# ENGINEERING CMS SOFTWARE FOR HIGH-EFFICIENCY / MANY CORE ARCHITECTURES

DAVID ABDURACHMANOV (UNIVERSITY OF NEBRASKA-LINCOLN), 13 DEC 2016

MOTIVATION THE ELECTRICITY COSTS

‣ Gartner reported in 2010 that energy-related costs are the **fastest-rising cost in the data center**

‣ New report in 2016 June: **Annual growth for server shipments have fallen significantly, same applies for power consumption, yet we continue to have "a drastic increase in demand for data center services"**
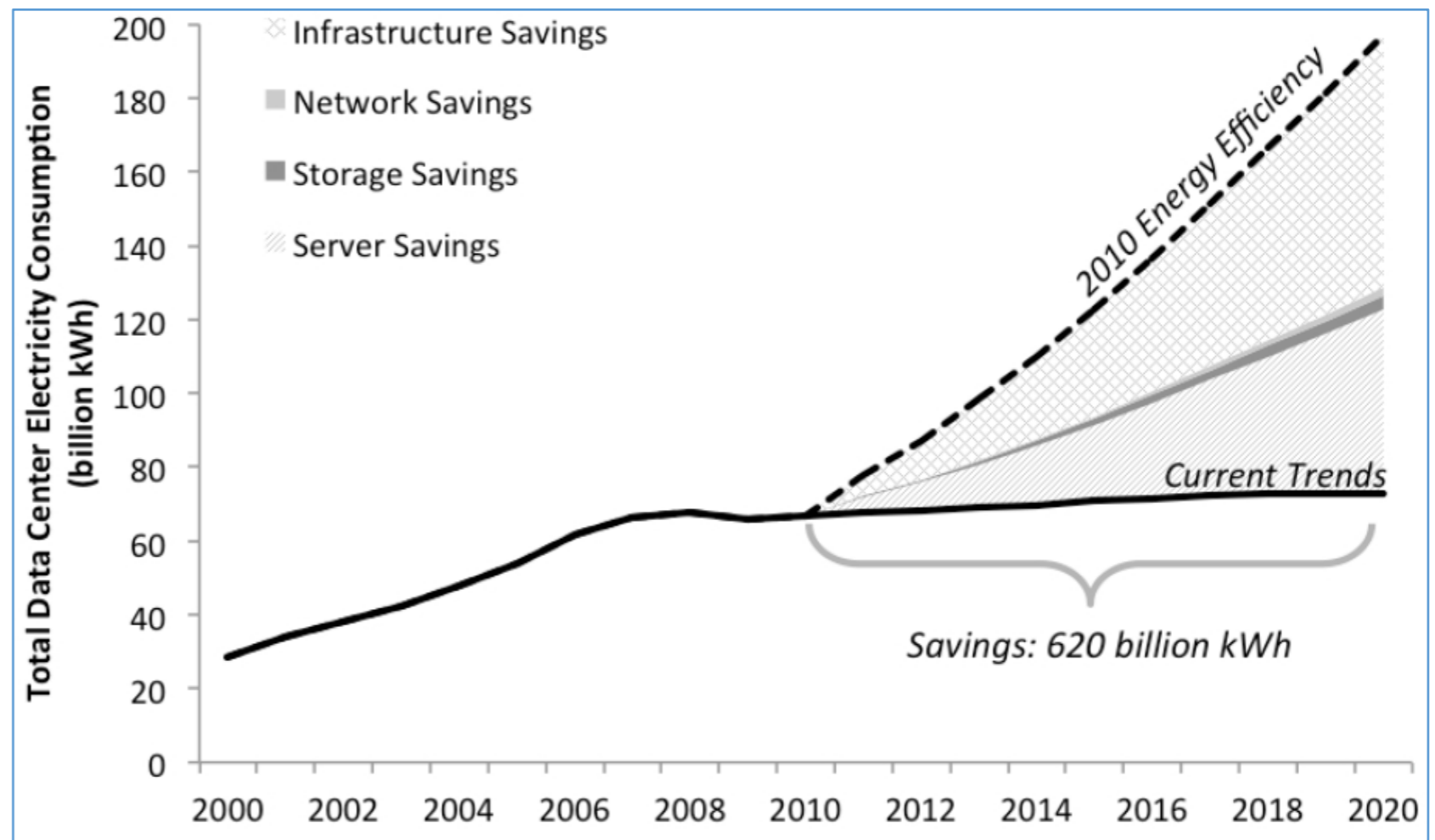
United States Data Center
Energy Usage Report (LBNL-1005775),
June 2016

MANY/MULTI CORE RISE

‣ The power dissipation (watts) have stopped increasing in the last decade and the same applies to the frequencies.

