HPC - CFD and its Engineering Applications

Ashoke De

Air-force Chair Professor & Associate Dean of Academic Affairs

Computational Propulsion Lab, Dept. Of Aerospace Engineering, Dept. of Sustainable Energy Engineering, IITK

9-1-2024

NSM Workshop on
High **P**erformance **C**omputing
(HPC) **M**ethods for **C**omplex
and **M**oving **G**eometries
December 01-02, 2023

Outline of the talk

- Introduction
- Brief History of HPC
- Thinking in Parallel
- CFD-HPC Essentials
- CFD-HPC Examples
- Conclusions

What is HPC?

High Performance Computing is the use of aggregated computer systems (supercomputers) to deliver performance and throughput that cannot be achieved with regular computers.

- Aggregated: A supercomputer is made of a large number of individual computers (the "nodes" in a "cluster")
- Performance: Running the same computation faster
- Throughput: Running more computations for a given time

When do you need HPC?

- Your lab computer or laptop is unable to perform a calculation in a reasonable time frame.
- You want to increase the complexity of an existing calculation but your lab computer or laptop is not sufficient for this new task.
- You need specialised resources not available on your lab computer or laptop: large memory, large storage, GPU accelerators.

HPC - Supercomputer

High Performance Computing, i.e. Supercomputer (https://en.wikipedia.org/wiki/Supercomputer):

- A supercomputer is a computer with a high-level computational capacity compared to a general-purpose computer.

- Introduced in 1960 (Cray): from a few computing nodes to current Massively Parallel Processors (MPP) with 10^4 "off-the-shelf" nodes

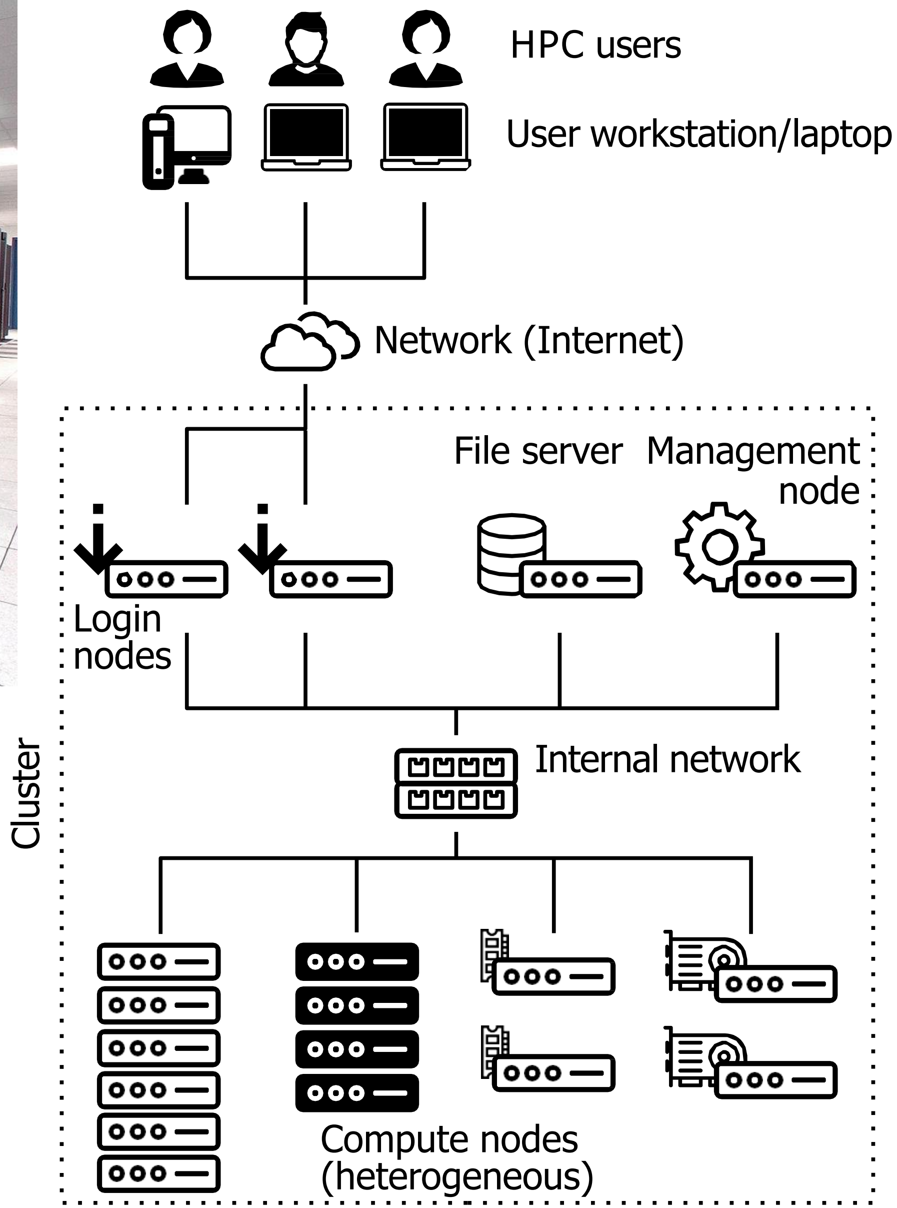
- It's motivated by the search for solutions of *grand challenges* computing applications in many research fields
  - quantum mechanics, weather forecasting, climate research, oil and gas exploration, molecular modeling, physical simulations...

- Performance of a supercomputer is measured in floating-point operations per second (FLOPS) instead of million instructions per second (MIPS).
  - T(era)Flops (10^{12}), P(eta)Flops (10^{15}), ExaFlops (10^{18}), Z(etta)Flops (10^{21})
  - Today $n \ast$ 10PFlops vs single workstation less than 1 TFlops...

- Parallelism is the key implemented with different approaches:
  - hundreds or thousands of discrete computers (e.g., laptops) distributed across a network (e.g., the Internet) devote some or all of their time to solving a common problem;
  - huge number of dedicated processors are placed in proximity and tightly connected to each other working in a coordinated way on a single task and saving time to move data around.

Supercomputer Cluster Setup

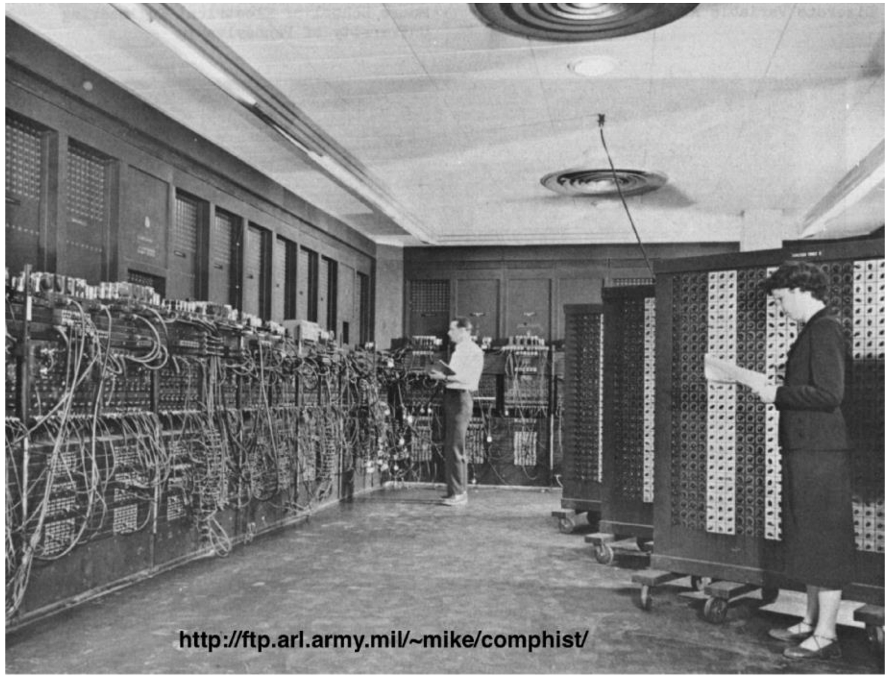
IBM Blue Gene P “Intrepid”
Argonne National Laboratory

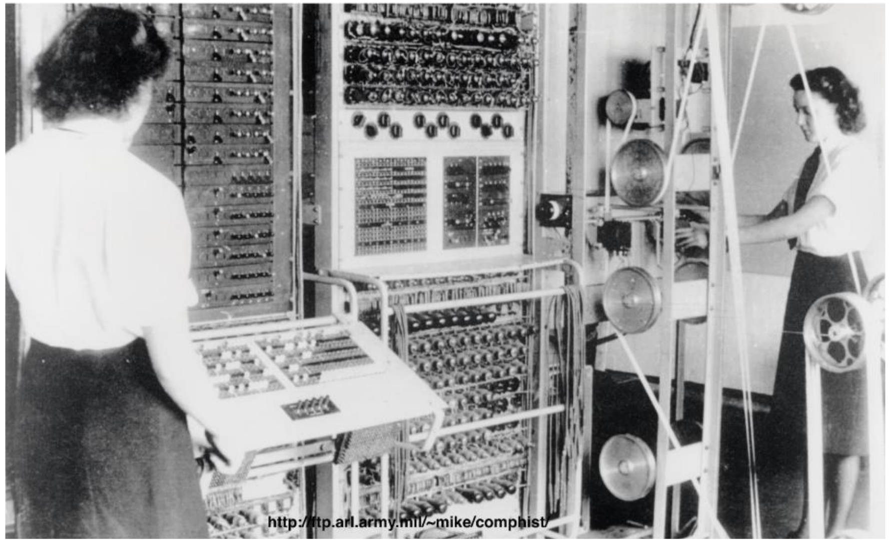


A very brief history of Supercomputing: the beginning..

