IT – the enabling technology for cutting-edge research
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
Outline

• **A photon factory (a bit of physics)**
• Science examples
  • Life Sciences
  • Catalysts
  • Glass
• SR science and HPC
• EU collaborations
• Outlook/Conclusions
What is a Synchrotron?

A super microscope to examine condensed matter:

- Radiotelescope
- Eye & Microscope
- Synchrotron
- Colliders
Influence of the wavelength

Infrared

Visible

Ultraviolet

Extreme ultraviolet

X-ray

Gamma rays

Influence of the wavelength
Explore matter and materials

Synchrotrons are tools allowing to understand matter at the molecular and atomic scale (size and time) to solve numerous problems of applied and fundamental science.
From the X-ray tube to ESRF

Brilliance = photons/s/mm²/mrad²/0.1%BW

The ESRF X-rays are **1000 billion times** brighter than a typical hospital X-ray source.
Synchrotrons

- Storage ring
- LINAC and booster
- Beamline
- Insertion devices
The Storage Ring

Bending magnet

Insertion Devices (undulators)

Focusing magnets
Bending magnet

Electron bunch

Emission spectrum

Number of photons vs. Wavelength (Å) vs. Energy (eV)

- Bending magnet
- X-ray tubes
- Sun

Energy levels:
- 1 eV
- 10 eV
- 100 eV
- 1 keV
- 10 keV
- 100 keV
**Insertion devices**

- **Electrons**
- **Bending magnet**
- **Wiggler**
- **Undulator**

**Brilliance**

$$\text{(photons/s/mm}^2\text{/mrad}^2/0.1\%\text{BW)}$$

**Graph:**

- **Energy (keV):**
  - **2**
  - **10**
  - **50**

- **Brilliance:**
  - **10^{14}**
  - **10^{15}**
  - **10^{16}**
  - **10^{17}**
  - **10^{18}**
  - **10^{19}**
  - **10^{20}**
Closed orbit feedback system

- Electron bunches circulate at ~297 397 700 m/s
- Storage ring circumference = 844.4 m
- 1 bunch makes 352 200 turns/s
- 224 Libera DSPs to correct the orbit in real-time
100 nm vertical stability routinely achieved
1 μm horizontal stability routinely achieved

Feedback OFF

Horizontal OFF
Vertical OFF

Horizontal ON
Vertical ON
Elements of a beamline

- Control hutch
- Experiment hutch
- Optics hutch
Sample Mounting Diffractometer

Detector

Diffractometer
More than 50 synchrotron light sources worldwide
20 new synchrotrons in the last 15 years
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
Progress of mankind is often linked to a better understanding of matter.

Stone age

Bronze age

Oil age

Silicon age

Steel age

R. Dimper - eResearch Australasia 2012
Societal Challenges

- Understanding the world around us
- Solve societal challenges in
  - Health
  - Energy
  - Materials
- Allow for sustainable living conditions on earth
Outline

• A photon factory (a bit of physics)
• Science examples
  • Life Sciences
  • Catalysts
  • Glass
• SR science and HPC
• EU collaborations
• Outlook/Conclusions
Bragg’s law: \( 2d \sin \theta = n\lambda \)

- \( \lambda \): wavelength
- \( \theta \): incident angle
- \( d \): distance between 2 atomic planes
- \( n \): integer

First X-ray diagram obtained by Max von Laue in 1912
Life Sciences – Protein Crystallography
ever higher resolution and shorter exposure times

- Diffraction pattern
- Electron density cloud
- 3D structure
Screening & Automation for Macromolecular Crystallography

- Available protein crystals become smaller and smaller
- Interesting complexes become larger
- 2009 figures (>150,000 samples tested) are expected to increase by more than a factor of 10
- Massif automation and nano beams are needed
Ribosomes translate DNA information into proteins.

Knowing the structure of the ribosome allows to understand protein synthesis.

Without functional ribosomes, bacteria cannot survive. This is why ribosomes are an important target for new antibiotics.

The Nobel Prize in Chemistry for 2009 awards studies on the structure and functioning of ribosomes to Ramakrishnan, Steitz, et Yonath.
Placement of ions in large structures such as the ribosome shown here previously required the use of HPC clusters for calculation but can now be performed on GPUs in just a few minutes.