‣ Thus CPUs instead started adding weaker cores, but higher number of them.

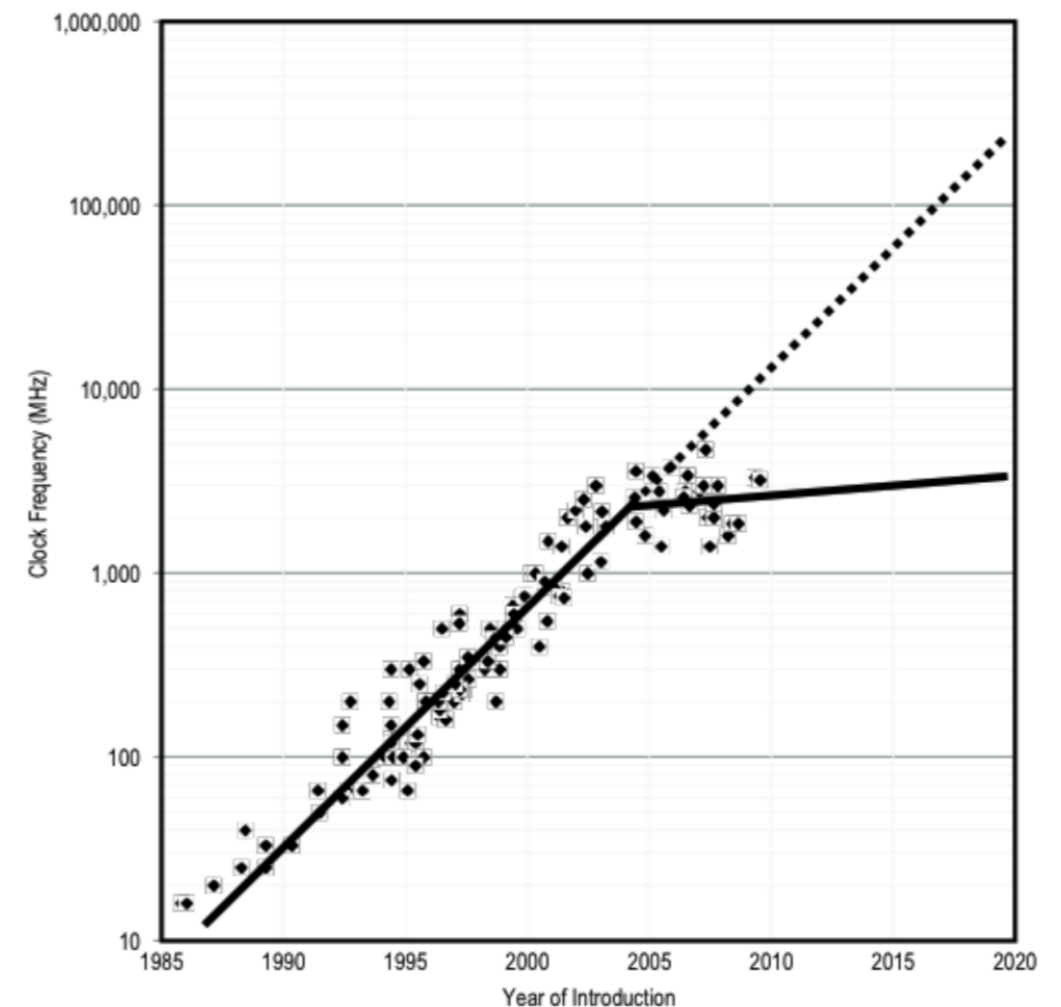