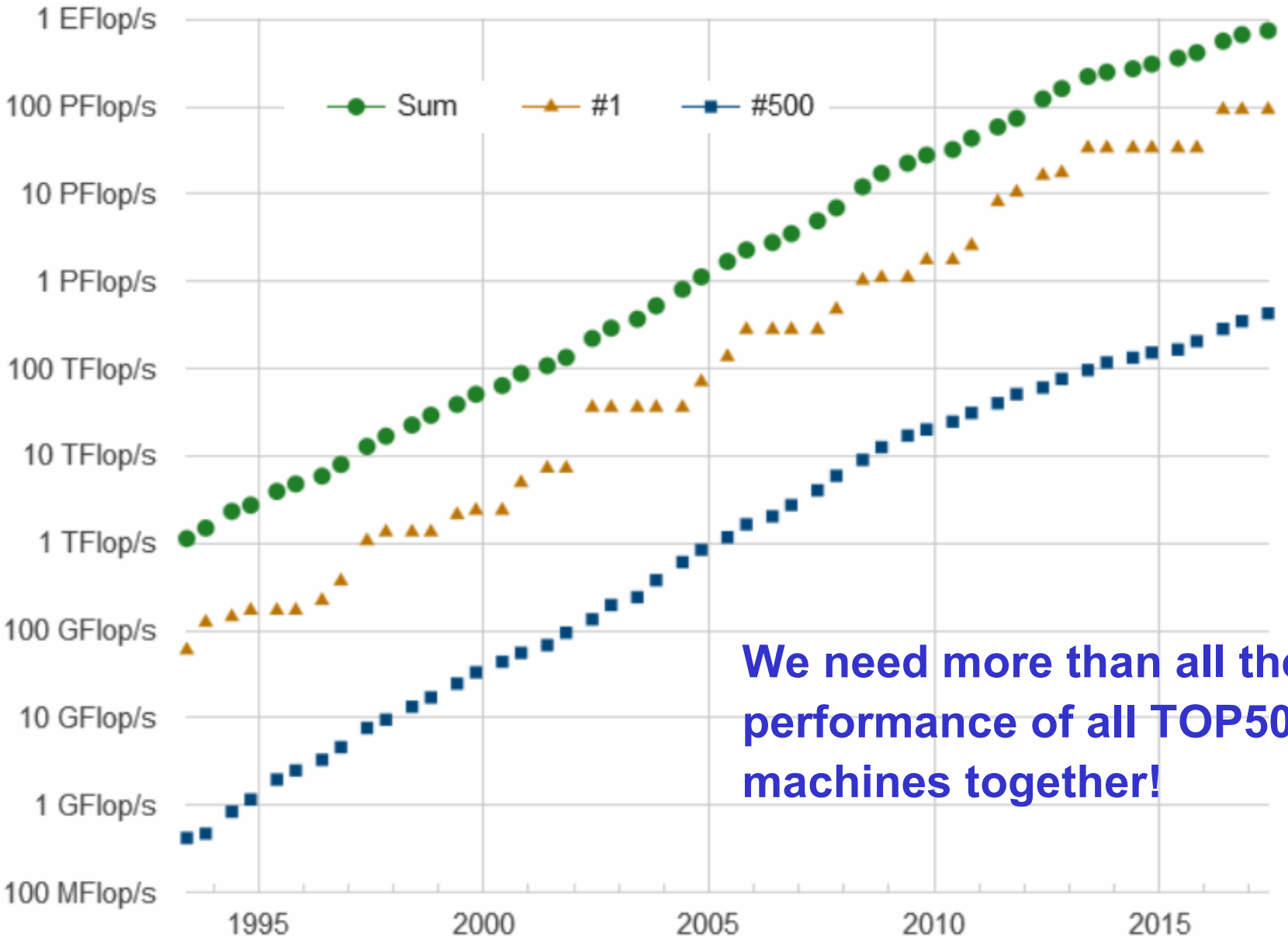
Sunway TaihuLight Supercomputer

- 2016 (fully operational)
- 12,543,6000,000,000,000,00 FLOPS (125.436 petaFLOPS)
- Architecture Sunway SW26010 260C (Digital Alpha clone) 1.45GHz 10,649,600 cores
- Power "The cooling system for TaihuLight uses a closed-coupled chilled water outfit suited for 28 MW with a custom liquid cooling unit"*

  *https://www.nextplatform.com/2016/06/20/look-inside-chinas-chart-topping-new-supercomputer/
- Cost US$ ~$270 million



TOP500 Performance Development



TaihuLight for Exascale Computing?

We need 8x the worlds fastest supercomputer:

- Architecture — Sunway SW26010 260C (Digital Alpha clone) @1.45GHz: > 85M cores
- Power — 224 MW (including cooling)
  costs ~ US$ 40K/hour, US$ 340M/year
  from coal: 2,302,195 tons of CO2 per year
- Cost — US$ 2.16 billion

We have to get at least

10x better in energy efficiency

2-3x better in cost

Also: scalable programming models

Alternative: Green500

Shoubu supercomputer (#1 Green500 in 2015):

- **Cores:** 1,181,952
- **Theoretical Peak:** 1,535.83 TFLOPS/s
- **Memory:** 82 TB
- **Processor:** Xeon E5-2618Lv3 8C 2.3GHz
- PEZY-SC accelerators (GPU-like that use OpenCL) (theoretical 6-7 GFLOPS/W)

- → **150 MW for Exascale** (very optimistic)

- Good, but not good enough

GPUs?

