French Technological Watch Group: General overview and first results

gabriel.hautreux@genci.fr

FRENCH TECHNOLOGICAL WATCH GROUP

Led by GENCI and its partners

FRENCH TECHNOLOGICAL WATCH GROUP

Led by GENCI and its partners

- **Share and mutualise** expertise of partners at the national level

- **Anticipate upcoming** Exascale architectures and provide orientations for future procurements
  - ⇨Manycore/heterogeneous processors
  - ⇨Deeper and more complex memory hierarchies
  - ⇨Fault Tolerance
  - ⇨Energy optimisation

- **Organise code modernization** for scientific communities
  - ⇨Ease migration to these new platforms
  - ⇨Preserve code legacy by using standards – OpenMP
  - ⇨Enable potential specific optimisations with low level languages

TECHNOLOGICAL WATCH GROUP
Led by GENCI and its partners

Link to developers of representative applications

- **Climate** : DYNAMICO, MesoNH
- **Astro & Geophysics** : RAMSES, EMMA, SPECFEM3D
- **Combustion** : YALES2, AVBP, TRUST
- **Fusion** : GYSELA5D
- **Materials** : METALWALLS
- **Particle Physics** : SMILEI, CMS-MEM
- **Kernels** : PATMOS, HYDRO, QR_MUMPS, QMC=Chem
- **Deep/Machine Learning**

Link to developers of tools and libraries

- System (runtime, checkpoint/restart …)
- Profiling perf, energy
- Numercial solvers
- Data management/analysis

Core of the group : HPC expertise

- National centers: CINES, IDRIS and TGCC
- Inria
- Maison de la Simulation
- « Groupe Calcul »
- GENCI

- Vendors support

Direct link between communities

FRENCH TECHNOLOGICAL WATCH GROUP
Led by GENCI and its partners

- Deployment of 2 early technology platforms
  - A **Bull sequana platform at CINES based on Intel® KNL**
    - **48 KNL nodes => 146 Tflop/s peak performance**
  - An **IBM OpenPOWER platform at IDRIS based on P8+ Nvidia GPU**
    - 4*P100 GPU (latest generation) per node
    - Nvlink between P8 and GPU
    - **12 nodes => 254 Tflop/s peak performance (GPU only)**
    - **More than 400GB/s bandwidth per node**
    - *Software stack is not at its highest maturity*

- Tight collaboration between :
  - Vendors, integrators, developers, HPC centres

- Evaluation of real time performance/energy profiling tools
  - Allinea - Performance Report / MAQAO / Others
  - Intel® Vtune™ Amplifier and MPI Performance Snapshot (MPS)
  - Nvprof for the OpenPOWER solution

SOFTWARE ENVIRONMENT
Programming Options for OpenPOWER

- Programming on Intel Knight Landing
  - Straightforward:
    - **Intel Compiler, Intel MPI, OpenMP**

- Programming options for OpenPOWER
  - Not as straightforward as Intel KNL
    - General availibility for OpenMP is scheduled for the end of 2016
    - Full support is scheduled for 2017

- **High-level pragma languages**
- Easy porting by adding pragmas to legacy code
- Not necessary to deal with low-level details
- **Performance portable**
- Some low-level abstractions are exposed
- Complex code synthesis performed by compiler
- Code transformations available
- Requires compiler support
- Some low-level details are unnaccessible

- Low-level languages
- **Full control over architectural details**
- Many application-specific optimizations shown in literature
- Cumbersome programming
- Not portable to other architectures
- Limited support