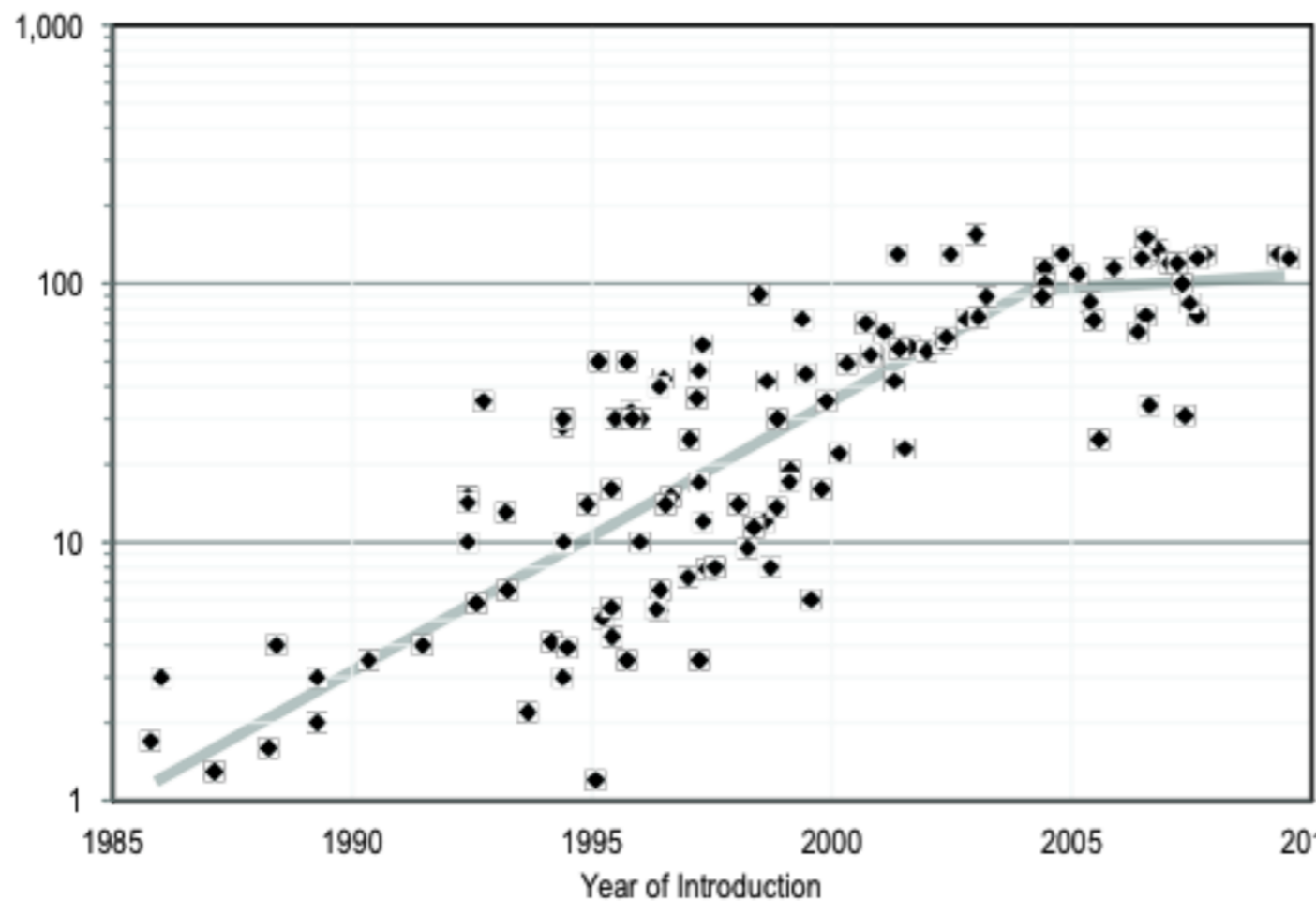‣ Also recent CPUs are very dynamic, e.g., AVX frequencies in Haswell Xeon.