Mainly motivated by military needs

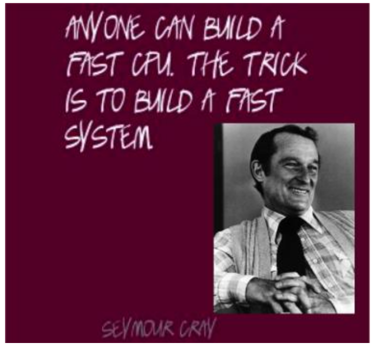
- ENIAC (USA 1943), first stored-program electronic computer
- Colossus (UK 1943), successor of Bombe (designed by Alan Turing) and built in Bletchley Park to crack Enigma nazist codes.
- Supercomputers to design nuclear weapons





A very brief history of Supercomputing: 2nd generation

1964: Control Data Corporation releases CDC 6600, the first *supercomputer*



- Single CPU@40MHz, 1-3 MFlops, 4 racks Freon cooled, 10x faster than the powerful computer ever built (IBM 7030)...
- 8 M$ cost ($\sim$ 60M$ in today's money)
- designed by the *supercomputers guru* Seymour Cray...

1975-1990: *vector processors* vs multiple *scalar*

- The "One Million Dollars" question: few, big and fast, or many, small and slow?
- *"If you were plowing a field, which would you rather use? Two strong oxen or 1024 chickens?"* (S.C.)

Cray-1

vector processor @80MHz, 133 MFlops

5-8 M$ (25M$ in today's money), 20-ton compressor for Freon cooling

innovative shape for short wiring and fast clock

Cray-1 at Deutsches Museum (by Clemens PFEIFFER)

A very brief history of Supercomputing: The CRAY era...

Cray X-MP, 1982

- 2 vector processors @105MHz, 400 aggregated MFlops
- memory shared

Cray X-MP on display at EPFL
(by RAMA)

Cray Y-MP, 1988

Cray Y-MP at Lawrence Livermore Nat. Lab
(by LLNL)

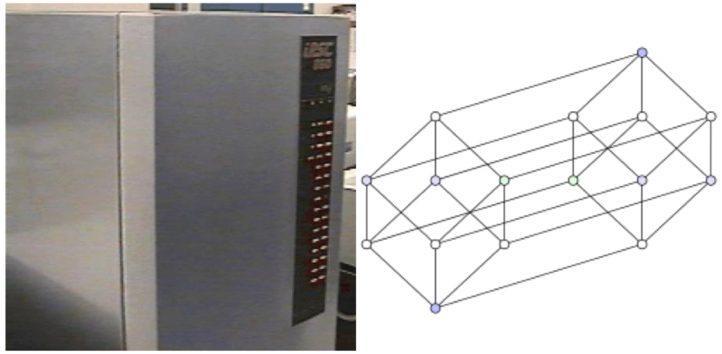
- 2, 4, or 8 vector processors with a peak of 333 MFlops each (–> 2.6GFlops!!!)
- memory shared
- dedicated OS: *UNICOS*
- followed by successful C90 series

A very brief history of Supercomputing: The MPP era...

1990 –> *The attack of Killer Micros*, Eugene Brook's talk at Supercomputing90

- Massively Parallel Processor (**MPP**) era had begun, BUT vector processing do not allow to scaling to hundred processors systems
- from shared memory to *Distributed memory*, *Message passing* and "exotic" network

Intel iPSC Hypercube, 1985



- 32-128 nodes (286 + 287 coprocessor)
- 8 Eth ports per node –> 5-Dim *hypercube* topology
- iPSC860 –> Intel Paragon systems...

Connection Machine

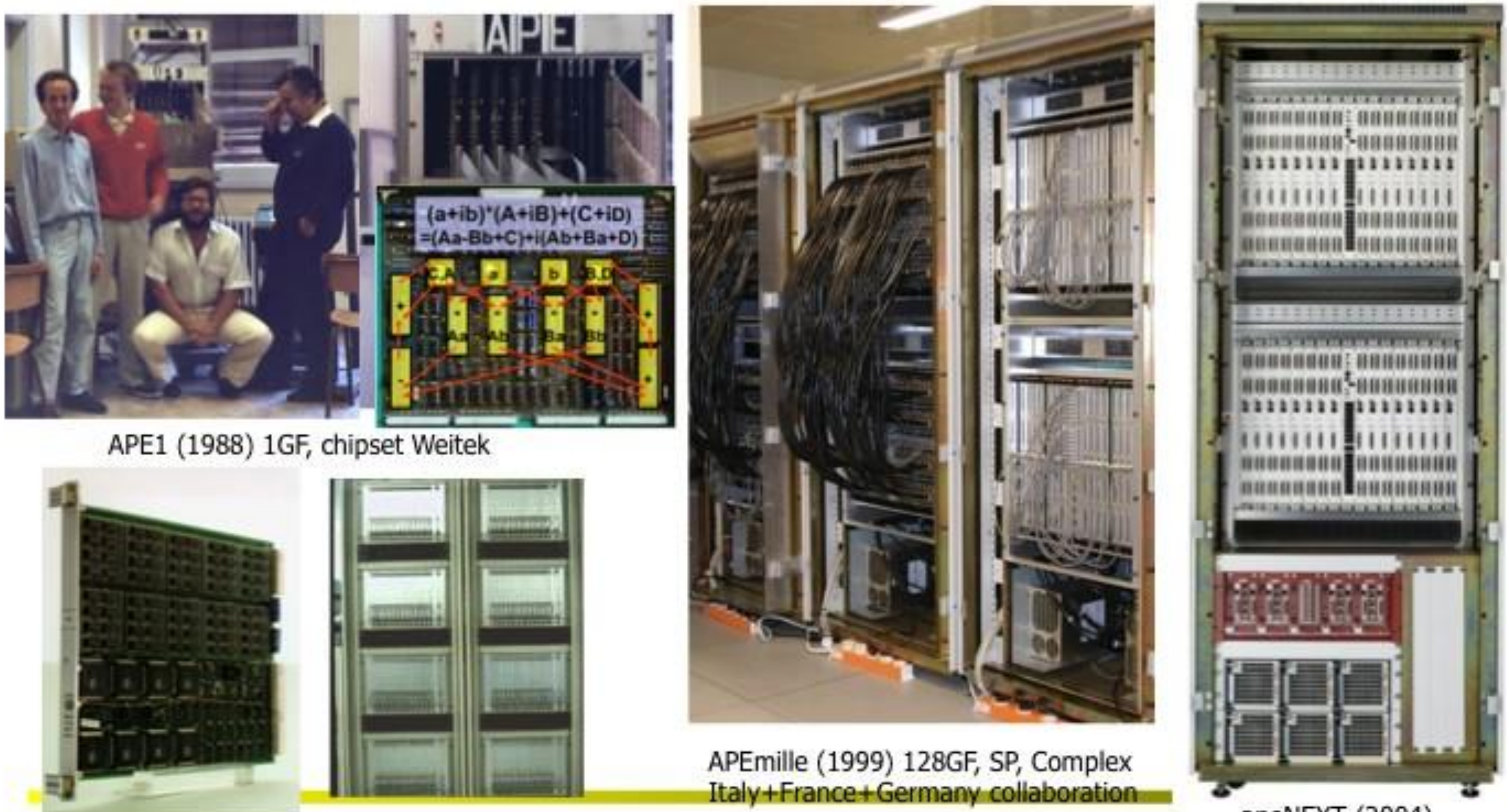
CM-1 (1985), CM-5 (1991)

- CM-1: 65k SIMD processors arranged in Hypercube
- CM-5: MIMD, 1024 processors, 59.7 GFlops, 1st TOP500 leader...

A very brief history of Supercomputing: The MPP era…

INFN APE (*Array Processor Experiment*) is a 30 Years old project…

- Several generations of MPP systems (APE1, APE100, APEmille, apeNEXT)
- Custom FloatingPoint and 3D Torus interconnect optimized for LQCD application



APE1 (1988) 1GF, chipset Weitek

APEmille (1999) 128GF, SP, Complex
Italy+France+Germany collaboration

APE100 (1992) 25GF. SP. REAL"Home made" VLSI processors

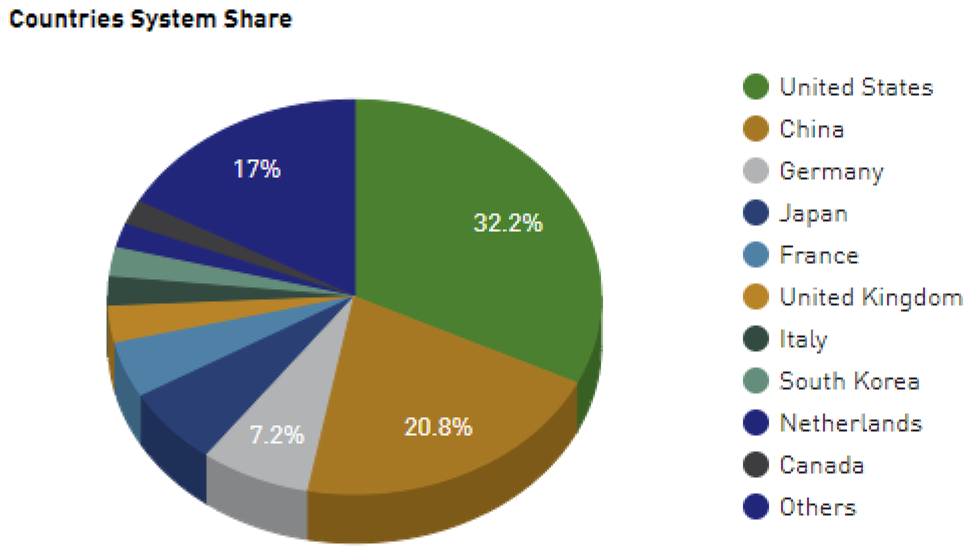
apeNEXT (2004)

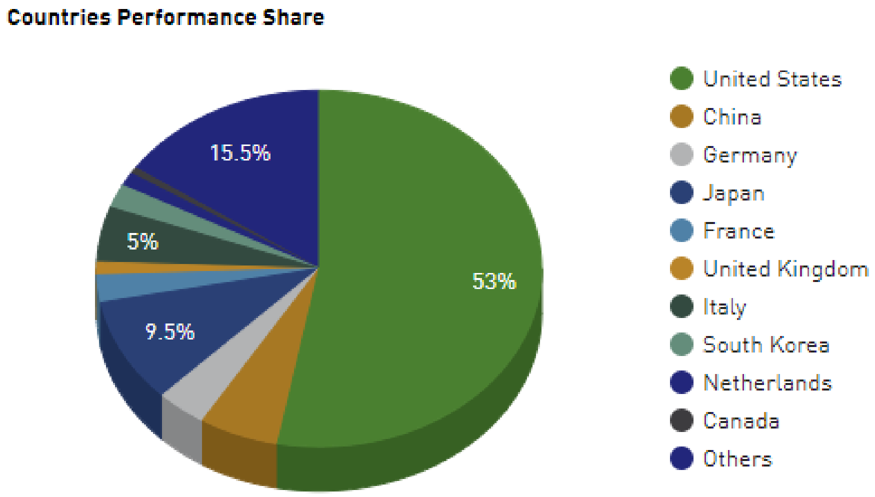
A very brief history of Supercomputing: The Cluster era...

