Horizon 2020 ICT 04-2015 Call

- Information and Communications Technologies
  - Research and Innovation
  - 14 April Call: **Customized and low power computing**

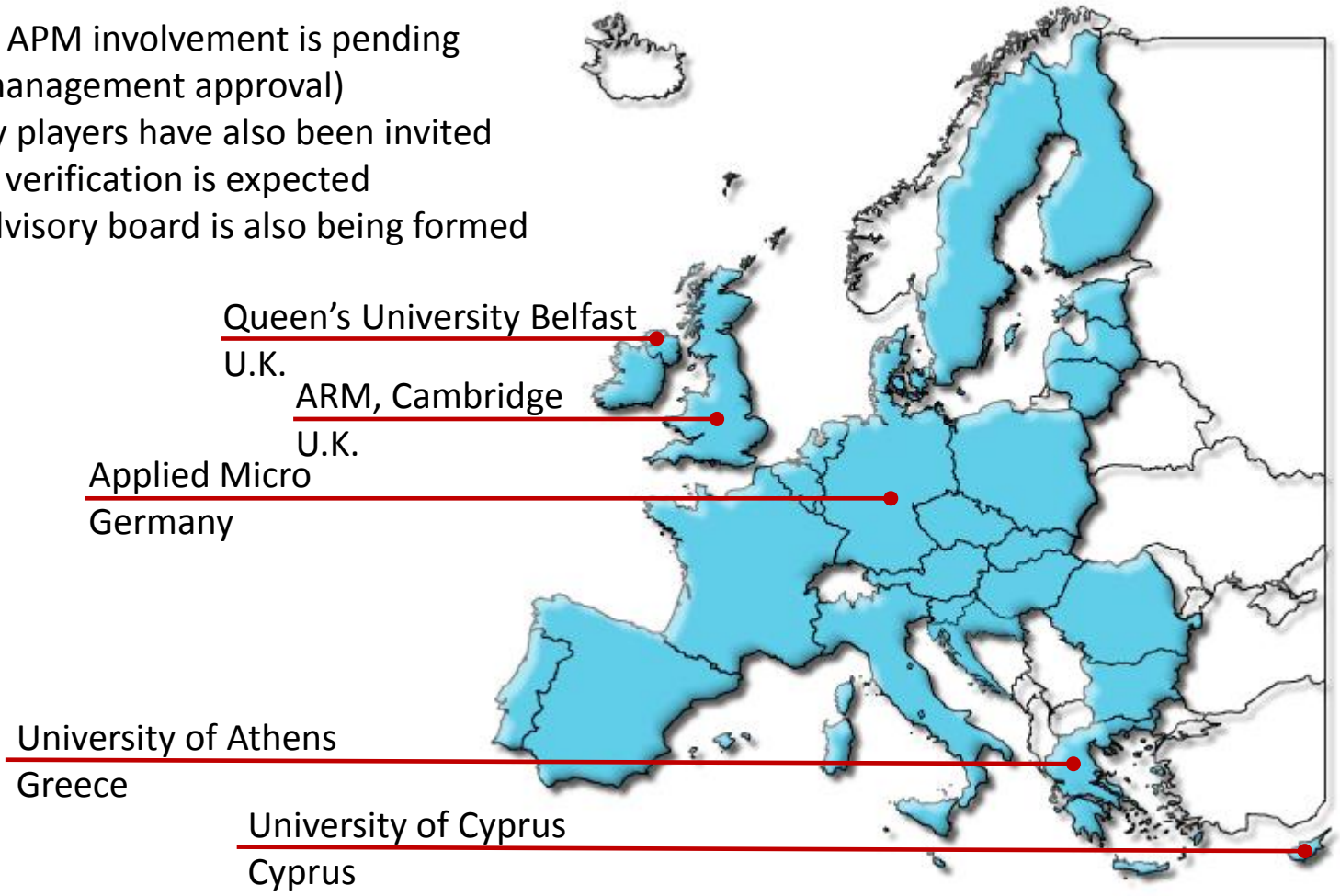
Next generation servers, micro-server and highly parallel embedded computing systems based on ultra-low power architectures. **The target is highly performing low-power low-cost micro-servers, using cutting-edge technologies** like, for example, optical interconnects, 3D integrated system on chip, **innovative power management**, which can be deployed across the full spectrum of **home, embedded, and business applications**. **Focus is on integration of hardware and software components into fully working prototypes** and including validation under real-life workloads from various application areas. Specific emphasis is given on **low-power, low-cost, high-density, secure, reliable, scalable small form-factor datacentres** ("datacentre-in-a-box"). Proposals requesting a Large contribution are expected.

