High-Performance Computing (HPC)

What is it and why do we care?

Partners

Funding

Horizon 2020
European Union Funding
for Research & Innovation

Reusing this material

Defining HPC

Q: What is high-performance computing?

Defining HPC

Q: What is high-performance computing?

A: Using a high-performance computer (a supercomputer)…

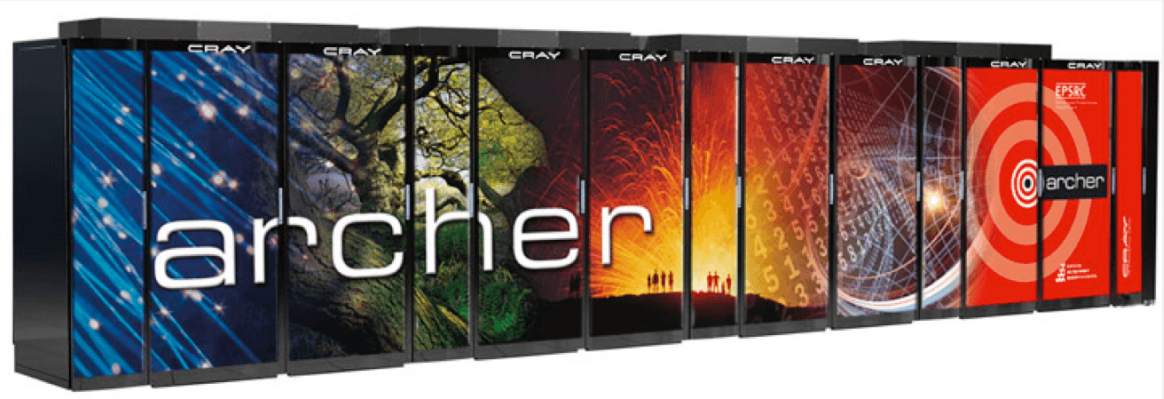
Defining HPC

Q: What is a high-performance computer?

Defining HPC

Q: What is a high-performance computer?

A:



Defining HPC

Q: What is a high-performance computer?

A: a machine that combines a large number* of processors and makes their combined computing power available to use

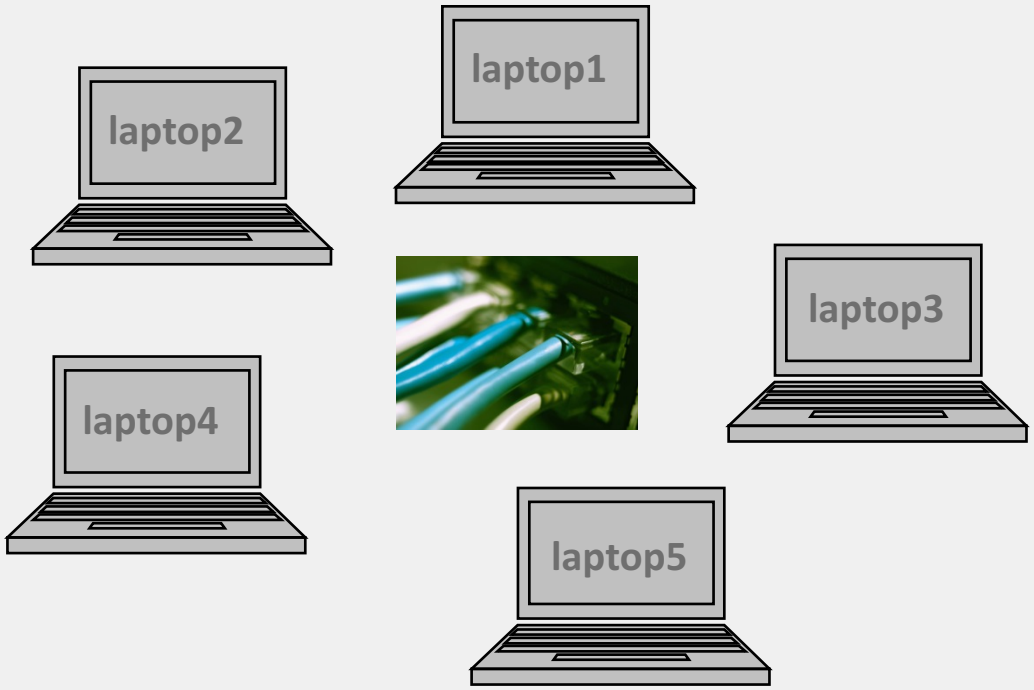
Based fundamentally on *parallel computing*: using many processors (cores**) at the same time to solve a problem