- In 1994, Becker and Stirling built the first *PC Cluster*, with standard PCs and commodity network
- Each PC has its own private memory space and, in principle, different OS
- Low cost, scalable, leverages on CPU improvements, MP programming,...

- From mid 2000 *a Cray inside my cellphone*: introduction of powerful *GPGPU* used as accelerators.
- Most of current TOP500 ranking list are **Hybrid Supecomputers** based on clusters. Main differences due to:
  - CPUs (x86 multi-core, Power, custom)
  - FP accelerators (GPGPU, MIC, FPGA,...)
  - network technology (ethernet, Infiniband, Myrinet,...)
  - network topology (Fat-tree, Torus,...)

TOP500 supercomputer sites (Nov 2023)



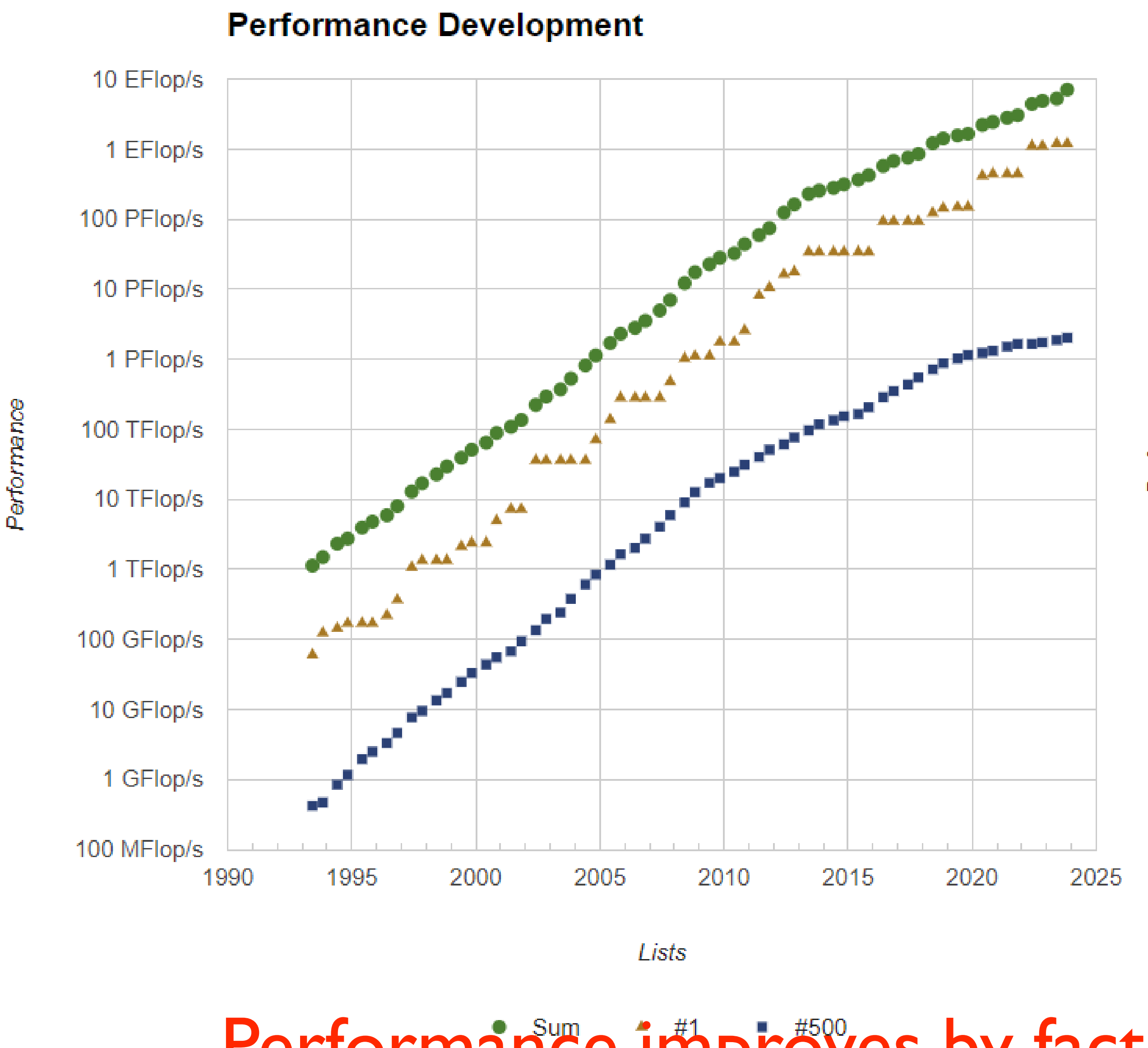


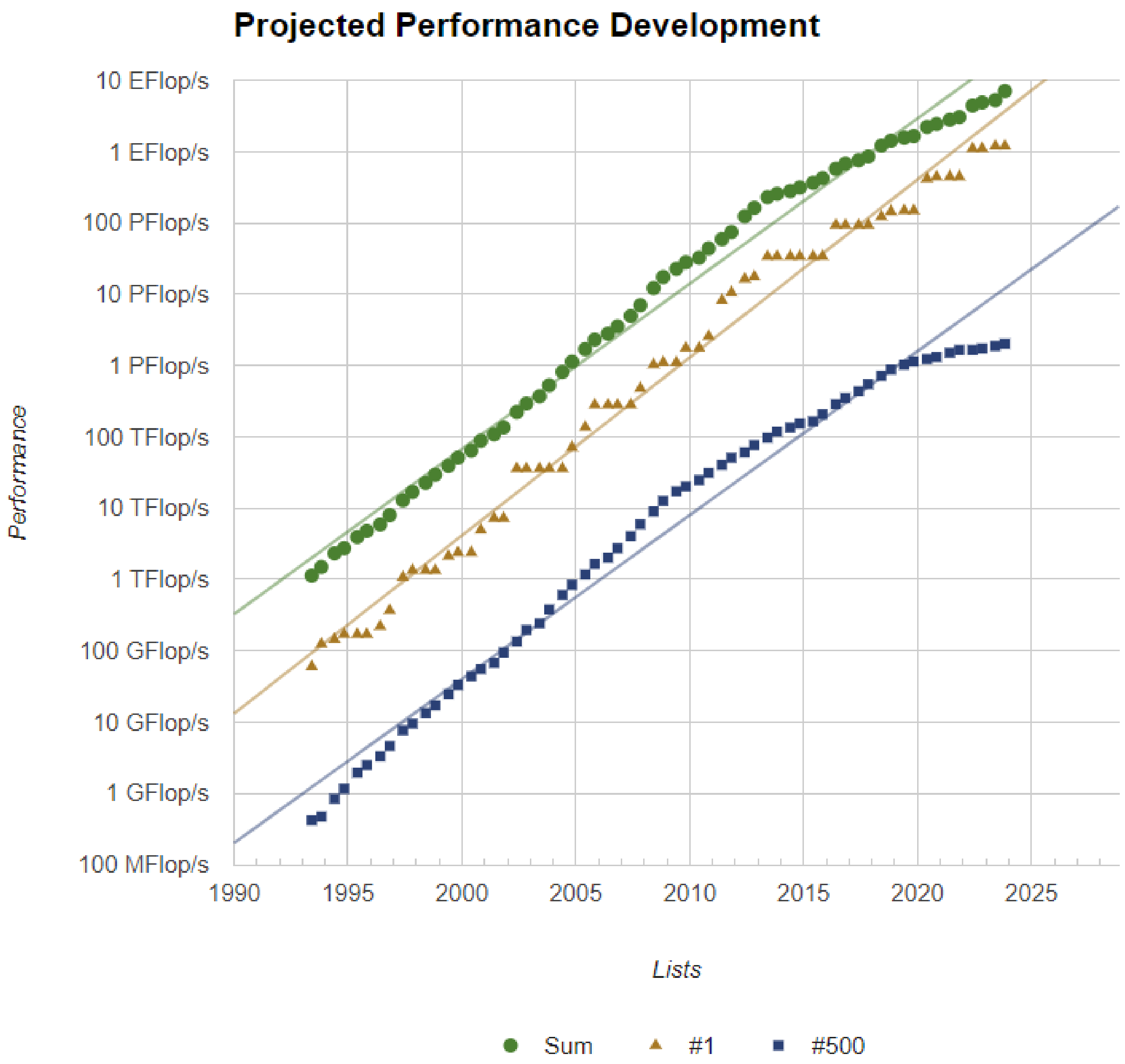
Operating system: Since November 2017, all TOP500 sites use Linux

Applications: astronomy, nuclear physics, biomolecular simulations, quantum mechanics, climate research, weather forecasting, oil/gas exploration, aerodynamics, cryptography…

www.top500.org

HPC measure & compare

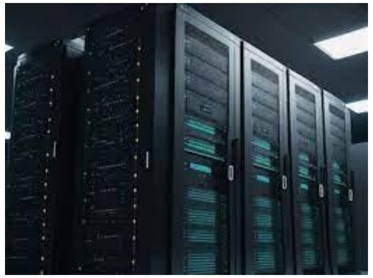




Performance improves by factor of ~10 every 4 years!

