# EXDCI-2 – WP3
# “Scientific Case and Industrial Collaboration”

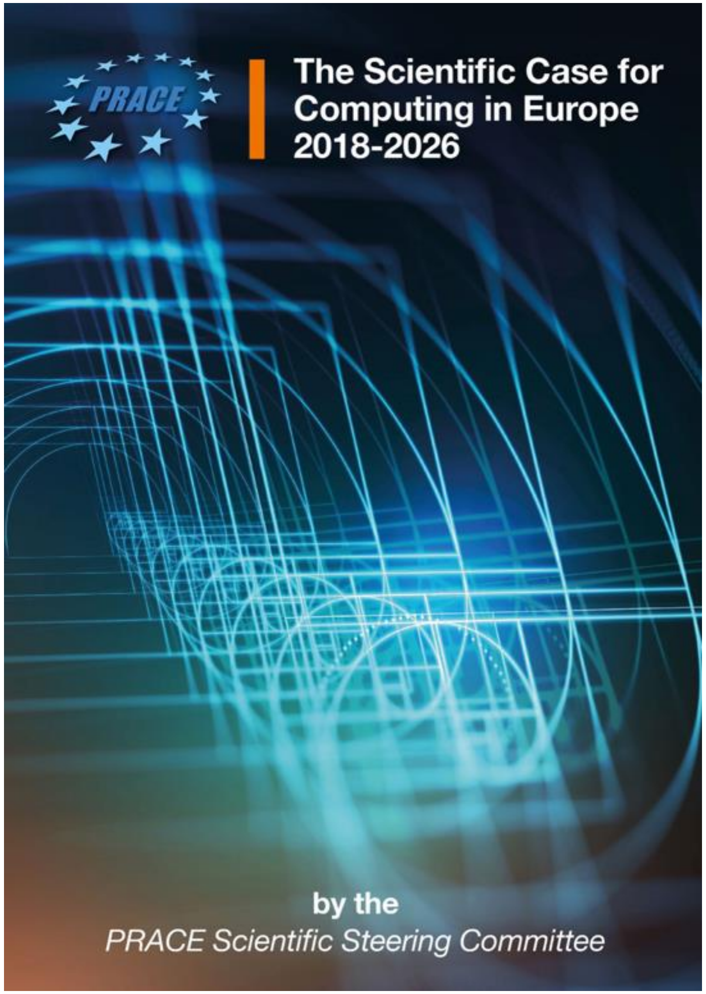
M. Bode (RWTH Aachen) and S. Requena (GENCI)



The EXDCI-2 project has received funding from the European Unions Horizon 2020 research and innovation programmed under grant agreement no. 800957.

# 3rd Edition of the PRACE Scientific Case

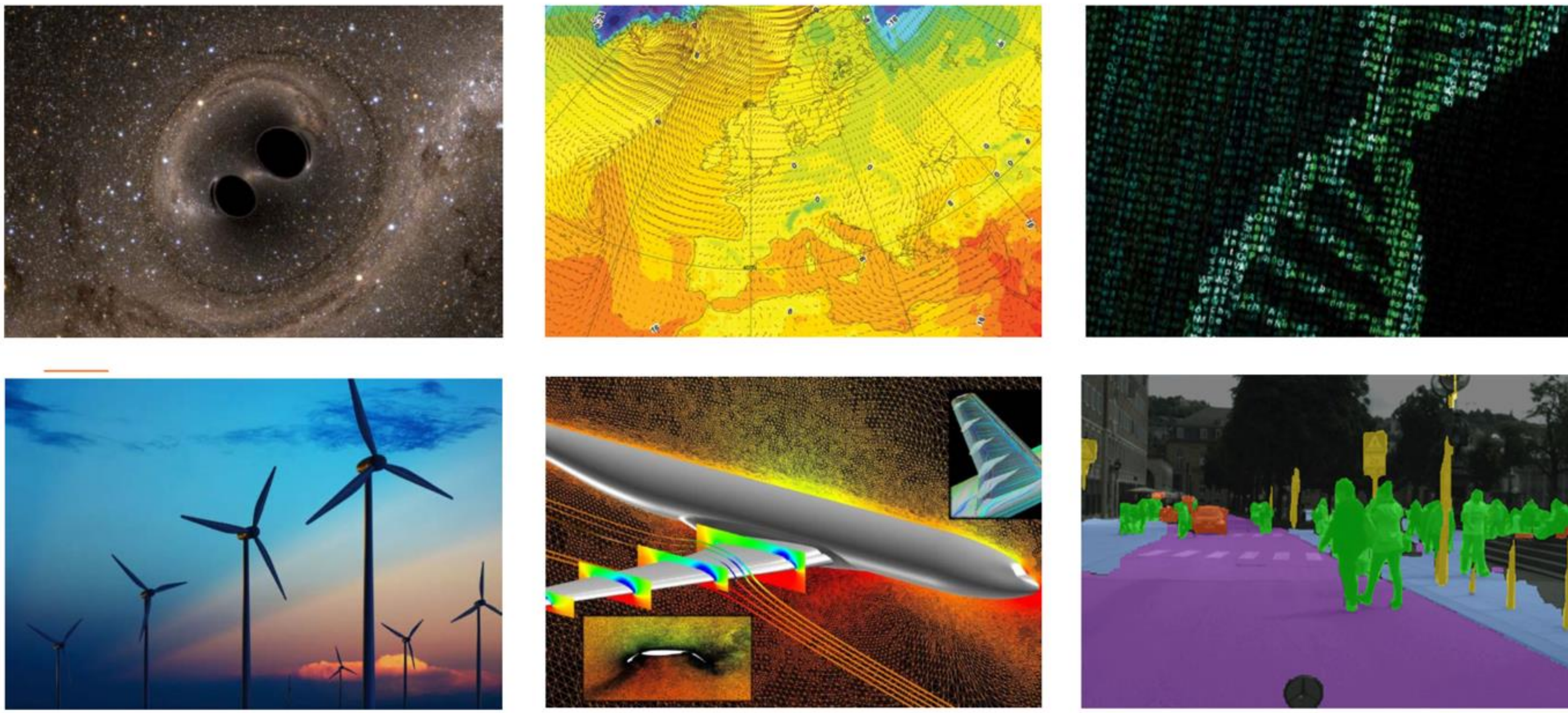
- Elaborated by the PRACE Scientific Steering Committee in 2018
- Support from EXDCI WP3
- Focus on some disciplines
  - Physical Sciences
  - Climate, NWP and Earth Sciences
  - Life Sciences
  - Energy and industrial apps
  - Materials
- And key requirements in terms of data management, infrastructures and services