Source: Communications of the ACM, October 2009, Vol.52, No. 10
Professor Ada Yonath with some members of her team at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Jacob Halfov (far left), Haim Rozenberg, Ella Zimmerwain, Itai Wekselman, Chen Davidovich, Anat Bashan and Yehada Halfon (far right).
G-protein–coupled receptors

The Nobel Prize in Chemistry 2012 is awarded to Robert J. Lefkowitz and Brian K. Kobilka for studies of G-protein–coupled receptors.

G-protein–coupled receptors (GPCRs) form a remarkable modular system that allows transmission of a wide variety of signals over the cell membrane, between cells and over long distances in the body. Today, we understand the molecular mechanism of how these receptors work in intricate detail, in large part because of the studies by Kobilka and Lefkowitz.
G-Protein Receptors

A

B

C

D

Hormone

GTP

GDP

cAMP

ATP
Computed Tomography

• X-ray computed tomography (CT) is an imaging technique to produce cross-sectional images (previously also known as CAT scans (Computed Axial Tomography)
• Used for diagnostics and therapy purposes
• Many slices form a volume
• CT is known as a moderate- to high-radiation diagnostic technique

The Grenoble team in the control room of the medical research laboratory. From left to right: Paola Coan, Alberto Bravin and Emmanuel Brun. Credit ESRF/Blascha Faust.
CT scans – Breast Cancer

• 3D – diagnostic computed tomography
  • The typical dual view digital mammography is limited and does not detect 10-20% of breast tumors
  • Hospital CT scans can not be used – radiation dose too high

• Synchrotron CT scans:
  • High energy X-rays
  • Phase contrast imaging
  • Novel mathematical algorithms: “equally sloped tomography”
  • Spatial resolution 2-3 times higher than in a hospital
  • Radiation dose 25 times lower

In the US alone an estimated 40 000 persons/year die of breast cancer!
3D Phase Nano-tomography

- Sample (S) positioned on a translation-rotation stage
- Monochromatic beam is used
- By varying D1 and D2, different magnification factors are obtained
- Images are recorded at different focus-to-sample distances over a complete rotation of the sample (3000 different angles)
- Images were used for phase retrieval (PR), used as input for the tomographic reconstruction algorithm
Bones are hierarchically organised
- Structure on the nano-scale
- Dynamical tissue, self repairing
- Bone fragility disease generally associated with a disturbance of the remodeling process.

- Tiny sample of 0.4x0.4x5mm used
- Different densities in grey scale
- The blue cylinder depicts virtual reconstructed volume from the sample
3D Bone ultrastructure
3D Bone ultrastructure
3D Bone ultrastructure

Principal publication and authors
X-Ray Phase Nanotomography Resolves the 3D Human Bone Ultrastructure, M. Langer (a,b), A. Pacureanu (a,b,c), H. Suhonen (b), Q. Grimal (d), P. Cloetens (b), F. Peyrin (a,b,e), *PLoS ONE, 7: e35691 (2012).*
(a) CREATIS, CNRS 5220, INSERM U1044, INSA Lyon, Université de Lyon (France)
(b) ESRF
(c) Current address: Centre for Image Analysis & SciLifeLab, Uppsala University (Sweden)
(d) LIP, CNRS, Université Paris (France)
(e) Labex PRIMES, Université de Lyon (France)
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
**More science examples**

**Understanding Catalysts**

X-ray diffraction computed tomography of a metal/metal oxide catalyst


μ-XRD-CT

Ni Precursor  Support
Biocatalysts

Understanding biocatalysts for enzymatic fuel cells

A ribbon representation of the three-dimensional structure of membrane-bound hydrogenase (MBH). The [NiFe] catalytic centre and the three Fe-S clusters involved in electron transport are shown as spheres. The spatial arrangement of the Fe-S clusters is also shown, as is a depiction of the cellular localisation of MBH.