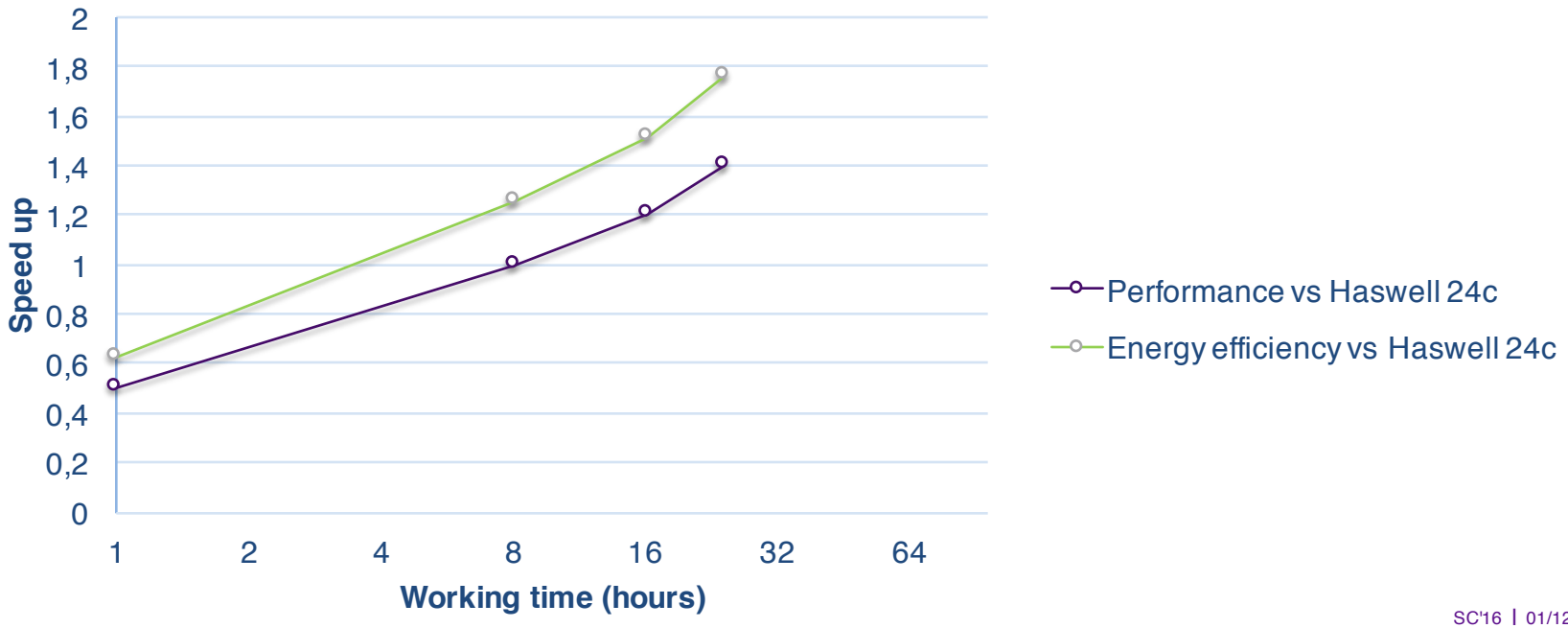
	OpenMP	OpenACC	CUDA
Language or Pragmas	**Pragmas**	Pragmas	Based on C/C++ or Fortran
High-Level Constructs	**Parallel, loop-like, simd, tasking, etc.**	Parallel, loop-like, simd	None
Companies/Organizations Actively Involved	**IBM, AMD, Intel, Cray, Pathscale, Texas Instruments, etc.**	Cray, PGI, gcc	NVIDIA, PGI, IBM
Features	**Architecture-independent Flexible Parallelism**	Targeted for GPU-like accelerators	Only available on GPUs