# Why HPC matters ?

PRACE
PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

## European Computing Solves Societal Challenges



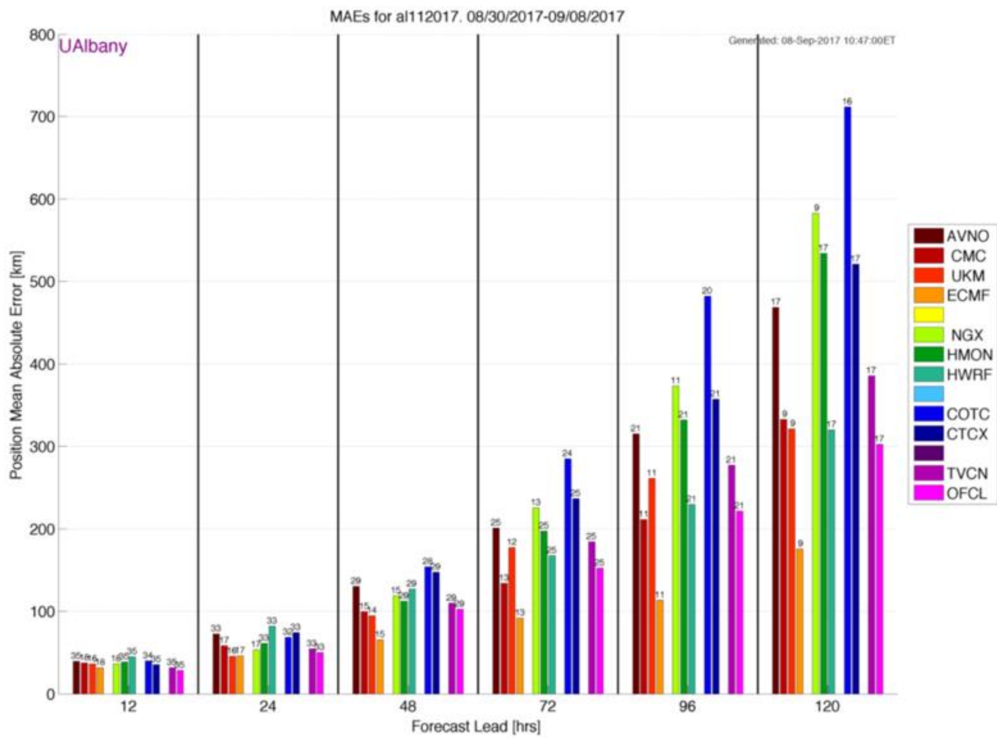
*PRACE's goal is to help solve these challenges. The days when scientists did not have to care about the hardware are over, and so are the days when compute centers did not have to worry about the scientific application!*

# Europe leads international code development


Predicting evacuation needs is a life/death matter – but avoiding it saves €250M

Maintaining our European lead on software should be one of our investment priorities

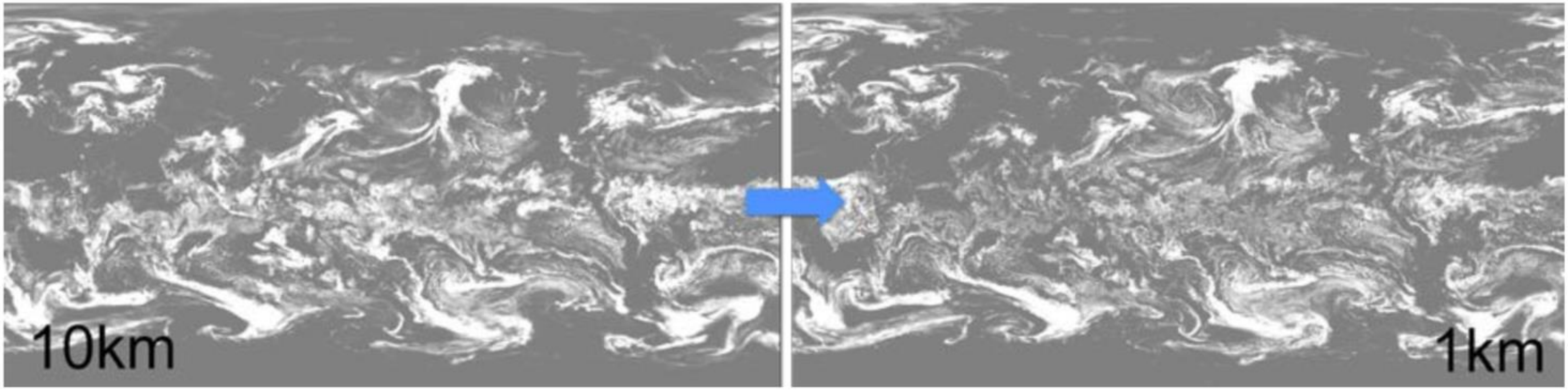
The best forecast codes are European
Arguably, the US would have done better by investing less in machines and more in software



# Impact of scientific collaborations

PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

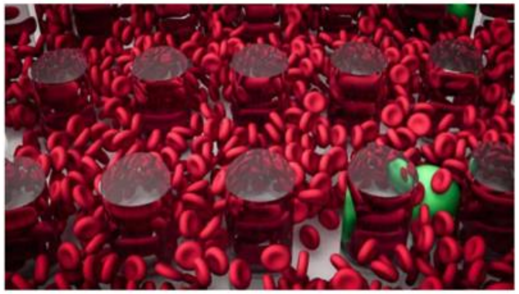
## The Centers of Excellence are Working



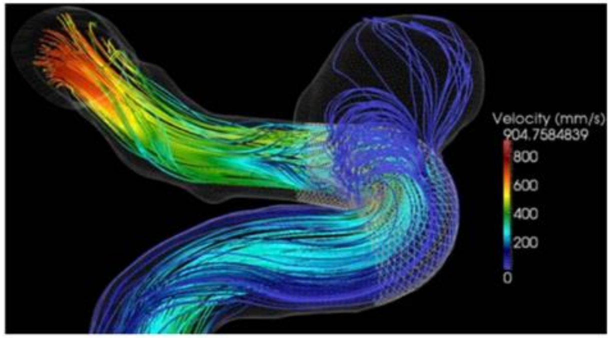

*Accurate 20 day weather forecasts would translate to higher revenues in agriculture, tourism and production*

- Excellent to outstanding impact in a number of research & code areas
- Important not to be short-sighted: Changing codes is a *longer* process than building an Exascale machine, but with greater long-term European impact. Give them room to focus on the real goal.
- Substantial impact on European competitiveness through workforce education & degrees
- But: Beware that the present CoEs do not cover all areas of science!

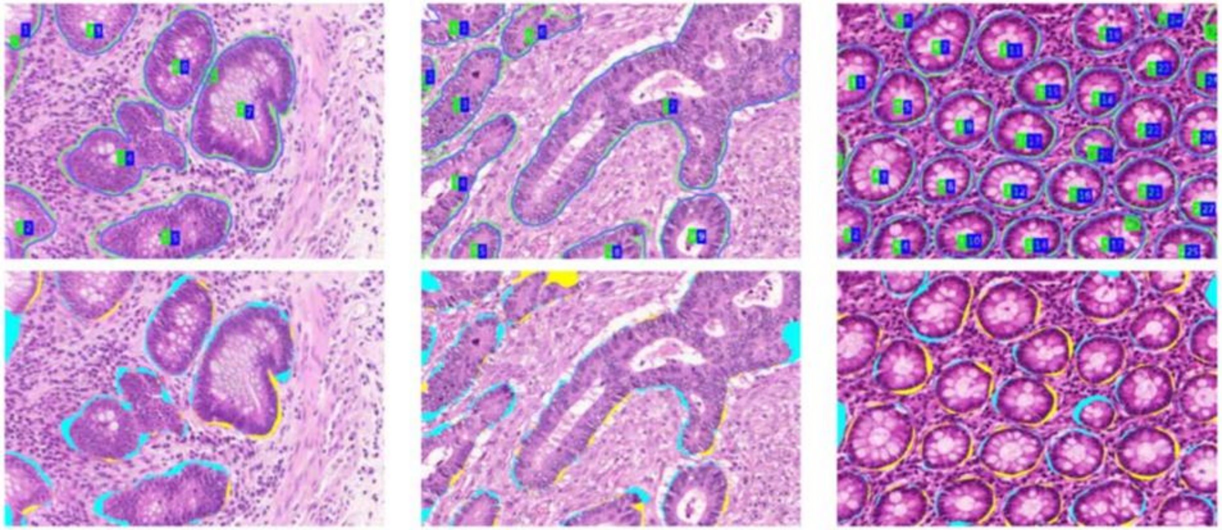
# Human Health Computing is Translational



It will become possible to design microfluidics & nanobiotechnology devices to target e.g. specific cancer cells by modeling flow/interactions




Lattice-Boltzmann enable simultations of blood flow, e.g. effects of using a stent to treat brain aneurysm. On-demand computing!