  - **Key requirements:** Interdisciplinary, Cross Layer, Spread out in Europe, Industrial partners and Small scale companies (SME) act as drivers

http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/9080-ict-04-2015.html#tab1

Project Consortium

- ARM and APM involvement is pending (upper management approval)
- Other key players have also been invited and final verification is expected
- Strong advisory board is also being formed



Project Motivation

- The miniaturization of transistors lead to:
  - Static parameter variations
  - Dynamic parameter variations (due to temperature, workload, aging)

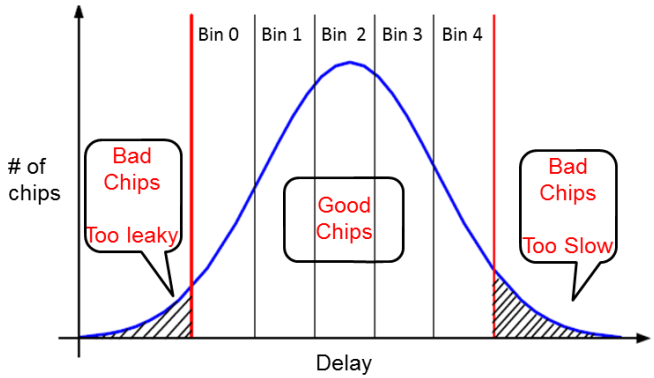
- Such variations might lead to:
  - Timing failures (various control and data instructions failing)
  - Memory failures: - Cache - SRAM read/write failures
    - Data retention time failures in DRAM – main memory

  Manifest as : Errors Correctable, Detectable, Undetectable (SDC), system crash

- Traditionally avoid such failures by:
  - Guardbands/Margins by running at higher **V**oltage or lower **F**requency based on the worst case datapath within all the pipeline stages
  - Overdesign of memory cells by sizing-up, allowing a higher Vddmin
  - Refreshing at a high rate based on the retention time of the worst case cell

Project Motivation

- Such approaches lead to:
  - High performance penalty and power overhead which affect ALL manufactured chips (CPUs, memories) even the good ones



- Vendors and ITRS start doubting the viability of such approaches
  - Soon the margins and resulting power/performance overheads put in doubt the continuation of technology scaling

- Alternative design paradigms have been proposed
  - Better than worst case (insert special FFs in each pipeline reg to detect errors and operate at lower voltage, in case of an error re-execute)
  - Fault tolerant SW (catch HW errors at SW and correct them)
  - Approximate computing (exploit the resilience of applications and allow errors induced by lower voltages or approximate units)

Project Motivation

- All such methods although very promising they are very intrusive requiring substantial changes in the HW design flow/tools, modification of the SW stack, re-design of the HW
- Therefore commercial processor and middleware vendors have not adopted them not willing to change their proved infrastructure/tools/flows
- New commercial platforms (e.g. APM x-gene, ARM Juno, AMD Seattle) may allow to play with V,F but the ranges are very small still based on the worst case path/cell
- More than 40% performance penalty due to margins (e.g. INTEL ISSCC 2011)

- WHY 'bad' guys to set the limits of today systems?
- Anyway we are learning to cope with heterogeneous systems (commercial examples big.little ARM, cpu-gpu products)

