From High-Throughput Experimentation (HTE) to High Output Experimentation

Laurent Allan, BAUMES
Instituto de Tecnologia Quimica (UPV-CSIC)

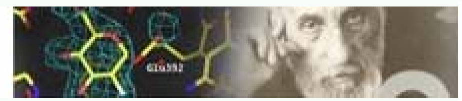
Royal Society of Chemistry Symposium 2009 “Catalysts for Change”
17-18 June 2009, Barcelona, Spain

Instituto de Tecnologia Quimica

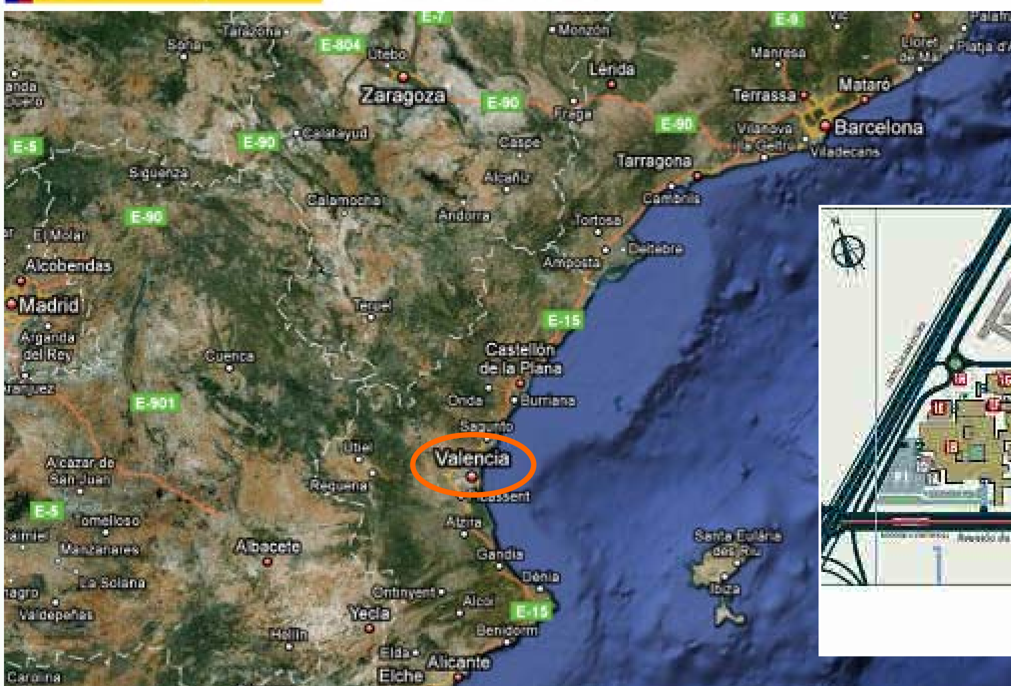
- Joint research centre
- CSIC **(Spanish Council for Scientific Research)** and UPV **(Polytechnic Univ. Of Valencia)**
- Director: Prof. Avelino Corma
- Budget 2008: 4 millions Euros

GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA E INNOVACIÓN

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS



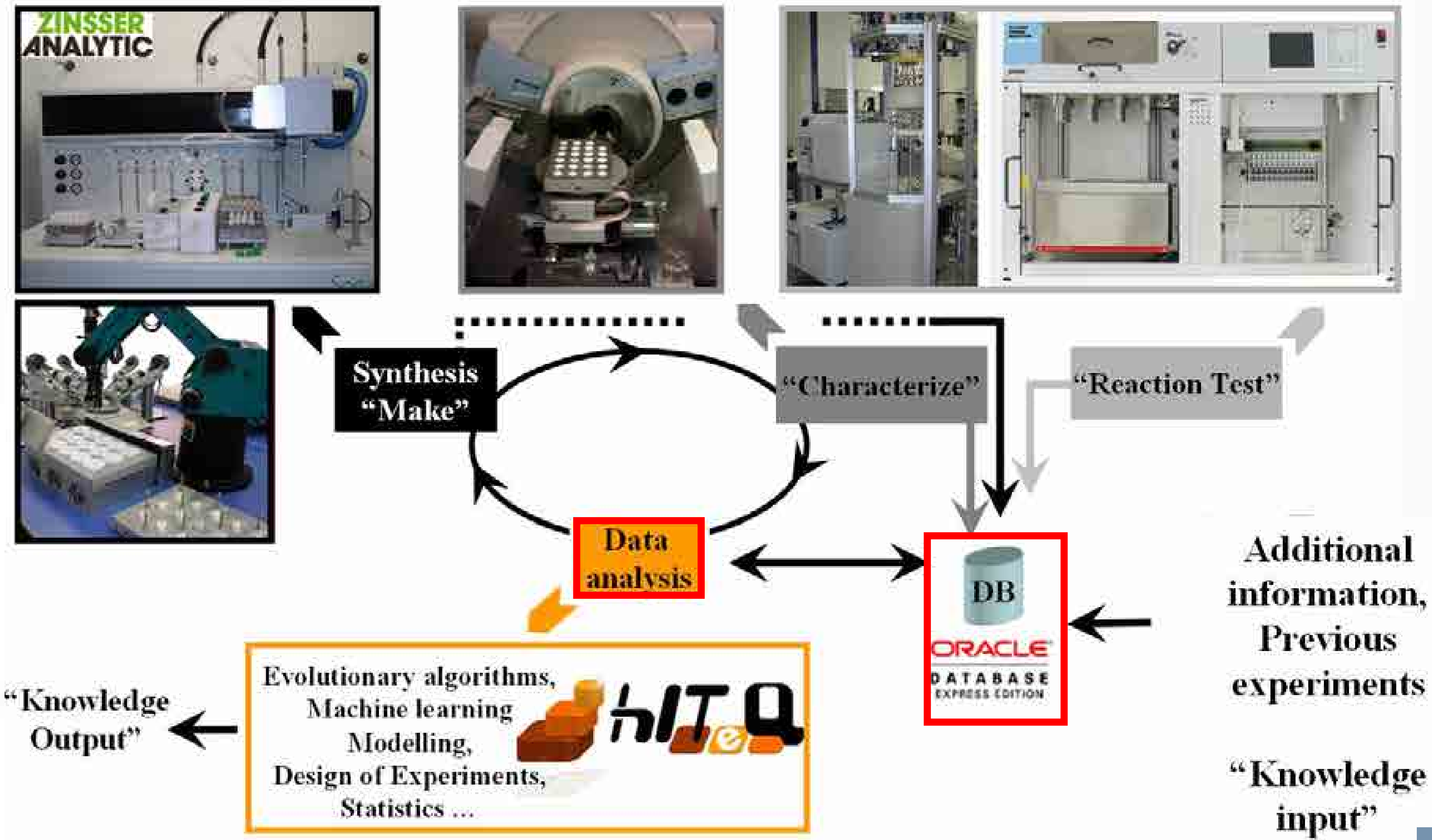
UNIVERSIDAD POLITECNICA DE VALENCIA



Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

HTE for Crystalline Materials

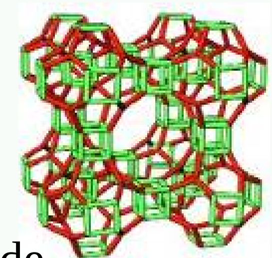
High-throughput technology



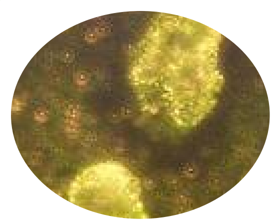
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Database (e-notebook) for materials

- A Database for various materials (solid catalysts, zeolites, MOFs, ceramics…)

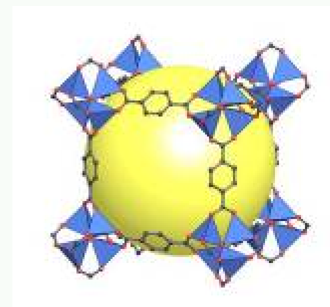


- A simple but effective two-side concept (synthesis/characterization)

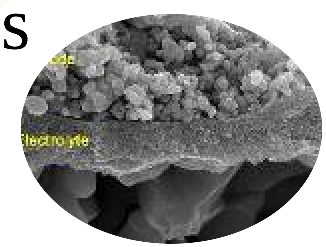


- Task flow chart (e.g. e-notebook) for synthesis
- Integration for new robots and treatment platform
- User-supervised knowledge integration (meta-data…)

- Zeolites
- MOFs



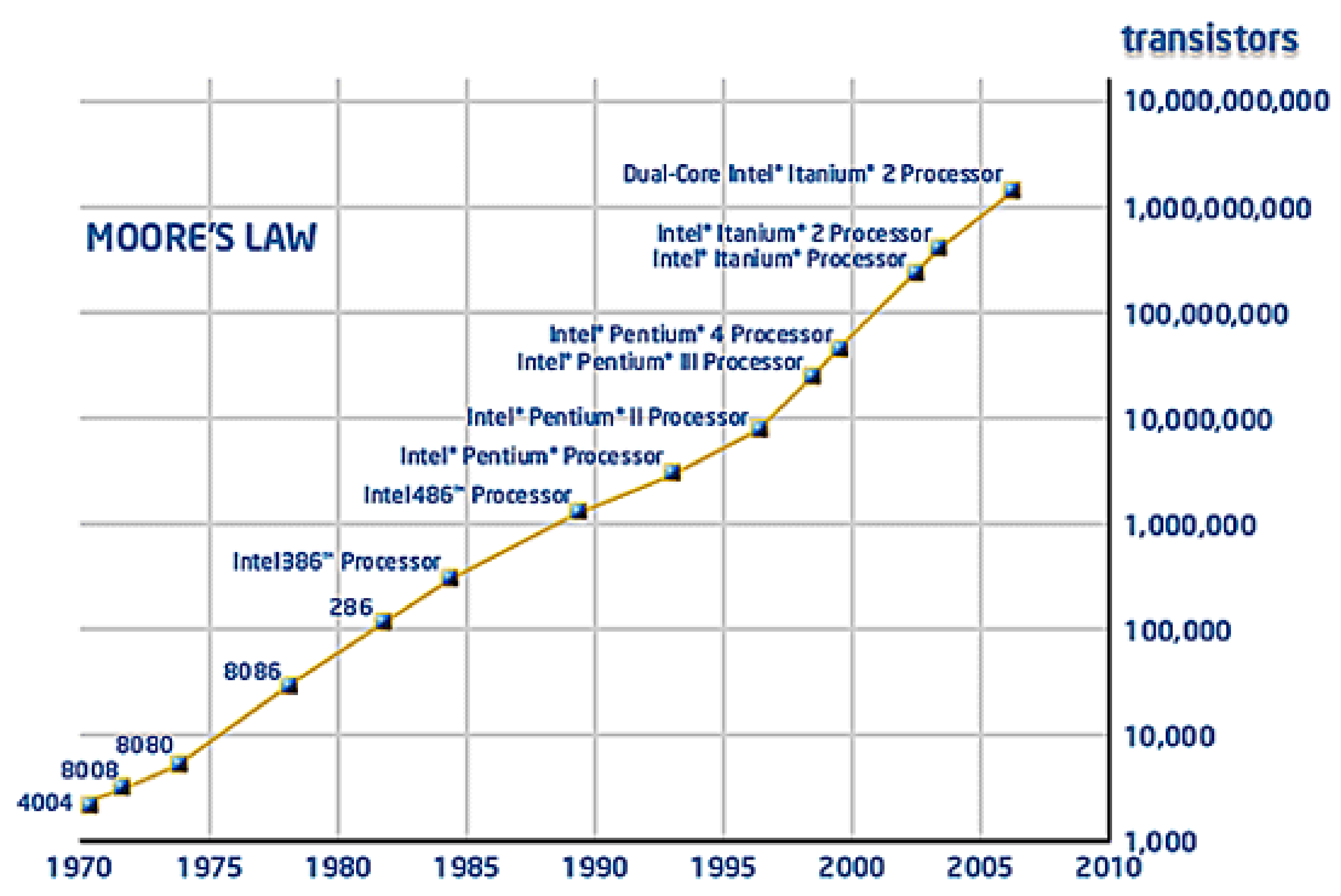
- Solid catalysts
- Membranes (ceramics)



Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Moore's law

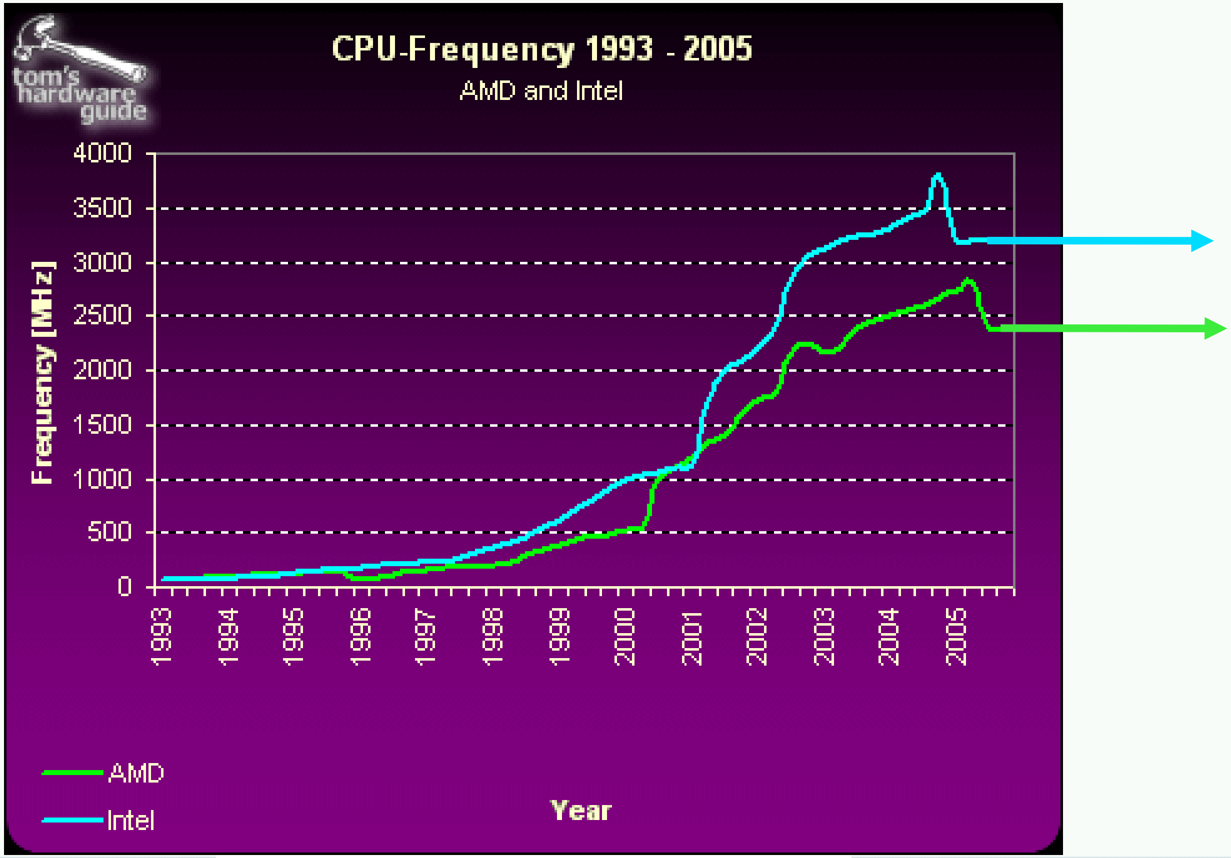
Moore's law applies since the 1970 's, and according to a 2005 interview, he predicts that « we have another 10 to 20 years before we reach a fundamental limit. »



CSIC UNIVERSIDAD POLITECNICA DE VALENCIA

Moore's law and CPU clock speed

- Unfortunately, Moore's law (doubling of the number of transistors on integrated circuits every two years) does not apply to CPU frequency ...