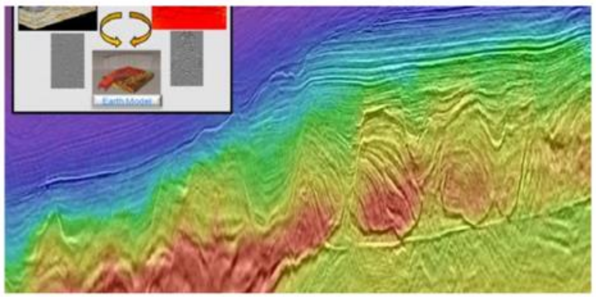
(a) benign (b) malignant (c) benign

Tumor classification with deep learning neural networks now beats the best pathologists

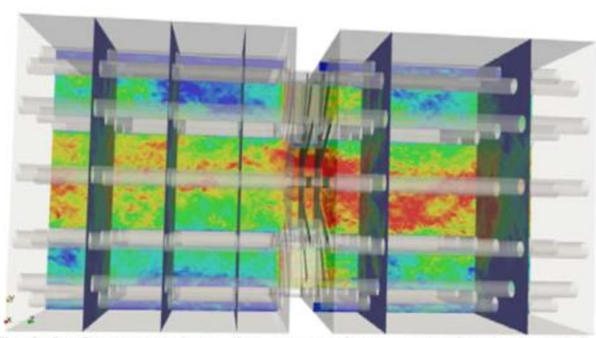
# HPC is also impacting EU industrial competitiveness

PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

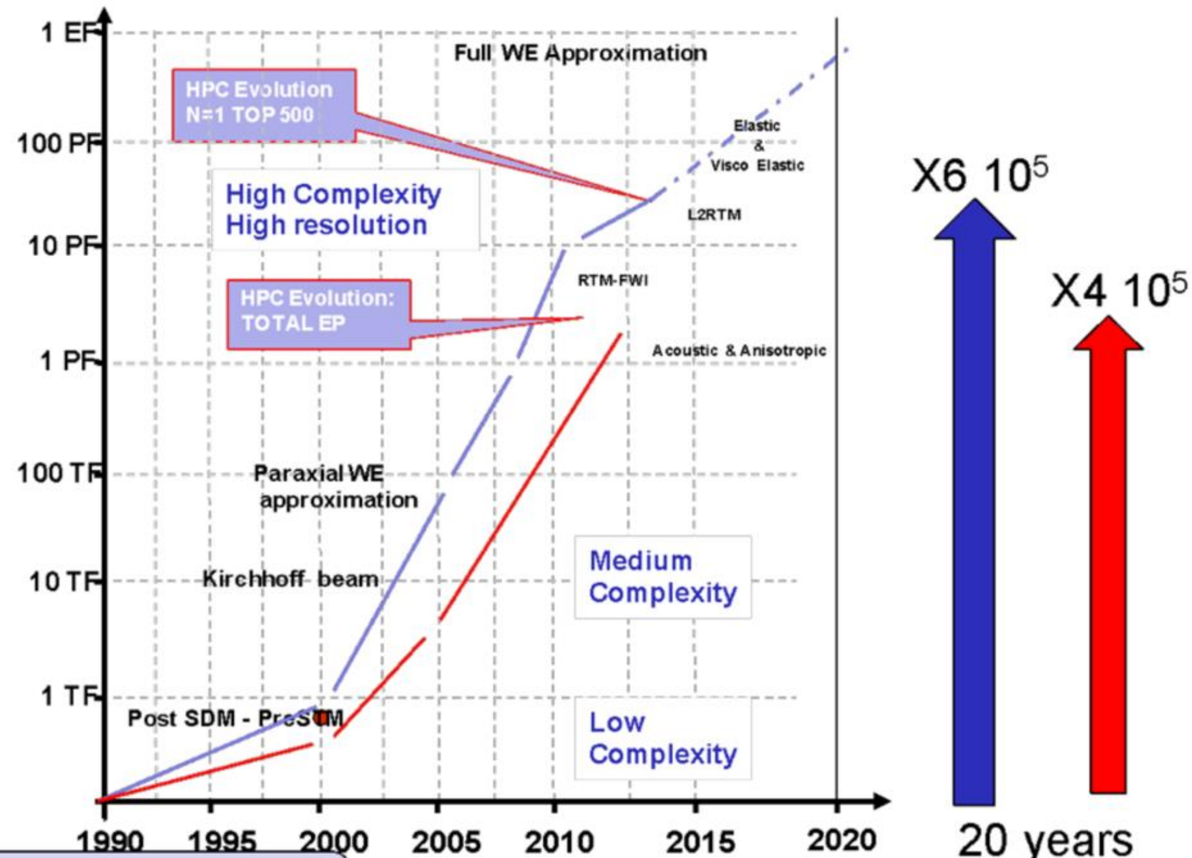
## Energy applications need Exascale



Oil & gas is one of the world's largest users of HPC – because computing saves time & money



Fluid dynamics is used to understand heat & flow in nuclear reactors



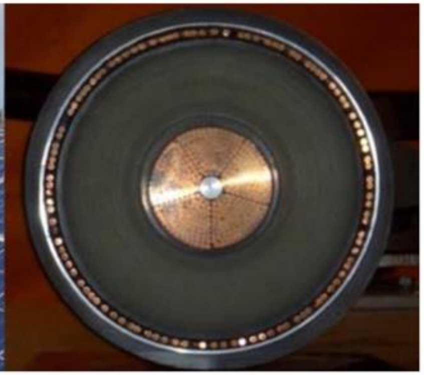

1 TF = $10^{12}$ operations / s
1 PF = $10^{15}$ operations / s
1 EF = $10^{18}$ operations / s

# And fosters development of new energies

PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

## Computing is Driving Renewable Energy

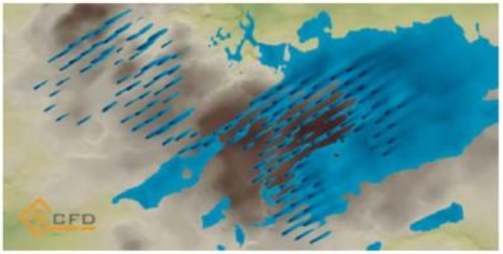


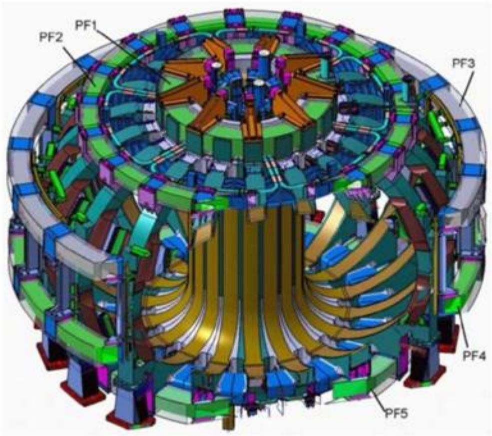



HPC is creating new generations of insulators that enable higher voltage cables, which reduces losses

Optimisation of wing shapes placement of wind farms

New generations of solar cells (materials)



