Curso de Doctorado
Universidad del Centro de la Provincia de Buenos Aires, Argentina, 24 al 27 de Abril, 2006

Evolutionary Computation and Machine Learning for the Optimisation and Design of Physical, Chemical and Biological Systems

Dr. Natalio Krasnogor

Automated Scheduling, Optimisation & Planning Research Group
School of Computer Science and Information Technology

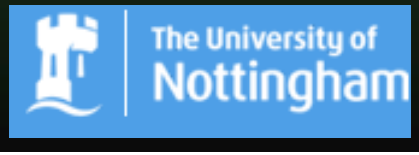
Centre for Integrative Systems Biology
School of Biology

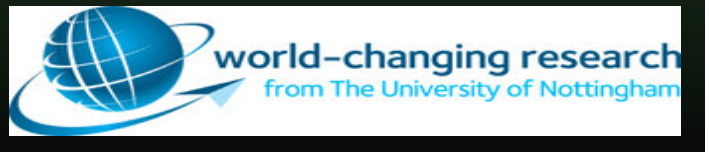
Centre for Healthcare Associated Infections
Institute of Infection, Immunity & Inflammation

University of Nottingham

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk

Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril





Content

- Motivation & Overview of Problems, Methodologies and Computational challenges
- Introduction to Evolutionary Computing and Machine Learning

5 hours

- Bioinformatics
- Systems Biology

7 hours

- Chemoinformatics and Computational Physics
- Back to computing….

4 hours

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Motivation

Major advances in the rational de

This has happened before in other research and industrial disciplines,e.g:

- VLSI design
- Space antennae design
- Transport Network design/optimisation
- Personnel Rostering
- Scheduling and timetabling

designs/optimisations.

That is, complex systems are plagued with NP-Hardness, non-approximability, uncertainty, etc results

tems by modern AI

Yet, they are routinely solved by sophisticated optimisation and design techniques, like Evolutionary algorithms, etc

We anticipate that as the numbe r of research challenges and applications in these domains (and their complexity) increase we will need to rely even more on automated design and optimisation based on sophisticated AI & machine learning

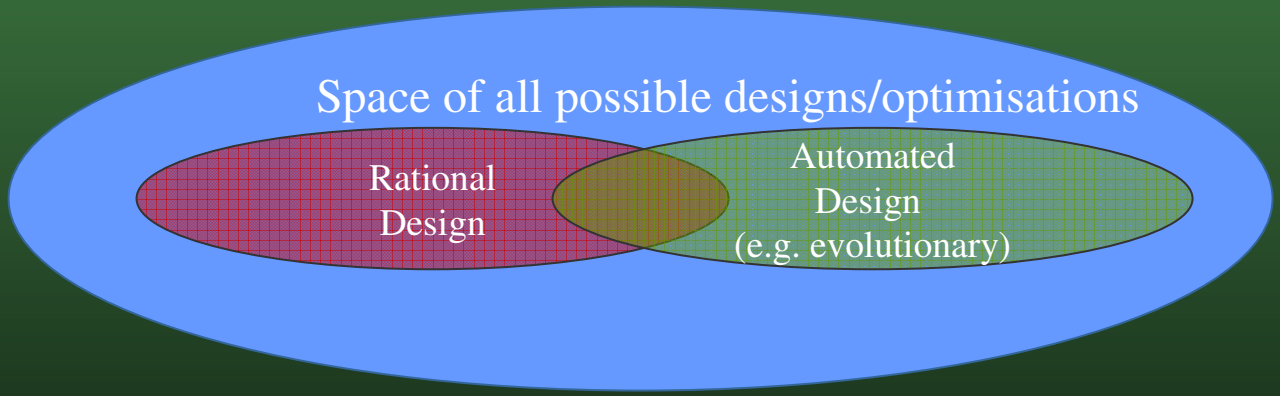
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Automated Design/Optimisation is not only good because it can solve larger problems but also because this approach gives access to different regions of the space of possible designs (examples of this abound in the literature)



www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

The Research Challenge

- For the Engineer, Chemist, Physicist, Biologist, etc :
  - To come up with a relevant (MODEL) SYSTEM *M**
- For the Computer Scientist:
  - To develop adequate sophisticated algorithms -beyond exhaustive search- to automatically design or optimise existing designs on *M** *regardless of computationally (worst-case) unfavourable results of exact algorithms.*

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

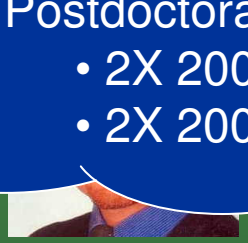
world-changing research from The University of Nottingham

My PhD students and Postdocs Team:

Cuatro Nuevas Vacantes de Postdoctorando:
- 2X 2008
- 2X 2009

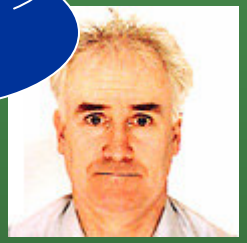
Lin Li

Jaume Bacardit



Danny Barthel

Itziar Fr[illegible]



Mike Stout

Peter Siepman

Pawel Walera

German Terrazas

Scott Littlewood

Evolutionary Computation & Machine Learning
N[illegible]omputation, Complex systems
Bioinformatics
Physics
Chemistry

~ £14M Funded by EPSRC/BBSRC/EU/DTA

-Adam Sweetman (with P. Moriarty @ Physics)

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Protein Structure Problems

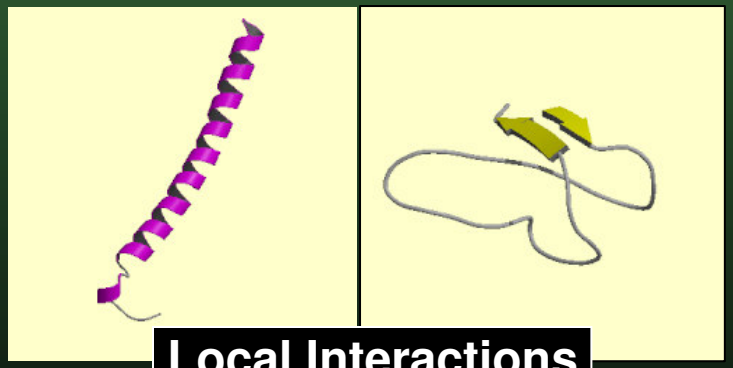
Primary Structure = Sequence

MKYNNHDKIRDFIIIEAYMFRFKKKVKPEVDMTIKEF
ILLTYLFHQQENTLPFKKIVSDLCYKQSDLVQHIKVL
VKHSYISKVRSKIDERNTYISISEEQREKIAERVTLFD
QIIKQFNLADQSESQMIPKDSKEFLNLMMYTMYFK
NIIKKHLTLSFVEFTILAIITSQNKNIVLLKDLIETIHHK
YPQTVRALNNLKKQGYLIKERSTEDERKILIHMDDA
QQDHAEQLLAQVNQLLADKDHLHLVFE

Quaternary or Native Structure

Secondary Structure

Tertiary



Local Interactions

Global Interactions

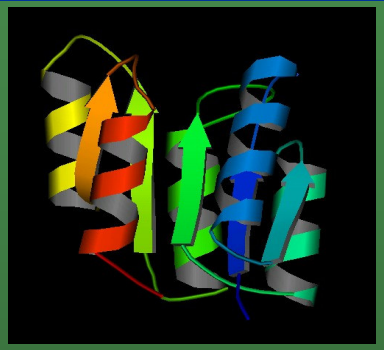
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

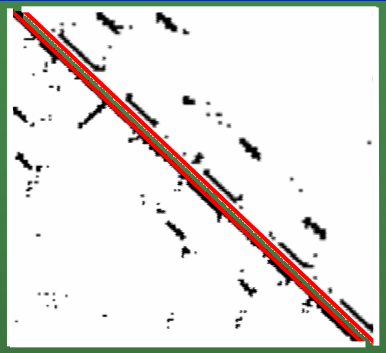
The University of Nottingham

world-changing research from The University of Nottingham

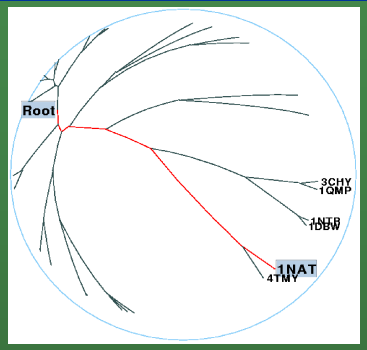
Similarity Comparison of Proteins



One protein structure



Pairwise comparison of proteins



Comparison of multiple proteins

- In the native state atoms that are far away in the chain come close to each other and form *contacts*.

- These can be represented in a two-dimensional *contact map* (middle)

- CMs are used to compare pairs of proteins according to their USM *similarity*.

- Taking a set of proteins, a *similarity matrix* is computed and used to cluster proteins accordingly to their similarity (right).