KNL PORTING AND OPTIMISATION
First results

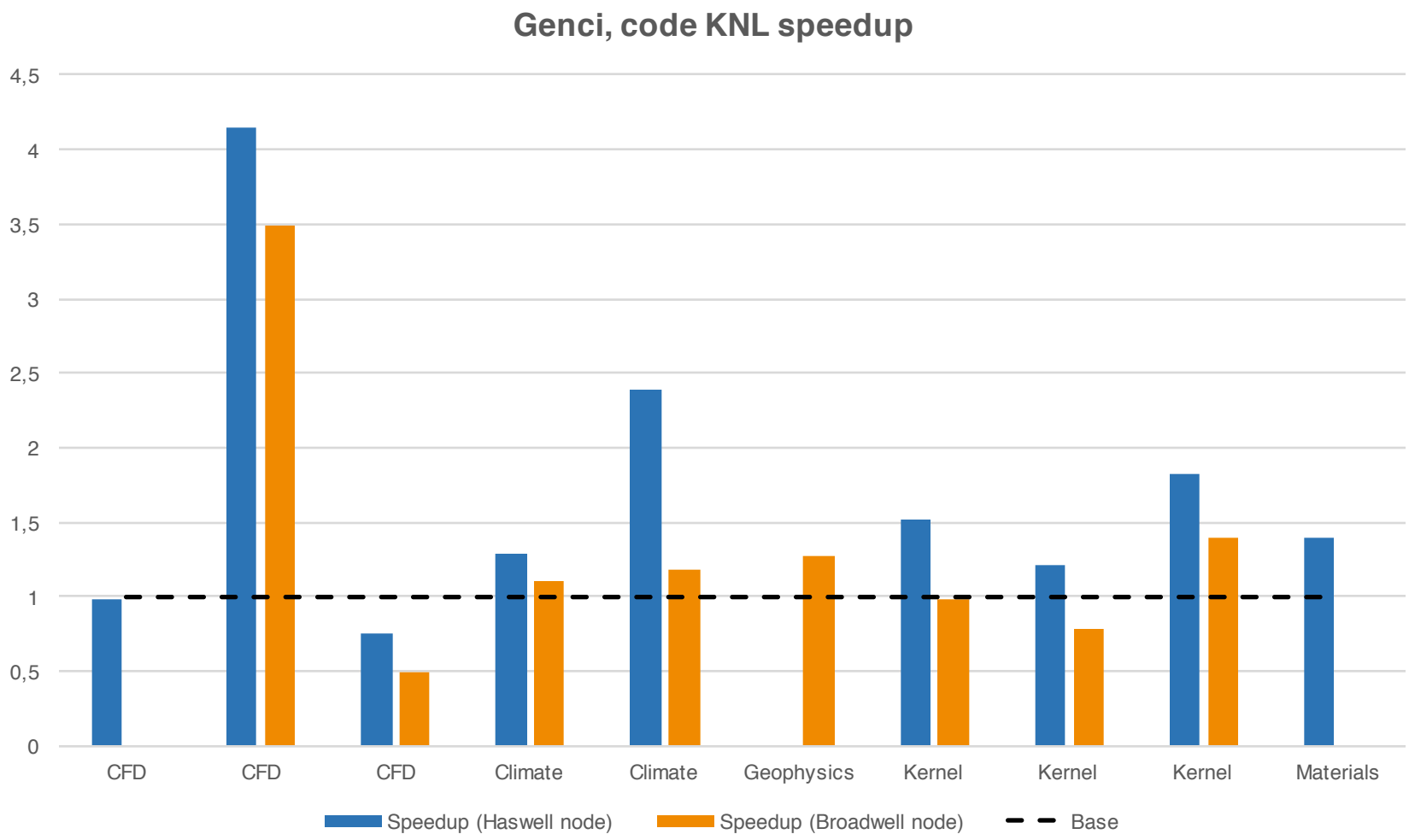
- Porting & optimisation workshops with Atos & Intel support
  - Workshop on Colfax® Ninja developer platforms (Intel processor Xeon Phi 7210)
  - Workshop using the 48 Intel Xeon Phi 7250 nodes machine “ Frioul ”

- Porting efficiency:

Effective mean speed up obtained after 2 workshops (Haswell vs KNL)