* this number keeps on increasing over time

** define cores vs processors clearly in lecture on hardware building blocks

Generic Parallel Machine
(computer cluster)

- Rough conceptual model is a collection of laptops
- Connected together by a network so they can all communicate

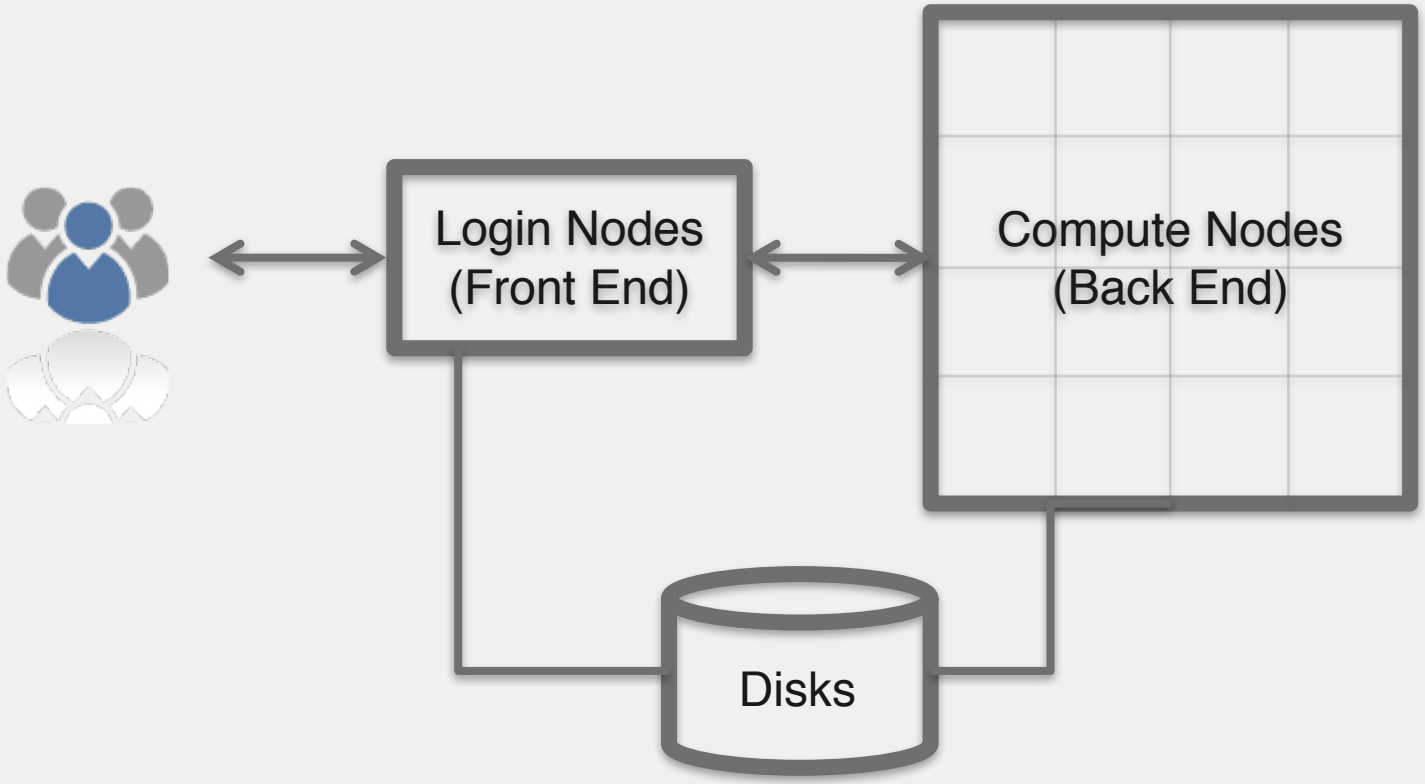


- Each laptop is a *compute node*
  - has a processor, hard disk, memory, …
  - Each runs a copy of an operating system (Linux)
  - If each processor has 4 cores, total system has 20 cores