The translocation of water molecules away from the active site of MBH. The gas tunnel funnelling hydrogen and oxygen towards the catalytic centre is shown as a grey surface, the water (red spheres) filled cavities observed in the crystal structure are shown as blue surfaces.
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
Insights into the SODA–LIME–GLASS glass production process

Collaboration between: Saint Gobain/CNRS, U. of Toulouse, U. of Grenoble, INRIA Saclay and ESRF

- **SiO$_2$**
  - $+$ Na$_2$CO$_3$ (lowers glass transition)
  - $+$ CaCO$_3$ (improves chemical durability)

- Complex system
  - strongly out-of-equilibrium system
  - many chemical reactions occurring at HT
  - micro-structural transformations

- Industrial glass synthesis at $\sim$1500°C
  - soda-lime glass melts at $\sim$1050°C
  - removal of chemical gradients and defects

- Role of microstructure in final glass homogeneity?
  - many XRD studies
  - no previous in-situ tomographic imaging

Study chemical reaction and rearrangements of grains
Experiment:
Fast time resolved tomography on the High Energy Beamline ID15

Sample composition
- SiO\textsubscript{2} 64% wt (160-200µm)
- Na\textsubscript{2}CO\textsubscript{3} 17% wt (250-320µm)
- CaCO\textsubscript{3} 15% wt (80-100µm)

Temperature ramp:
- 20°C → 700°C at 70°C/min
- 700°C → 900°C at 5°C/min

Detector pixel size 1.2µm

Acquisition time 15s
Results

- $\text{Na}_2\text{CO}_3$ grains (in red) change morphology interacting with sand grains (yellow and blue)
- Redistribution of $\text{Na}_2\text{CO}_3$ accelerates the solid state reaction (not diffusion driven)
- Sodium silicates (in white) are produced (probably $\text{Na}_2\text{SiO}_3$ and $\text{Na}_2\text{Si}_2\text{O}_5$)

- Some $\text{CaCO}_3$ grains dissolve into larger $\text{Na}_2\text{CO}_3$ grains to form a Na-Ca double carbonate
- Liquid phase at 780°C

- $\text{CaCO}_3$ grains not in contact with $\text{Na}_2\text{CO}_3$ transform into refractory grains of $\text{CaO} \rightarrow$ defects
What is a nanometer?

- **1m**
- **1mm**
- **7µm**
- **2nm**
1nm relates to 1m like:

We talk about Nano-Science < 100nm, i.e. the distance of less than 1 000 atoms!

A striking application of Nano-Science is the micro electronics industry with transistors soon as small as 10nm (100 atoms!)

The trend of scaling for NAND flash memory allows doubling of components manufactured in the same wafer area in less than 18 months.
The new frontier: Nanoscience

- Nano beams to examine nano samples
- Unprecedented challenges in
  - experimental stability (temperature, vibrations, positioning)
  - sample positioning
  - simulation, data storage, and data analysis
ESRF Publication Record

PUBLICATIONS IN PEER REVIEWED JOURNALS

20,276 PAPERS IN THE 1994-2012 PERIOD!
(AS OF 16-OCT-2012)

HIGH IMPACT

by 16 Oct 2012
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
The role of HPC

Surviving the Data Deluge

A data centric view of the ESRF (2011 figures)

- 40 Beamlines (30 ESRF, 10 CRGs)
- 6500 user visits
- >15 500 eight hour shifts scheduled for experiments
- >1800 peer reviewed publications
- Many distinct scientific fields with a large variety of different computing needs
- An average of ~4 TB/day and ~1 000 000 files/day
Micro tomography studies in 2004 of Toumai (Sahelanthropus tchadensis) to reveal the 3D structure of the skull and allow remodeling of missing fragments

- 1 scan = volume of 160 x 160 x 200 mm, pixel size of 45 μm = ~ 1*10^{11} Bytes = ~100 GB
- The raw data set is in the order of 400 GB
- The processed data represents >4 TB
- Processing took several 1000 h CPU time on 2.2 GHz AMD Opteron systems

Today it is likely that we would scan Toumai at <4.5 μm resolution

4 TB would become 400 TB
New methods: nanobeams & raster scans

### Micrometer
- 3D, 2 modalities
- Low sampling
- “Few” small files

### Nanometer
- 3D, Many modalities + in-situ
- High sampling
- Many big files

<table>
<thead>
<tr>
<th>KiloBytes</th>
<th>MegaBytes</th>
<th>GigaBytes</th>
<th>TeraBytes</th>
<th>PetaBytes</th>
</tr>
</thead>
</table>

- Store
- Process
- Reduce
- On-line Analysis

- Transfer
- Off-line Analysis
- Long term storage
New Detectors – new, data intensive science