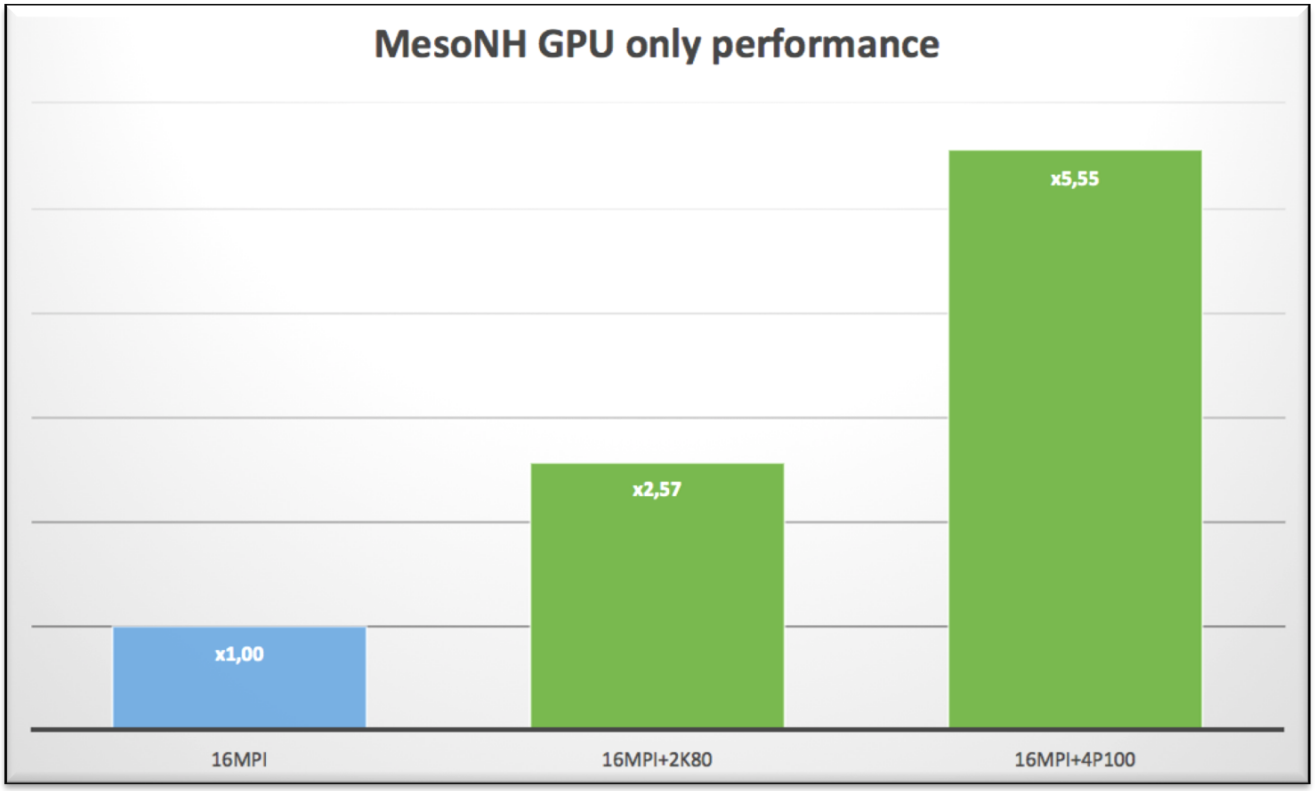
KNL PORTING AND OPTIMISATION
First results



OPENPOWER PORTING AND OPTIMISATION

First results

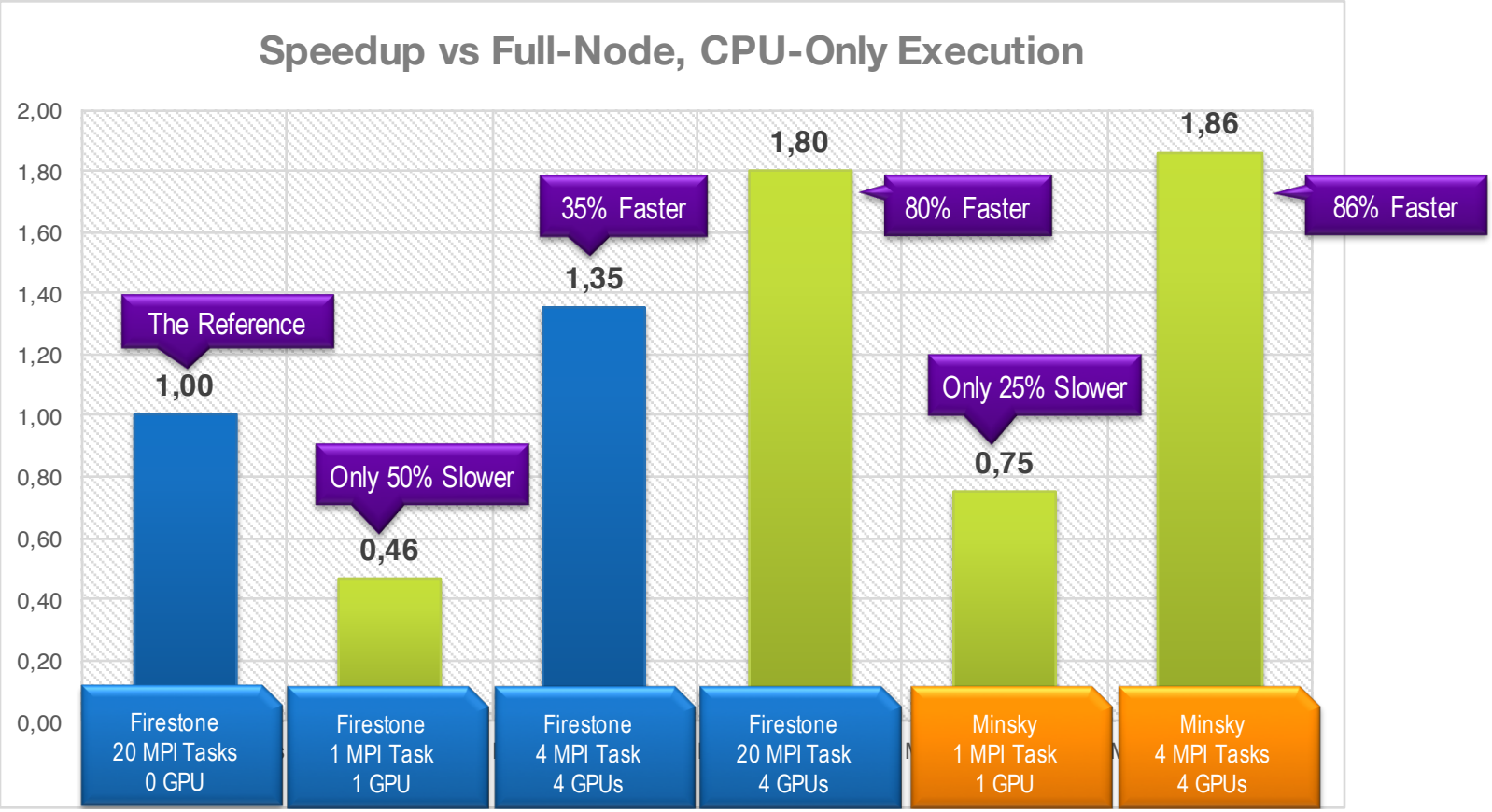
- MesoNH: weather forecasting code
  - 60% of the code has been ported using OpenACC
  - Results with 16 MPI processes and MPS (multi-process service)