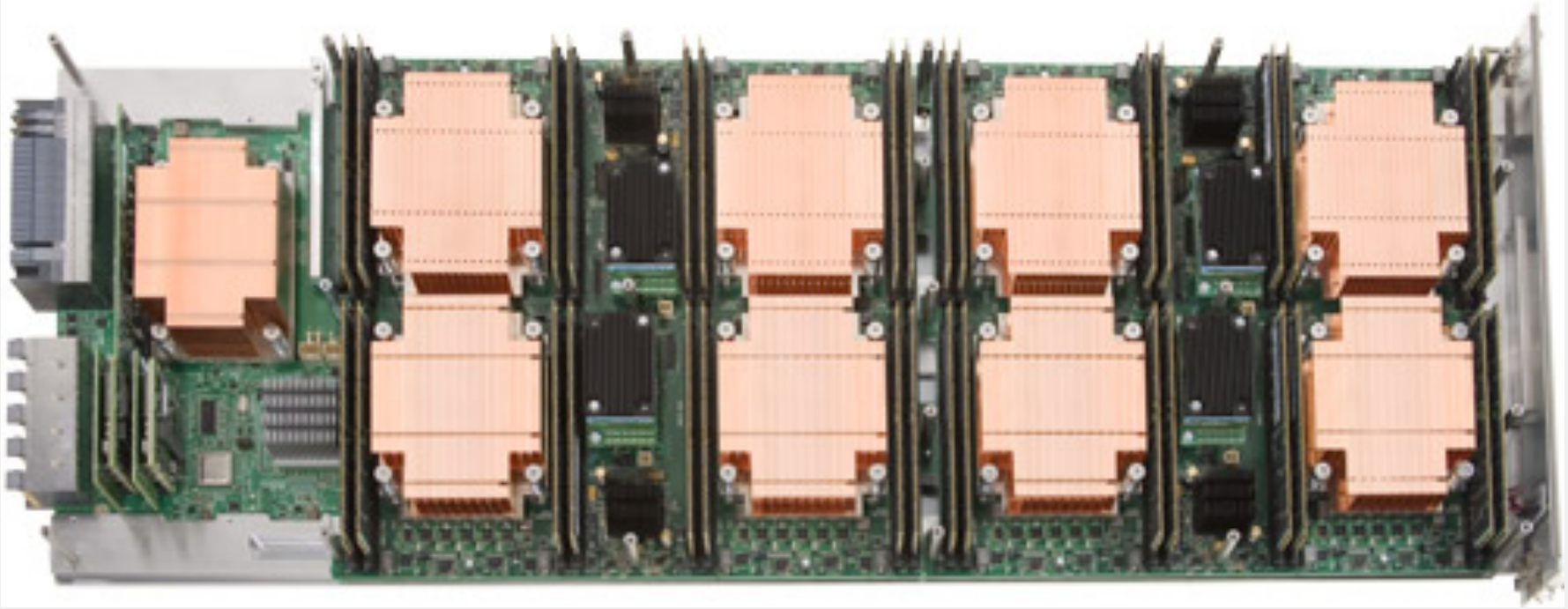
HPC architectures

- Majority of HPC machines follow this generic conceptual layout for a computer cluster:
  - many compute nodes connected together by a network
  - each compute node has separate, independent memory
- Some smaller HPC machines allow many processors to all access the same shared memory
  - allows some software to run in parallel with fewer modifications
  - convenient for many data-intensive applications
    - including bioinformatics/genomics
  - difficult / expensive to scale this approach to very many processors

Typical HPC system layout



HPC hardware (ARCHER - Cray XC30)



- One blade has four compute nodes
- Each node has two processors and 64GB of memory
- Each processor has 12 cores

HPC hardware (ARCHER - Cray XC30)

A blade being slotted into a cabinet.