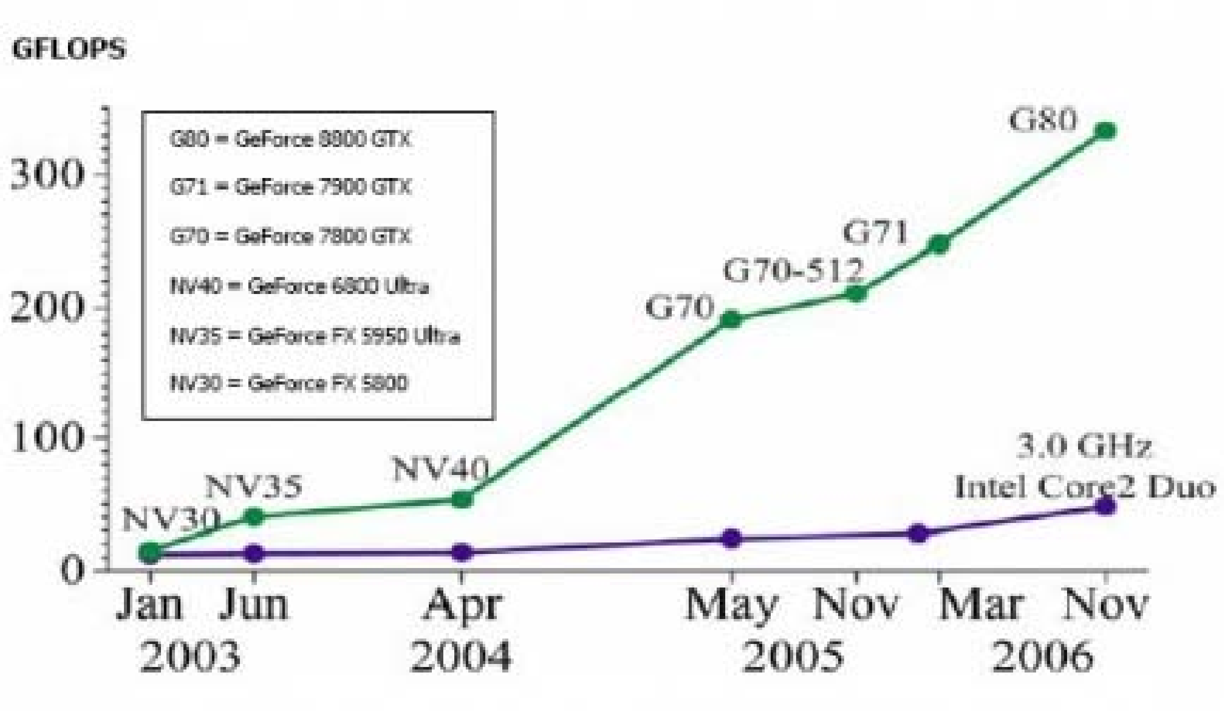
Current limit: heat dissipation

- The latest AMD Phenom II X4 940 BE will run at 6Ghz, but in a bath of liquid helium at -232°C...

NVIDIA GPGPU cards

Due to the simplified architecture (multi-core with several stream processors per core all executing exactly the same instruction at the same time), computing power of GPGPU has increased much faster than CPU computing power.

In Nov 2006, the first NVIDIA G80-based full GPGPU card (8800GTS) featured 8 cores containing16 processors each (=128 processors) for a top speed of 320 GFLOPS, to be compared with 10GFLops for the fastest 2006 Intel.



GPGPU based EAs

- $500 GTX295 card (480 processors) is supposed to run at 1.8 TeraFlops ! (250x speedup /CPU).

- It is possible to insert up to 4 such cards into a single PC !!!
- 1000x speedup over a standard PC !

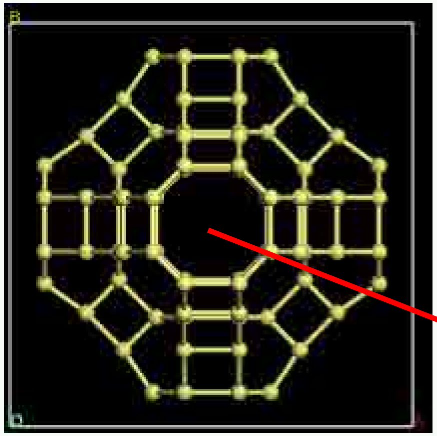
  - *1 day* on a $2000 (+host PC) configuration is equivalent to...
    2 years and 9 months on a conventional PC !

- Opens new perspectives to evolutionary algorithms, allowing to explore much wider search spaces than with current computers.

- GPGPUs directly benefit from Moore's law, meaning that their power will still increase in the future (pushed by millions of dollars of the gaming industry).