OPENPOWER PORTING AND OPTIMISATION

First results

- Vasp (an IBM ported app)
  - Results reflect the first results, the whole application still have to be ported



FRENCH TECHNOLOGICAL WATCH GROUP

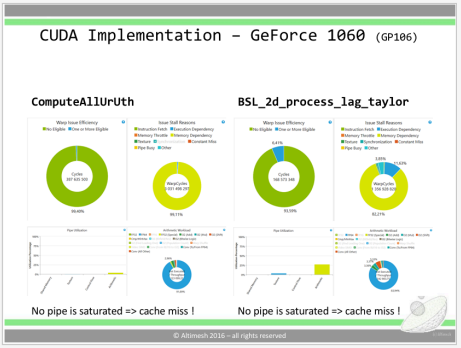
Focus on Gysela

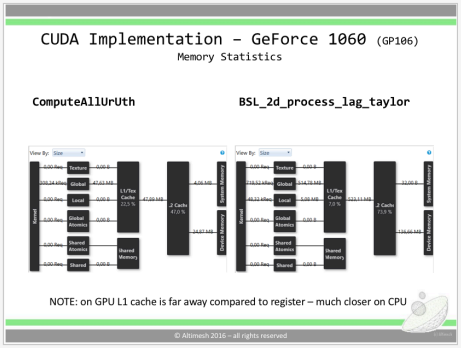
- Work has just began on GPU with Altimesh
  - Very first results

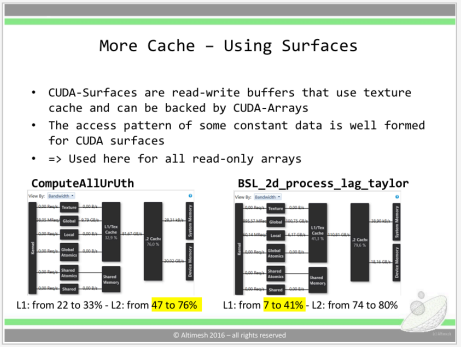
- First lessons (= general good practices)
  - Avoid buffers:
    - That store addition/multiplication
    - That are reused once

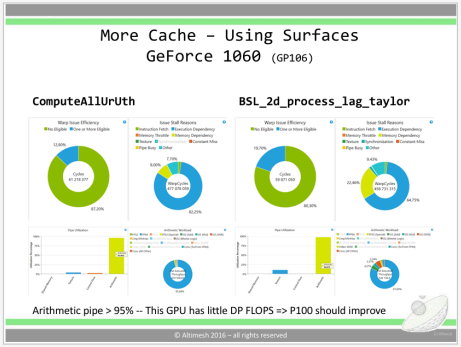
  →Prefer recomputing than precomputing

  - Use stack vs heap
    - Malloc is very expensive (especially with many cores)
  - Use "zero-copy" for data that you acces/write once