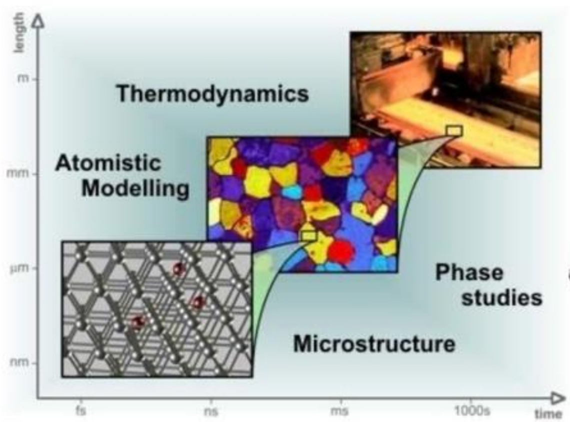


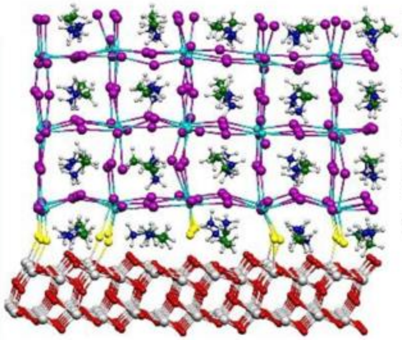

Advanced MHD simulations are critical to make the plasma stable enough to make Tokamaks like ITER useful for energy production through fusion

# Or future materials

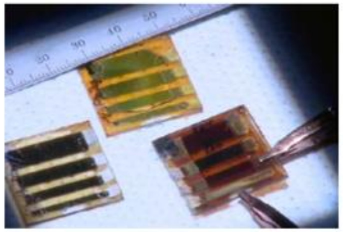
PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

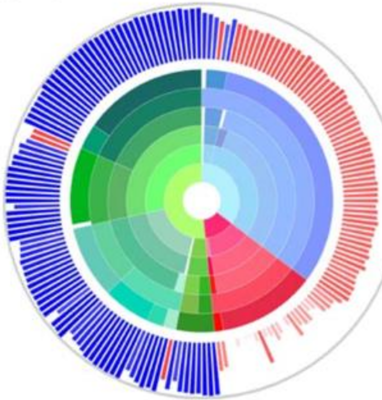
## Future Materials: From Molecules to Machines



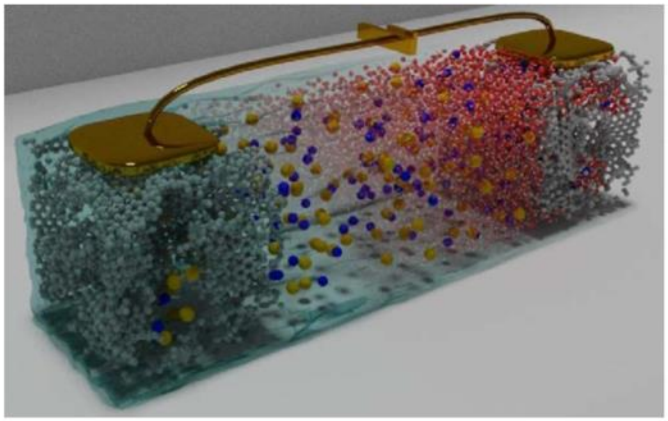



Perovskite solar cells have outstanding efficiency, but contain lead. Simulations used to develop lead-free versions





Modern materials design is increasingly data-driven, using AI/deep learning methods to rapidly predict properties without full DFT calculations



Supercapacitors are used to store energy e.g. in braking kinetic-energy recovery systems in cars/buses. Molecular simulations are used to understand the charge build-up to improve their capacity and efficiency

# Key Infrastructure Requirements

Urgent need for more compute cycles - but also high demands for memory bandwidth & I/O

Scientific Applications are not only ready to use Exascale, but have specific high-impact research goals that can only be realized with access to next-generation resources

However: Exascale will require new approaches to scaling by using ensembles, deep learning, and statistical models. High-end computing will dominate, but tomorrow's solutions will be different than yesterday's

Exascale systems will need to be able to handle tens of thousands of active jobs and large I/O requests

Software & algorithms take longer to change than hardware - first-generation systems must be based on present concrete needs, rather than hopes about being able to co-design

Linux rules the modern computing world: All systems will need to support a full Linux stack of development tools

Centers & their staff need to engage directly in software development together with scientific communities

# Requests/Recommendations from the PRACE SSC

To be competitive with the rest of the world, *Europe* needs a *European* infrastructure & community for computing, support and research – with the agenda focused on solving the most challenging problems

Future infrastructures need to have involvement of leading academic and industrial researchers in direct government, not merely providing advice - because they are the ones that vouch for the requested funding

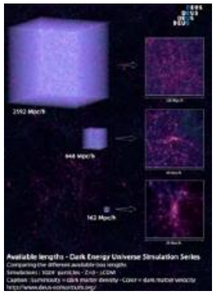
Future Infrastructures and operations will need to be much more diverse to support HPC, Data Science and different types of accelerators - but we also need to avoid fragmentation. Suggested to focus on **TWO** architectures.

# TECHNOLOGICAL CONTEXT

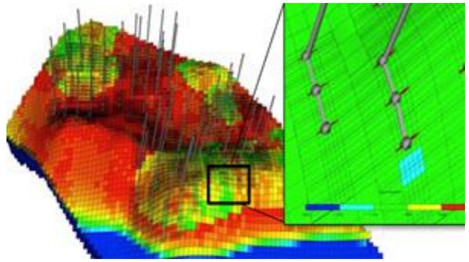
## Convergence HPC / DATA / AI



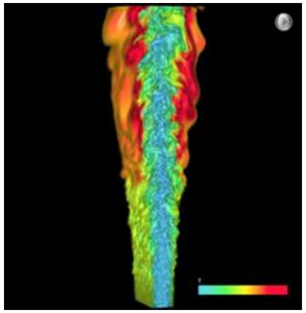
## ❑Explosion of computational data



Cosmology
DEUS project
150 PB raw data



Reservoir modelling
of giga models 350 TB/run



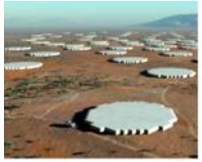
HiFi turbulent
DNS combustion
S3D : 1PB / 30mn

Climate CMIP exercises

**Status CMIP5 data archive:**
- **1.8 PB for 59000 data sets stored in 4.3 Mio Files in 23 ESGF data nodes**
- **CMIP5 data is about 50 times CMIP3**

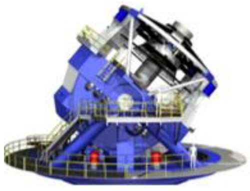
**Extrapolation to CMIP6:**
- **CMIP6 has a more complex experiment structure than CMIP5.**
- **Expectations: more models, finer spatial resolution and larger ensembles**
- **Factor of 20: 36 PB in 86 Mio Files**
- **Factor of 50: 90 PB in 215 Mio Files**