Quantuum ADSC Q315
- 6k x 6k pixels = 72 MB/image;
- ½ fps = 36 MB/s

Frelon2k
- 16 frames/s = 66 MB/s

Dalsa Pantera TF1M60
- 60 fps = 150 MB/s

Sarnoff CAM512
- 512x512, 400 fps, 200MB/s
- 1kx1k = 500MB/s, PCIe camera link

Medipix2/Pilatus/Maxipix
- 1000 fps of 256x256 – 14 bit = 130MB/s, 5 chips grouped together: 650 MB/s, PCIe interface

Eiger
- 24 000 fps, up to 4x4 kpixels, multiple 10GE interfaces, >1GB/s
The role of HPC

Computing is (and will remain) on the critical path for data intensive science

We are in a race condition on two fronts:

**Lower latency** ⇒ diminish the time to measure, store, analyse data
- On-line data analysis
- Network, Storage, and Clusters

**New functionality** ⇒ new ways to measure, store, analyse data
- On-line and off-line data analysis
- Remote access
- Data archiving and data curation
- Data mining
Fighting Latency

- **Data communication network**
  - Not a bottleneck right now
  - 1, 10, 40 Gbps FO network links
  - Aggregation of links for the backbone

- **Storage**
  - Never big enough
  - Never fast enough
  - Difficult to manage

- **Data Analysis**
  - Never fast enough
  - Difficult to manage
Data Storage

- 1.8 PB NAS Storage, 2500 disks, 16 Controllers
- Largest file system: 1 PB
- Easy to manage
- NFS based – large client compatibility
- Limited in performance
- Does not scale well

- Individual dedicated storage units
- Scalable performance
- Flexible investment
- More difficult to manage
Why parallel computing/programming?

Basic problem: process data “fast”

Why?

• It takes 1-3 days to collect the data
• Access to a synchrotron is difficult hence time is precious
• It is important to “judge” the quality of the experiment “on the fly”
• It is increasingly difficult to take data “home”
• It is increasingly difficult to analyse data at home

Brute-force solution $\rightarrow$ buy bigger and faster computers

• Moore’s law: transistor density doubles every 18 months
• limits: mono-atomic layers
• higher clock speed $\rightarrow$ exponential temperature rise

Brute-force no longer works
**Improve overall throughput:**
- split task in several (many) independent jobs
- distribute jobs to several (many) different processors
- select “most appropriate” processor for each job
- scales well with number of processors
- no change to program code needed
→ simple parallel processing

**Reduce elapsed time for each job:**
- distribute each job over several processors
- (re-)structure program into independent loops
- optimize data access for each processor
- needs change (possibly restructuring/rewrite) of program code
→ task for parallel programming
“Classical” Parallel Programming

OpenMP (Open Multi-Processing, www.openmp.org):
- shared memory model: single node with several cores
- mainly directives based (C: #pragma, Fortran: C$OMP)
- compiler does data distribution, computation details

- supports many nodes with heterogeneous architecture
- function call based
- user needs to organize data and task distribution

MPI and OpenMP can be combined within an application
In theory speed-up could scale linearly with the number of processors

**In practice:**
- inefficient serial program remains inefficient when parallelized
- overhead due to communication between parallel processes
- unequal execution time of parallel processes causes wait times
- Amdahl's law: speedup is limited by non-parallelizable part of program (e.g. if this takes 5%, then maximum speedup is 20). Adding more processors does not help.
Parallelization has practical limits!

For instance I/O bandwidth
OAR scheduler

- OAR – derivative of PBS
- developed by INRIA (Research Institute for Applied Informatics)
- OAR is a resource manager:
  - Batch and interactive jobs
  - resource distribution
  - Resource and activity monitoring
  - job queuing mechanism
  - scheduling policy
  - priority scheme interactive or batch mode
  - request resources (cores, memory, time...)
  - Dynamic insertion/deletion of jobs
  - Logging

Job schedulers are not very good for small jobs!
Lowering latency with GPUs

Massively multithreaded shared memory architectures
- Optimised for large scale data-parallelism
- More transistors dedicated to processing
  - No need for branch prediction and large caches
- Very power efficient for suitable algorithms
- Scalable systems

General purpose GPU programming
- CUDA, OpenCL, DirectCompute
- OpenACC, OpenHMPP, C++ AMP, PGI C99 & Fortran
- Bindings with libraries:
  - Thrust C++, OpenCL C++, ArrayFire, pyOpenCL, pyCUDA, JOCL, JCUDA, OPENCL.NET, CUDAfy (CUDA.NET)
  - Matlab Parallel Processing Toolbox, Accelereyes Jacket