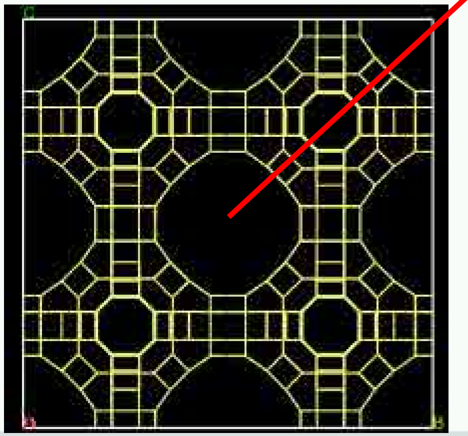
Hypothetical zeolite structures

Unit Cell

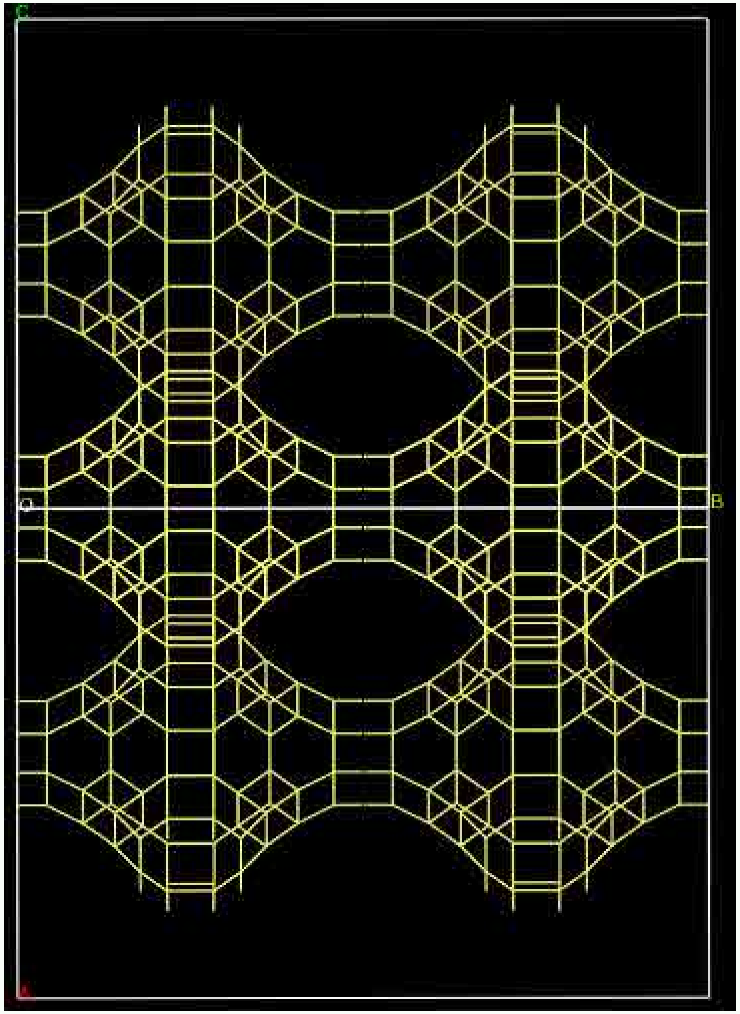


8 MR
Tri-directional

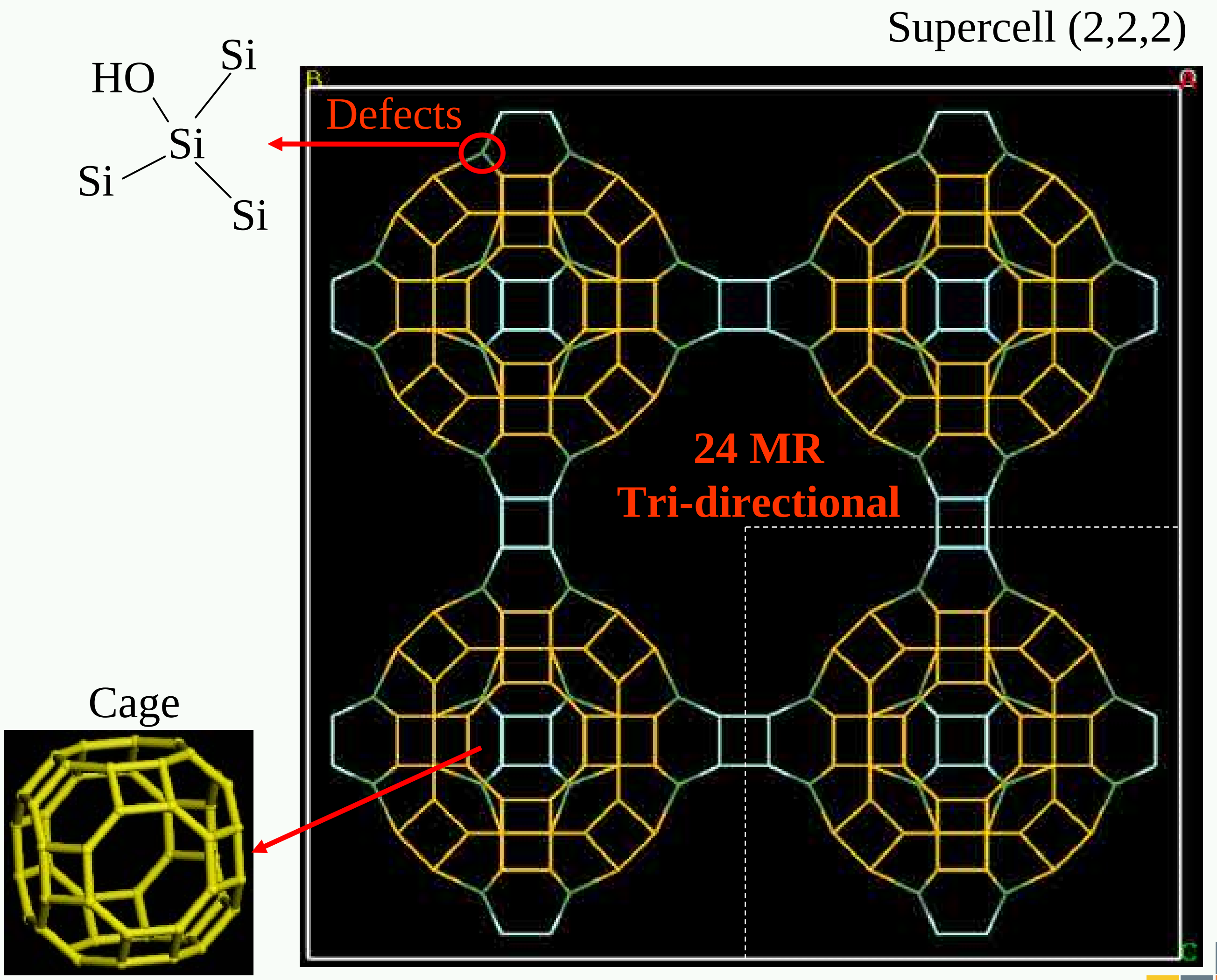
Supercell (2,2,2)