FIGURE A.4 Microprocessor power dissipation (watts) over time (1985-2010).

THE NEW ERA THE CHANGE

- ARM announced ARMv8.{0,1,2,**3**} which now covers **server market.**
    - Designs with up to 64 cores announced & to be used in supercomputers.
- ARM recently announced Scalable Vector Extension (SVE) for HPC.
    - Up to 2K vector sizes & binary compatibility.
- ARM provides IP (building blocks for SOC and ISA) licenses, but not the product.
- Number of server grade SOCs announced from **APM** (early CMS partner), **Cavium**, **AMD**, **Broadcom**, **Qualcomm**, **Huawei**, **Phytium Technology Co.**, etc. based on ARM's and custom IP.

- All powered by **a single ISA, but completely different**.
- IBM announced OpenPOWER Foundation ⇒ other companies now can build PowerPC CPUs.
- POWER8+ with NVLink 1.0 is the first one after OpenPOWER Foundation was launched.

- Intel recently released KNL (**many-core and vector CPU**) with legacy support for x86_64.
    - Also recently announced KNM with improvements for machine learning.
- Xeon + **FPGA** (former Altera) products from Intel
- Lake Crest - **deep learning accelerator silicon**

COMMODITY STILL COMMON STUFF

‣ Both offer **LP64 data model and LE support**, **RHEL/CentOS/Fedora**, almost no porting efforts.

‣ Both have announced NVIDIA GPGPU (CUDA) support.

‣ The idea is to be **extremely boring**, i.e. administrators and users shouldn't notice that they are running on different architecture ⇒ it's yet another PC.

‣ Benchmarking is an activity/**a tool** to measure **technology** and **software progress over time**.

‣ We are moving away from homogeneous to heterogeneous computing:

  ‣ Rapid development of ARM server SOCs and platforms from multiple vendors, e.g. X-Gene 1 (**8-core**) ➡ X-Gene 3 (**32-core**) ➡ X-Gene 3XL (**64-core**) while also increasing single-threaded performance (estimated).

  ‣ IBM is moving forward with OpenPOWER Foundation and POWER9.

  ‣ IBM POWER8 already is competitive to Xeon in terms of performance.

  ‣ CPUs with SMT 2 (Xeon), SMT 4 (Xeon Phi, Broadcom Vulcan), SMT 8 (PowerPC).

  ‣ Long-vector (e.g. AVX512) machines (Xeon Phi, ARMv8 SVE)

  ‣ GPGPU power and efficiency is rising fast and supports all major architectures: Intel, ARM and IBM.

  ‣ Code modernisation projects in HEP focusing on Xeon Phi, GPGPU acceleration (CUDA, OpenCL) and vectorisation

‣ Our code base has to be **flexible to adapt** to the future technologies.

WHY NEW ARCHITECTURES?

‣ Distributed computing in **HEP before ~2000** had multiple vendors involved, and incl. special workstations and heterogeneous computing

‣ **Hight Throughput Computing (HTC)** converged on x86/Linux at **~2000**

   ‣ Commodity hardware enabled the current model of WLCG:

Build Once, Run Everywhere

   ‣ **Two vendors:** Intel (dominating) and AMD

‣ The on-chip power density limitations are driving the computing market towards a greater variety of solutions, i.e. workflow optimised

‣ Specialised processors and heterogeneous computing rise up

‣ Incl. heterogeneous worker nodes
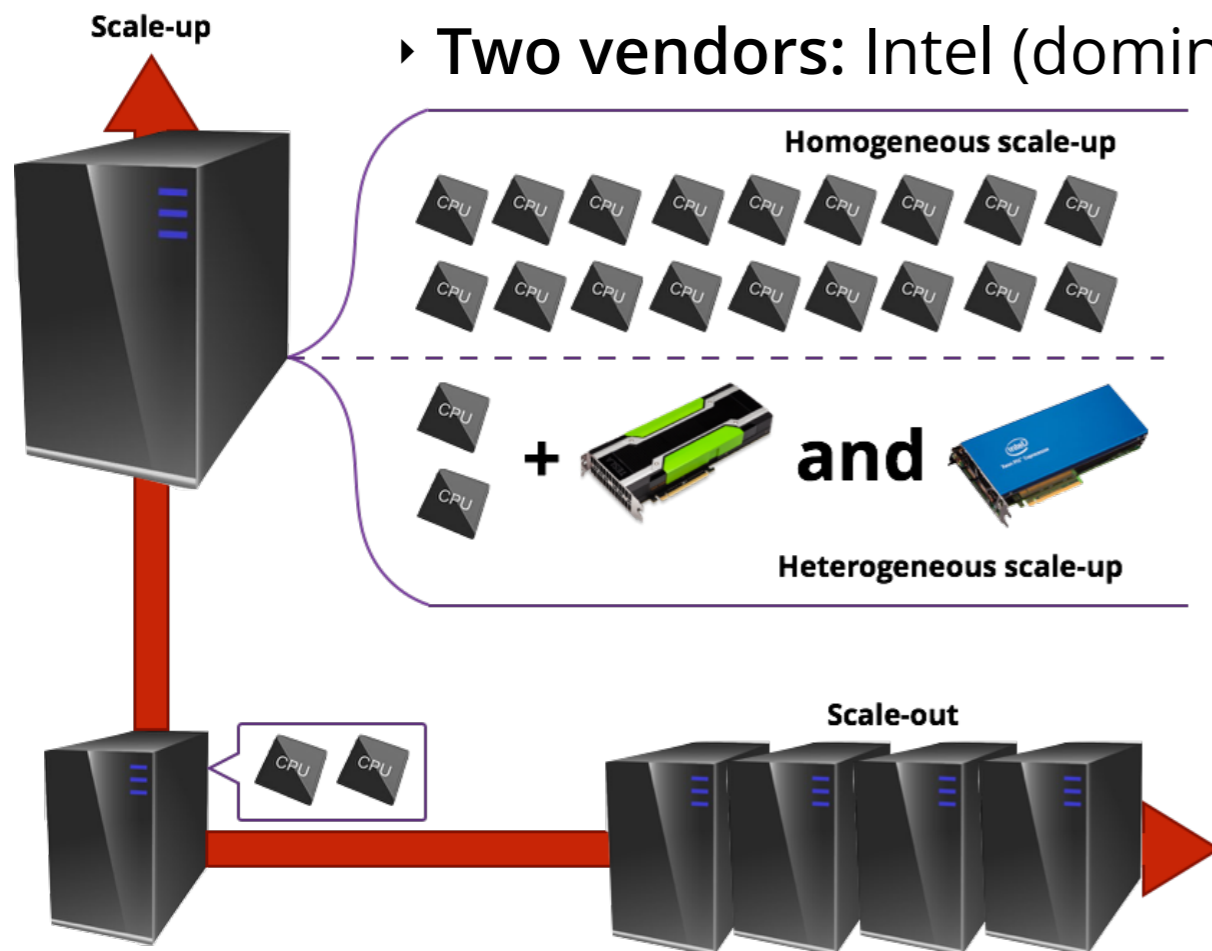
CMS SOFTWARE BUDNLE