- Energy efficiency is the key for performance computing
  → high integration → high memory / I/O throughput
  → There is strong need to process more data faster at less power!
- Energy efficiency is the most important technology driver
  (from mobile to datacenters)
- Many people consider GPUs for energy efficient computing, **but**

M2050 GPU

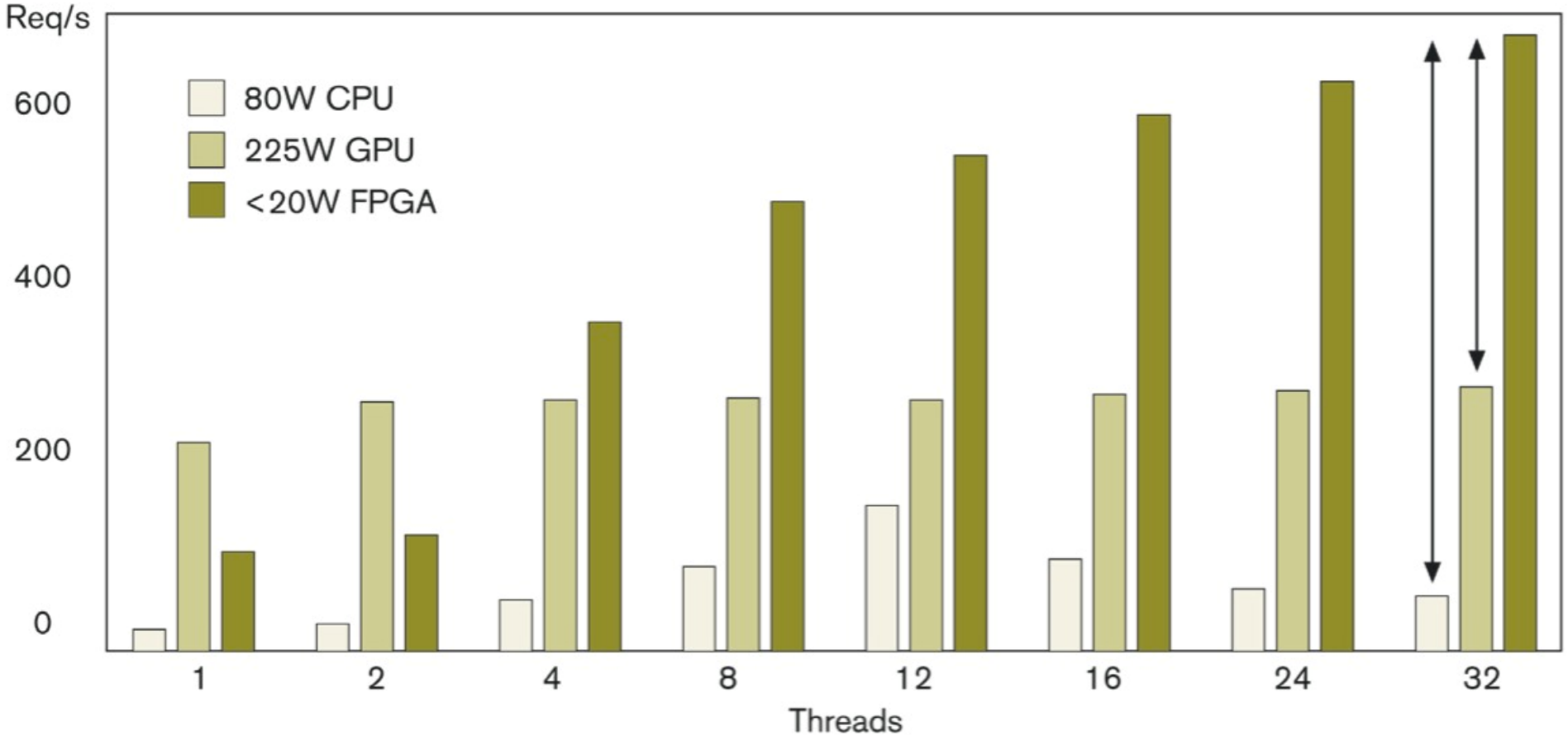
TPD: 225 W (5.4 kWh/day)

Equipment cost: 2400 US$
(energy cost exceeds in less than
5 years considering 0.25 US$/kWh)

1 TFLOPs peak
→ **225 MW at Exascale** (very optimistic)

Alternative: FPGAs

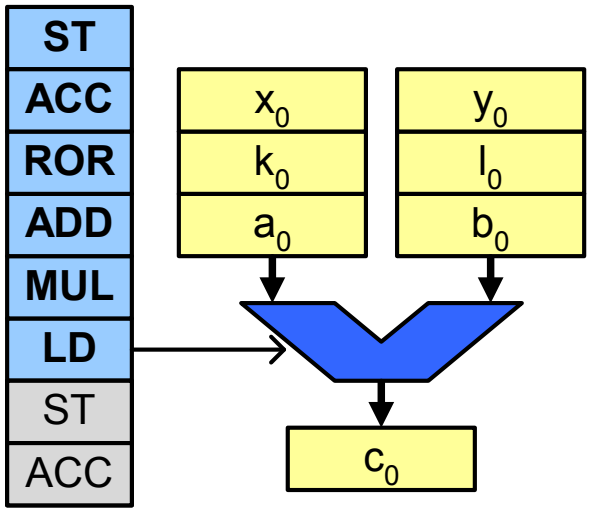
Baidu's analysis on predictive search algorithms:



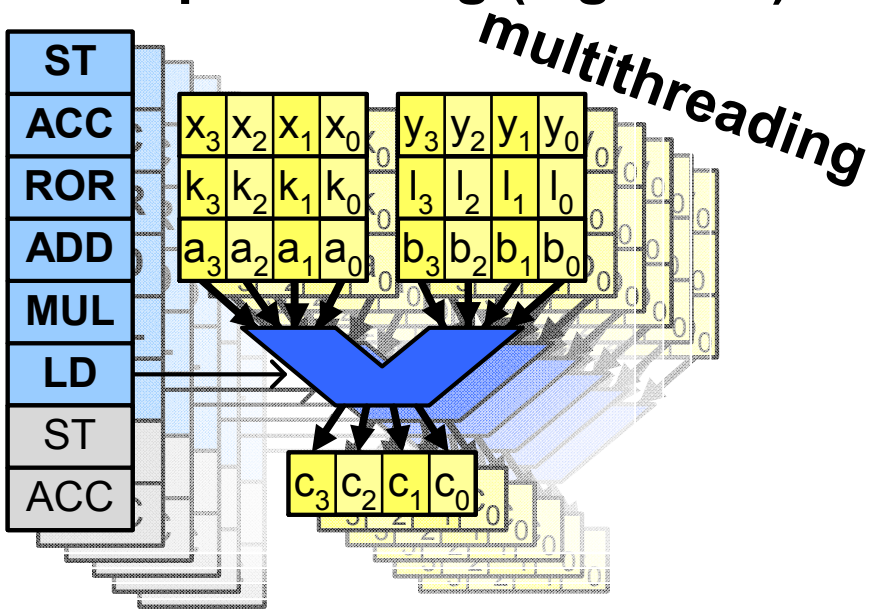
- FPGA ~500 GFLOPS @ 20W
- **40 MW at Exascale** (optimistic)
  → close to the **green** zone

CPUs vs. GPUs vs. FPGAs

Scalar processing

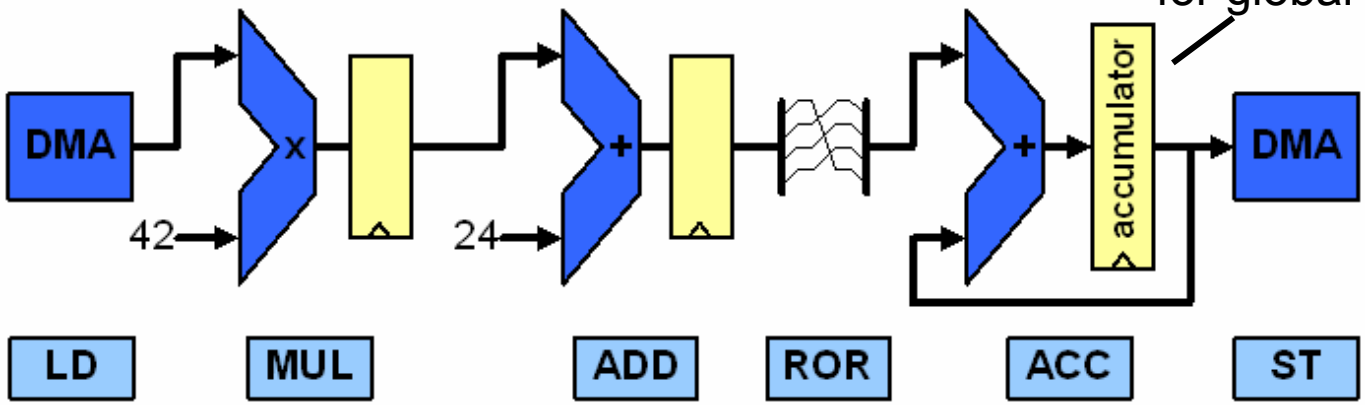


SIMD processing (e.g. GPU)



Dataflow processing / pipelining (FPGAs)

Pipelining often better for global states



CPUs vs. GPUs vs. FPGAs

- **CPUs are ideal for control flow dominated tasks (e.g., an OS, compilation)**
- **GPUs are ideal for double precision number crunching**
- **FPGAs are ideal for number crunching problems that fit dataflow processing model**
- **FPGA advantages**
  - **Customized processing**
  - **Optimized data movement**
  - **High integration (entire system including memory, mass storage, networking, acceleration and CPU)**
  - **High performance at low power**

We are at the end of CMOS scaling

- **Compute demand still exploding**
  (Health, Climate, AI, Big data)

- **No alternatives (optical, quantum computing) at the horizon!**
  → We have to make the best out of CMOS!
- **Issue is not the number of transistors but energy efficiency**
  → Energy efficiency translates into compute performance!
- **Objectives are hard to meet with CPUs or GPUs**
  → Custom Computing using FPGAs!
- **Needs rethinking**
  - **Programming models**
  - **Operating systems (runtime environments)**
  - **The "ecosystem"**

HPC FPGA Research in Manchester

FPGA for HPC computing

- **Manchester participates with three joined H2020 projects:**
  - **ECOSCALE** http://www.ecoscale.eu/
    (Scalable programming environment and hardware architecture tailored to current and future HPC applications)

    Manchester: FPGA virtualization and run-time management, automatic generation of reconfigurable module libraries from OpenCL kernels, support for all recent Xilinx FPGAs (including Ultrascale+)