## ❑And instrumental data as well

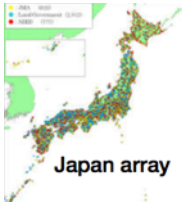


LOFAR/SKA
16 TB/s raw

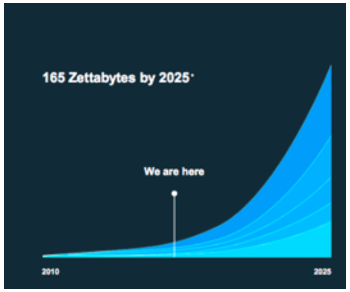
COPERNICUS/SWOT
4 PB/day raw



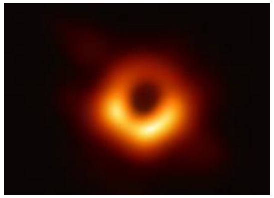
LSST/EUCLID
20 TB/night raw




Seismic sensors
100 TB/yr




Internet & **IoT**

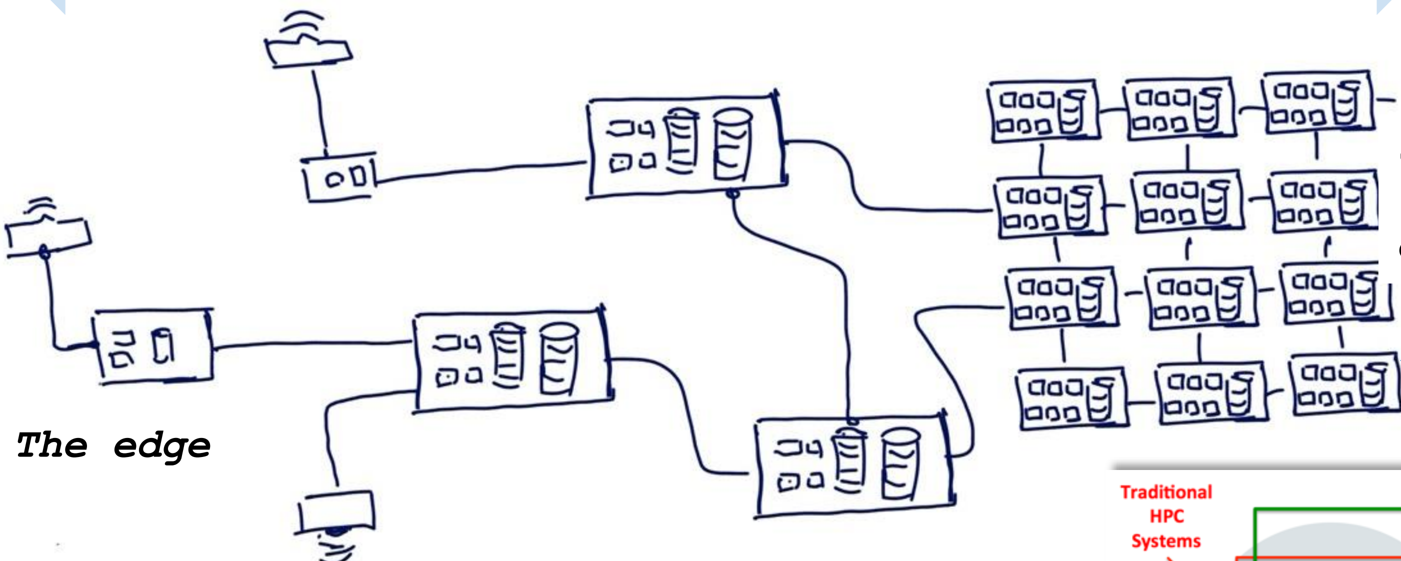


First picture of black
hole (M87) = 15 PB

# TECHNOLOGICAL CONTEXT

## Toward Advanced Cyberinfrastructure Platforms

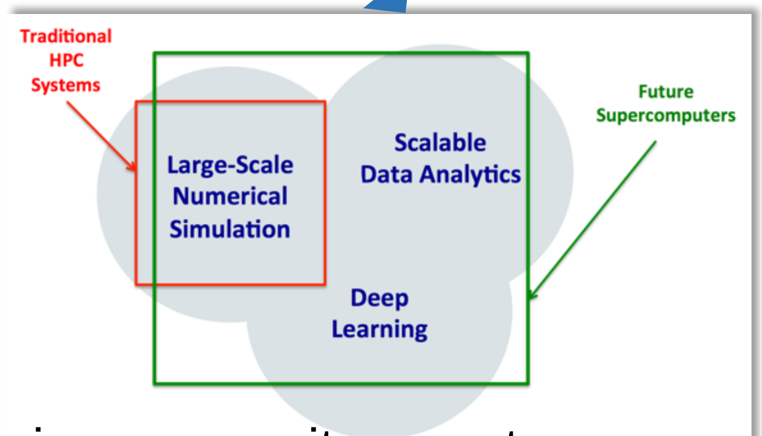
a new digital continuum



Strongly coupled no
(Supercomputers,
data centers, etc.)

*The edge*

Intermediate nodes
(CDN, data centers, etc.)




- Challenges : cohabitation of SW stacks, containers, security, smart resource managers, end to end workflows (from the edge to the tape), …
- Development of **new services, co design** and **user support**

# CONVERGENCE LARGE SCALE INSTRUMENTS – HPC/HPDA

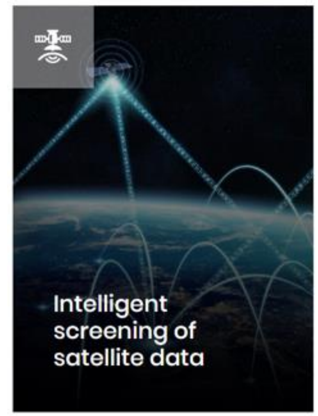
## Exemple of the CEF PHIDIAS Project

PHIDIAS goal is to build **a prototype for Data/High Performance Computing (HPC) services based on Earth sciences Satellite Data use cases**

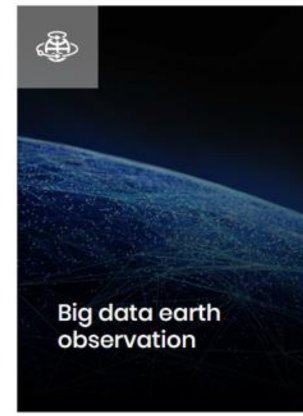
### PHIDIAS Goals

- Building a prototype for earth sciences satellite open data and HPC services
- Optimising and industrialising treatment workflows for extensive reusability
- Ensuring open access to standardised HPC services
- Improving FAIRisation processes, cross-use and open access
- Developing new data-processing models coupled with HPC capabilities
- Deploying data-processing methods as a service for scientific communities, public authorities, private entities and citizen scientists


Improve the use of numerical services for marine data management, service and processing, considering EOSC challenge and the DIAS




Improve efficiency and genericity of the intelligent screening of environmental satellite data




Enhance EO data processing chains scalability for environmental monitoring from the end-users needs of THEIA land data centre network

***Enabling HPC and data capacities of the European Data Infrastructure for Scientists & Researchers***

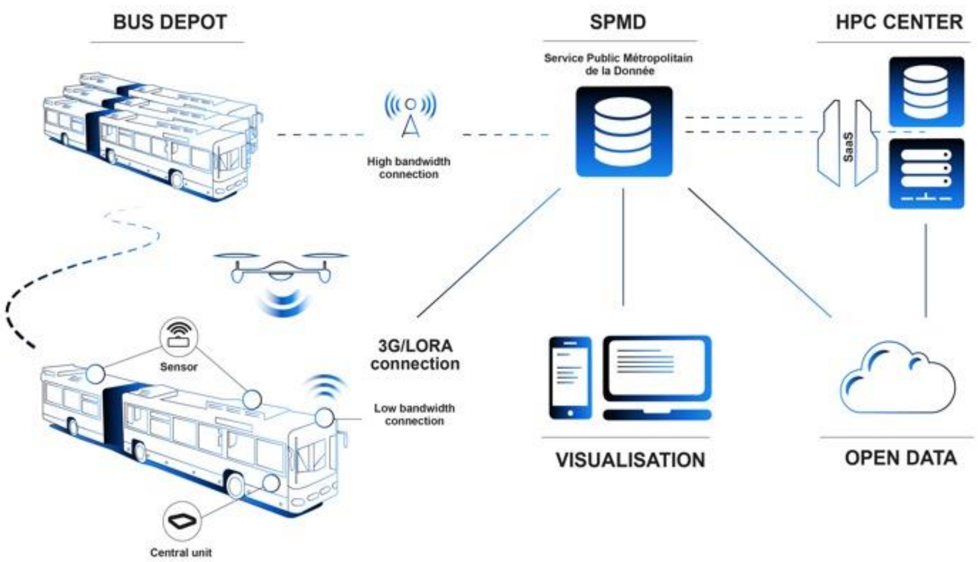
# CONVERGENCE WITH IOT– HPC/HPDA

## Exemple of the CEF AQMO Project

**Cofinancé par le mécanisme pour l'interconnexion en Europe de l'Union européenne**

A Q M O
Air Quality and Mobility

### Transports en commun & mesure de la qualité de l'air




**La maîtrise de la qualité de l'air passe par :**

- La mesure sur une longue période
- La simulation numérique
- La mise à disposition des données auprès des citoyens

Le projet AQMO aborde ces problématiques par une approche de bout en bout.