Supercomputer Indian status (Nov 2023)

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
90	**AIRAWAT - PSAI** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR, Netweb Technologies Center for Development of Advanced Computing (C-DAC) India	81,344	8.50	13.17	
163	**PARAM Siddhi-AI** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, EVIDEN Center for Development of Advanced Computing (C-DAC) India	41,664	4.62	5.27	
201	**Pratyush** - Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE Indian Institute of Tropical Meteorology India	119,232	3.76	4.01	1,353
354	**Mihir** - Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE National Centre for Medium Range Weather Forecasting India	83,592	2.57	2.81	955



www.top500.org

How does HPC increase throughput?

- ✓ Clusters have multiple compute nodes (e.g. about 100 on *Machine-X*).
- ✓ Each compute node has many CPU cores (e.g. typically 16 or 40 on *Machine-X*).
- ✓ In total, *Machine-X* offers about 2600 CPU cores, compared to 4-32 on a personal computer.
- ✓ Users submit jobs that run on the compute nodes.
- ✓ A user can submit multiple jobs at the same time.

Example:

- Alice runs 800 simulations that each take 1 hour to complete.
- On her 4-core laptop, 4 simulations can run at a time. The whole experiment takes around 200 hours (over 8 days).
- On *Machine-X*, using 10 compute nodes with 16 cores each, 160 simulations can run at the same time. The whole experiment takes around 5 hours.

How does HPC increase performance?

- The performance of a single CPU core is not enough for many computations.

- Some computations can be performed using multiple CPU cores on one or more compute nodes.

- Example:
  - Bob runs a simulation that takes 80 days to complete on a single CPU core.
  - On his 8-core workstation, the simulation takes around 10 days to complete.
  - Using 2 40-core Plato compute nodes, the simulation takes a single day to complete.

✓ Parallel algorithms are necessary…

✓ Serial programs cannot simply be "made parallel" without effort.

DEFINITION Concurrency: The property of a parallel algorithm that a number of operations can be performed by separate processors at the same time.

Concurrency is the key concept in the design of parallel algorithms:

- Requires a different way of looking at the strategy to solve a problem
- May require a very different approach from a serial program to achieve high efficiency

DEFINITION Scalability:The ability of a parallel algorithm to demonstrate a speedup proportional to the number of processors used.

DEFINITION Speedup:The ratio of the serial wallclock time to the parallel wallclock time required for execution.

$$S = \frac{\text{wallclock time}_{serial}}{\text{wallclock time}_{parallel}}$$

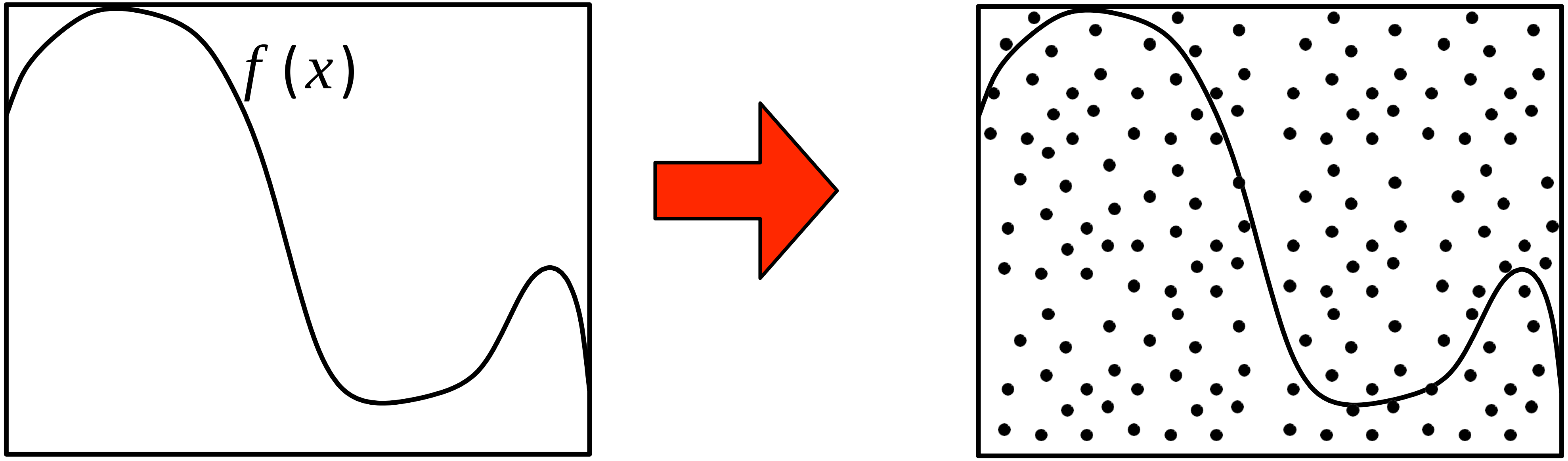
- An algorithm that has good scalability will take half the time with double the number of processors

- Parallel Overhead, the time required to coordinate parallel tasks and communicate information between processors, degrades scalability.

Example: Numerical Integration

Numerical Integration: Monte Carlo Method

- Choose N points within the box of total area A
- Determine the number of points falling below $f(x)$
- Integral value is $I = \frac{n}{N} A$

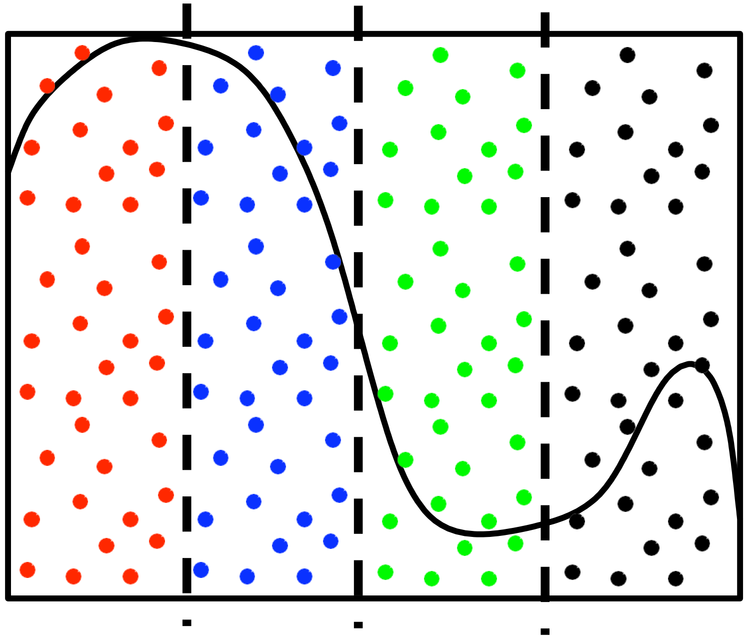


How do we do this computation in parallel?

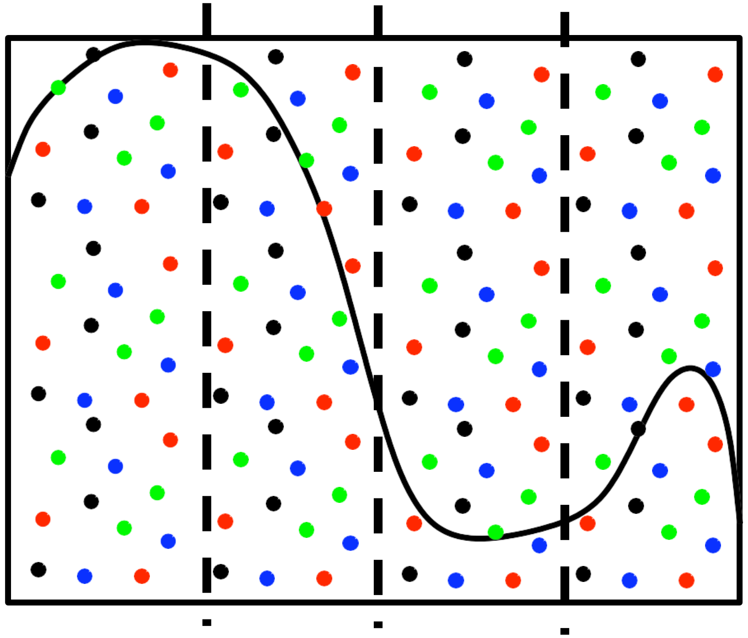
Example: Numerical Integration

Strategies for Parallel Computation of the Numerical Integral:

1) Give different ranges of to different processors and sum results



2) Give $N/4$ points to each processor and sum results



Example: Fibonacci Series

The Fibonacci series is defined by:

$$f(k+2) = f(k+1) + f(k) \quad \text{with } f(1) = f(2) = 1$$

The Fibonacci series is therefore $(1, 1, 2, 3, 5, 8, 13, 21, \ldots)$

The Fibonacci series can be calculated using the loop

```
f(1)=1
f(2)=1
do i=3, N
     f(i)=f(i-1)+f(i-2)
enddo
```

How do we do this computation in parallel?

This calculation cannot be made parallel.

- We cannot calculate $f(k+2)$ until we have $f(k+1)$ and $f(k)$

-This is an example of data dependence that results in a non-parallelizable problem

Example: Protein Folding

- Protein folding problems involve a large number of independent calculations that do not depend on data from other calculations

- Concurrent calculations with no dependence on the data from other calculations are termed Embarrassingly Parallel

- These embarrassingly parallel problems are ideal for solution by HPC methods, and can realize nearly ideal concurrency and scalability

Unique Problems Require Unique Solutions

- Each scientific or mathematical problem will, in general, require a unique strategy for efficient parallelization

Thus, each of you may require a different parallel implementation of your numerical problem to achieve good performance.

- Flexibility in the way a problem is solved is beneficial to finding a parallel algorithm that yields a good parallel scaling.

