From Data to Information to Knowledge to Decision making

- Data: Things that are measured
  - New technologies lead to new data:
    - Competition to have the latest technology
    - Focus on storage needs to store yet more data
  - How do we leverage data? Integratable? Scalable?
- Information: Processed data
  - work cannot be done without required information.
  - By improving information supply and its processing, the whole process usually can be made more efficient. Input and its processing
  - Reduce unnecessary complexity of information processing systems; protect against information overload.
  - information management can result in better profitability
    - Data Acquisition, data storage, Access to data, automation
- Knowledge: Processed data plus meaningful relationships between measured entities Power of graph modeling


Big Data:
Expanding on 3 fronts
at an increasing rate.
Data
Velocity
Real Time
Near Real Time
Periodic
batch
Data: Things that are measured
• Information: Processed data
• Knowledge: Processed data plus meaningful relationships between measured entities
Power of graph modeling
table
MB
GB
TB
PB
Data
Volume
data base
photo
Social
Web
Audio
Video
unstructured
Mobile
Data
Variety

Creation of information model

- To understand, and possibly control or operate something, a general strategy is to build a model of it.
- Models can be built on paper, or even made of wood; but computer modelling is superior to these approaches in several respects. Complete, large, elaborate, easily modifiable information models form the core of many important information systems.

Web-based Process

- Data and applications can be accessed from any location
  - Data and applications can easily be shared through a common platform
  - Clouds need not be public; companies
  - can introduce private cloud computing solutions

A large database

- **Jigsaw** is an online business directory of companies and business professionals (This information consisting of what is commonly found on a business cards.
- Credibility of provided information is a concern
- it has also raised questions of privacy as most of the site's database is entered without permission from the person being listed

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

InfoWare 2013: ICCGI Panel

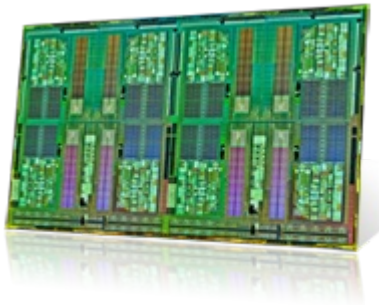
Parallel Scientific Computing:

Which technology is best suited for my problem?

Arno Leist, PhD

Parallel Trends in CPU Architectures

- Core count increases
  - 2-8 cores most common
  - 15 cores in the next high-end Xeon (Ivy Bridge-EX)



- Vector units are getting wider
  - AVX 256-bit float, 128-bit integer (Sandy Bridge, Bulldozer)
  - AVX2 256-bit (Haswell)
  - Xeon Phi vector instructions 512-bit (Knights Corner)
  - Next generation: AVX3.x with 512-bit? (Knights Landing, Skylake)
- Superscalar on steroids: simultaneous multi-threading (SMT)
  - Two threads per core/module (Sandy Bridge, Bulldozer)
  - Four threads per core (Knights Corner, POWER7)

x86-Based Architectures

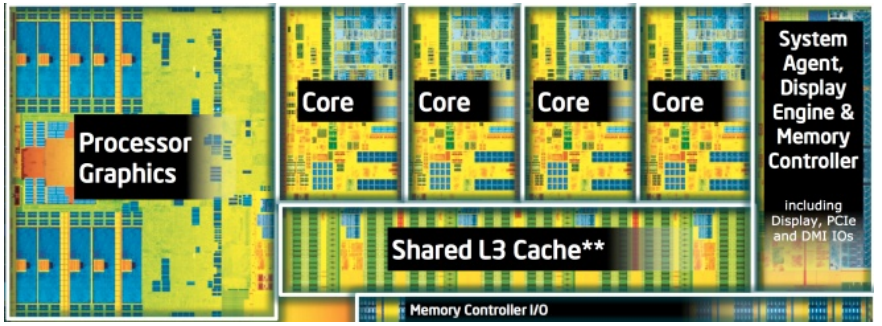
- General purpose CPUs
  - Intel Xeon (Sandy/Ivy Bridge architecture)
  - AMD Opteron (Piledriver architecture)
- Co-processors
  - Intel Xeon Phi (Knights Corner architecture)
    - 61 cores
    - 244 threads
    - 512-bit vector units
    - 352 GB/s bandwidth

The GPU as Compute Accelerator

- NVIDIA Tesla K20X
  - 2688 CUDA cores
  - 14 Streaming Multiprocessors (SM)
  - 250 GB/s bandwidth
  - Up to 2048 resident threads per SM
- AMD FirePro S9000
  - 1792 Stream Processors
  - 28 Compute Units
  - 264 GB/s bandwidth

Heterogeneous Architectures

- AMD APU
- Intel 3rd generation Core architecture CPUs with HD Graphics 4000/2500 and newer
- NVIDIA Project Denver
  - Custom ARMv8 compliant 64-bit CPU
  - Maxwell GPU
  - Release in 2015?



Parallel Software Frameworks

- Multi-threaded code
  - Pthreads and other low-level threading libraries
  - Multi-tasking libraries (e.g. TBB)
  - OpenMP
  - …
- Vectorised code
  - CUDA (NVIDIA GPUs; x86 compiler from PGI)
  - OpenCL (CPU and GPU)
  - OpenMP 4.0 (CPU and GPU)
  - OpenACC (so far mainly NVIDIA GPUs)
  - Intel Cilk Plus (x86 CPUs)
  - …

Discussion

Which technology is best suited for my problem?

Discussion: Things to Consider

- Not all combinations are possible
  - What to choose first, hardware or software?
- Existing software
  - Is it feasible to re-implement it? Or will only new algorithms be parallelised?
  - What language is it written in?
  - Is some of it already parallelised?
- New developments
  - How much of it can be parallelised?
  - What kind of parallelism can be extracted? (threading, vectorisation)
  - Data access patterns: random, sequential, localised?
  - Are the algorithms compute bound or bandwidth bound?
  - How many computations are performed per data element?

Are the models powerful enough for information handling? Missing Jigsaw Pieces

Jean-Denis MATHIAS

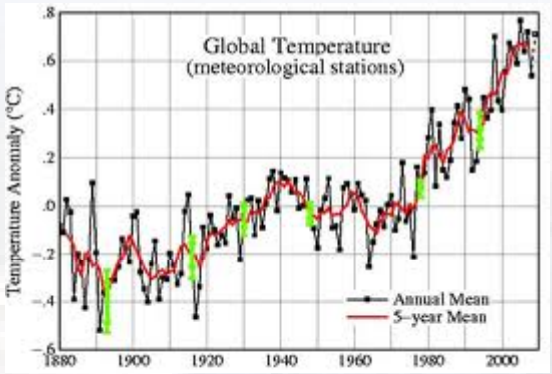
Laboratoire d'**I**ngénierie pour les **S**ystèmes **C**omplexes (**LISC**)

IRSTEA – Clermont-Ferrand, France

Example

Climate change

Experimental data



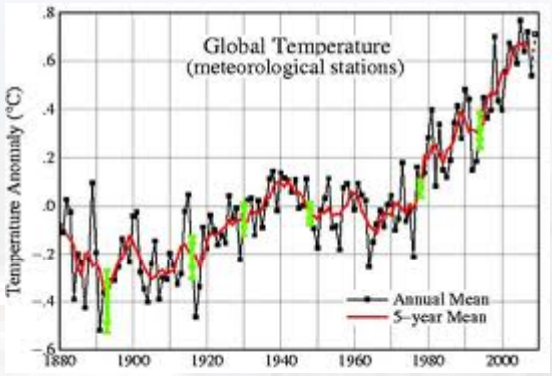
agreenliving.org

Global warming?

Example

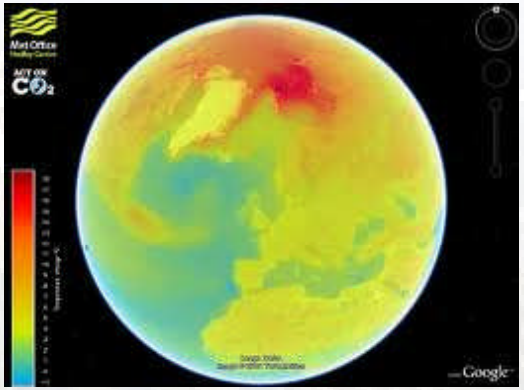
Climate change

Experimental data



agreenliving.org

Model



gearthblog.com

Complexity…

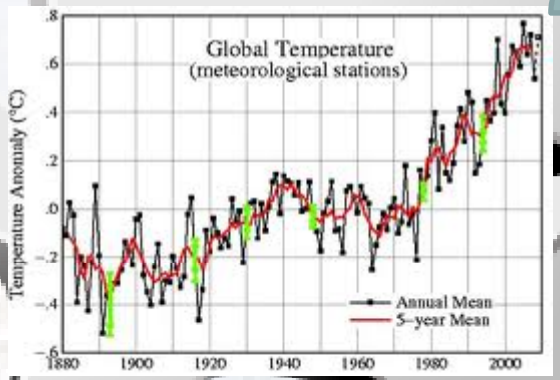
Global warming?????????

Better understanding

Example

Climate change

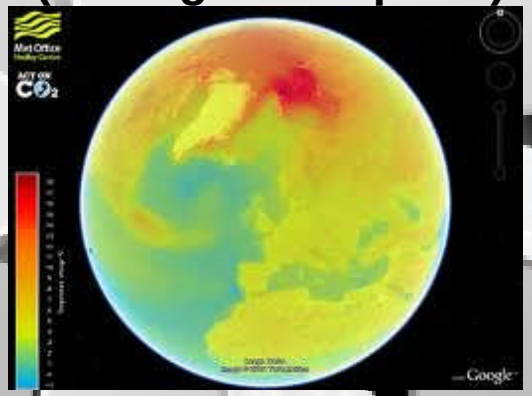
Experimental data



agreenliving.org

Data

Model (testing assumption)



gearthblog.com

Better understanding

Information handling depends on the system complexity…!!