Déploiement dans la métropole rennaise

**AQMO aborde les questions suivantes:**

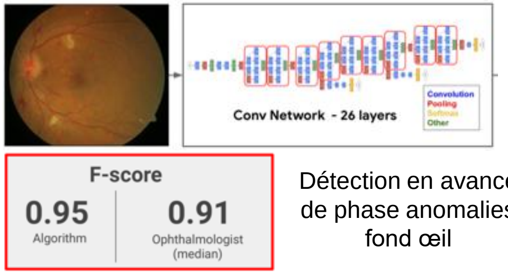
- Etat de la qualité de l'air?
- Comprendre les origines?
- Impact des mesures mises en place?
- Pouvoir répondre à des questions « *et si* ?»

# CONVERGENCE HPC – AI

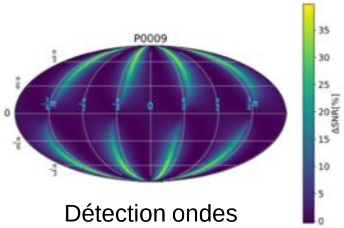
## Context

### ❑Why HPC/HTC need AI ?

- For inferring data flows from large scale scientific instruments
  - Stream access, support of end to end workflows (*edge to tape*)



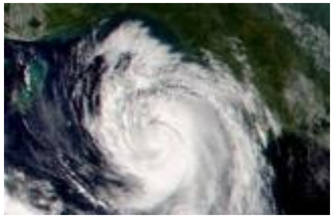

Détection en avance de phase anomalies fond œil




Détection ondes gravitationnelles LIGO

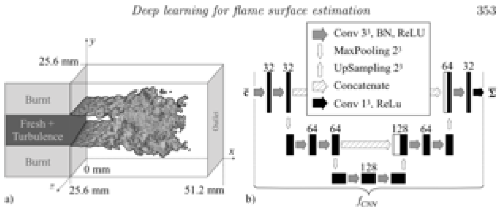
Prévention séismes

- For coupling learnt models and simulation codes (cf Gordon Bell’18) – toward cognitive simulation
  - Interpolation and extrapolation of long trajectory in MD methods, integration of reduced/surrogate models for multiscale , optimisation / UQ (reducing parameter space), forcing of de models (climate), acceleration the convergence of iterative methods, mesh tuning …



Prédication évènements extrêmes climat

Couplage simulation LES combustion et IA



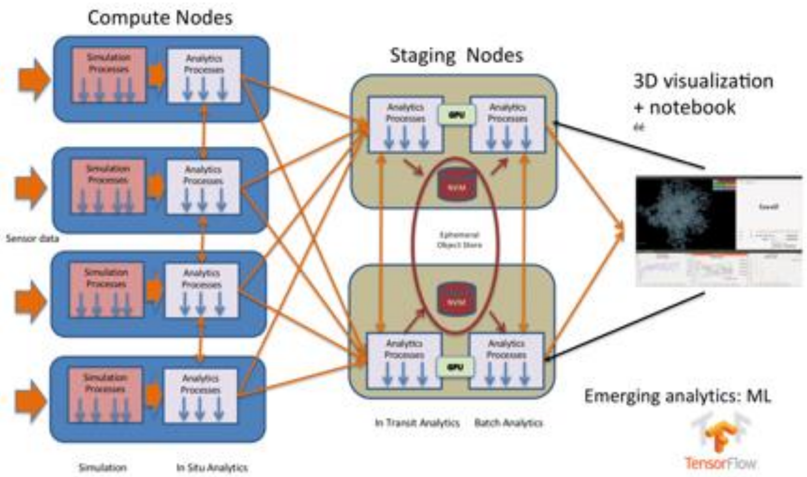

Optimisation crash auto

# CONVERGENCE HPC – AI

## Context

## ❑Why HPC/HTC need AI ?

- For (in-situ, in-transit) post processing of numerical simulations
  - simulation ➡ (learning, inference) ➡ storage of pertinent data
  - *smart computational steering*




  - **Optimise data movement -> minimise energy**

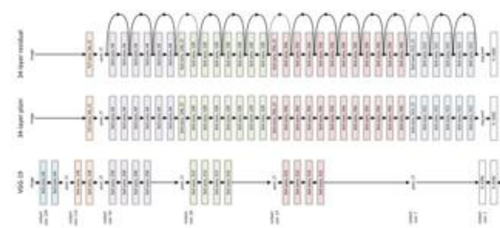
- For better exploiting systems and computing centers
  - AI driven schedulers, improved security, preventive maintenance, optimisation of the infras, …

# CONVERGENCE HPC – AI

## Context

❑ And AI needs HPC ?

- HPC generates huge amounts of data suitable for AI training
  - Ex : last IPCC campaign in France -> 14 PB of data



- Scale up of the learning phase of neural networks (networks + complex/deep, more data, more classes, …)
  - HPC provide unique 4 levels of parallelism for massive scaling DNN training (© S. Matsuoka)

  1. Hyper parameter search
  2. Data parallelism : different batch data
  3. Domain decomposition : // layer calculations in propagation
  4. Intra chip ILP, vector : // convolutions

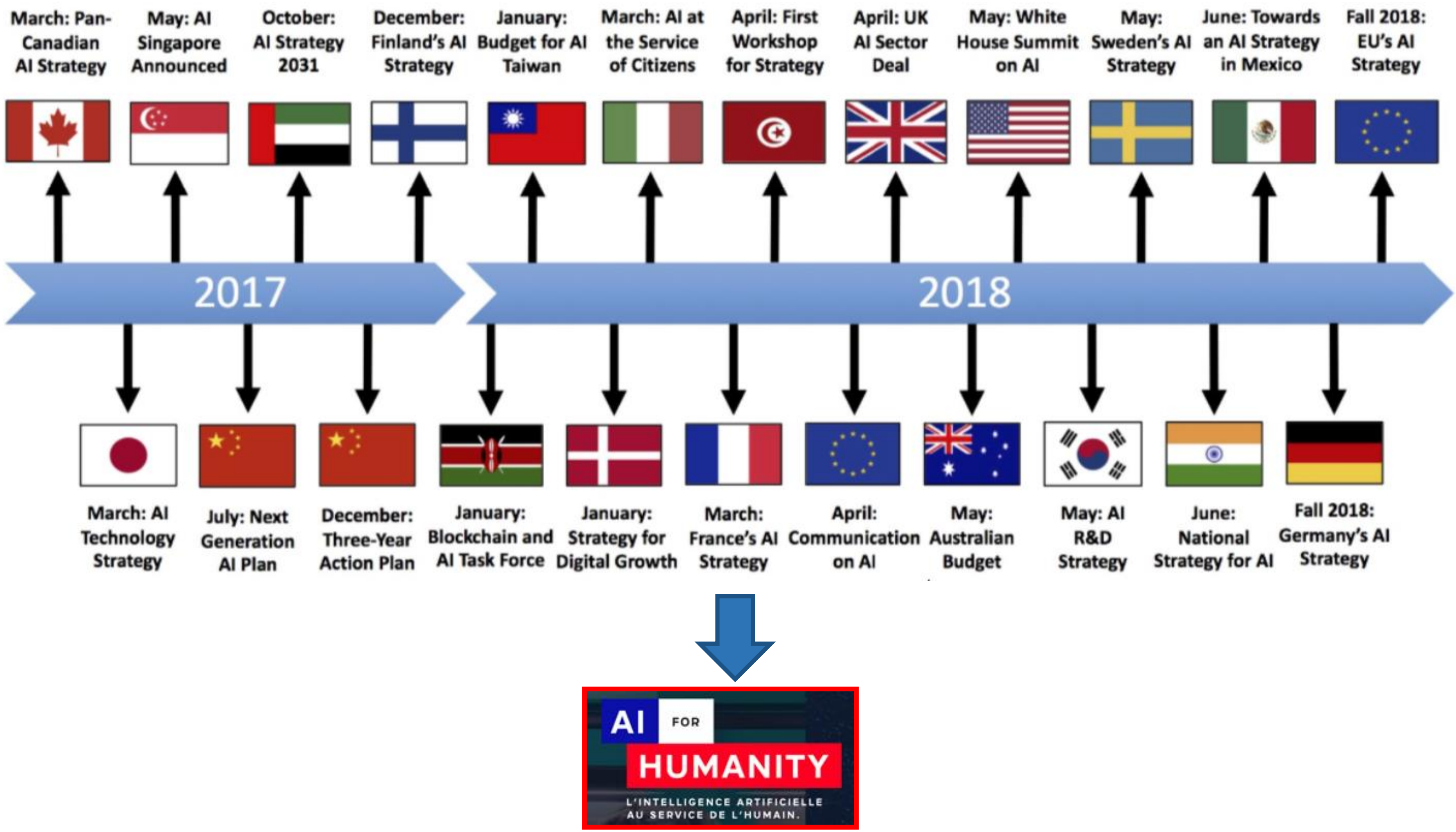
- Auto-tuning of the choice of models (Auto DL/ML, AutoAI), use of federated/transfer learning, new methods, …