HPC FPGA Research in Manchester

FPGA for HPC computing

- **Machester participates in further joined H2020 projects:**
  - **ExaNeSt** `http://www.exanest.eu`
    HPC interconnection networks, storage and cooling

  - **ExaNode** `http://www.exanode.eu`
    Investigates and develops a highly integrated, high-performance, hetero-geneous System-on-a-Chip (SoC) aimed towards exascale computing

  - **Budget: over 12M, ~2M for Manchester**

HPC FPGA Research in Manchester

FPGA for HPC computing

- **Follow-up project EuroEXA**

  “Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon”

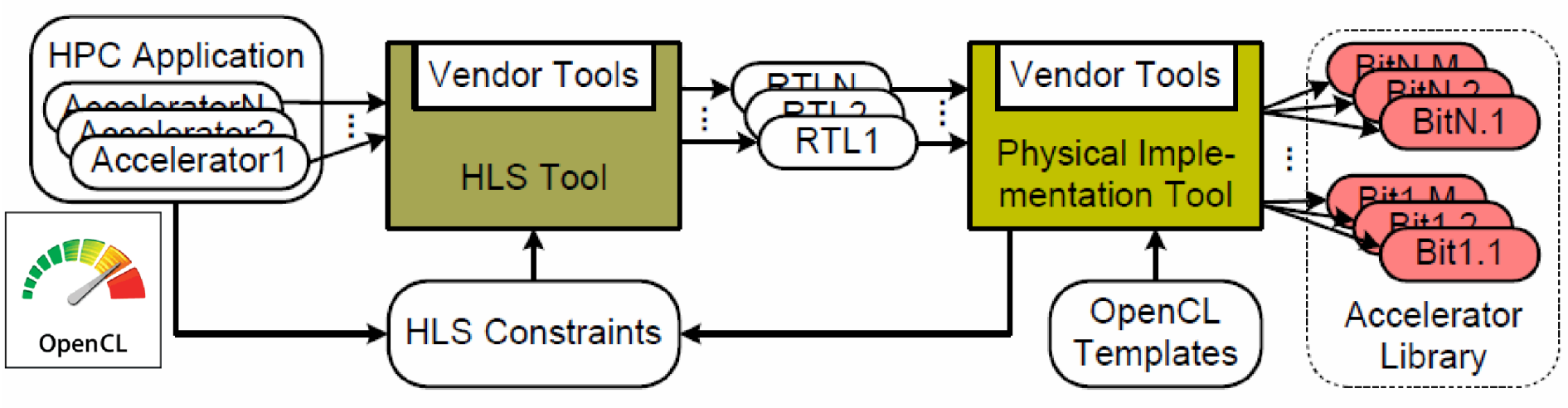
  - **Budget: 20M, over 6M for Manchester**

HPC FPGA Research in Manchester

- **Follow-up project EuroEXA — partners:**
  - **Spain — Barcelona Supercomputing Center**
  - **United Kingdom — ARM-UK, Iceotope, Maxeler Technologies, The University Of Manchester, The Hartree Centre of STFC, ECMWF (European Centre for Medium-Range Weather Forecasts)**
  - **Greece — FORTH, Synelixis Solutions Ltd**
  - **Belgium — IMEC**
  - **Sweden — ZeroPoint Technologies**
  - **Netherlands — Neurasmus**
  - **Italy — INFN (Istituto Nazionale Di Fisica Nucleare), INAF (Istituto Nazionale Di Astrofisica)**
  - **Germany — Fraunhofer-Gesellschaft**