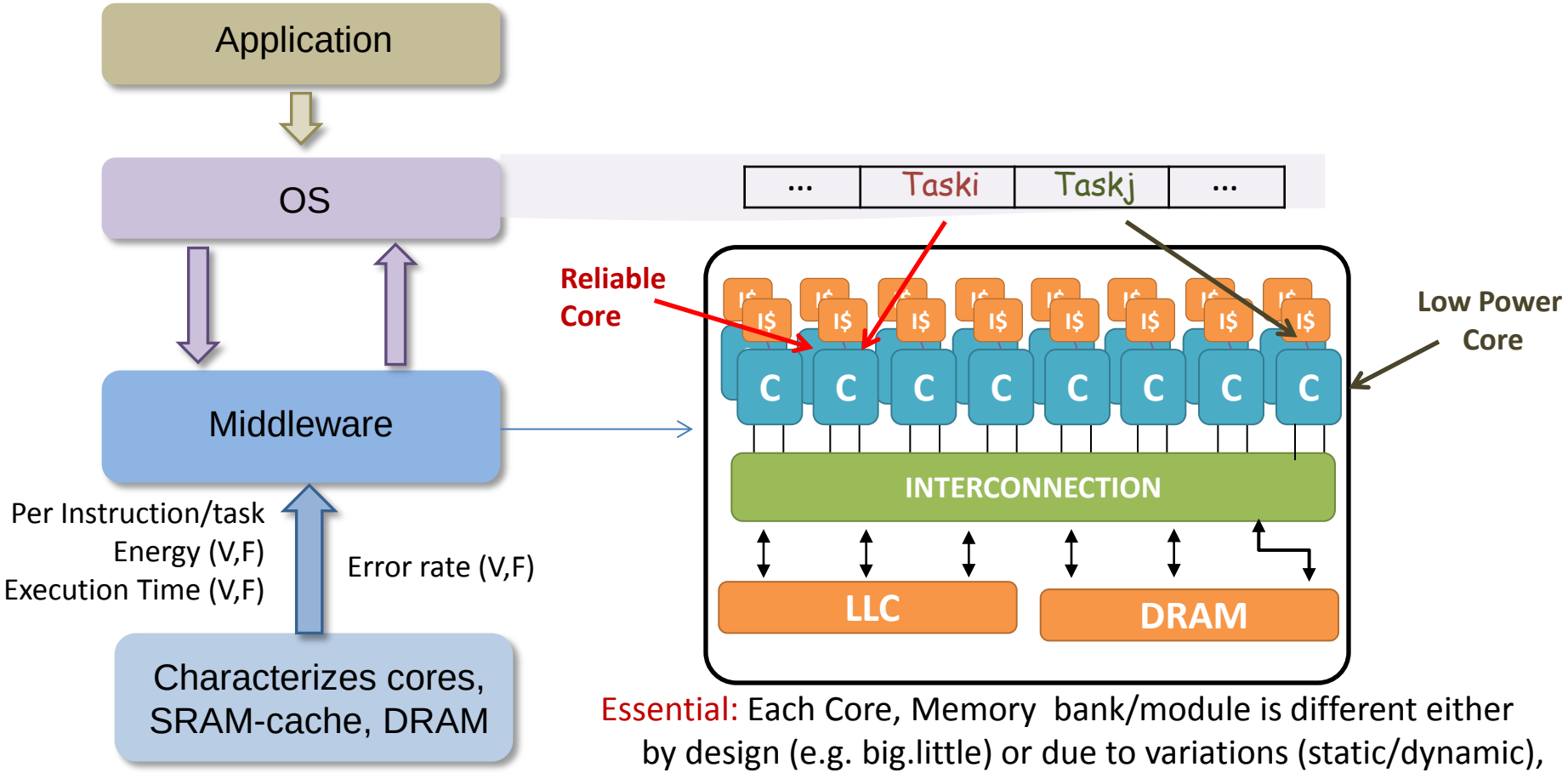
  => Why not learning to cope with not so good/very good cpus/memories in the same way ?
- Application resource utilization is low and non-uniform (embedded, home, DC)
  - low load, strict QoS, time of day

Project Key Idea

Develop an automatic methodology for characterizing and controlling the margins of commercial platforms in different market segments when running any application to improve energy efficiency while maintaining QoS and high availability

- Characterize the F,V, retention time limits of existing platforms (i.e. APM x-gene) and reveal them to the SW
- Learn by exploring operation outside margins through data mining and correlation analysis. Predict potential failures
- SW based control methods for logging/filtering data and controlling margins (firmware/OS/Applications)
- Challenge: Harness Unreliable Operation through whatever available on platform
  - Lower availability, Lower quality, Data corruption (SDCs), Wear-out

Project Key Idea – Top Down



Essential: Each Core, Memory bank/module is different either by design (e.g. big.little) or due to variations (static/dynamic), multiple voltage domains

- ✓ Characterize Energy - Error profiles of cores and memories in commercial platforms and develop models that could easily be integrated easily (i.e. as SW accessible LUT)
- ✓ Exploit the energy, performance trade offs at the firmware/OS/runtime/hypervisor
- ✓ Adapt energy/reliability of each core based on user requirements and conditions
- ✓ Ensure robust execution of OS

Project Pillars

- Tune/set voltage and frequency (firmware and OS layer)
- Define V-F states of processor (BIOS ROM or on the fly)
- Voltage monitoring for droops
  - coarse off-chip (10KHz)
  - fine grain off-chip and on-chip (GHz)
- Different resources on different voltage domains (offer isolation)
  - single-chip: big, little, gpu, LLC
  - multiple sockets, nodes
- Clock gating/Stall
- Memory refresh rate
- Performance counters (proxy of behavior, e.g. system calls)
- Error logs (e.g. correctable errors)
- Temperature monitoring
- Log information in ROM (SSD) (track failures across crashes etc.)
- Software drivers/handlers that give access to observe/control the above