16 MR
Tri-directional

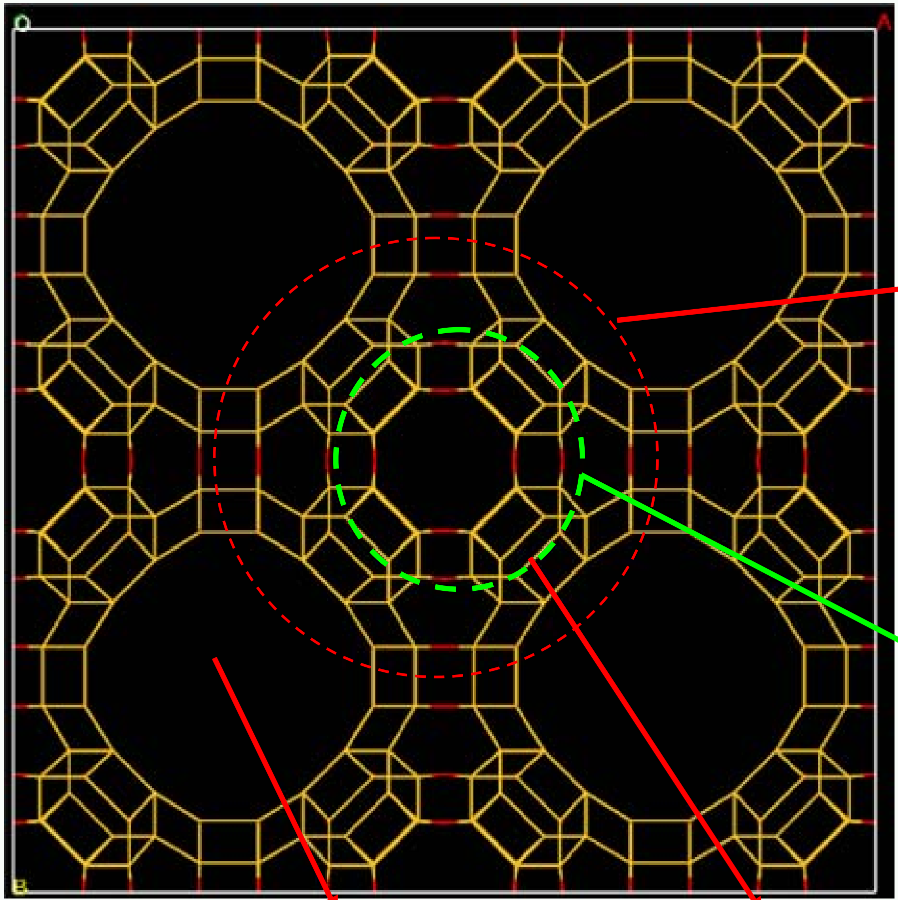


Hypothetical zeolite structures



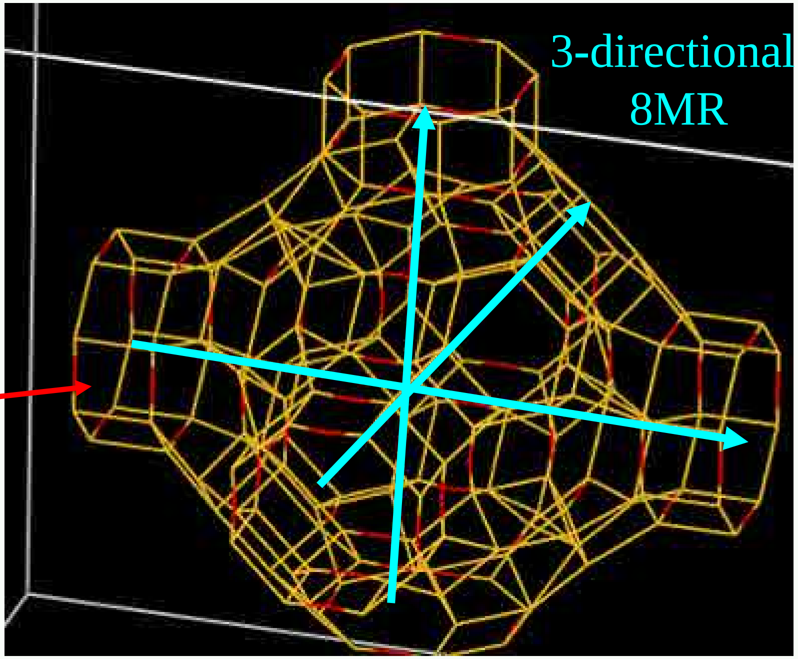
Hypothetical zeolite structures

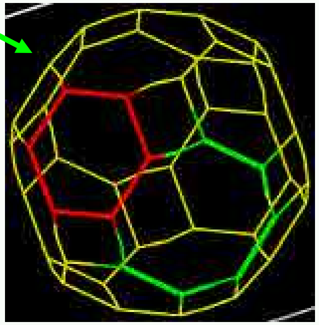
Supercell



3-directional 16MR connected with 8MR

Centre





4MR
6MR
8MR

Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Combinatorial methodology applied to zeolite synthesis