HPC FPGA Research in Manchester

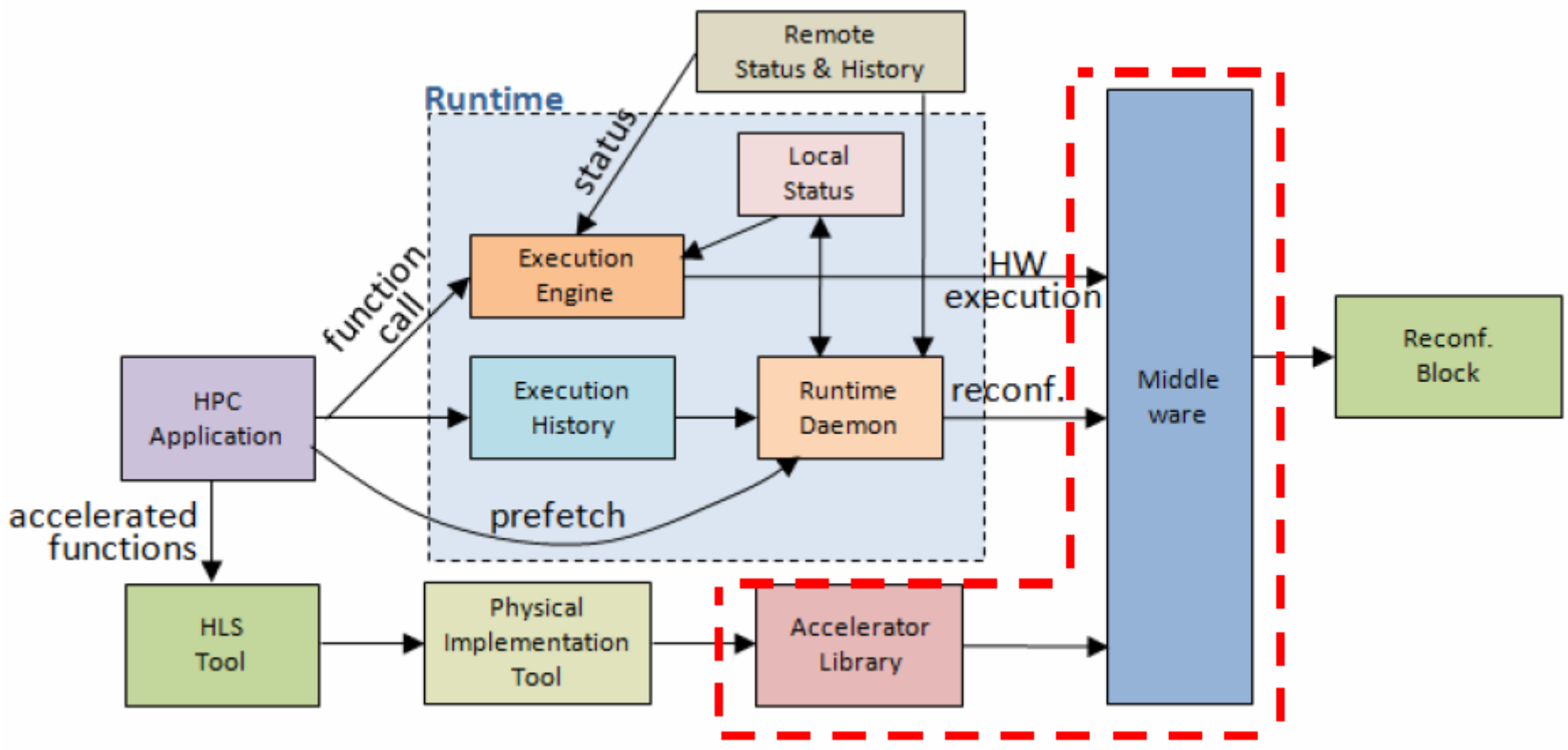
Manchester contribution: from OpenCL to configurable modules



- **Design and implementation of the tool flow backend (usable by software domain experts)**
- **Develop the middleware for the management of reconfigurable blocks at runtime**
- **Provide partial reconfiguration**

HPC FPGA Research in Manchester

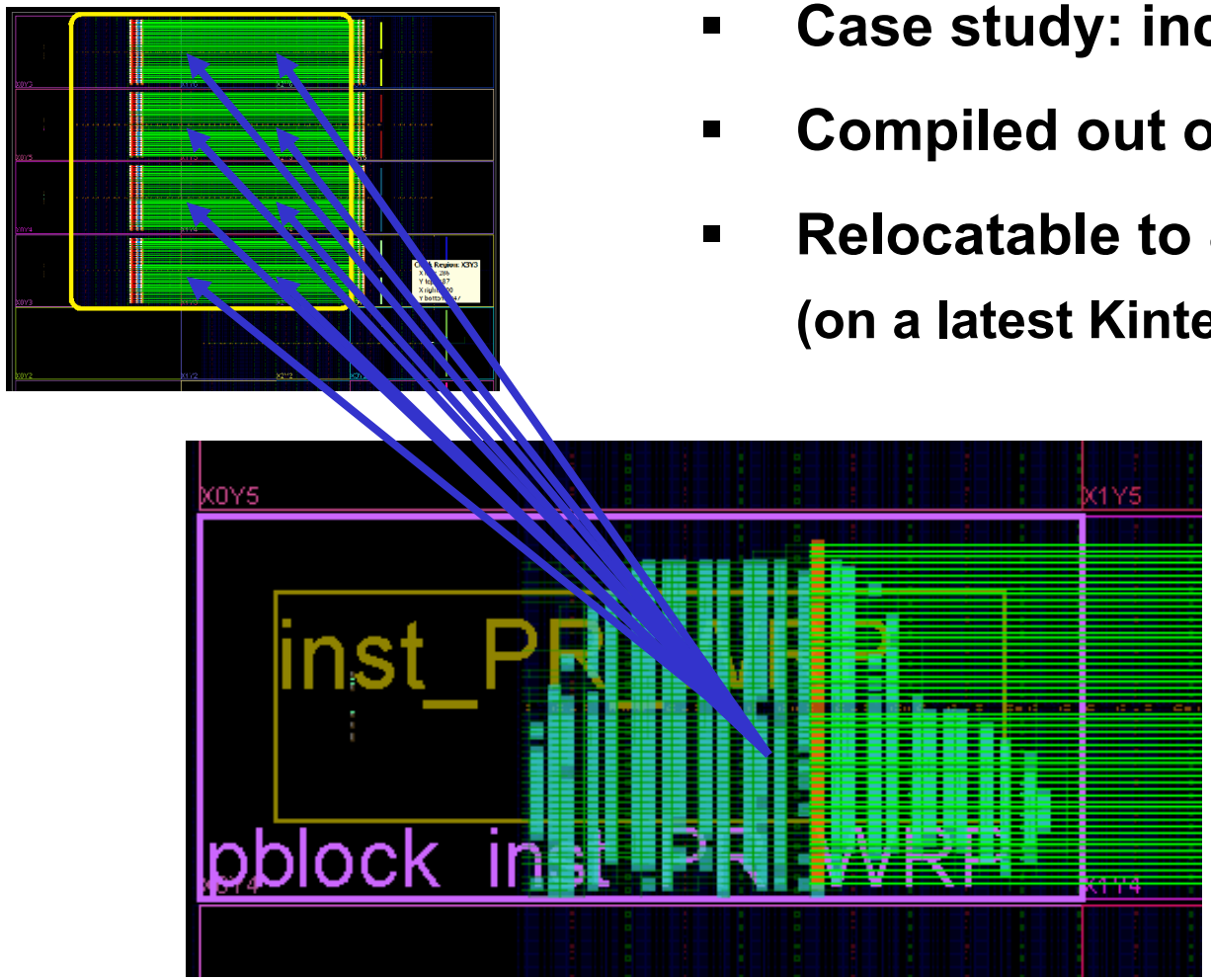
Manchester contribution:



- **Runtime management of reconfigurable resources (i.e. operating system services)**

HPC FPGA Research in Manchester

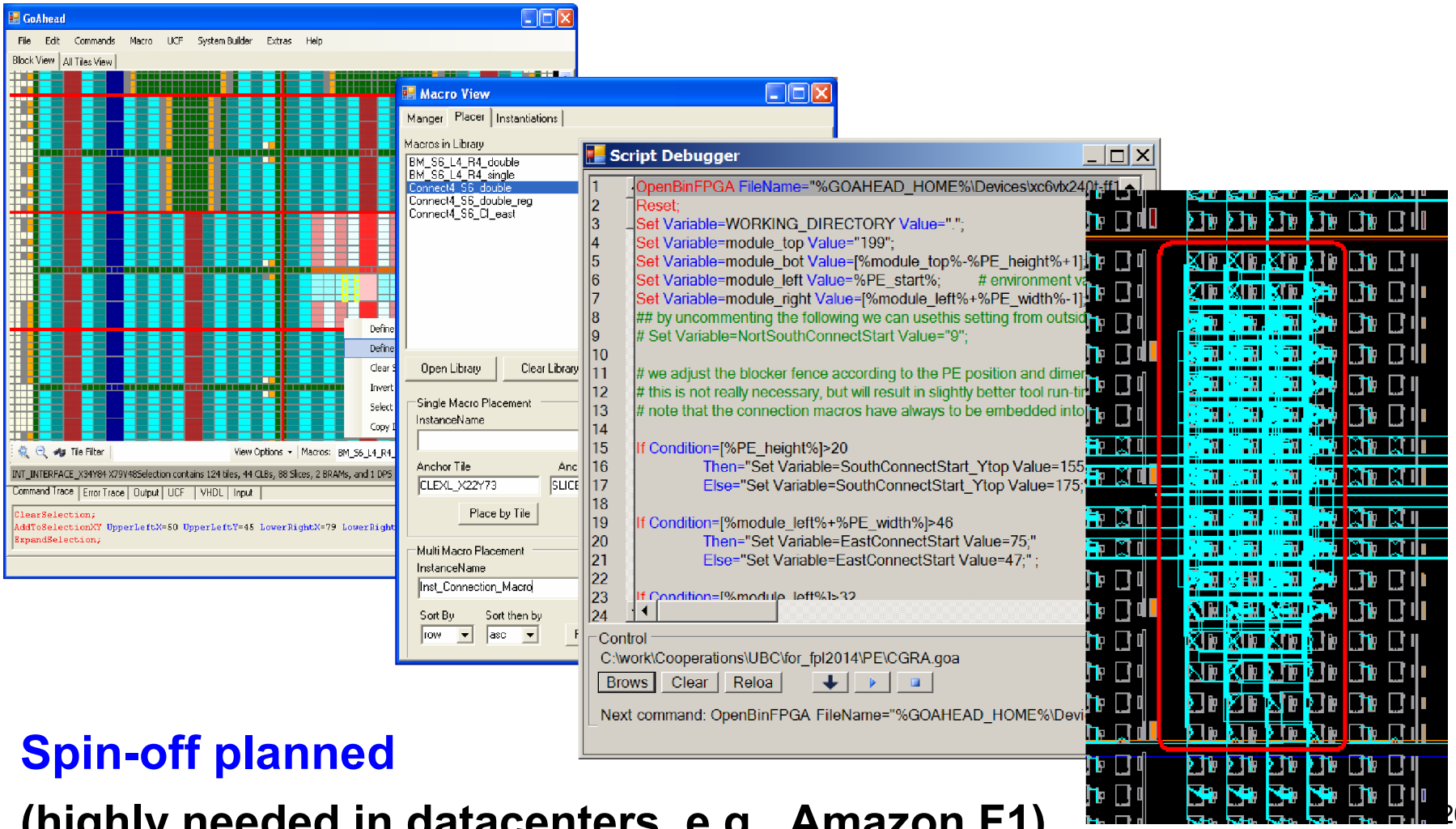
Manchester contribution:



- **Case study: increment3d**
- **Compiled out of the box**
- **Relocatable to 8 positions**
  (on a latest Kintex UltraScale+ FPGA)

GoAhead FPGA PR Tools from Manchester

FPGA design tools for implementing partial reconfiguration (most advanced tools currently available)