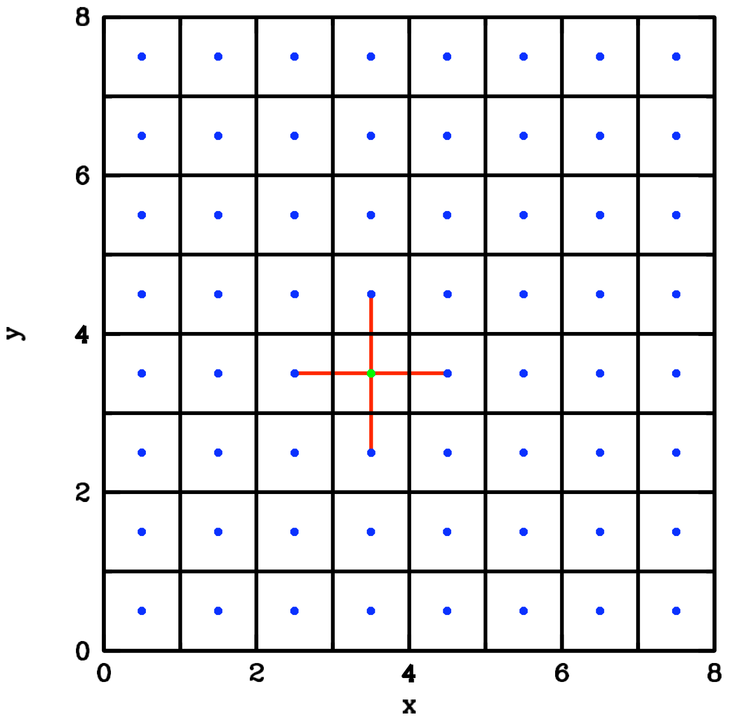
- Often, one has to employ substantial creativity in the way a parallel algorithm is implemented to achieve good scalability.

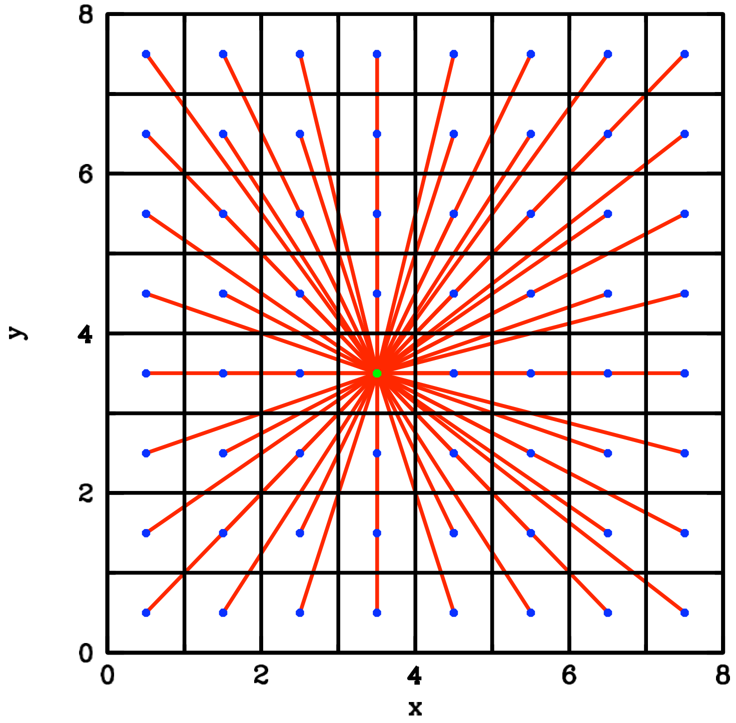
Understanding the Dependencies

•One must understand all aspects of the problem to be solved, in particular the possible dependencies of the data.

•It is important to understand fully all parts of a serial code that you wish to parallelize.

Example: Pressure Forces (Local) vs. Gravitational Forces (Global)





Rule of Thumb

When designing a parallel algorithm, always remember:

Computation is FAST

Communication is SLOW

Input/Output (I/O) is INCREDIBLY SLOW

Other Issues

In addition to concurrency and scalability, there are a number of other important factors in the design of parallel algorithms:

Locality

Granularity

Modularity

Flexibility

Load balancing

We'll learn about these as we discuss the design of parallel algorithms.

CFD-HPC Essentials

CFD-HPC Examples

Role of CAE (CFD)

- **Computer-Aided Engineering** (CAE) is the broad usage of computer software to aid in engineering analysis tasks.

  - ✓ Finite Element Analysis (FEA)
  - ✓ Computational Fluid Dynamics (CFD)
  - ✓ Multi-body dynamics (MBD)
  - ✓ Optimization

- Software tools that have been developed to support these activities are considered CAE tools.

- The term encompasses simulation, validation, and optimization of products and manufacturing tools.

In the future, CAE systems will be major providers of information to help support design teams in decision making !!

CFD

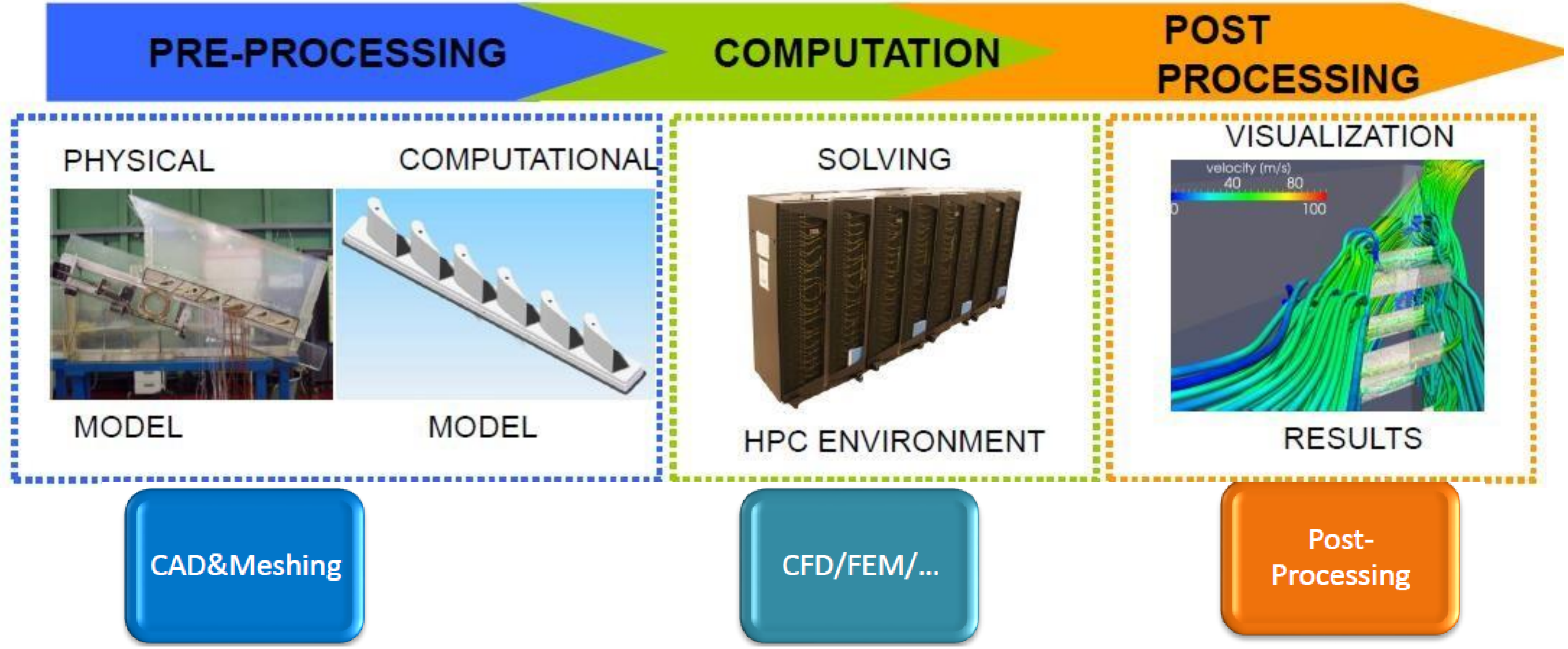
- **CFD is the "science" of predicting fluid behaviour**
  - **Flow field, heat transfer, mass transfer, chemical reactions, etc…**
    - By solving the governing equations of fluid flow using a numerical approach (computer based simulation)

- **The results of CFD analyses**
  - **Represent valid engineering data that may be used for**
    - Conceptual studies of new designs (with reduction of lead time and costs)
    - Studies where controlled experiments are difficult to perform
    - Studies with hazardous operating conditions
    - Redesign engineering

- **CFD analyses represent a valid**
  - **Complement to experimental tests**
    - Reducing the total effort required in laboratory tests

Phases of CFD

- **Pre-processing** – defining the geometry model, the physical model and the boundary conditions
- **Computing** (usually performed on high powered computers (HPC))
- **Post-processing of results** (using scientific visualization tools & techniques)



Iterative process !!

Image Source: internet

Why use CFD ?

- Analysis and Design
  - ❖ Simulation-based design instead of "build & test"
    - ✓More cost effectively and more rapidly than with experiments
    - ✓CFD solution provides high-fidelity database for interrogation of flow field
  - ❖ Simulation of physical fluid phenomena that are difficult to be measured by experiments
    - ✓Scale simulations (e.g., full-scale ships, airplanes)
    - ✓Hazards (e.g., explosions, radiation, pollution)
    - ✓Physics (e.g., weather prediction, planetary boundary layer, stellar evolution)
- Knowledge and exploration of flow physics

Why use CFD ?

	Simulation(CFD)	Experiment
Cost	Cheap	Expensive
Time	Short	Long
Scale	Any	Small/Middle
Information	All	Measured Points
Repeatable	All	Some
Security	Safe	Some Dangerous

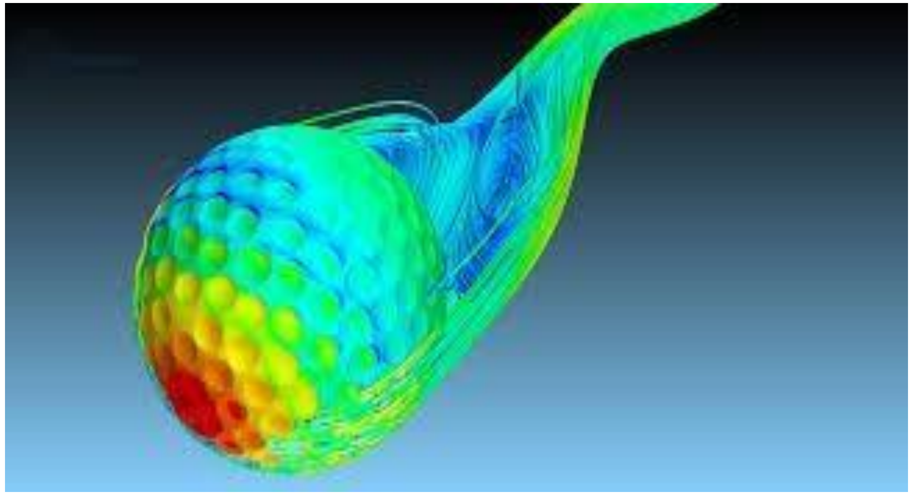


Image Source: internet

How an engineering problem is addressed using CFD?

I. There is a Physical or Engineering Problem to solve

From the Continuum and Real Problem to a Simplified Mathematical model

I. Problem Identification & Pre-processing

From Simplified Mathematical model to the Discrete and Virtual Numerical Model

CFD is a tool that require you switch on your brain… and your fantasy!

III. Model setting and solver execution

Consider quality of the numerical model: convergence

IV. Results analysis and post-processing

Consider quality of the numerical and mathematical model (as VIRTUAL model of a REAL problem): Accuracy

Image Source: internet

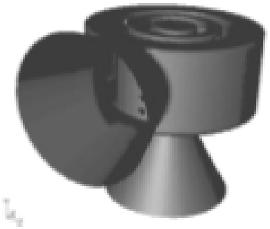
CFD steps

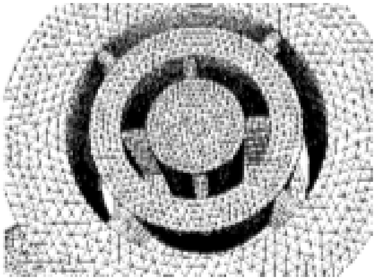
1. Problem Identification: Pre-processing

1. Clarify the GOALS and the CONSTRAINS of the project
2. Identify the domain you will model.

3a. Design the domain

3b. Generate the mesh



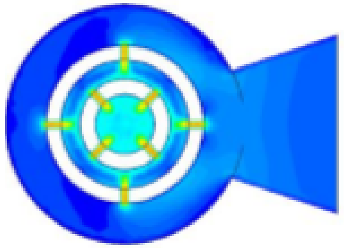


2. Model setting and Solver Execution

4. Identify the equations required to the model
   - Define the BC, the materials properties and the numerical setting
   - Set up a solution monitoring
5. Initialise and (ask the solver to) compute solution

3. Result analysis and review: Post-Processing

6. Result analysis from qualitative and quantitative points of view
7. Critical revision to the model



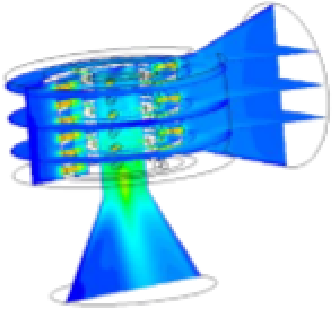
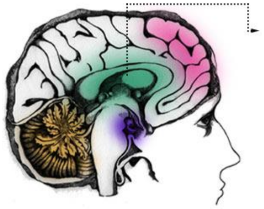


Image Source: internet

Numerical methods



Analytical Equations —Discretization→ Discretized Equations

- Discretization Methods
  - ✓ Finite Difference

    Straightforward to apply, simple, sturctured grids
  - ✓ Finite Element

    Any geometries
  - ✓ Spectral methods

    Using truncated Fourier series or Chebyshev polynomials
  - ✓ Lattice Boltzmann methods

    Using discrete Boltzmann equations
  - ✓ Finite Volume

    Conservation, any geometries

Image Source: internet

CFD solvers are based on numerical methods

- **Commercial CFD. General purposes**
  - FLUENT (ANSYS, USA)
  - CFX (ANSYS, UK)
  - STAR CD (Computational Dynamics Ltd., UK)
  - COMSOL Multiphysics
- **In house code**
  - Confidentiality
  - Internal know how
  - Specific applications
- **Academic Code**
  - Natural location in which CFD commercial codes were born
  - Main Interests: development of new models
- **Open Source code**
  - OpenFoam (www.openfoam.org/com)
  - …. probably the future

❑ Advantages

- ✓ Costs
- ✓ Time to market
- ✓ Capability to Simulate Real Conditions
- ✓ Ability to Simulate Ideal Conditions
- ✓ Comprehensive Information

❑ Limitations

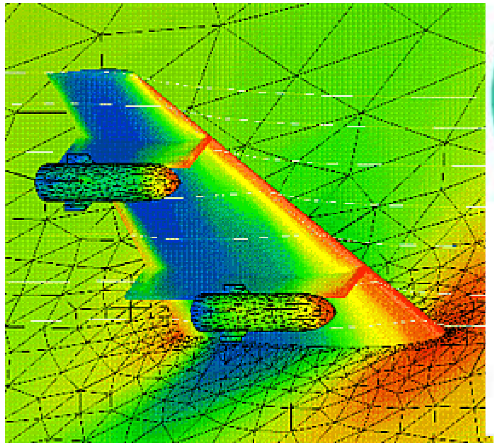
- Physical models
- Numerical Error
- Boundary conditions

Image Source: internet

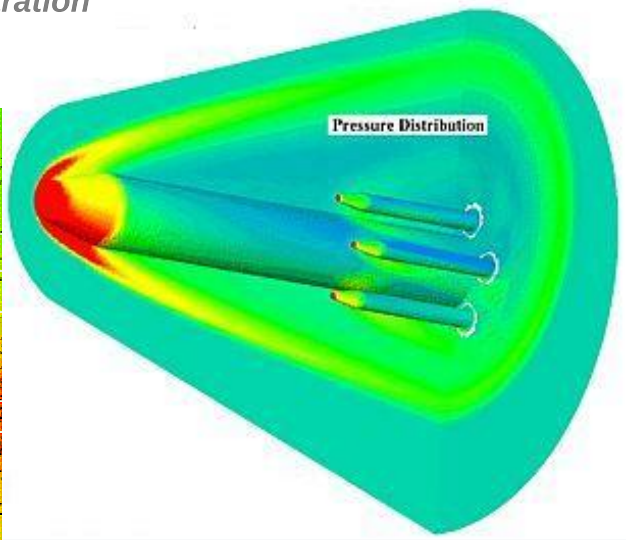
(Aerospace)

- Where is CFD used?
  - ***Aerospace***
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports

F18 Store Separation



Wing-Body Interaction

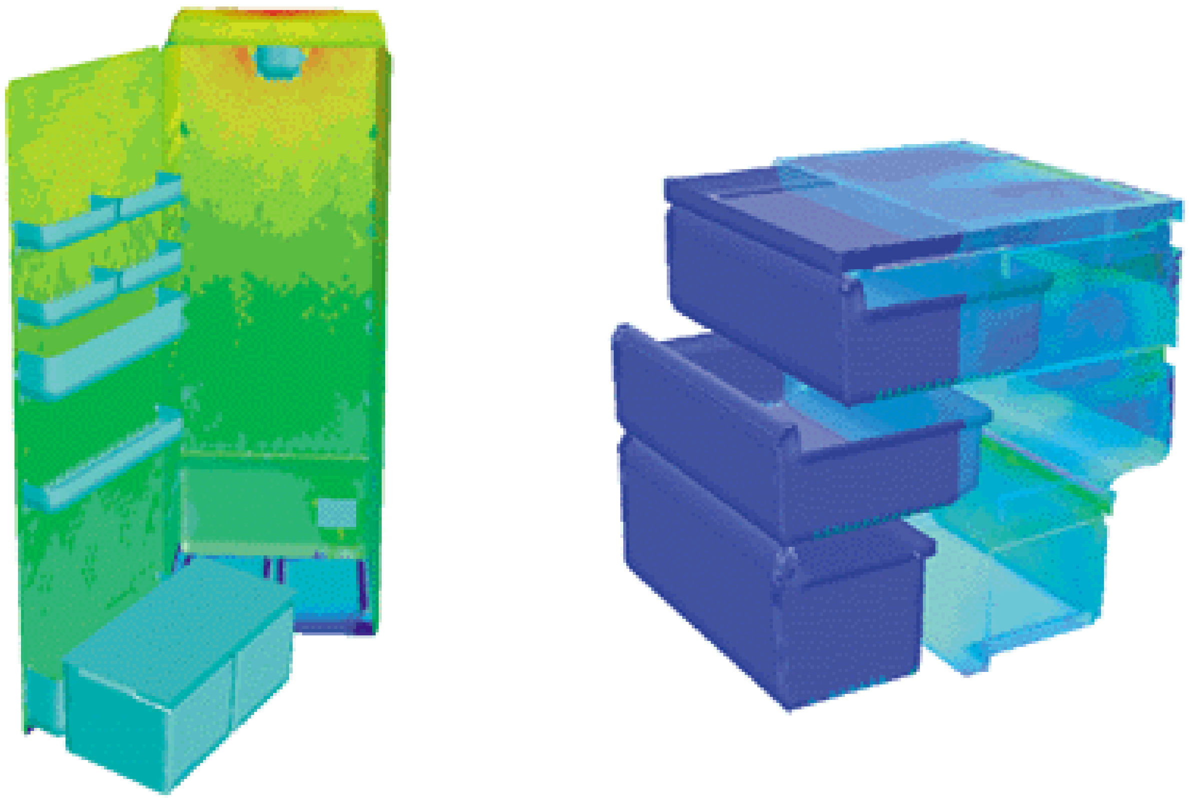


Hypersonic Launch Vehicle

Image Source: internet

(Appliances)

- Where is CFD used?
  - Aerospace
  - ***Appliances***
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



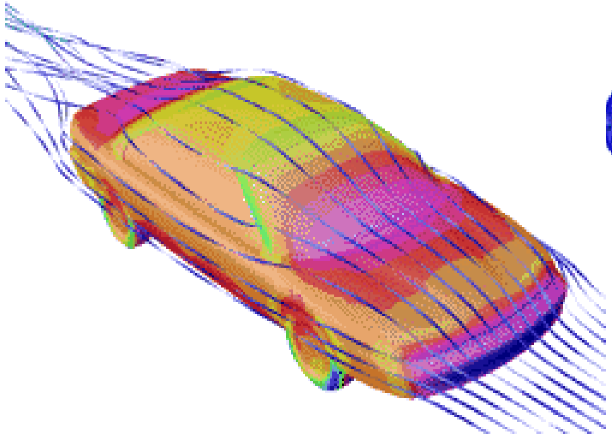
Surface-heat-flux plots of the No-Frost refrigerator and freezer compartments helped BOSCH-SIEMENS engineers to optimize the location of air inlets.