- Explicability and trust on AI (XAI as a glass box), coupling between formal methods and neural networks, …


*Figure 3 – Les apports de l'IA explicable*

# CONVERGENCE HPC – AI

## AI roadmaps since 2017



# INTRODUCING JEAN ZAY @ IDRIS

## One of the biggest converged system in Europe


### ❑ Objectives

- **Support** with **sovereign** and **leading edge** HPC facilities the French AI research community
- **Foster** synergies between AI and HPC communities
- To be **integrated** into the French AI plan

### ❑ **Converged** system ?

- HPC + HPDA + AI

### ❑ New dynamic access modes

- Elastic pool of resources
- Support of containers, notebooks, …


### ❑Big Milestones

- March 2018 : French "**AI for Humanity**"
- Jan 2019 : Contract between HPE and GENCI announced
- May 2019 : Installation
- July 2019 : Grands challenges
- October : 2019 Full production
- *Start 2020 : 2nd upgrade following ongoing contract of progress*
- *End 2020 : Major upgrade planned*

16PF


### ❑A balanced architecture

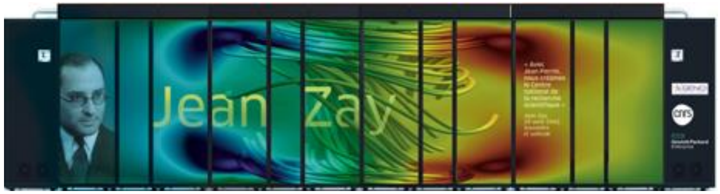
- HPE SGI 8600
- Scalar partition (HPC): 1528 nodes, 3056 CPU CSL 6248, **61 120 cores**, OPA
- Converged partitions – **1292 GPU**
  - 261 thin nodes, **1044 GPU V100 32GB**, 4xOPA
  - 31 fat nodes, **248 GPU V100 32 GB**, 4xOPA

### ❑Storage

- **1.3 PB @ 360GB/s** full flash (L1)
- **30 PB @ 150GB/s** HDD (L2)
- SpectrumScale parallel filesystem

# INTRODUCING JEAN ZAY @ IDRIS

## Some recent updates




## ❑Grand challenges projects

- Partial opening to few teams during final acceptance (3 months)
- More than 300% oversubscription
- 32 projects finally accepted
  - 18 HPC **and 14 AI**
  - From academia and industry
  - HPC on GPU : astrophysics, combustion, material, electromagnetism, medicine

## ❑1st look on the AI Challenges

- From 4 to 1044 GPUs
- medicine, vision, detection activity & weak signals, autonomous driving and crash modelling, in-situ post processing + simulation, image, finance, cryptography and security, adversarial attacks, game theory, agents, video, NLP, autoML and new AI models, energy, geology, plant recognition, …

## ❑ Full opening of the system : November 2019

- Already more than 800% oversubscription of the GPU partition for HPC and AI workloads !

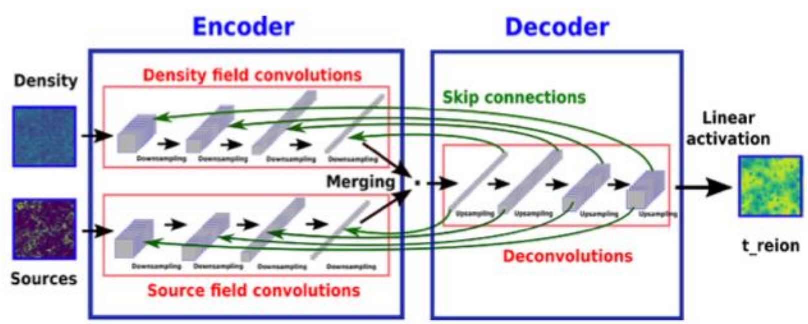
# SOME EARLY RESULTS

## Coupling learn models and simulation models in cosmology

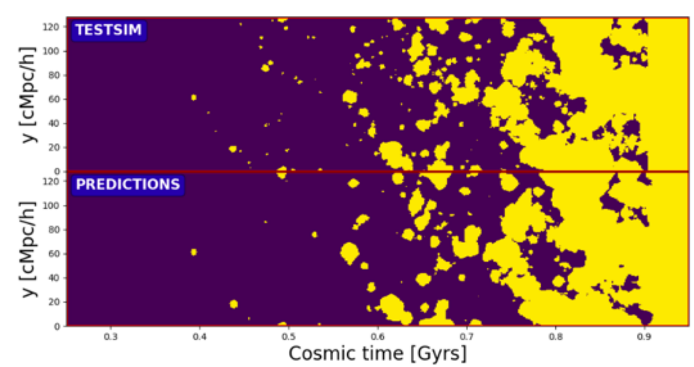
- ❑Principal Investigator : D. Aubert from Obervatoire de Strasbourg

Observatoire astronomique
de Strasbourg

- ❑Study of the **reonization** of the Universe = 1 Gyrs after Big Bang
  - Will be observed soon by instruments like EELT, JWST or SKA
  - For the moment only based on massive and costly simulations coupling gravitation, hydrodynamics and radiative transport
  - Idea : couple gravitation/hydrodynamics numerical models AND learnt radiative transport models