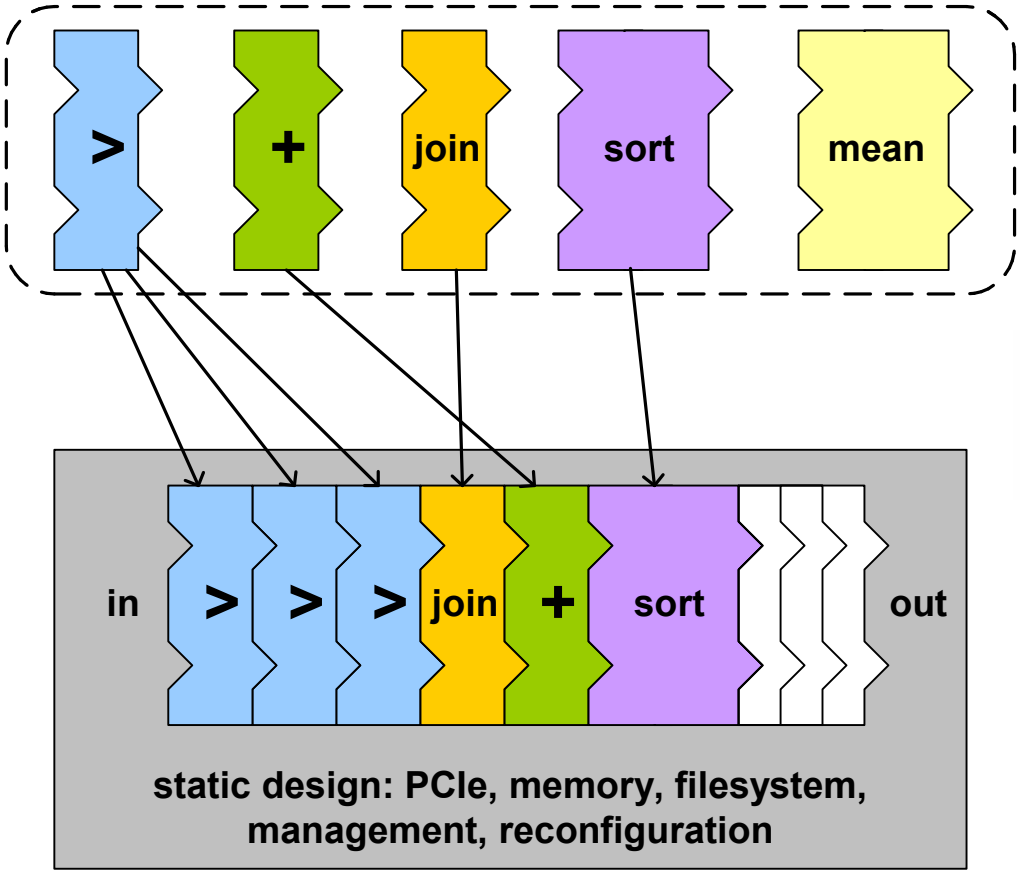


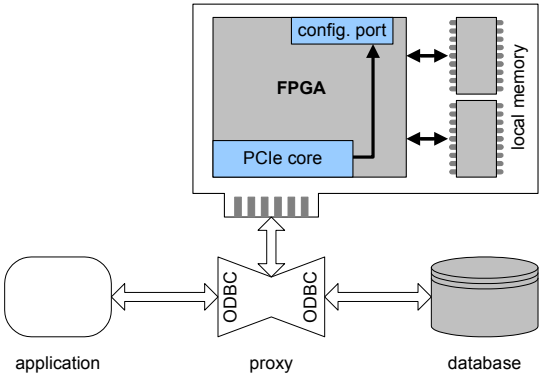
Spin-off planned

(highly needed in datacenters, e.g., Amazon F1)