Image Source: internet

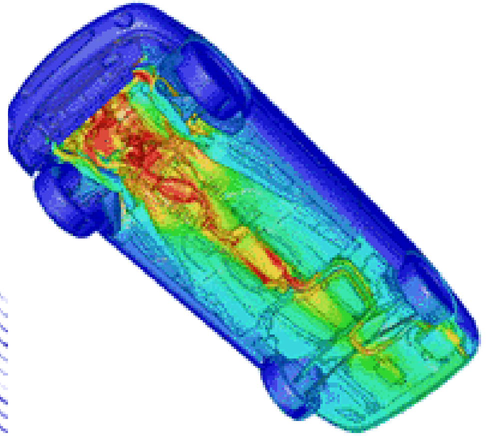
CFD-HPC Applications

(Automotive)

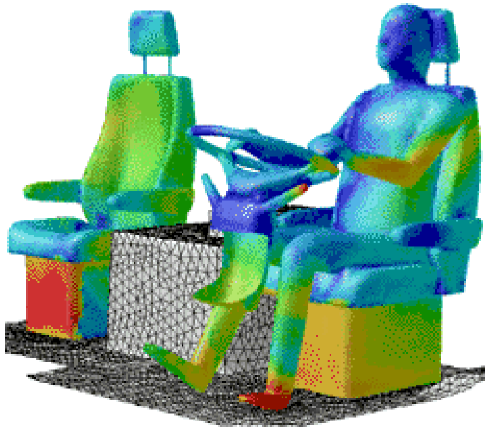
- Where is CFD used?
  - Aerospace
  - Appliances
  - ***Automotive***
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



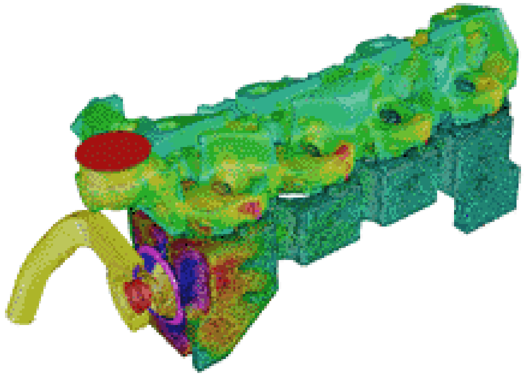
External Aerodynamics



Undercarriage Aerodynamics



Interior Ventilation



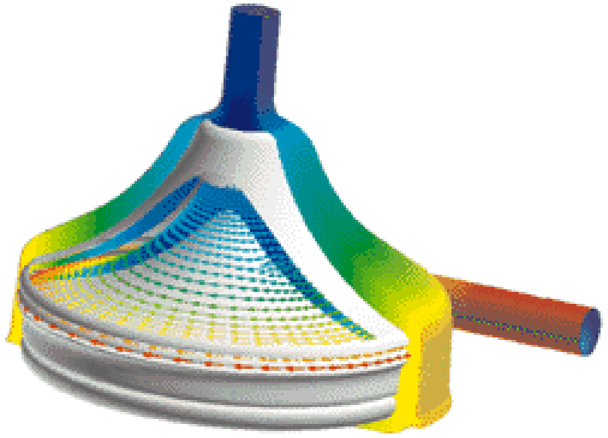
Engine Cooling

Source: internet

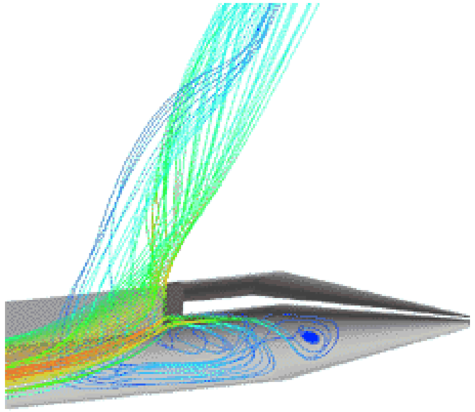
Image Source: internet

(Biomedical)

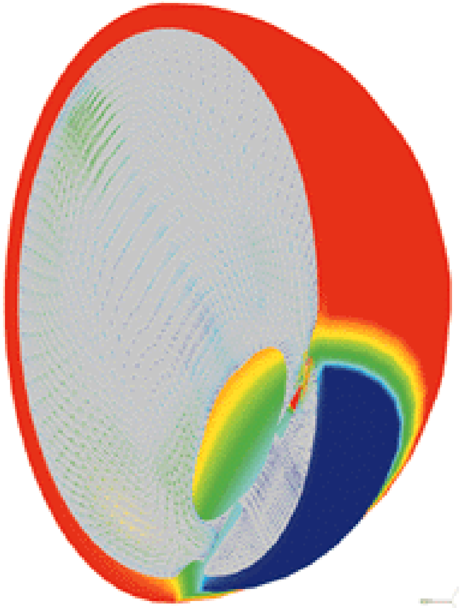
- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - ***Biomedical***
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



Medtronic Blood Pump



Spinal Catheter

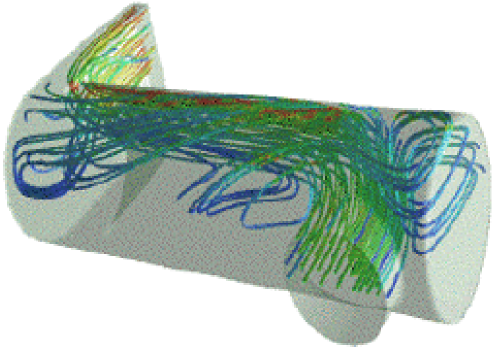


Temperature and natural convection currents in the eye following laser heating.

Image Source: internet

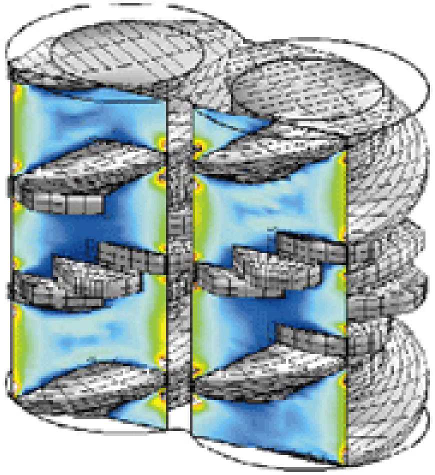
(Chemical Processing)

- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - ***Chemical Processing***
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



Polymerization reactor vessel - prediction of flow separation and residence time effects.

Twin-screw extruder modeling

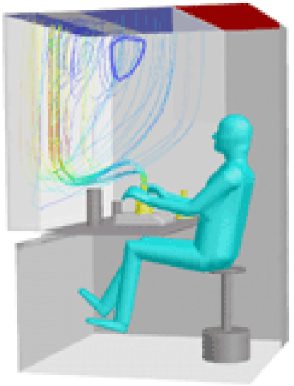


Shear rate distribution in twin-screw extruder simulation

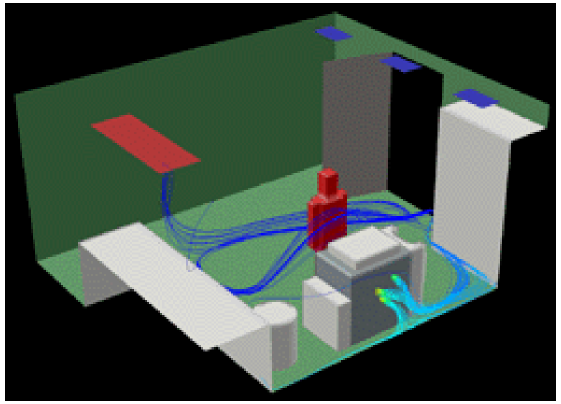
Image Source: internet

(HVAC&R)

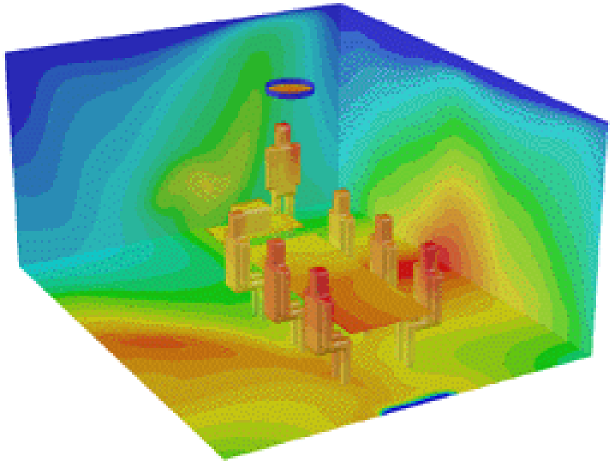
- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - ***HVAC&R***
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



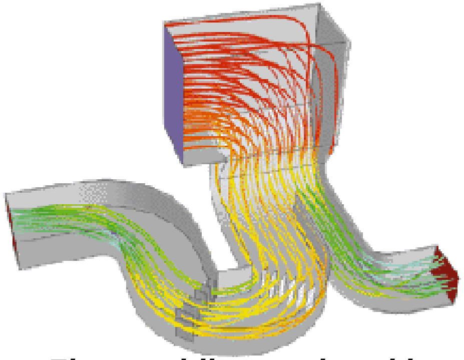
Streamlines for workstation ventilation



Particle traces of copier VOC emissions colored by concentration level fall behind the copier and then circulate through the room before exiting the exhaust.