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Protein Structure Feature Prediction using Learning Classifier Systems

- PSP can be divided in several sub problems:
  - Secondary structure
  - Coordination number prediction
  - Solvent accessibility
  - Disulfide bonding prediction
  - etc

- The coordination number of a protein is a simplified profile of a proteins 3D structure

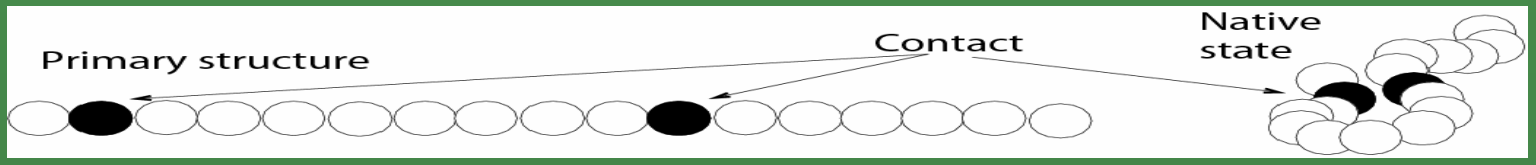
- CN indicates, for each residue in the protein, the number of other residues that are closer than a certain threshold to it

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

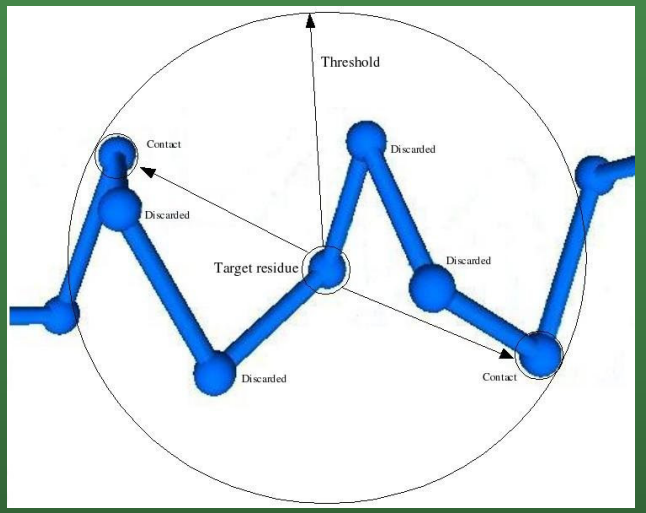
asap automated scheduling optimisation & planning research

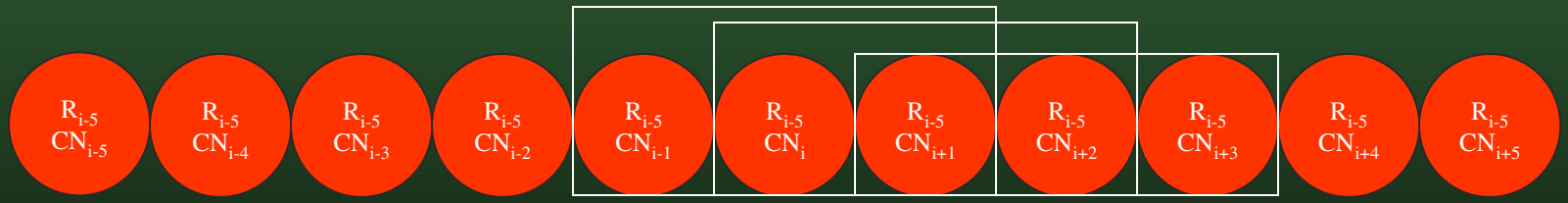
The University of Nottingham

world-changing research from The University of Nottingham



- Given the AA sequence of a protein chain we would like to predict the coordination number of each residue in the chain
- We have to transform the data into a regular structure so that it can be processed by standard machine learning techniques