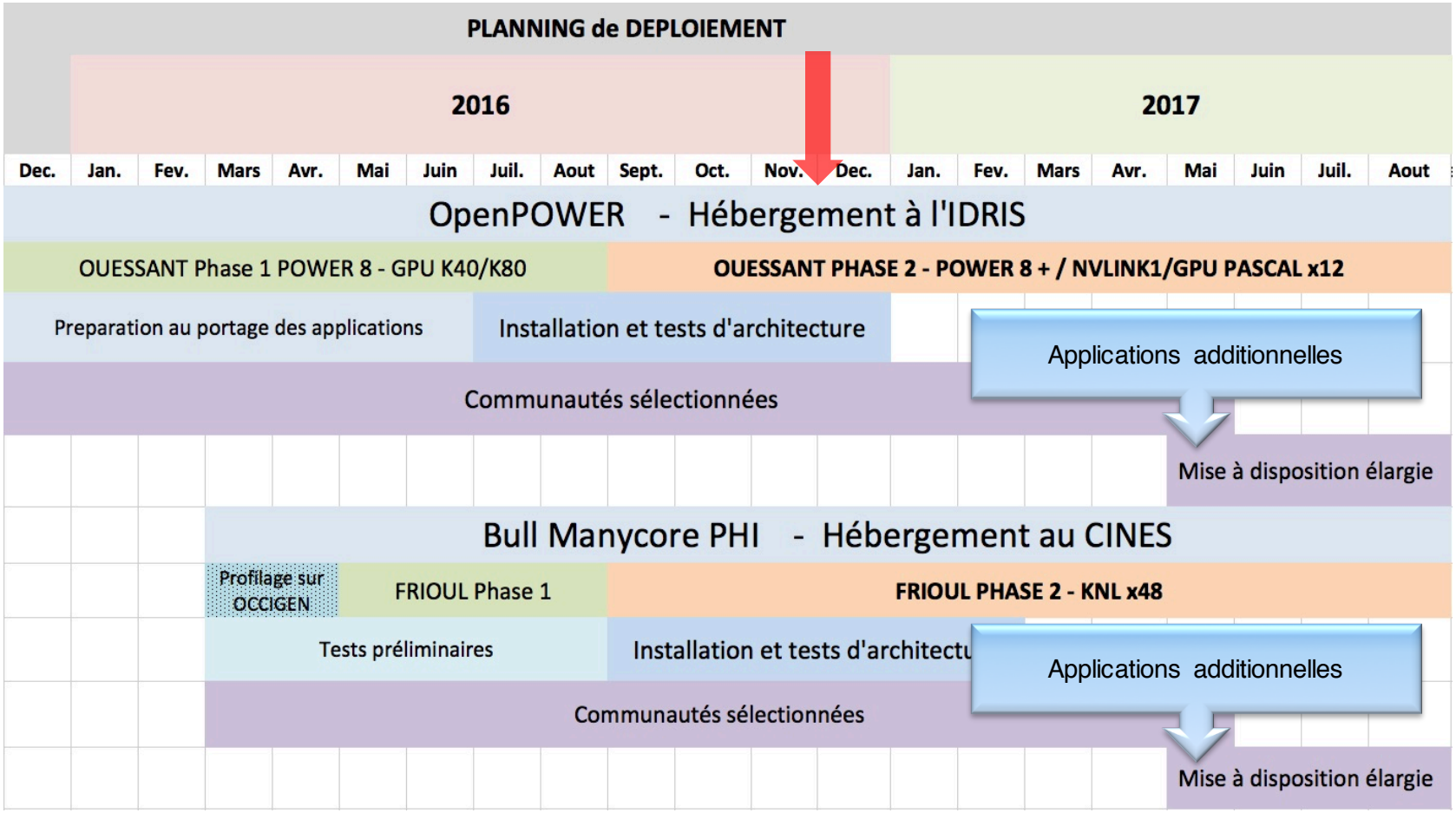




TECHNOLOGICAL WATCH GROUP OPENING

Phase 3 – Q2 2017

- Opening to a wider panel of applications in May 2017
  - Applications will be reviewed through the « DARI » preparatory access



FRENCH TECHNOLOGICAL WATCH GROUP

Questions?

Thank you for your attention