Mean age of air contours indicate location of fresh supply air

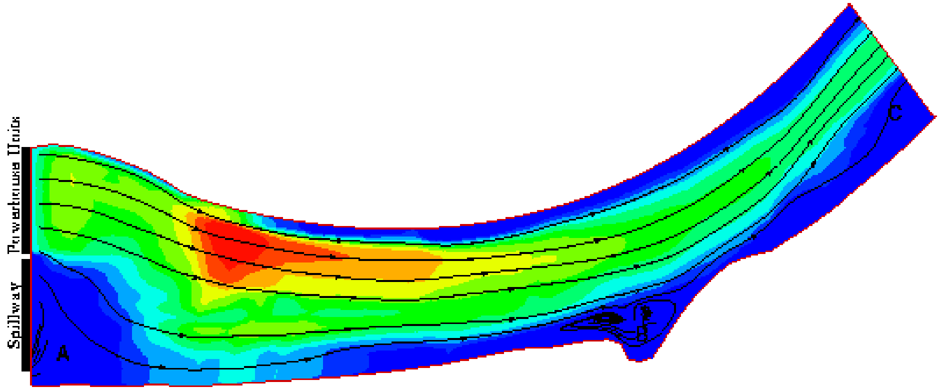


Flow pathlines colored by pressure quantify head loss in ductwork

Image Source: internet

(Hydraulics)

- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - ***Hydraulics***
  - Marine
  - Oil & Gas
  - Power Generation
  - Sports



Total Discharge = 125,000 cfs (no flow through spillway)

Image Source: internet

(Marine)

- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - ***Marine***
  - Oil & Gas
  - Power Generation
  - Sports

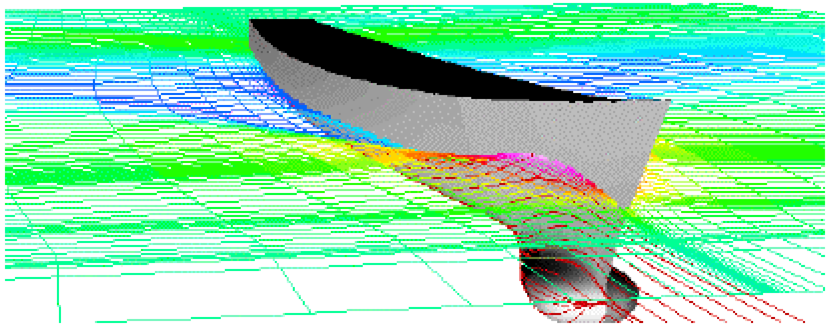
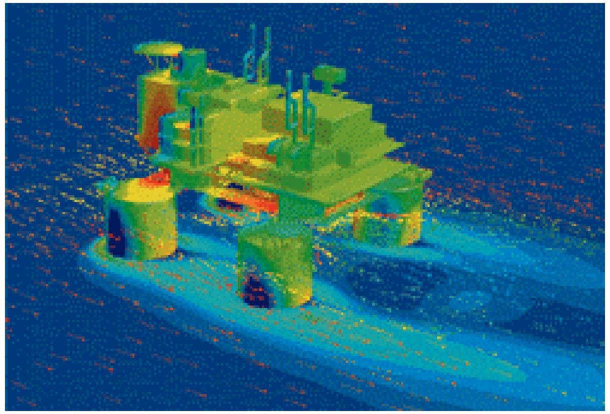


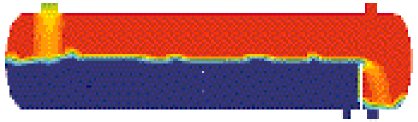
Image Source: internet

(Oil & Gas)

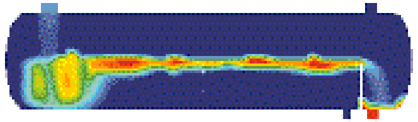
- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - ***Oil & Gas***
  - Power Generation
  - Sports



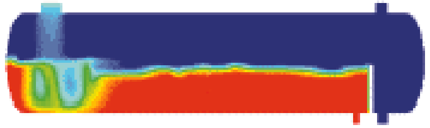
Flow vectors and pressure distribution on an offshore oil rig



Volume fraction of gas

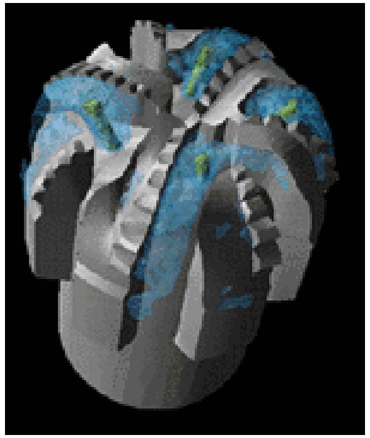


Volume fraction of oil



Volume fraction of water

Analysis of multiphase separator



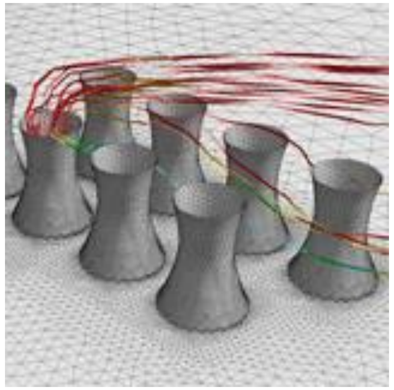
Flow of lubricating mud over drill bit

Image Source: internet

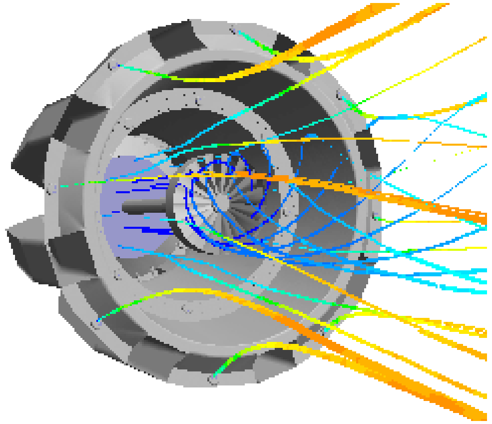
(Power Generation)

- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - ***Power Generation***
  - Sports

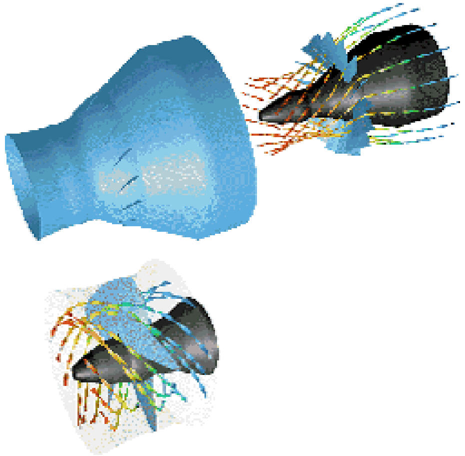
Source: internet



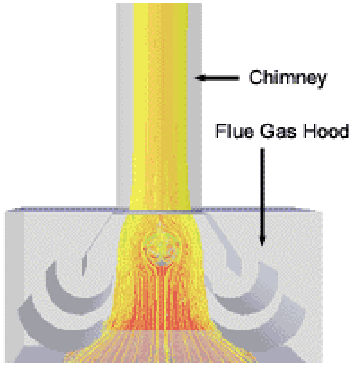
Flow around cooling towers



Flow in a burner



Flow pattern through a water turbine.



Pathlines from the inlet colored by temperature during standard operating conditions

Image Source: internet

CFD-HPC Applications

(Sports)

- Where is CFD used?
  - Aerospace
  - Appliances
  - Automotive
  - Biomedical
  - Chemical Processing
  - HVAC&R
  - Hydraulics
  - Marine
  - Oil & Gas
  - Power Generation
  - ***Sports***

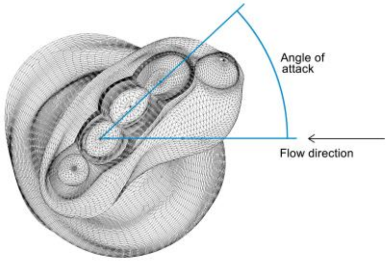


Image Source: internet

Conclusions

Common pathology of CFD man!

- ✓ *(I calculate, therefore I am):*
  To be not fully conscious, false feeling of facility, trend to neglect quantitative results in qualitative results' favor, trend to underestimate the discretization procedure, to accept not converging results!

- ✓ *(I am God):*
  Trend to full the resources, risk to calculate a lot, more and more, and understand few, less and less, risk to accept as "true" a solution because the simulation converge!

However you have the more robust and efficient code it is always important to estimate each point of view of the project if you want to obtain a result.

Therefore take CARE and switch always your brain and your judging capability on when evaluating results form CFD codes.

Conclusions

CFD: Computational Fluid Dynamics or Colorful Fluid Dynamics?

❖ In solving fluid flow problems we need to be aware that the underlying physics is complex and the **results** generated by a CFD code are **at best as good as the physics embedded in it** and **at worst as good as its operator.**

✓ *No one believes the CFD results except the one who performed the calculation, and everyone believes the experimental results except the one who performed the experiment. (P.J. Roache)*

✓ *The greatest disaster one can encounter in computation is not instability or lack of convergence but results that are simultaneously good enough to be believable but bad enough to cause trouble. (J.H. Ferziger)*

✓ *Essentially, all models are wrong, but some are useful*

George E. P. Box (1919–2013)

Conclusions

Last suggestion…..

Look at..

your (possible) competence ……
the weakness of your stakeholder (the company)… the "smart and fascinating"" tools available….

and **challenge yourself**

If we wait until we are ready,
We'll be waiting for the rest of our lives !

If it doesn't challenge you,
It doesn't change you ..

Image Source: internet


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
спасибо
gracias
dank je
bedankt
dziękuje
obrigado
sukriya
kop khun krap
go raibh maith agat
mochchakkeram
arigato
dakujem
terima kasih
grazie
мерси
감사합니다
merci


Q&A
HOW
Why
Who
When
What
Where