Majority: CUDA and OpenCL
Fast Azimuthal Integration (pyFAI)

- On-the-fly calibration of Powder Diffraction Experiments and SAXS
- Real-time data analysis on high-resolution experiments

**Processing rates (pyFAI on CPU, OpenCL on GPU)**

<table>
<thead>
<tr>
<th></th>
<th>Pilatus 1M</th>
<th>H. Frelon (2M)</th>
<th>Frelon (4M)</th>
<th>Pilatus 6M</th>
<th>Fairchild (16M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5520 CPU</td>
<td>16.89</td>
<td>8.98</td>
<td>4.10</td>
<td>2.57</td>
<td>1.09</td>
</tr>
<tr>
<td>X5690 CPU</td>
<td>24.27</td>
<td>12.89</td>
<td>5.98</td>
<td>3.71</td>
<td>1.55</td>
</tr>
<tr>
<td>C2075 GPU</td>
<td>133.33</td>
<td>68.03</td>
<td>28.41</td>
<td>19.57</td>
<td>9.58</td>
</tr>
<tr>
<td>GTX580 GPU</td>
<td>151.52</td>
<td>77.52</td>
<td>33.67</td>
<td>23.26</td>
<td>9.93</td>
</tr>
</tbody>
</table>
BigDFT – Wavelets for nanoscience

- Based on the density functional theory (DFT), a code for quantum properties of systems at nanoscale based on wavelets
- Used within PRACE to benchmark HPC systems
- Flexible, easy to optimise
- Simple numerical operations
  - Short convolutions
  - Linear algebra (BLAS)
- Optimal for massive parallel environments up to thousands of processors

- Prize BULL-Fourier 2009: Luigi Genovese (ESRF)
  - Collaboration INRIA-CEA-UJF
  - Benchmark code PRACE WP8
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
EU projects

100 million euro EC investment

www.icordi.eu

Origin: Carlos Morais-Pires
The PaNdata initiative

- Photons and Neutrons are complementary investigation tools
- Cross discipline experiments are increasing in number
- Synergy is essential for the project

Five P+N sites in Europe are in PaN-Data:
- ISIS + DIAMOND
- SINQ + SLS
- ILL + ESRF
- HMI + BESSY, now the HZB
- LLB + SOLEIL
- (+ DESY, ELETTRA, MAXLAB, and ALBA)

- 30 000+ Scientists in Europe
The PaN-data RIs
Research Infrastructures and Participants

11 Research Infrastructures & 16 Participating Institutions
Goal: a shared data infrastructure for Photon and Neutron RIs

- Harmonise data policies in laboratories
- Harmonise authentication and authorisation
- Standardise data formats and annotation of data
- Allow transparent and secure remote access to data
- Establish sustainable and compatible distributed data catalogues
- Allow long term preservation of data
- Provide tools/interfaces for curating data
- Provide compatible open source data analysis software

- A very ambitious work programme!
ESRF paleontological microtomographic database

15 Most visited [15]

(171) Qatzeh 10 maxilla
(137) Engis 2 child
(102) Spider Orchid
(80) Novispathodus cluster T084C30 PIMPZ 28766
(71) Trinil 11621 upper molar
(59) AMNH-68253

(65) Qatzeh10 mandible
(57) Trichomyia isangeli
(50) boatie Scotidiae
(47) 15523 000
(47) Qatzeh 15 mandible
(45) Trinil 11620 upper molar

(45) Novispathodus sp T084C30 12
(44) 4F4
(44) AMNH-55901
Outline

- A photon factory (a bit of physics)
- Science examples
  - Life Sciences
  - Catalysts
  - Glass
- SR science and HPC
- EU collaborations
- Outlook/Conclusions
Conclusions/Outlook

✓ Synchrotrons are analytical tools to examine matter with unprecedented detail

✓ They have an enormous potential for applied (industrial) and fundamental science

✓ IT underpins our scientific endeavour everywhere

✓ The complexity of IT is increasing quickly

✓ Research must attract the brightest IT staff to enable science and help tackling the societal challenges of the future

✓ Policy and decision makers should realise that it would be an enormous error to diminish science funding
Thank you for your attention!