



eScience

# *Content*

- **introduction**
- **previous revolutions**
- **common misunderstandings**
- **present web 2.0 developments**
- **useful steps towards eScience**

# *Introduction ...*

- **introduction**
- **previous revolutions**
- **common misunderstandings**
- **present web 2.0 developments**
- **useful steps towards eScience**

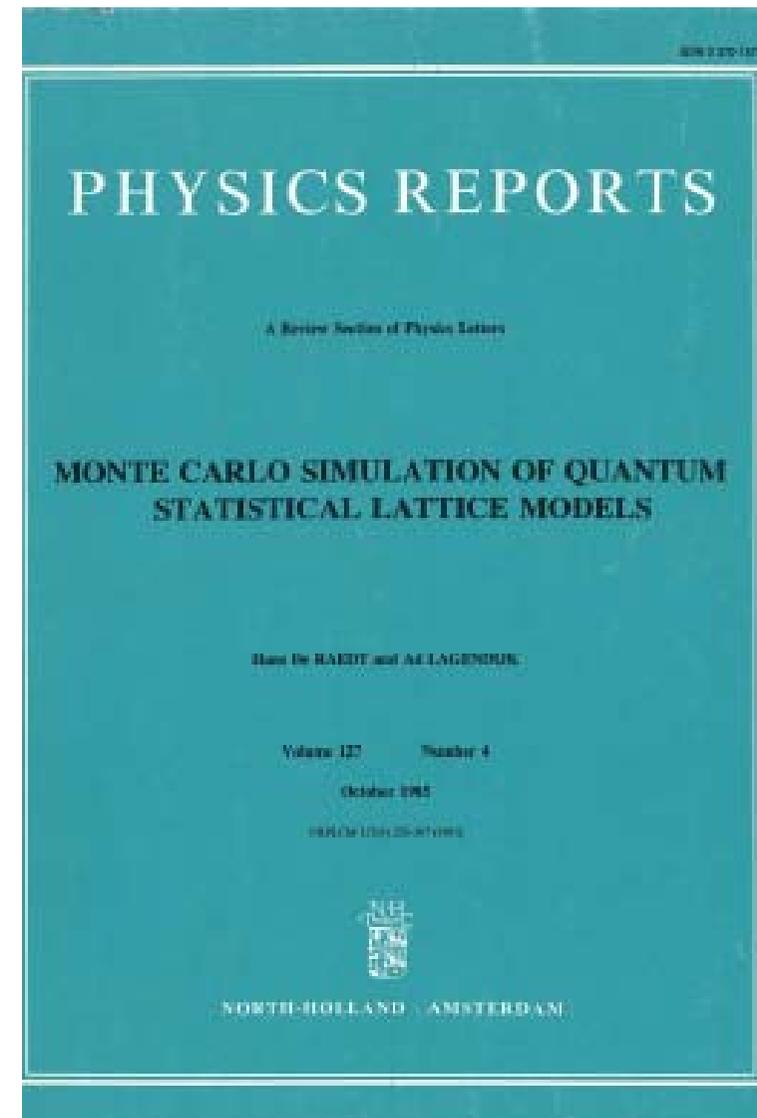
# *My experience*

I am an experimental physicist  
(chemist by training)  
with a lot of (30%) theory papers



# *My experience with large data sets*

**I have extensively worked on massive computer simulations (parallel algorithms)**



# *Hard-core programming*

I develop programs  
(from number crunching to GUI) using  
Fortran, C, C++, perl, php,  
Mathematica, Maple



I regularly give advice to my  
students that speed up their  
programs orders of magnitude



*Am I old-fashioned and disqualified?*





AD LAGENDIJK

# SURVIVAL GUIDE FOR SCIENTISTS

Writing - Presentation - Email

AMSTERDAM UNIVERSITY PRESS

# *My experience: physics*

- Why am I invited?
- Is physics not just one of the many science disciplines?
- No, it is the most successful empirical science ever.  
Explanation and predictive power

OUT OF THE WAY, YOU SWINE!  
A PHYSICIST IS COMING!



# *Are physicists smarter?*

**No,**

**but their discipline has  
been further developed**



# *Physicists have been there*

Physicists have seen it al:

- international collaborations
- massive calculations  
(QCD, molecular dynamics)
- massive datasets  
(Navier-Stokes and high-energy physics)
- global data exchange



# *eScience is hype*



**Much of eScience is hype**  
**It attracts the wrong people:**  
**wrong scientists**  
**operators**  
**managers**

# *Your multiple choice*

**1. Is there no science to eScience?**

**or**

**2. Is there no science to eScience yet?**

**or**

**3. May there be some future to eScience,  
if we can keep out hype-evangelists?**

# *Previous revolutions ...*

- ✓ introduction
- previous revolutions
- common misunderstandings
- present web 2.0 developments
- useful steps towards eScience

# *Is there anything new?*

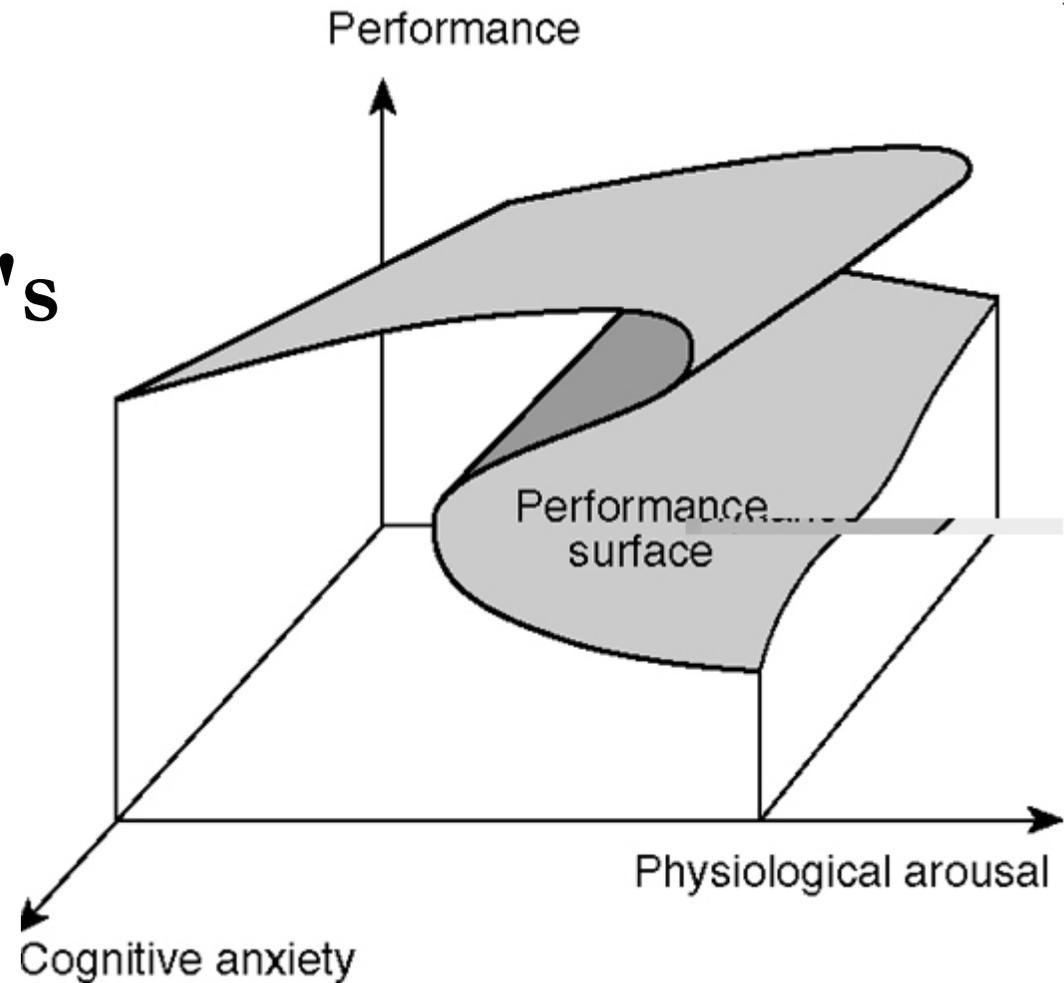
I will show you some wrecks on the  
wrecking yard of promised revolutions  
or new paradigms



# *French wreck*

## Catastrophy theory

René Thom in the 1960's



# *Chaos theory*

**Joseph Ford (1983):**

**The third revolution after  
quantum mechanics  
and  
relativity theory**



# *You cannot trust evangelists*

**In 1990 a Danish scientist gave a talk in our institute telling us we would soon need to "extend human rights to computers"**



# *New York wreck*

In the beginning of the 1980's I heard many talks about the New York

**Ultimate Computer**

that would solve all problems

# *eWreck: digital repositories*

**Any university, any publisher, any librarian has started ten years ago their own digital repositories:**

**non-refereed, uncorrected, incompatible file formats (even in pdf) collection of trash (with a few exceptions)**



# *Eternal promise: data mining*

**Hype in business schools 10 years ago**



# *Elsevier's wreck*

**Busy more than 20 years distilling  
semantic information out of scientific  
articles (they are SGML pioneers)**



**ELSEVIER**

# *Common misunderstandings ...*

- ✓ introduction
- ✓ previous revolutions
- **common misunderstandings**
- present web 2.0 developments
- useful steps towards eScience

# *eScience claims*

## **Missions:**

- for communication and dissemination
- for analysis and data mining
- for data visualization and exploration

## **Missing:**

- scientific literature mining
- structuring of new scientific literature

# *Popular misunderstandings (1)*

Progress of present-day science would be (dramatically) accelerated if scientists would collaborate more



Politicians, CEO's, science policy makers, and public at large:



They are *\*not\** active scientists.



# *Power of natural science*

Science is very tough

If you prove your boss wrong, your boss is wrong and you will become his boss

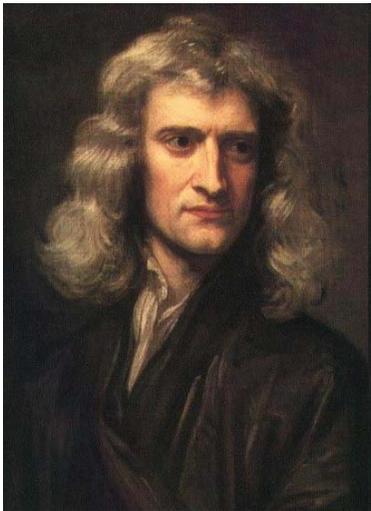
If you work in a bank and you prove your boss to be wrong, you are fired



# *History of science*

Scientists are in it for

- curiosity how Nature works
- recognition, ego, winning, fame



# *Losers*

Science is strictly hierarchical

Many scientists - **especially those in the lower echelons** - want to bypass this hierarchy by

1. introducing other criteria (like applications)
2. introducing new organizational structures
3. forcing new fields

# *Politics*

Too large collaborations lead to politics

IPCC



# *Popular misunderstandings (2)*

Progress of present-day science would be (dramatically) accelerated if scientists would have many more data available (higher resolution, larger span, ... )

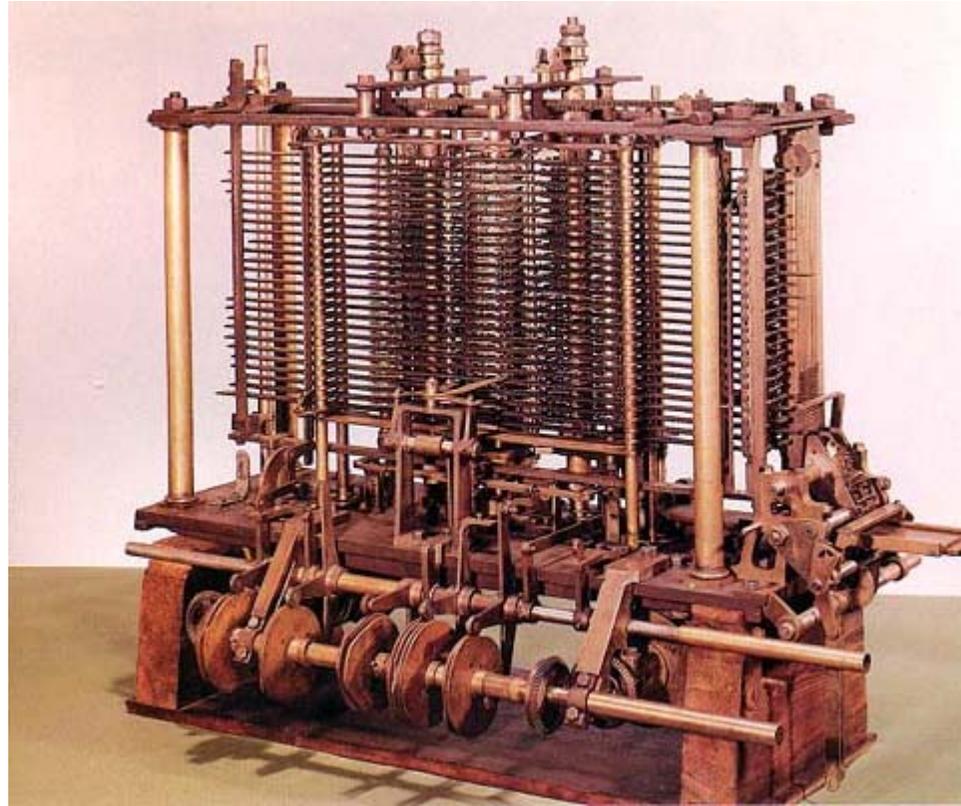


All interesting problems in science are in NP complete.

Their solution can be written in polynomial time, but the calculation of the solution not.

Computational power does not solve anything

# *Turing machine*



# *Popular understanding (3)*

**Progress of science would be increased if the speed of communication would be increased:**

**RSS feeds as an example**



**Authors fight for years to get their paper in a prestigious, rapidly-publishing, letter journal. And all the time they keep its content secret.**

# *Brute force solutions*

**Many of the eScience approaches are brute-force solutions**

**Brute-force solutions have never been successful in science**



# *Present web 2.0 developments ...*

- ✓ introduction
- ✓ previous revolutions
- ✓ common misunderstandings
- present web 2.0 developments
- useful steps towards eScience

# *eScience claims*

## **Missions:**

- **for analysis and data mining**
- **for data visualization and exploration**
- **for communication and dissemination:  
blogs, wiki's, social sites, forums**

# *Science and movies*

**Science is not about making movies  
3D plots hardly ever give additional  
insight**



# *Blogging as science*

- 74 % of science blogging deals with popularizing science
- 24% of science blogging concerns the daily lives of people that happen to be scientists

Good for science, but if any contribution to science at all, only to "the context of discovery" and never to "the context of justification"

Scientific papers have a dull, prescribed structure for many reasons.

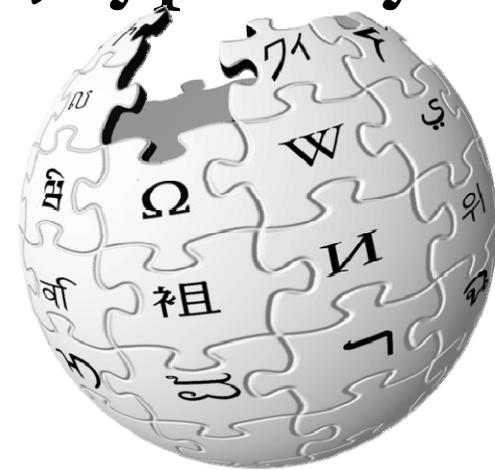
Blogging contributes nothing.

# *Wiki's for science dissemination*

Wiki's are interesting for sharing useful scientific knowledge that needs *\*not\** to be structured as a scientific paper:

manuals for equipment, calibration, typically what earlier went in lab journals

Terrible for scientific papers



**WIKIPEDIA**  
*The Free Encyclopedia*

# *Social sites and fora*

## *The Chronicle of Higher Education*

I am the only contributor, out of thousands,  
using his own name

*"drbeeper"*

*"immigrant"*

*"thisisme"*

*"onion"*





*"On the Internet nobody knows you are a dog!"*

# *Typical "computer" needs*

sharing data (**facilities present, but extremely slow**)

plotting software for plotting data

statistical software for analysing data

presentation software for presenting data

text formatter for non-articles (like grant proposal)

text formatter for articles (with version control)

# *In collaboration?*

- ✓ sharing data (**facilities present, but extremely slow**)
- ✗ plotting software for plotting data
- ✗ statistical software for analyzing data
- ✗ presentation software for presenting data
- ✓ text formatter for non-articles (**like grant proposal**)
- ✓ text formatter for articles (**with version control**)

# *Text formatters with sharing*

## Text formatting non-science articles:

- ✗** MS-Word      no-sharing
- ✗** OpenOffice etc.      no-sharing
- ✓** Google Docs

## Text formatting scientific articles:

- ✗** Wiki's and Google Docs very slow and clumsy
- ?** Where is the Open SVN server?

# *Useful steps towards eScience ...*

- ✓ introduction
- ✓ previous revolutions
- ✓ common misunderstandings
- ✓ present web 2.0 developments
- useful steps towards eScience

# *Advice*

*To: Microsoft, Google, Yahoo, Thomson, ...*

You have the power to make *\*the\** social site that is too say the

**LinkedIn or Facebook for scientists**

Forget the eScience for the time being. That will come later, and automatically

# *Step 1*

**Build the number 1 social site for scientists:**

- 1. Trust (not simple openID authentication)  
probably with PKI infrastructure**
- 2. Partition immediately in standard disciplines  
and get on "board" hot-shot scientists.  
Not eScience evangelists**

# *Step 2*

**Introduce high-quality, userfriendly, GUI for scientists to put their complete profile there:**

- **cv**
- **complete publication list (so many mistakes)**
- **expertise along standard discipline classification scheme**

**A lot of work, but it is essential**

# *Interface to science and scientists*

Communicate with scientists using  
standard internet protocols:

ftp (sftp, ftp over tls)

http (https, ssl)

smtp/pop3/imap (preferably over ssl)

Cross-platform compatibility and no  
browser monopoly



Never force on scientists, or even advertise, your  
proprietary server solutions (e.g. *MS Sharepoint* servers).

# *Privacy policy*

**With immediately a crystal clear  
privacy policy (that is too say whatever you  
do: it is always "opt-in" and never "opt-out")**

**Do not make the Facebook mistakes**

~~**Google**~~

# *Actors*

1. Individual incentives (end up in a mess)
2. Learned societies
3. Commercial parties

I would prefer an association  
between 2) and 3)

# *Company interest is commercial*

Much of the interest of companies in eScience (be it *Microsoft* or *Google*) is out of commercial reasons. That is not bad, but we scientists should be careful.



# *Learned societies*

## **Advantage:**

**prestige**

**persuasive power**

**open solutions**

## **Disadvantage:**

**lack of expertise**

**lack of (financial) resources**

# *Commercial parties*

**Advantage:**

**expertise**

**resources**

**Disadvantage:**

**lack of prestige and distrust**

**no persuasive power**

**proprietary solutions**



# *MPG joining the hype?*

**Buy many computers?**

**Buy many computer servers?**

**Force collaborations?**

**Redirect research?**

**Hire new directors?**

**Start an eScience department?**

**Contracts with major software companies?**



**Build a system of internet trust**

**European Research Council referees  
have to send in a copy of their  
passport every six months or so.  
A good trust system would  
make this redundant**



# *MPG Word*

Facilitate technically,  
collaborative  
writing of papers



# *MPG LinkedIn*

**Build the MPG social site. You have the power to force or to persuade any MPG employee to participate**



"I didn't know the 'old boys' network had a website."

# *Future*

eScience has a future

but it could also end up in a mess.

The computer world has already  
too much of that

# *Road to eScience*

**Cut the crap**

**Start with small steps**



# *The End*

- ✓ **introduction**
- ✓ **previous revolutions**
- ✓ **common misunderstandings**
- ✓ **present web 2.0 developments**
- ✓ **useful steps towards eScience**

# *Consultant?*

I have given a number of companies running social sites (like *Google, Nature, Science, SciLink*) a lot of unsolicited advice.

They have never listened to me

# *Authoring papers*

## **MS-Word**

**lost case, bugs, slow, legacy-visual basic  
bad history, html is a mess**

# *Challenge IT*

Use Web 2.0 and the Web as a Platform

Simple protocols supported by industry

Blogs, Wikis, RSS feeds, Tagging, Mash-ups ...

Computer Science community and the IT industry to deliver powerful

tools and technologies to support Data-Intensive research

Interoperability and open standards

Collaborative and multidisciplinary

Parallelism and Multicore

Client + Cloud: Software + Services