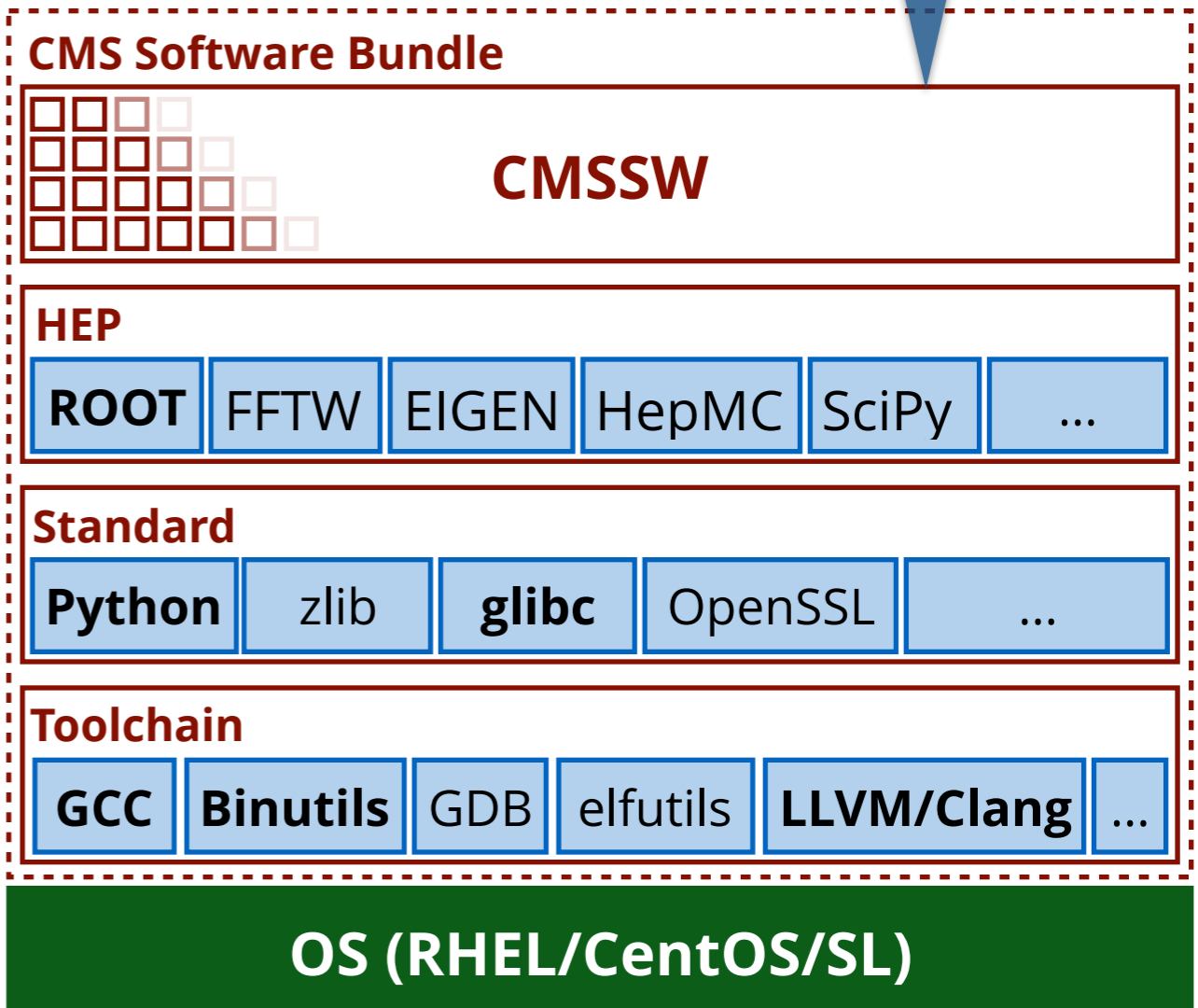The actual application software for "pattern recognition", "simulation", etc.