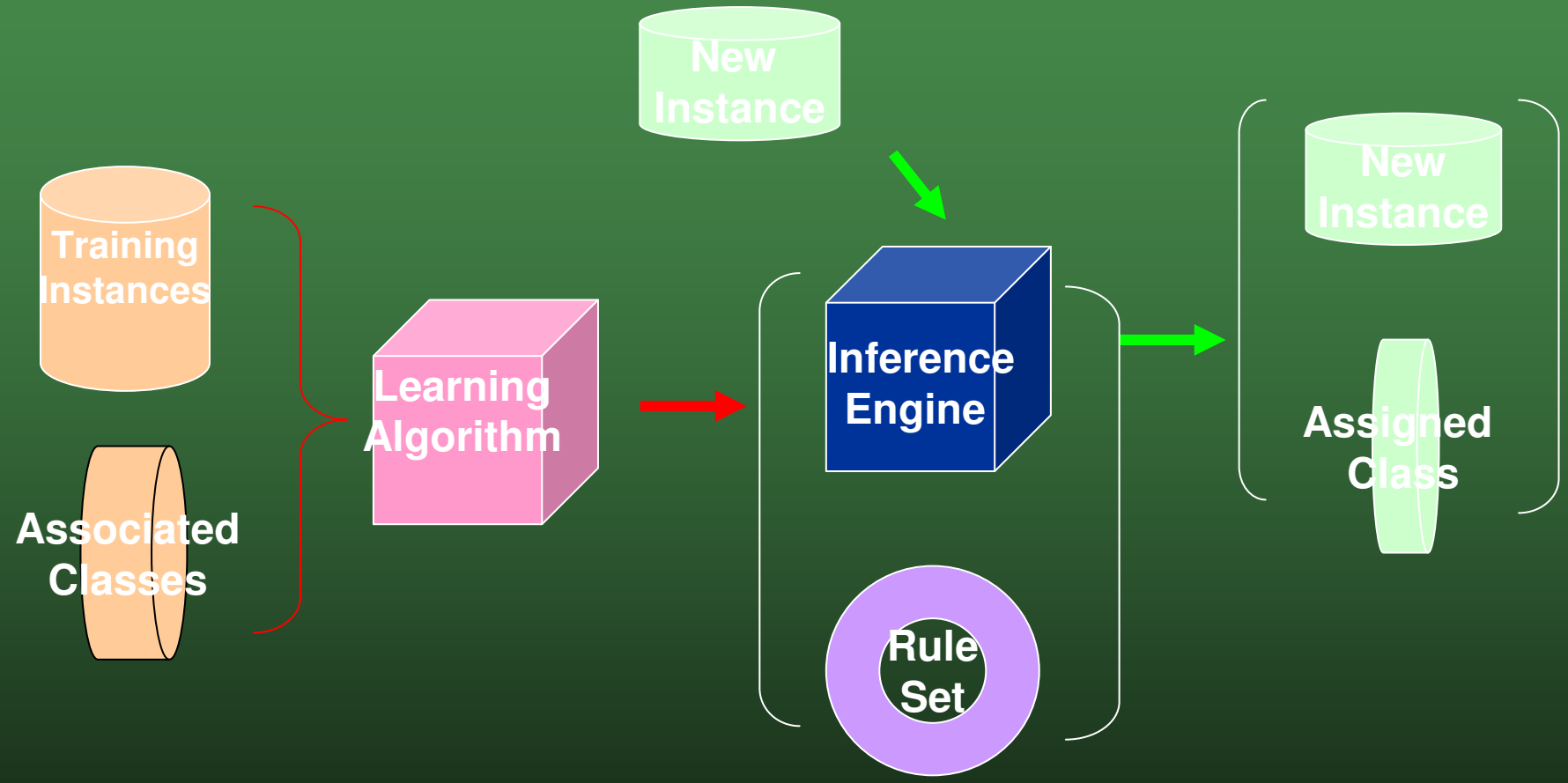
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Mechanics of Classification



www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

- We are developing cutting-edge Learning Classifier Systems (LCS) as the learning paradigm for these problems
- LCS are a very smart integration of evolutionary computation (robustness), reinforcement learning (quick convergence) and MDL (generalization)
- We also benchmark against other machine learning techniques e.g. Bayesian learning, decision trees, SVM, etc

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk

Protein Structure Resources Integration and Mining

- We are building an integrated database containing (under a relational model):
  - structural
  - physicochemical
  - functional
  - biological
  - evolutionary
  - as well as genetic information of protein data.

•Data is derived mainly from SCOP, PDB and DSSP databases and other web services out there.

•Data is extracted trough a variety of scripts that need to parse, compute, filter, etc gigabytes of data at a time

- Currently PDB and DSSP have about 34626 proteins. The above information requires monthly re-computation & updating and several tens of GB to run and store hence I/O is crucial here

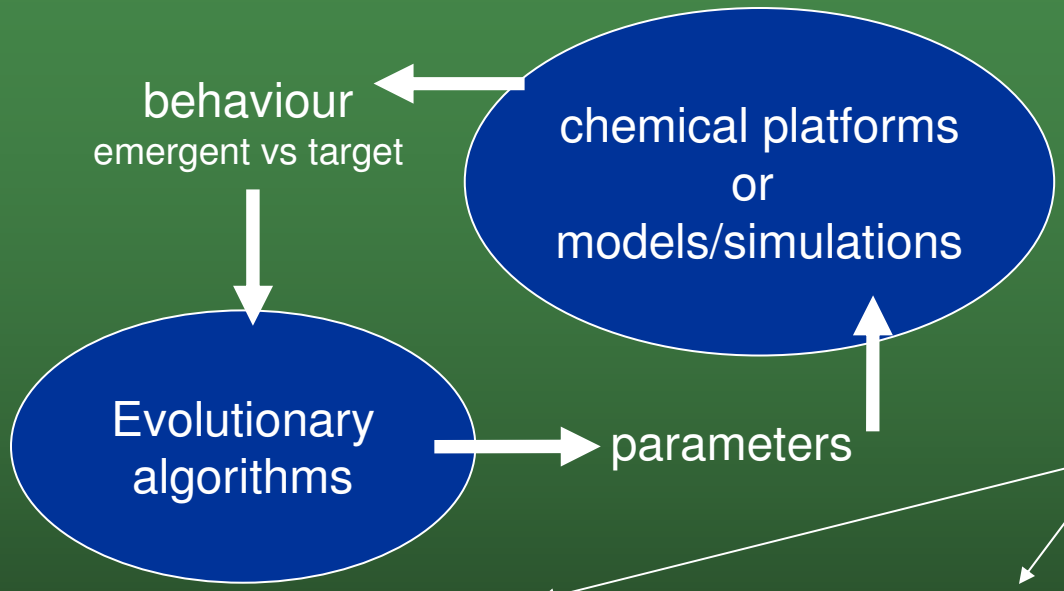
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

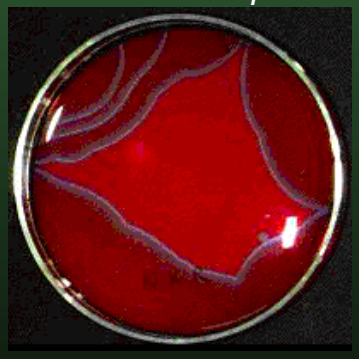
Complex Physico-Chemical Systems Design



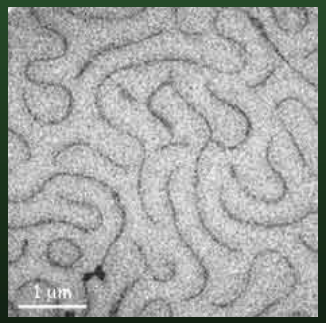
working with CHELLnet
http://www.chellnet.org
&
working with Prof. P. Moriarty's group

Our software evolves a set of parameters such that the Physico-chemical complex system produces a specified target behaviour.

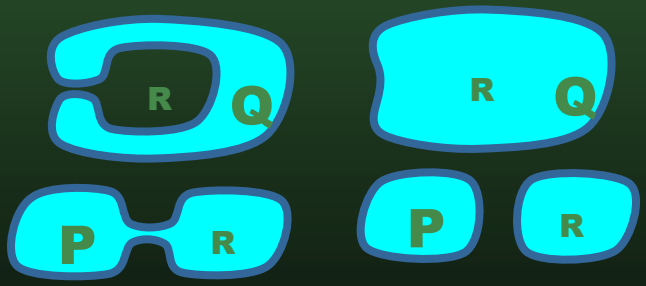
Patterns & *computation* in the BZ reactions



Nano-particle self-organisation



Vesicles/miscelles formation



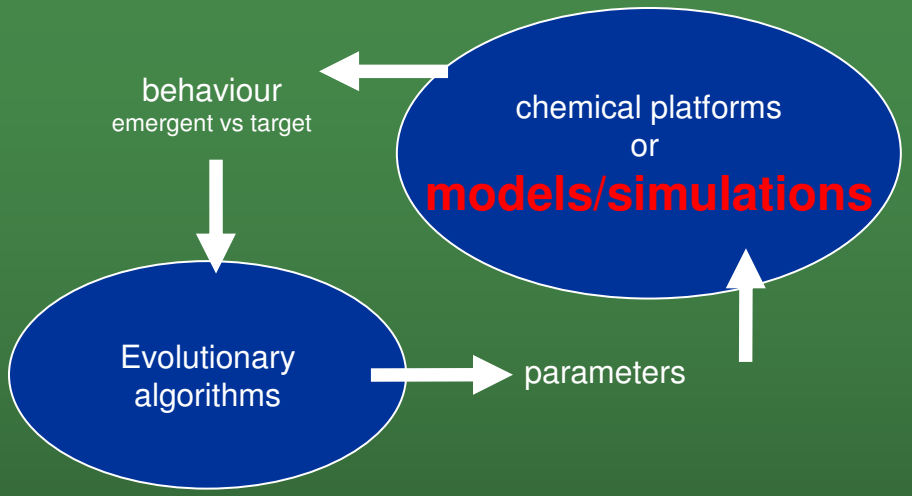
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Current Cluster use



- We use *simulations* to model the physico-chemical systems.
- These simulations can take a long time.
- Even a small population of 10 candidate solutions may take days to evaluate.

Solution?

- Parallelise the evaluation of a population – evaluate each solution on a separate processor/node.
- Produce multiple runs with different random seeds
- Subdivide parameter search space

Note that even once we tie the GA to the platform we will still require horse power to evaluate heavy objective functions

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Automated Software Self-Assembly Programming Paradigm (ASAP^2)

- Software self-assembly takes a set of human-made software components and integrate these components to yield a desired architecture to satisfy a given goal.
- New automatic paradigm for automatic program discovery. Aims: investigate, and analyze the behaviour of software self-assembly.
  - What and how software self-assembly can be affected by various factors.
  - How software self-assembly differs from other automatic programming approaches such as genetic programming.

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Current work

- Software must be embedded into a simulated physical world. We define the rules of the world as to make ASAP efficient & effective.
- Kinetic theory on perfect gas is used as a metaphor, i.e.. the embedding:

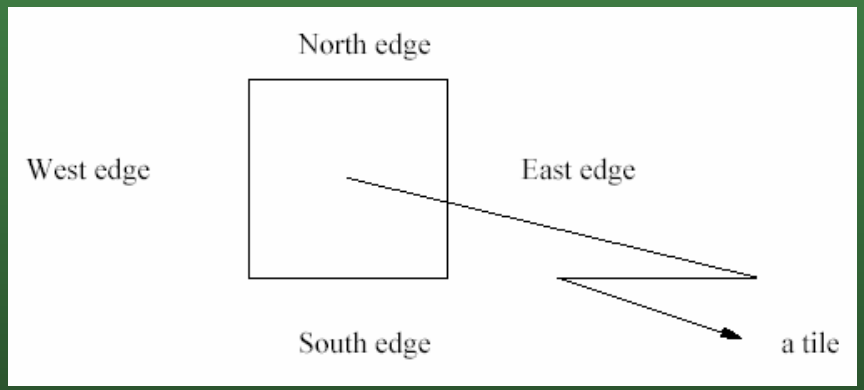
$$PV = nRT$$

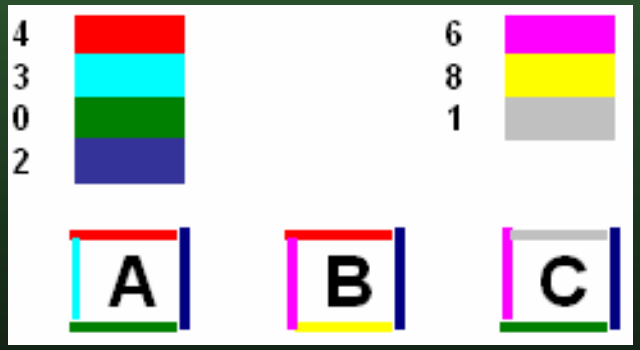
- Three sorting algorithms are used as initial software components repository.
- Components are put in the virtual world (V,T,n) and let to interact.
- We measure the diversity ($D\varepsilon$), Time to equilibrium ($T\varepsilon$) against three free environment parameters ($V \in [400, 500, 600, 700]$, $T \in [0.25, \ldots, 4.0]$ with an increment in value of 0.25, $n \in [1,2,3,4,8,16,24,32]$).
- Using components from the three different software repositories, we use cluster to run the experiments in (V, T, n) in distributedly
- We aim at further parallelizing each individual run.

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

Automated Software Self-Assembly Programming Paradigm (ASAP^2)

Programming Wang Tiles Self-Assembly





- Given a target shape (e.g. 10x10 tile square)
- Given a Wang Tile "world" model
- Goal: To evolve sets of tiles that self-assemble in the target shape under the dynamics of the model

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk

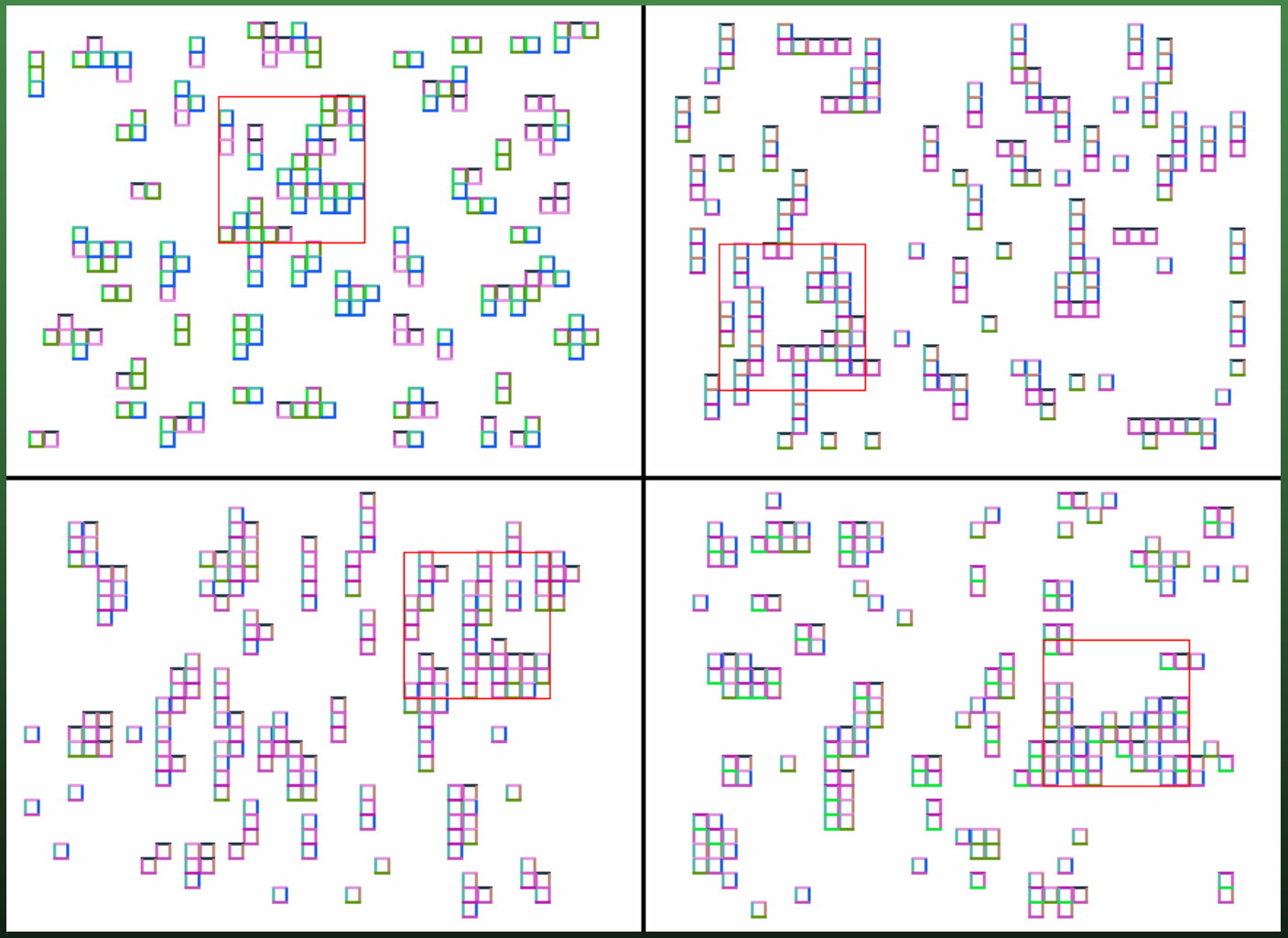
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Generations 0, 15, 30, 45 ...



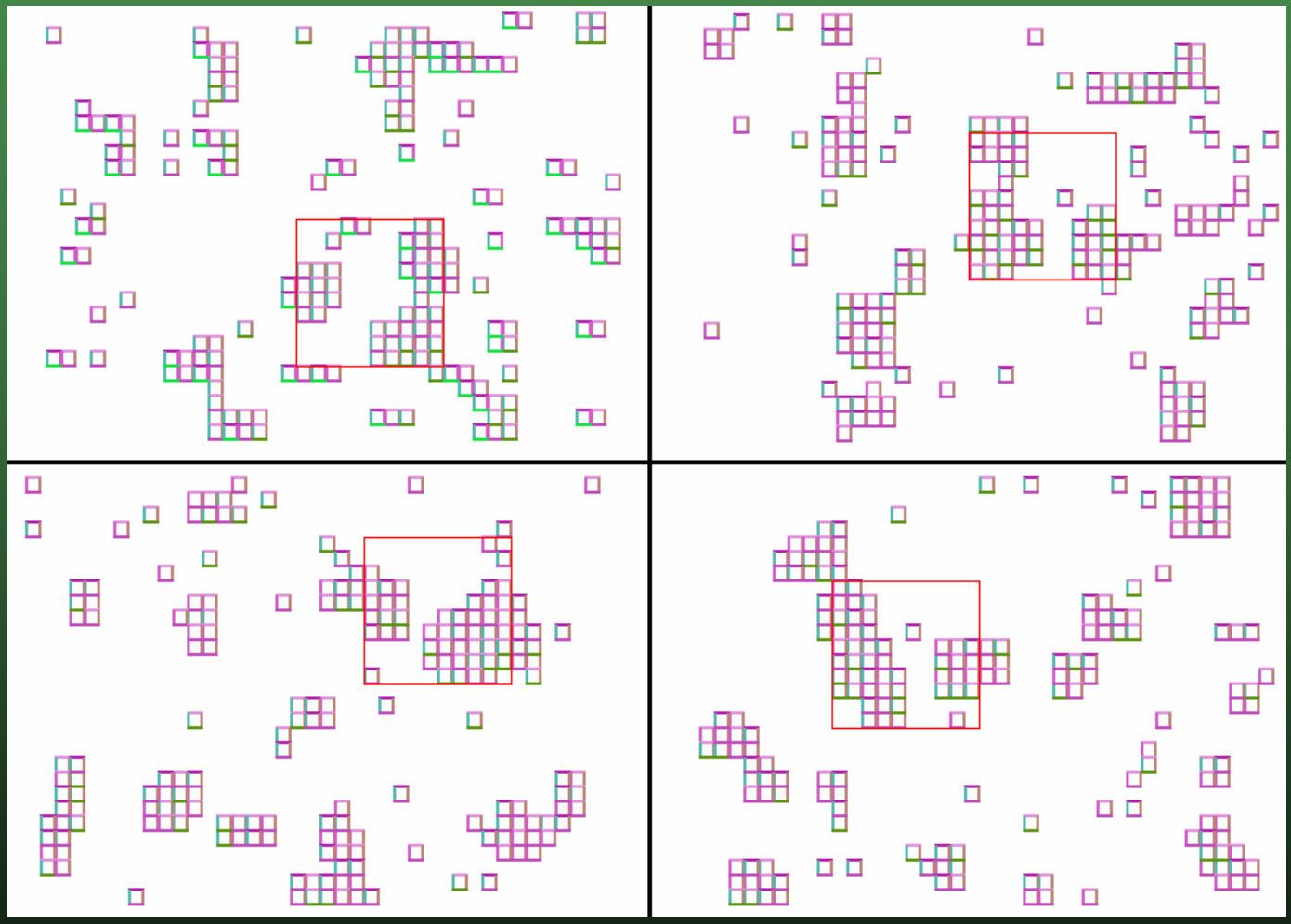
www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Generations ... 60, 75, 90, 99



www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Conclusions

- The best science today is interdisciplinary
- Computer science is at the heart of it all!
- Essentially all our work heavily depends on sophisticated algorithms and state of the art harware, e.g. cluster computers, STM/ATM microscopes, etc
- Today we can do science that was unimaginable just a few years ago
- Different types of jobs:
  - Purely distributed jobs (√)
  - Purely parallel jobs, e.g. vesicle formation (√)
  - Low CPU, Heavy I/O load (√)
  - Mixture of distributed & parallel (X)
  - Mixture of distributed & parallel & Heavy I/O (X)

All of these require state-of-the-art AI, ML, EA

√ already running
X to run within 6 months time

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Introduction to EC & ML

- Evolutionary Computing
- Evolutionary Machine Learning
  - Introduction to Learning Classifier Systems (Part 1)
  - Advanced LCS (Part 2)

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

Bioinformatics

- Conceptos Elementales de Biologia
  - Pelicula
  - Presentacion
- Introduction to Protein Structure Prediction
  - Simple Models and Algorithms
- Real-world Protein Structure Prediction Problems
  - Predicting Folding/Non Folding
  - Contact Number Prediction:
    - Part 1
    - Part 2
  - Post Synaptic Activity
- Protein Structure Comparison
  - Models, Measures, Metrics and Methods and the Procksi Server

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Systems Biology

- An (Unorthodox) Introduction
- Modelling Quorum Sensing
  - The biology of QS in P.*aeruginosa*
  - Mathematical Modelling
  - A Natural Computation Mechanism derived from QS
  - Towards a formal language based model

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Chemoinformatics and Computational Physics

- Evolutionary Design of Physicochemical Systems
  - Evolutionary Design of Complex Systems
  - Evolutionary Design for Surface nanoscience

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham

Back To Computing…

- The Automated Self-Assembly Programming Paradigm (ASAP²)
  - ASAP for programs dynamic synthesis
  - ASAP for Wang Tile's programming

www.cs.nott.ac.uk/~nxk
Natalio.Krasnogor@Nottingham.ac.uk
Curso de Doctorado, Universidad de La Laguna, 17 & 18 de Abril

asap automated scheduling optimisation & planning research

The University of Nottingham

world-changing research from The University of Nottingham