Each cabinet can hold up to 48 blades

ARCHER has 26 cabinets

The scale of HPC

- ARCHER has:
  - ~5000 nodes
  - ~118 000 cores
- like 30 000 quadcore laptops connected together (!)
- Largest systems globally currently (2017) have hundreds of thousands up to millions of cores
- HPC systems offer a large amount of computing power
- Dominant platforms for computational science

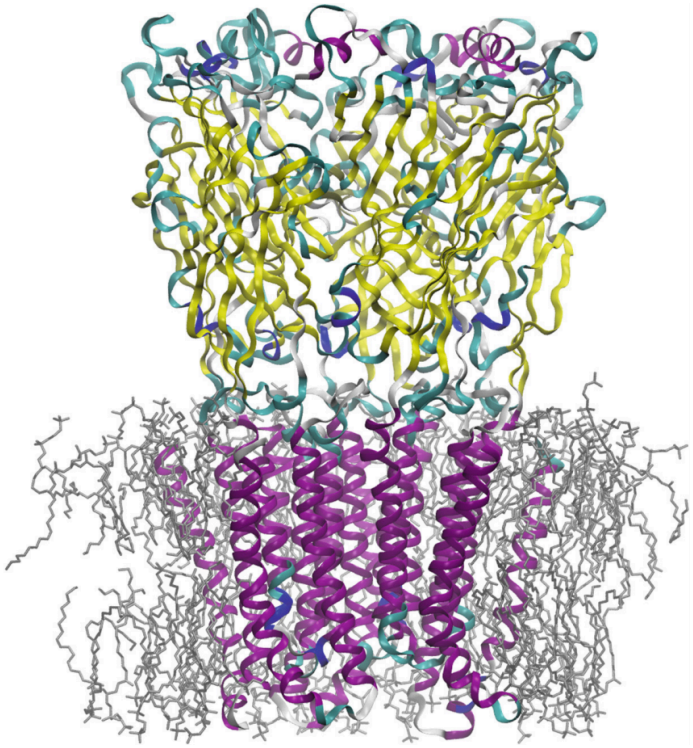
Who uses HPC?

- Traditionally used for:
  - materials science / solid state physics
  - computational chemistry
  - biomolecular simulation
  - particle physics
  - environmental modelling
    - weather & climate
    - geosciences
    - oceanography
  - many engineering applications

e.g. on ARCHER see http://archer.ac.uk/status/codes/

Examples

biomolecular simulation:



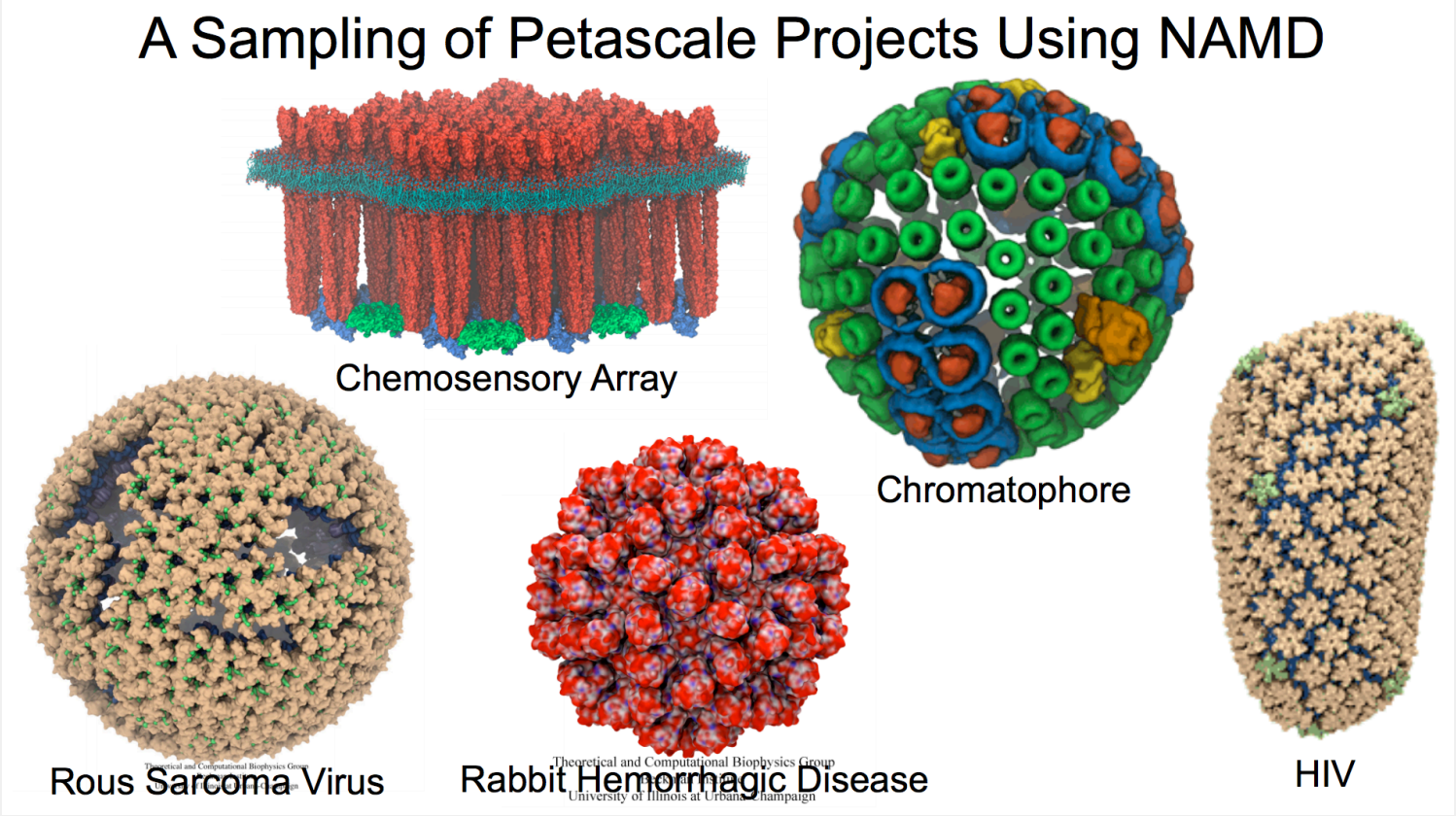
Ligand-gated ion-channel membrane protein GLIC (colored), embedded in a lipid membrane (grey), solvated in water (not shown)

145,000 atoms

Taken from:
https://doi.org/10.1007/978-3-319-15976-8_1
(Lindahl E. *et al.*) (2015) Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In: Markidis S., Laure E. (eds) Solving Software Challenges for Exascale. EASC 2014. Lecture Notes in Computer Science, vol 8759

Examples

biomolecular simulation:



HPC networking

- HPC relies on fast communications between nodes to allow applications to use many processors in parallel
- Networking hardware and software protocols need to be good enough to avoid becoming a bottleneck, and robust under heavy load
- Ethernet is cheap, used in small / lower performance systems
- High-end HPC systems use specialised network designs (e.g. Infiniband)
  - optimised communication protocols and connection topologies
  - special copper & fibre wiring
  - expensive!

HPC vs other types of computing

- HPC is one extreme in a continuum of computing:
  - Individual desktop/laptop
  - University research group / departmental machine (server or cluster)
  - University-wide, regional or national-level HPC machine
- Commercial datacentres (Amazon, Google, Facebook, etc.) have enormous computing clusters
  - These do not cater for scientific computing
- HPC machines optimised for traditional science applications:
  - strong floating-point performance (“number crunching”)
  - fast networking
  - software stack that includes scientific / maths libraries

HPC vs other types of computing

- HPC offers *capability computing*:
  - ability to solve very large / complex scientific problems quickly

- Can also use HPC for *capacity computing:*
  - many small / simple problems
  - this may be cheaper on generic computing clusters or cloud computing!
    - usage of ARCHER is charged starting at a minimum of 1 node (24 cores)

HPC and me

- HPC platforms try to cater for a broad range of computational needs, and have co-evolved with their user communities
  - Optimised for certain problem categories
  - Established usage models and conventions
- Scientific software and its users have also had to adapt to make use of HPC
  - Molecular dynamics software heavily optimised to use HPC machines
- HPC continues to evolve
  - Data-centric computing increasingly important (bioinformatics a big driver)