Objective: Optimization of known structures or search of new materials

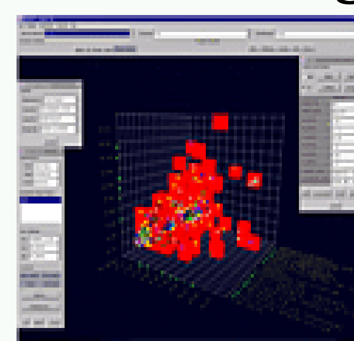
Global Strategy

Samples Design

HT Synthesis

HT Characterization

Data Mining



Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

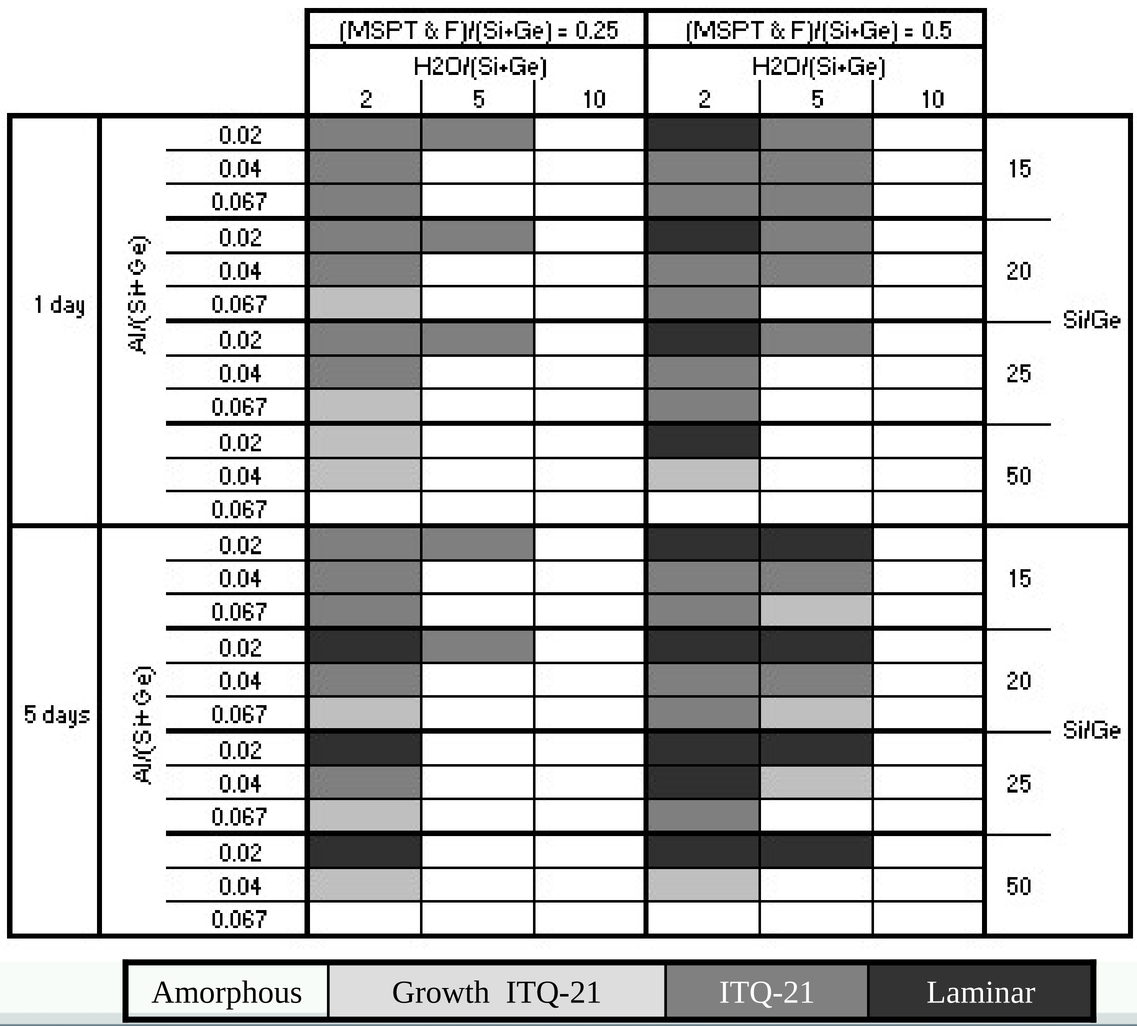
Study of zeolite ITQ-21

Experimental design: Factorial design 2x4x3x2x3 (144 exp)

	Time (days)	**Si/Ge**	**Al/ (Si+Ge)**	**(MSPT+F)/ (Si+Ge)**	**H_2O/(Si+Ge)**
Nr. Levels	2	4	3	2	3
Lowest value	1	15	0.02	0.25	2
Highest value	5	50	0.067	0.5	10