Project Objectives-Impact-Vision

- Improve energy efficiency by allowing operation at lower voltage (and same Frequency) for the low end/embedded applications
- Improve performance by allowing operation at higher frequency (and same voltage) for the high end/HPC applications

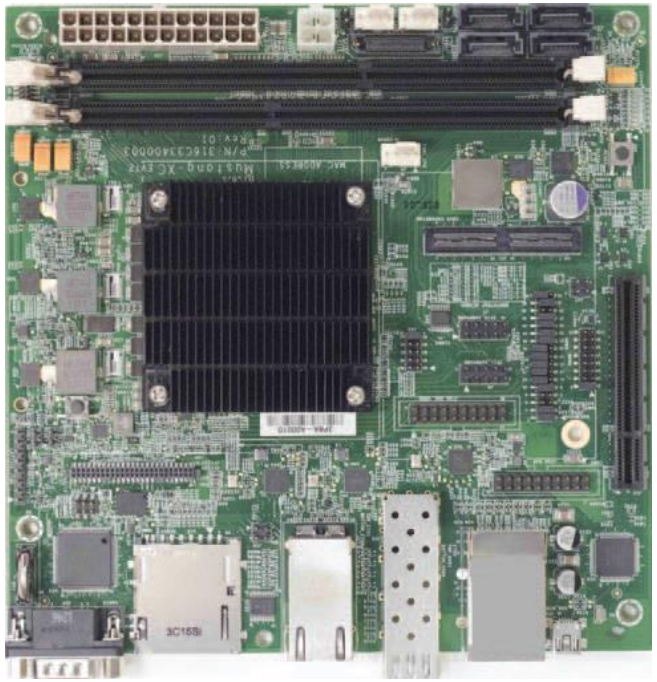
- Unique methodology for:
  - revealing the margins in any commercial platform
  - utilizing such margins at the SW level with minimum/no intervened HW techniques
  - exploiting the new V,F ranges through a unique middleware and new policies treating any underlying core and memory module as independent nodes
  - working at the limits if required by predicting potential failures through identified signatures in popular systems (e.g. ARM based)

In the end be able to execute any application at a lower energy or higher performance based on its requirements. Develop an innovative power management layer useful for platform in any market segment with minimum interventions and thus minimum cost.

One Project Platform: APM X-C1

- Can be used to built data center in a box
- 8 superscalar 2.4Ghz 64bit ARM-V8 cores
- 2x 8GB DDR3
- Supports Ubuntu, Fedora, and OpenSuse
- DVFS, clock gating, and C states
- RAS features – ECC, error isolation
- KVM is ported

- APM very interested on the project and agreed on participating
  (final verification pending)



http://www.hotchips.org/wp-content/uploads/hc_archives/hc26/HC26-11-day1-epub/HC26.11-4-ARM-Servers-epub/HC26.11.430-X-Gene-Singh-AppMicro-HotChips-2014-v5.pdf

APM Active Participation

- Develop a custom datacenter-in-a-box with enhanced capabilities housing the developed project technologies. The board will be equipped with the following: (not available to existing APM boards)
  - More extensive debug capabilities.
  - Explicit support for on-die voltage sense pins connecting to multiple voltage domains on chip which can be probed with a high-bandwidth oscilloscope and high-impedance active probe.
  - Explicit support for voltage setting.
  - Explicit support for frequency scaling - Overclocking is supported.
  - Error monitoring (what level/parts to be verified)

- APM offered to the project 4σ, 5σ chips (dies that are either going to be too fast and power hungry or too slow) in case it is difficult to cause failures

Partner in the Project - Role

- CERN
  - assess the benefits of the proposed power management layer in a commercial ARM platform running some of your service(s) (e.g. Filtering data and Data analytics)
    - port the applications on an ARM based commercial platform used in the project.
  - Very useful to tune your SW stack to become more energy aware based on the information provided by the layer we will develop in the project.
    - Enhance job scheduler to map jobs in a more energy aware manner

Partner in the Project - Gains

If **CERN** participates in the project then you will have the chance to

- Explore the margins existing in an ARM based platform and determine what could be saved in performance or energy for these platforms
- be the first to utilize and influence the developed technologies to meet your needs (and reduce operational costs)
- utilize the developed middleware within the platform and influence its development
- evaluate your workloads for achieving higher speed or lower energy via leveraging margins
- take advantage of the know how in various fields of world leading partners and direct them according to your needs

Participate in the development of an innovative sw based power management layer that enables the most energy/performance efficient datacenter-in-a-box based on 64-bit ARM processor, equipped with the latest hardware technologies and support for the overall software stack and latest linux distributions which we are going to further enhance.