CMSSW is **open-source** and available at GitHub

Mostly written in **C++14**, C, **Python** and **Fortran**

CMSSW is like **Software Collection** package or **Linux Container** without actually being any of them

Quick comparison:

CVMFS

CMS Software Bundle

CMSSW

HEP					
ROOT	FFTW	EIGEN	HepMC	SciPy	...

Standard					
Python	zlib	**glibc**	OpenSSL		...

Toolchain					
GCC	**Binutils**	GDB	elfutils	**LLVM/Clang**	...

OS (RHEL/CentOS/SL)

	CMSSW	**Firefox**
SLOCs	6M	7M
Initial Release	2005	2002
Contributors	>1300	>1200
Memory Footprint	~2GB	~0.3GB

Other CERN developed software would increase SLOCs

ROOT6 w/o Clang: 1.7M
GEANT4: 1.1M

PRODUCTS THE SPLIT

GENERAL PURPOSE (64-BIT)

Xeon Phi/MIC

PowerPC

ARMv8

RISC-V

‣ If architecture provides **LP64 data model and LE mode** it's mostly recompilation that is required to run (does not mean optimal performance).

‣ Supports "legacy" applications without maintainers as long as no assembly or/ and compiler intrinsics.

‣ Known toolchain (GCC/Clang/binutils) with same C and C++ support.

ACCELERATOR

Xeon Phi/MIC

GPGPU

FPGA

DSP

‣ Requires increased effort for new data formats, algorithms and data management to run and achieve optimal performance on given hardware.

‣ Might need to learn **CUDA**, **OpenCL**, **OpenMP**, **OpenACC** or **any other wrapper library** to "talk" to accelerator.

‣ Might require learning different language to exploit accelerator.

SERVERS THE DENSITY

COMMON

2U chassis with 4 nodes, each is 1U half-width with 2 sockets, e.g. Intel Xeon

Powerful accelerators (GPGPU, FPGA)

HIGH DENSITY

4.3U HP Moonshot with 45 micro-servers/blades with Intel Atom or APM X-Gene.

3U SuperMicro MicroCloud SuperServer with 24 nodes with Intel Xeon or Atom

Some accelerators (DSP, GPGPU, FPGA)

SUPER HIGH DENSITY

2U 128 node, hot water cooled, NXP PowerPC (BE) and ARMv8, SKA DOME micro-server

PowerPC BE version 1536 cores, 3072 threads and 6TB of RAM, ~6kW

10

SUMMARY

‣ CMS together with other LHC experiments works on the current and future computing chips from different vendors via CERN Openlab

‣ We are also involved with open source communities and other industry partners

‣ Application diversity could drive heterogeneity to aid in **{performance, power, cost}** optimizations

‣ Power constraints and market evolution **may drive change in the kinds of processors we use**

‣ The **race is heating up**, and **Intel/platform vendors are not sitting idle**

‣ We are constantly working on porting/rewriting/redesigning specific parts from software stack to use new software frameworks and future hardware

‣ We want to increase throughput density, lower power usage for computing and cooling, etc.

‣ **Contact:** davidlt <at> cern <dot> ch