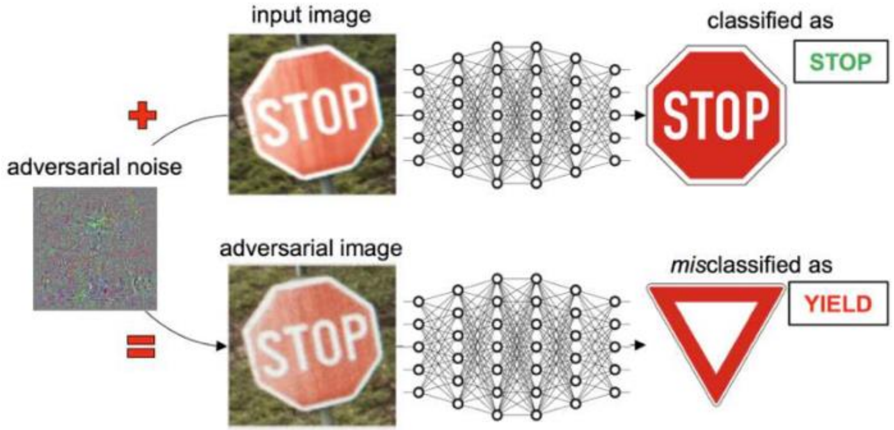



  - Use of auto-encoders based on TensorFlow and Keras
  - Methodology already validated for small cubes of 128 Mpc/$256^3$
  - Target = 128 Mpc on meshes of $1024^3$



# SOME EARLY RESULTS

## HPC for addressing adversarial attacks of neural networks




https://www.pluribus-one.it/sec-ml/wild-patterns

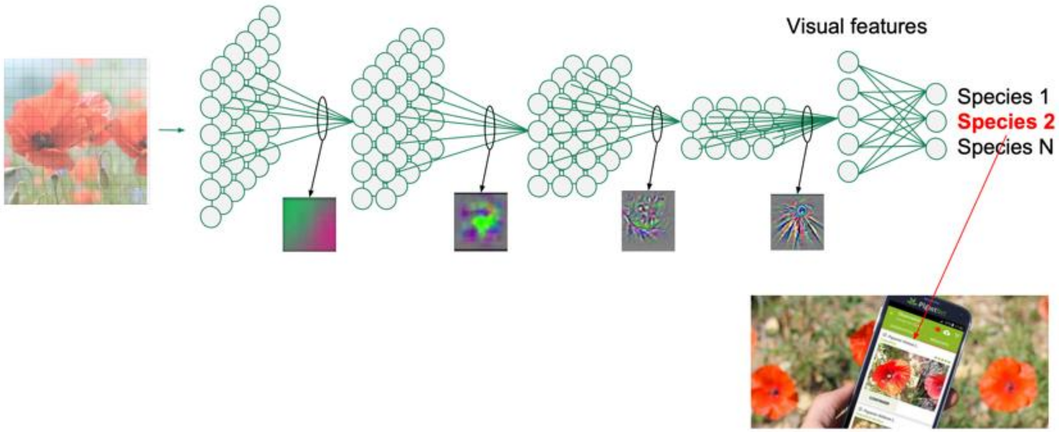
- Context
  - Neural Networks can be fooled using inputs crafted by malicious users
  - Various attacks exist: l1 , l2 , l∞
  - Adversarial Training is a technique to train networks that are robust against one specific attack

- **Goal: Train neural networks that are robust against all attacks**
- Use of Jean Zay
  - Large scale training using CIFAR10 and ImageNet on **up to 100 GPUs**
- Preliminary results from LAMSADE team
  - Randomized Adversarial Training (RAT) is robust against $l_\infty$ AND $l_2$ type attacks *(Araujo et al. Robust Neural Networks using Randomized Adversarial Training)*

# THE PL@NTNET USE CASE

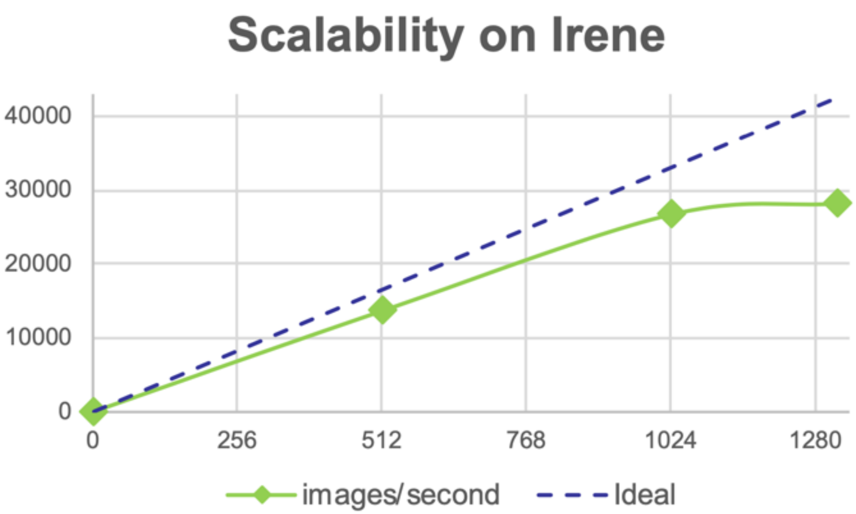
## Extreme scale up of DL models

Pl@ntNet




Joint collaboration between Intel, SurfSARA, GENCI, CINES and TGCC

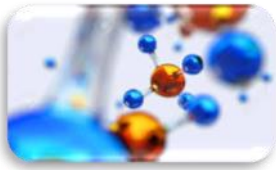
Up to 300k classes CNN
Use on Intel Caffe and specific optimisation on ResNet50




| Nodes | global batch size | images/second | Scaling efficiency |
|---|---|---|---|
| **2** | 64 | 64.5 | **100%** |
| **512** | 16384 | 13650 | **83%** |
| **1024** | 32768 | 26670 | **81%** |
| **1320** | **42240** | **28160** | **66%** |

# One more thing : Beyond Moore = Quantum based accelerators ?

- Potential uses cases



**Chemistry /materials** Hamiltonian Simulation **Polynomial speedup**

**Cryptography** Integer Factorization Shor algorithm **exponential speedup**

**Quantum database search** Grover algorithm **polynomial speedup**

**Statistical Analysis** Mathematical computation **exponential speedup**

- Strong R&D actors in Europe
  - ATOS, IBM, Pasqal, Quandela, …
  - CEA, CNRS, Univ Innsbruck, Univ Delft, Univ Cambridge, Fraunhofer, …
- Interest from Eu industry
  - Volkswagen for traffic optimisation with D-Wave
  - TOTAL and Airbus for assessing/developing QC algorithms
  - Thales on quantum captors …

- And start of massive investments
  - Quantum FET 1B€
  - German (0.7B€), UK (> 1B£ since 2014), French (soon) roadmaps

# PRACE Engagement (at date) with EU Industry with OpenR&D




European Extreme Data & Computing Initiative

# PRACE Engagement (at date) for supporting EU SMEs

PRACE

## SHAPE projects

To date, 50 SMEs have been helped
From 13 different countries
10th Call will close 1st Dec
11th call in May 2020

OPTIMA
Pharmacelera™
VORTEX Bladeless
CREO DYNAMICS
BAC ENGINEERING CONSULTANCY GROUP
OPTIMAD
airinnova Engineering Solutions and Innovations
JUAN YACHT DESIGN Juan Kouyoumdjian · Naval architecture
DESIGN METHODS™
INVENT MEDICAL
Algo'Tech INFORMATIQUE
nsilico SIMPLIFYING SCIENTIFIC PROCESSES
HYDROS INNOVATION
ALBATERN Wave Energy

PRACE SHAPE SME HPC Adoption Programme in Europe
www.prace-ri.eu


рахмат
danke
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
mersi
kia ora
barka
welalin
tack
spas
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
vinaka
спасибі
blagodaram
thank you
gracias
хвала
asante
manana
obrigada
tenki
murakoze
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
köszönöm
hvala
gràçie
bayarlalaa
bedankt
enkosi
akun
dankon
ačiū
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
sobodi
děkuji
obrigado
sagolun
sukriya
kop khun krap
ありがとう
arigatô
takk
dakujem
trugarez
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
terima kasih
tanemirt
rahmet
grazie
diolch
dhanyavadagalu
shukriya
mercé
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci