A BOINC-based platform for Scientific Computing

L.A. Fernández

A. Gordillo-Guerrero

V. Martín-Mayor

J.J. Ruiz-Lorenzo

Overview

- **Introduction**
  - Volunteer computing.
  - Previous experiences.
- **The IBERCIVIS Project**
  - Presentation.
  - How to use it?
- **Our experience as scientists**
  - Our scientific problem.
  - Pros and cons of the infrastructure.
- **Conclusions**

Overview

- **Introduction**
  - Volunteer computing.
  - Previous experiences.
- **The IBERCIVIS Project**
  - Presentation.
  - How to use it?
- **Our experience as scientists**
  - Our scientific problem.
  - Pros and cons of the infrastructure.
- **Conclusions**

Motivation

Are you <u>always</u> using 100% of your CPU power?

Motivation

Are you <u>always</u> using 100% of your CPU power?

Would not it be nice to use idle times of computers all around the world for Scientific Research?

Motivation

Are you <u>always</u> using 100% of your CPU power?

Would not it be nice to use idle times of computers all around the world for Scientific Research?

Volunteer Computing

Introduction

- **What is volunteer computing?**

  "Arrangement in which people (volunteers) provide computing resources to projects."

Introduction

- **What is volunteer computing?**

  "Arrangement in which people (volunteers) provide computing resources to projects."

  - Volunteers are anonymous.
  - Middleware: BOINC-based client. Screensaver.
  - Projects are:
    - Typically academic (university-based).
    - Trustful for the volunteer.

Introduction

- **Why does volunteer computing worth?**
  - Huge potential (>1 billion of PCs in the world).
  - Can provide computing power to project with limited funding but large public appeal.
  - Environmentally very positive:
    - Reducing equipment costs.
    - Optimizing consumed energy.
  - Promotes public interest in Science.

Introduction

- **Previous Experiences of Volunteer Computing**
  - SETI@home, MilkiWay@home
  - Docking@home, folding@home
  - Malariacontrol.net
  - Climateprediction.net, Quake-Catcher network
  - PrimeGrid
  - QMC@home, Einstein@home
  - ......

Overview

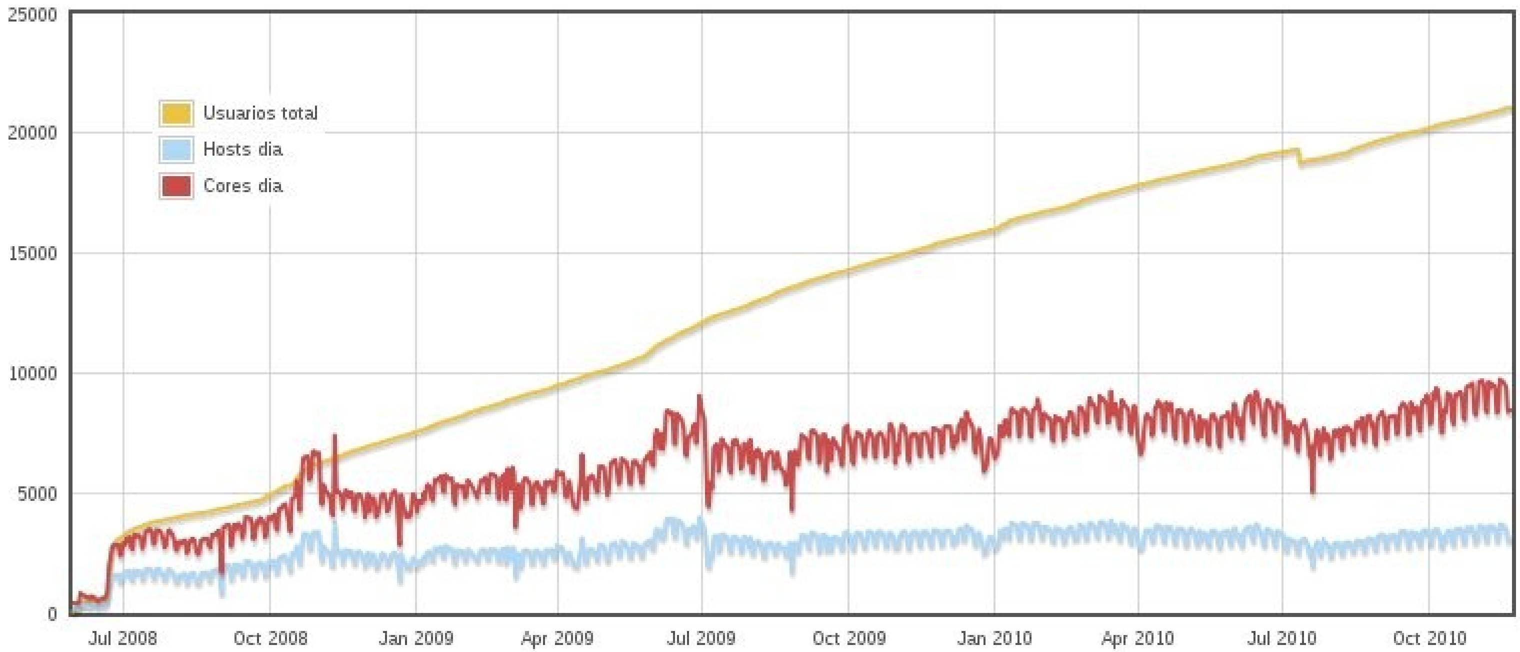
- **Introduction**
  - Volunteer computing.
  - Previous experiences.
- **The IBERCIVIS Project**
  - Presentation.
  - How to use it?
- **Our experience as scientists**
  - Our scientific problem.
  - Pros and cons of the infrastructure.
- **Conclusions**

The IBERCIVIS project

- **Main features**
  - Durable platform supported by public institutions.
  - Professional and stable development team.
  - Multidisciplinary projects:
    - Nuclear Fusion
    - Protein Docking
    - Phase transitions
    - Adsorption
    - Neuro-degenerative diseases
    - Ionizing radiation trajectories.
    - ....
  - Continuously opened to new projects.

The IBERCIVIS project

- **History:**
  - Launched in June 2008.
  - Around 20,000 registered volunteers.
  - More than 2000 nodes are working every day.



The IBERCIVIS project

- **How to compute in Ibercivis?**
  - 1st Step: Scientific proposal.
    - Desired characteristics:
      - – Millions of independent jobs to do.
      - – No need of communication between jobs.
      - – Free software (with no licenses).
      - – Less than 500MB of RAM use.
      - – Divisible jobs in parts of 30-60 minutes.
      - – Small I/O for each job (around 1MB).

The IBERCIVIS project

- **How to compute in Ibercivis?**
  - 2nd Step: Porting
    - Interaction with the development team.
    - Exhaustive evaluation of the converted code.

The IBERCIVIS project

- **How to compute in Ibercivis?**
  - 2nd Step: Porting
    - Interaction with the development team.
    - Exhaustive evaluation of the converted code.
  - 3rd Step: Production
    - Results.
    - Interaction with the communication team.
    - Possible optimization of the code.

Overview

- **Introduction**
  - Volunteer computing.
  - Previous experiences.
- **The IBERCIVIS Project**
  - Presentation.
  - How to use it?
- **Our experience as scientists**
  - Our scientific problem.
  - Pros and cons of the infrastructure.
- **Conclusions**

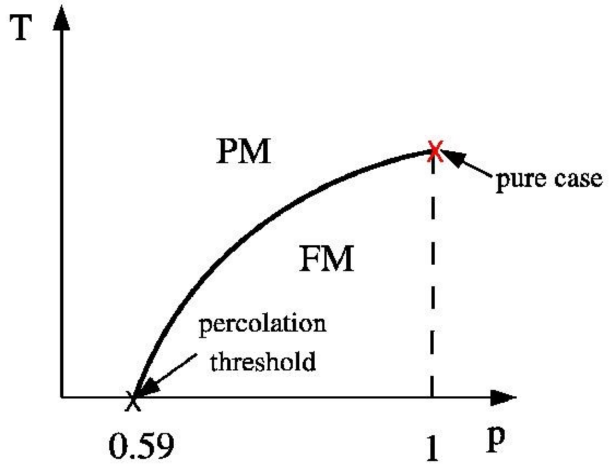
Our scientific problem

First Order Phase Transitions in Disordered Systems

- In 2D the problem is solved:

  Aizenman-Wehr theorem:

  Dilution ⇨ Second Order PT



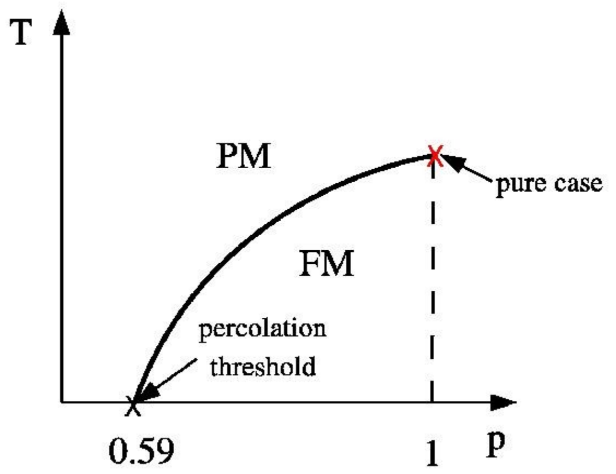
Our scientific problem

First Order Phase Transitions in Disordered Systems

- In 2D the problem is solved:

  Aizenman-Wehr theorem:

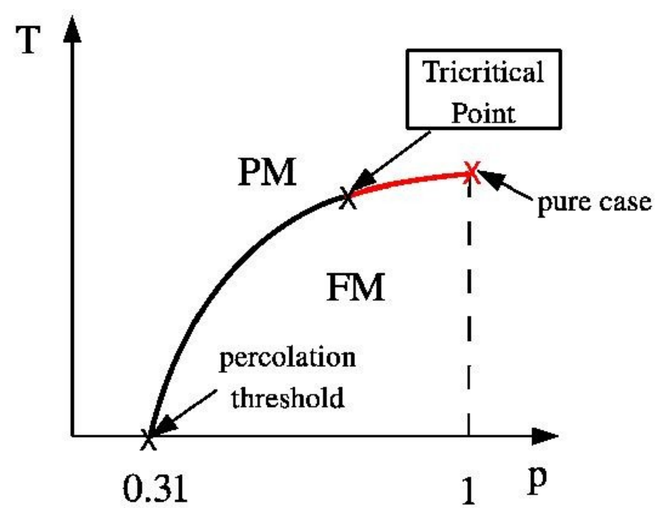
  Dilution ⇒ Second Order PT



- In 3D the behavior is different:

  Cardy-Jacobsen mapping:

  Tricritical point belonging to the RFIM Universality Class



Our scientific problem

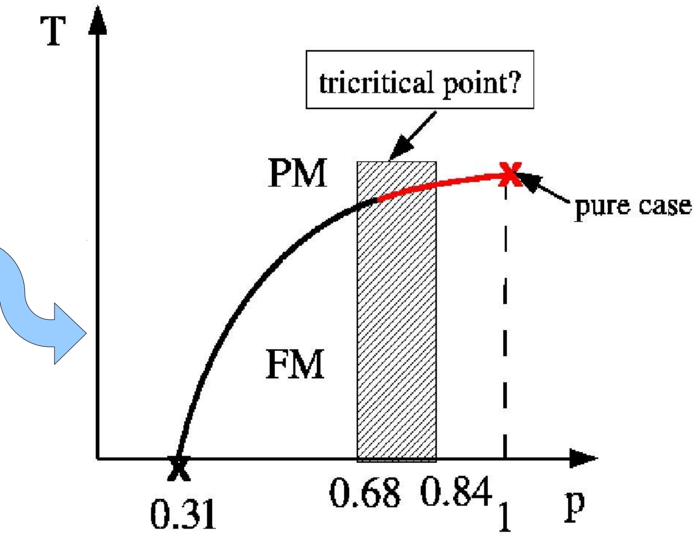
- **Previous results in 3D:**
  - Site-diluted Potts model Q=3

    *PRB **61**, 3215 (2000).*
  - Bond-diluted Potts model Q=4

    *W. Janke et al. Nuclear Physics B **719**, 275 (2005).*
  - Site-diluted Potts model Q=4

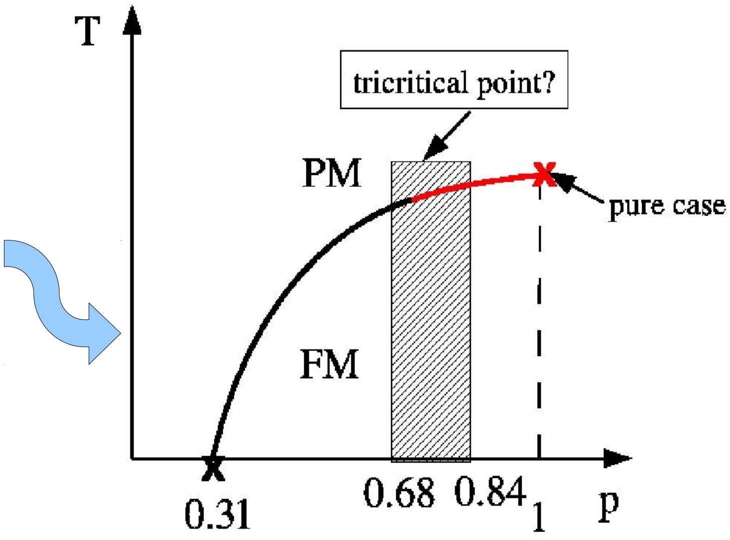
    *PRL **100,** 057201 (2008).*
  - Bond-diluted Potts model for large Q

    *F. Igloi et al. PRE **73**, 026126 (2006).*



Our scientific problem

- **Previous results in 3D:**
  - Site-diluted Potts model Q=3
    *PRB **61**, 3215 (2000).*
  - Bond-diluted Potts model Q=4
    *W. Janke et al. Nuclear Physics B **719**, 275 (2005).*
  - Site-diluted Potts model Q=4
    *PRL **100,** 057201 (2008).*
  - Bond-diluted Potts model for large Q
    *F. Igloi et al. PRE **73**, 026126 (2006).*



- **Our present approach:**
  - The model: Site-diluted Q=8 Potts model
    *We must find a **strong** first order phase transition.*
  - The method:
    *Extended Microcanonical Approach: PRL **98**, 137207 (2007).*

Our simulation code (advantages)

- C without exotic libraries: nice portability.
- No strong RAM requirements (max. 40 MB) nor disk storage for each configuration (max. 4 MB).
- Embarrassingly parallelizable problem:
  - System sizes
  - Dilutions
  - Disorder realizations (samples)
  - Energies

Our simulation code (disadvantages)

- Long simulation time
  - Grows exponentially with the system size
    - Around one week for a $L=64^3$ system
  - A continuity system is indispensable.

Our simulation code (disadvantages)

- Long simulation time
  - Grows exponentially with the system size
    - – Around one week for a $L=64^3$ system
  - A continuity system is indispensable.

- Huge amount of files to allocate
  - Example:
    - – L=64, one dilution, 500 samples, 20 continuity steps, 36 energies → *400,000 files*
  - Just one missed file is a corrupted sample.

Our experience as scientist

- **Advantages of the infrastructure**

Our experience as scientist

- **Advantages of the infrastructure**
  - Around $3x10^6$ hours of CPU time during approximately one year.
  - Easy adaptation of our original code.
  - Failure rate reduced to less than 0.5%.
  - Economic cost basically zero.
  - Opportunity to share our scientific problem with general public (web, blogs, radio, etc.).

Our experience as scientist

- **Disadvantages of the infrastructure**

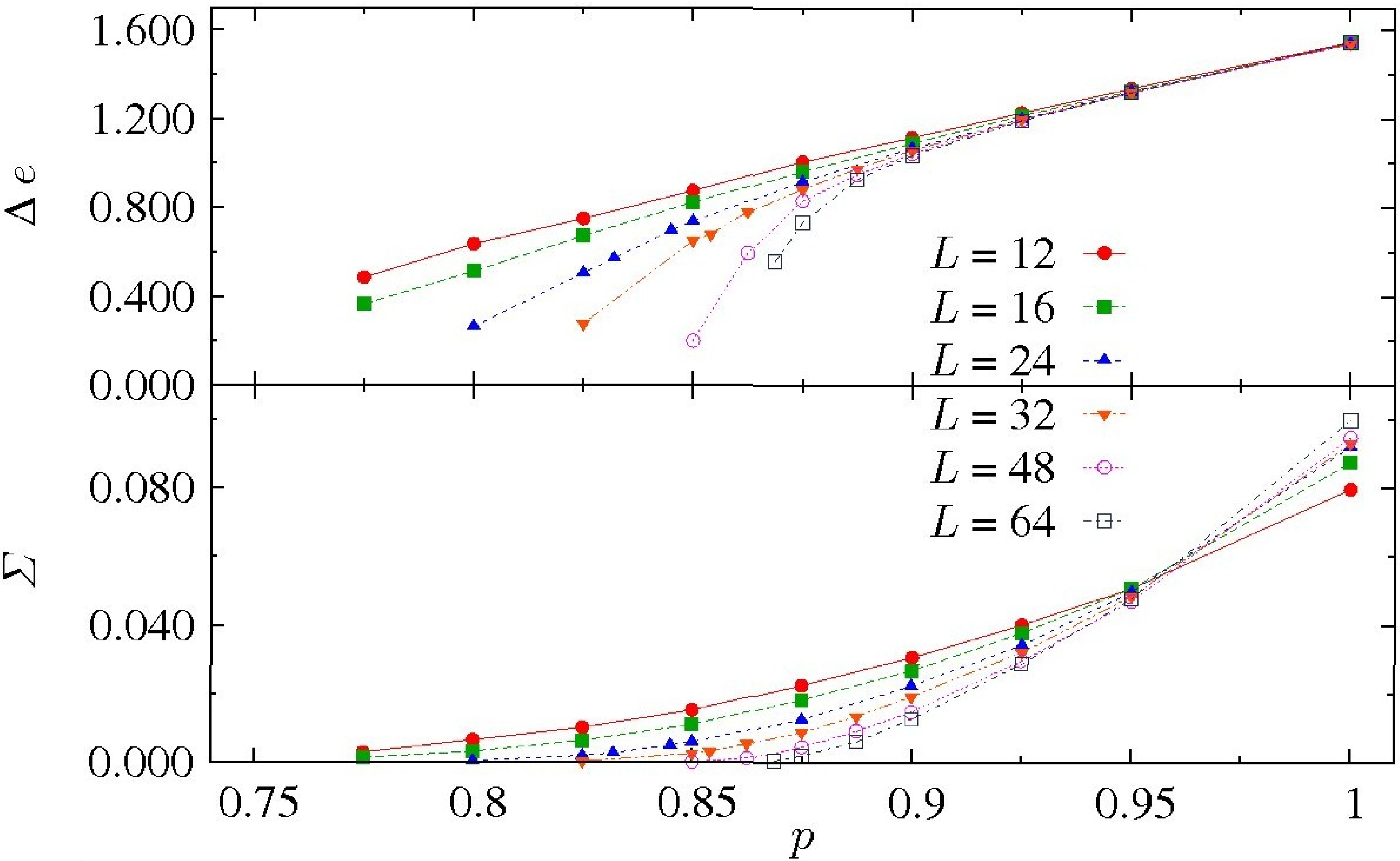
Our experience as scientist

- **Disadvantages of the infrastructure**
  - In principle less reliable than a non-distributed platform.
  - Crucial to detect corrupted samples.
  - Analysis times increase due to the continuity system.
  - Other infrastructures are more suitable for urgent simulations.

Preliminary Scientific Results

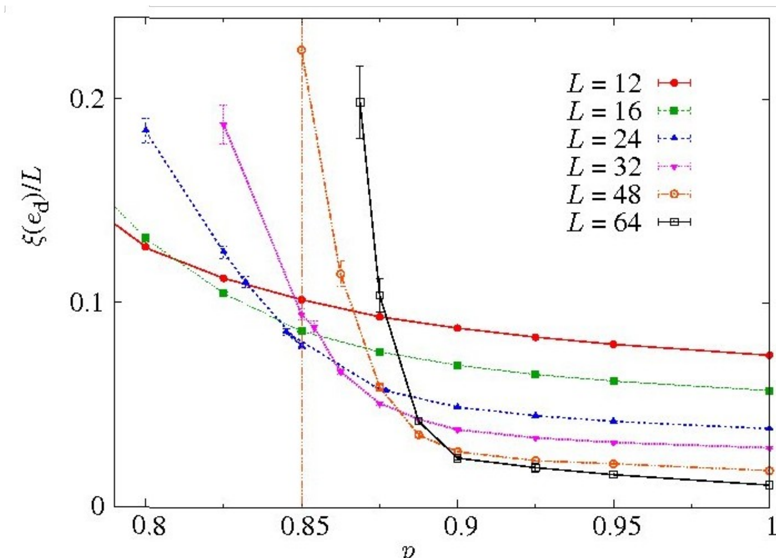
Preliminary Scientific Results

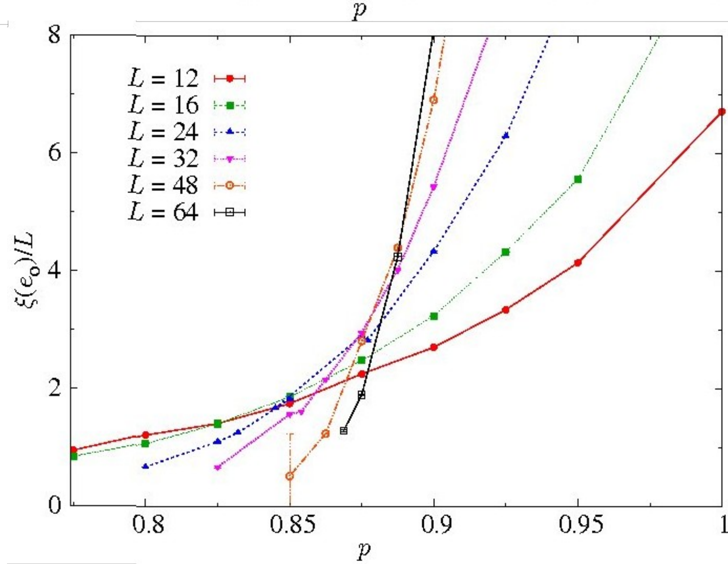
- **Latent Heat and Surface Tension**

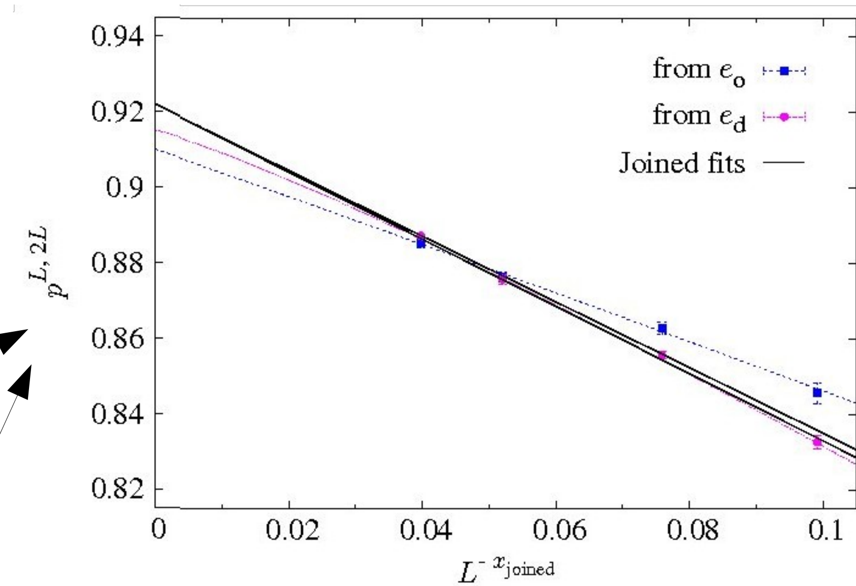


Preliminary Scientific Results

- **Tricritical Point location**







$$p_c = 0.922(1)$$

$$\frac{\chi^2}{\text{d.o.f}} = \frac{1.50}{2}$$

Overview

- **Introduction**
  - Volunteer computing.
  - Previous experiences.
- **The IBERCIVIS Project**
  - Presentation.
  - How to use it?
- **Our experience as scientists**
  - Our scientific problem.
  - Pros and cons of the infrastructure.
- **Conclusions**

Conclusions

- IBERCIVIS is a fine infrastructure to obtain a large amount of computation time.
- Easy access and porting of the code.
- Good development team.
- Really cheap (basically free).
- A new computing paradigm requires practice.
- Slower than non-distributed structures.

Thank you for your attention

www.ibercivis.net