FPGA Database Accelerator

- Build library with SQL operators
- Compose graph at run-time



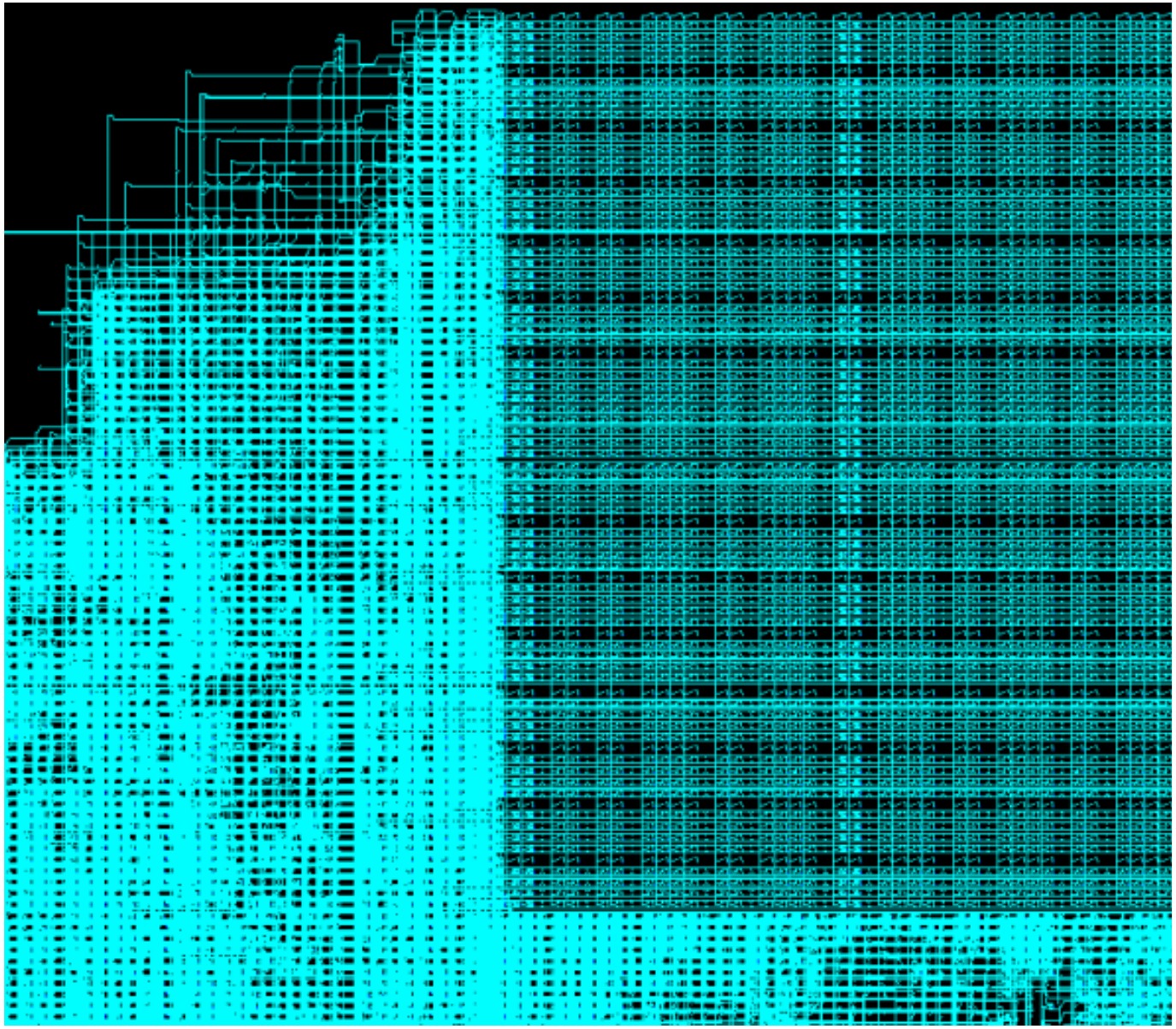


MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

implementation:

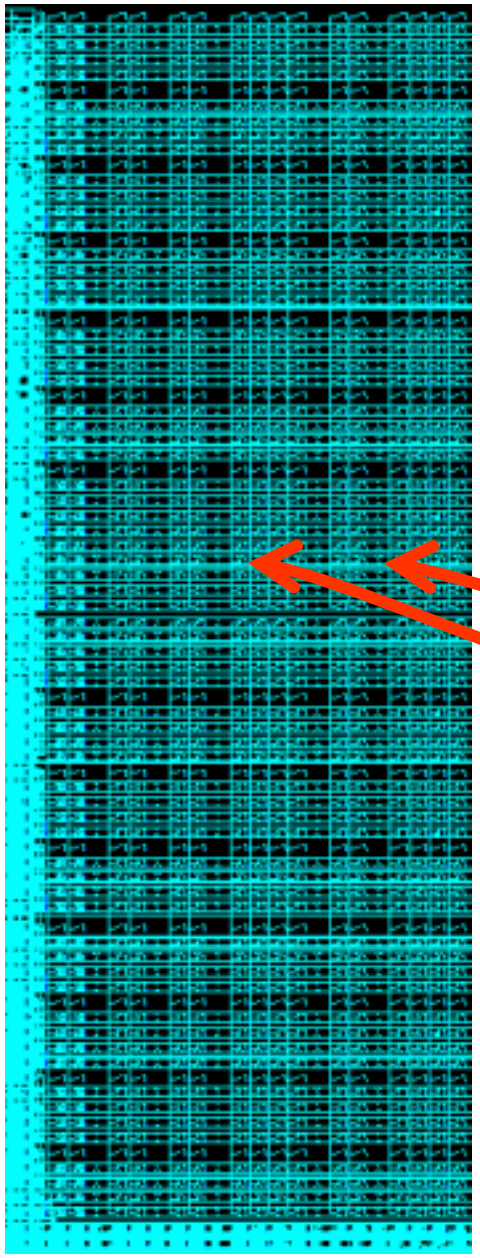
- 512 bit datapath
- 300 MHz (Virtex-6)
- Tables are stored in
  24 GB on board RAM
  (48 GB possible)

FPGA Database Accelerator



- Reconfigurable region
- Regularly routed
- 16 x 32-bit (512-bit total)
- @ 300 MHz
- ~40% of a MAX3 workstation is reconfigurable

FPGA Database Accelerator



- Build library with SQL operators
- Compose steam processing machine at runtime

- Integrated in:

PostgreSQL

Further Research

- **Big data analytics machine (80 SSDs connected to 4 FPGAs)**
- **Worlds fastest open source PCIe infrastructure**
- **Large problem sorting**
- **Resource elastic hardware**
- **Real FPGA virtualization (in the space domain)**
- **FPGA Interlays (FPGAs inside CPU cores)**

- **FORTE: Functional Oxide Reconfigurable Technologies (Manchester: explore memristors for future FPGAs)**

Contributors

- **Grigore Nicolae Bogdan (CDT)**
  Query optimization, resource management,
  FPGA virtualization
- **Athanasios Stratikopoulos (ARM)**
  FPGA virtualization
- **Malte Vesper (DSTL)**
  SSD stream processing infrastructure, applications
- **Raul Garcia (Conacyt)**
  Reconfigurable instruction set extensions
- **Christian Beckhoff** (hobbyist)
  GoAhead support (tool for building reconfigurable systems)
- **Dr. Edson Horta, Khoa Pham (H2020: ECOSCALE)**
  HLS support for PR and runtime management