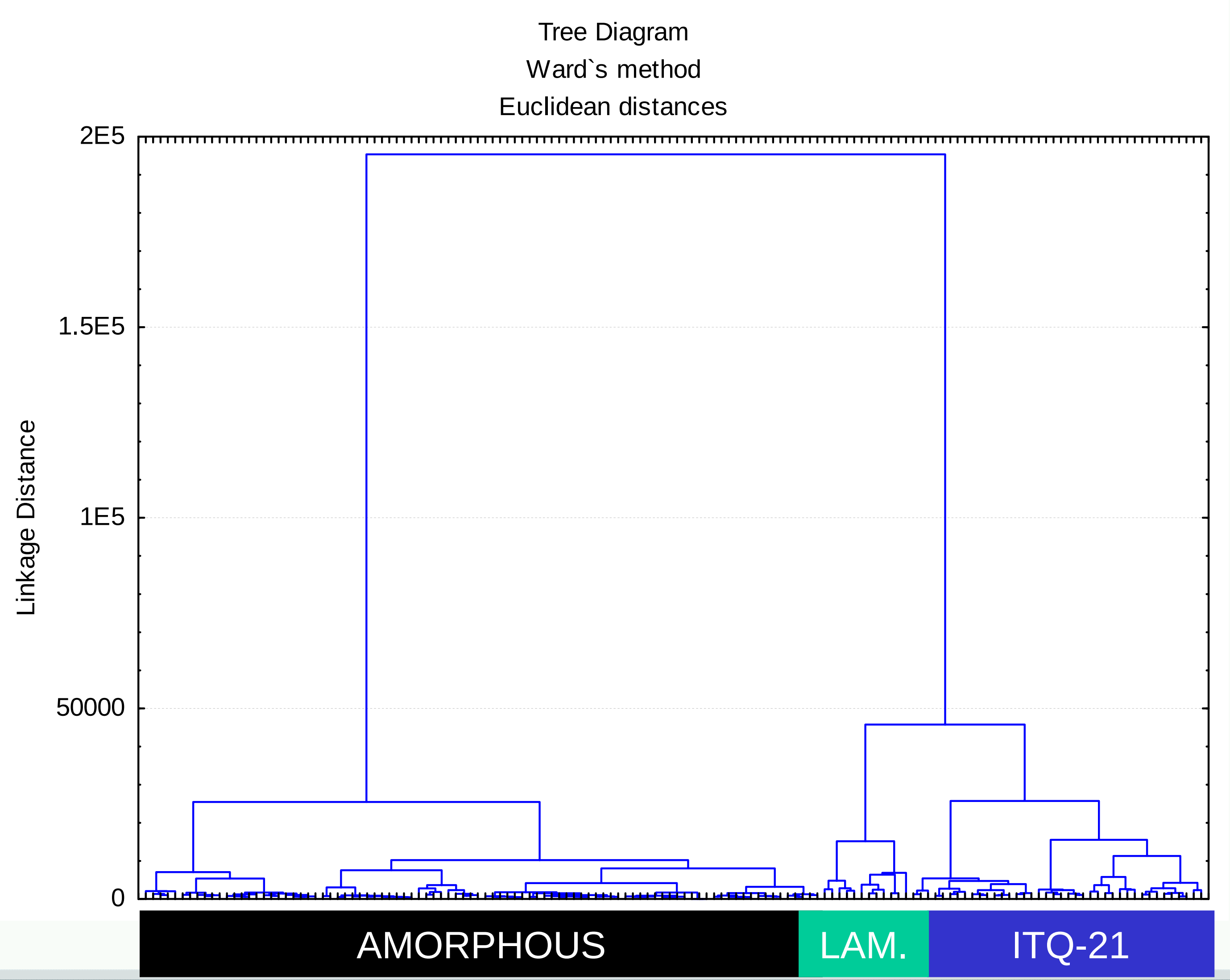
Study of zeolite ITQ-21

Screening Results: Phase Diagram

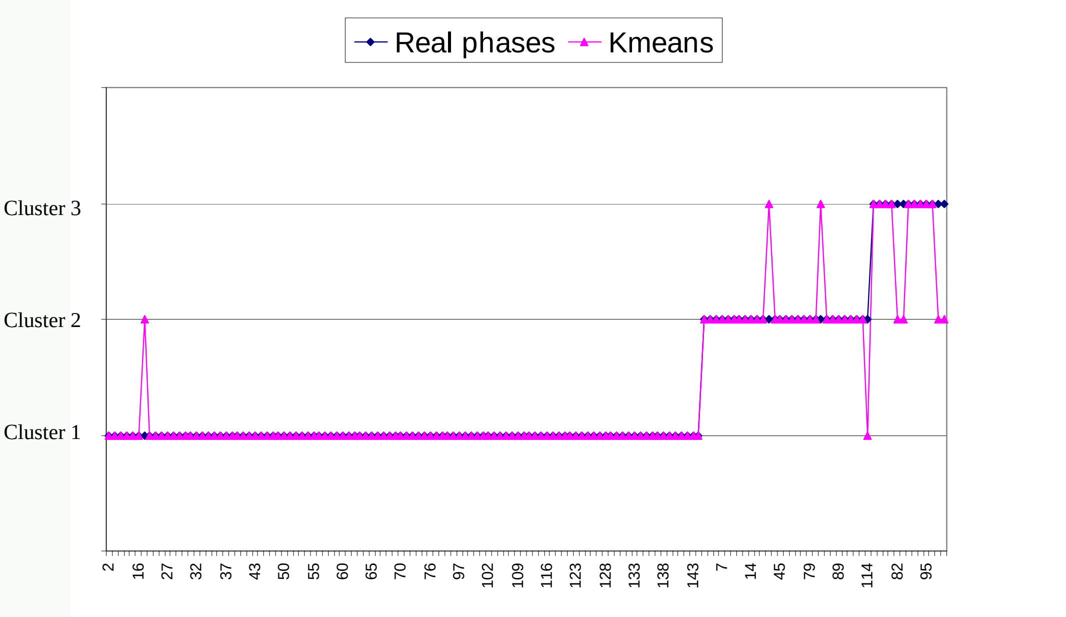


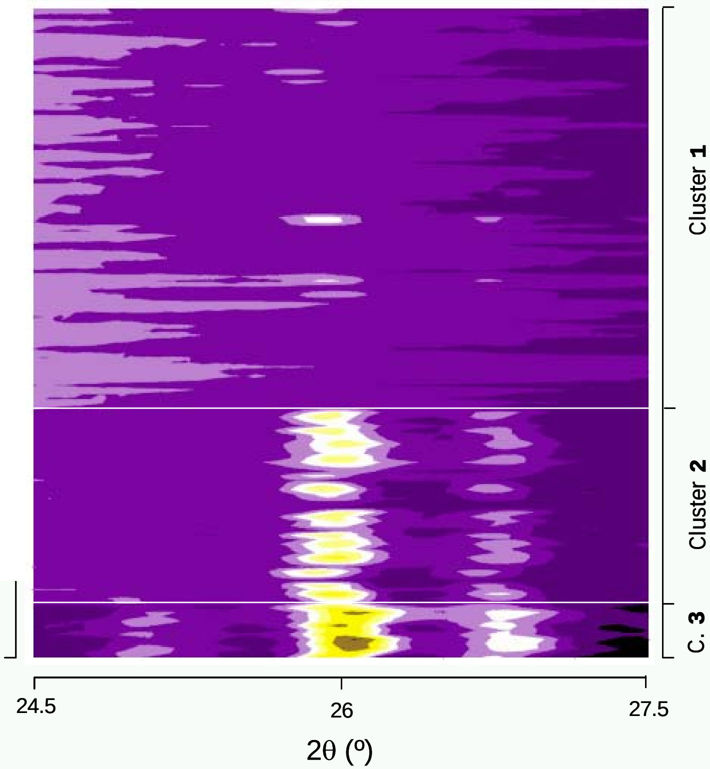
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Tree Clustering



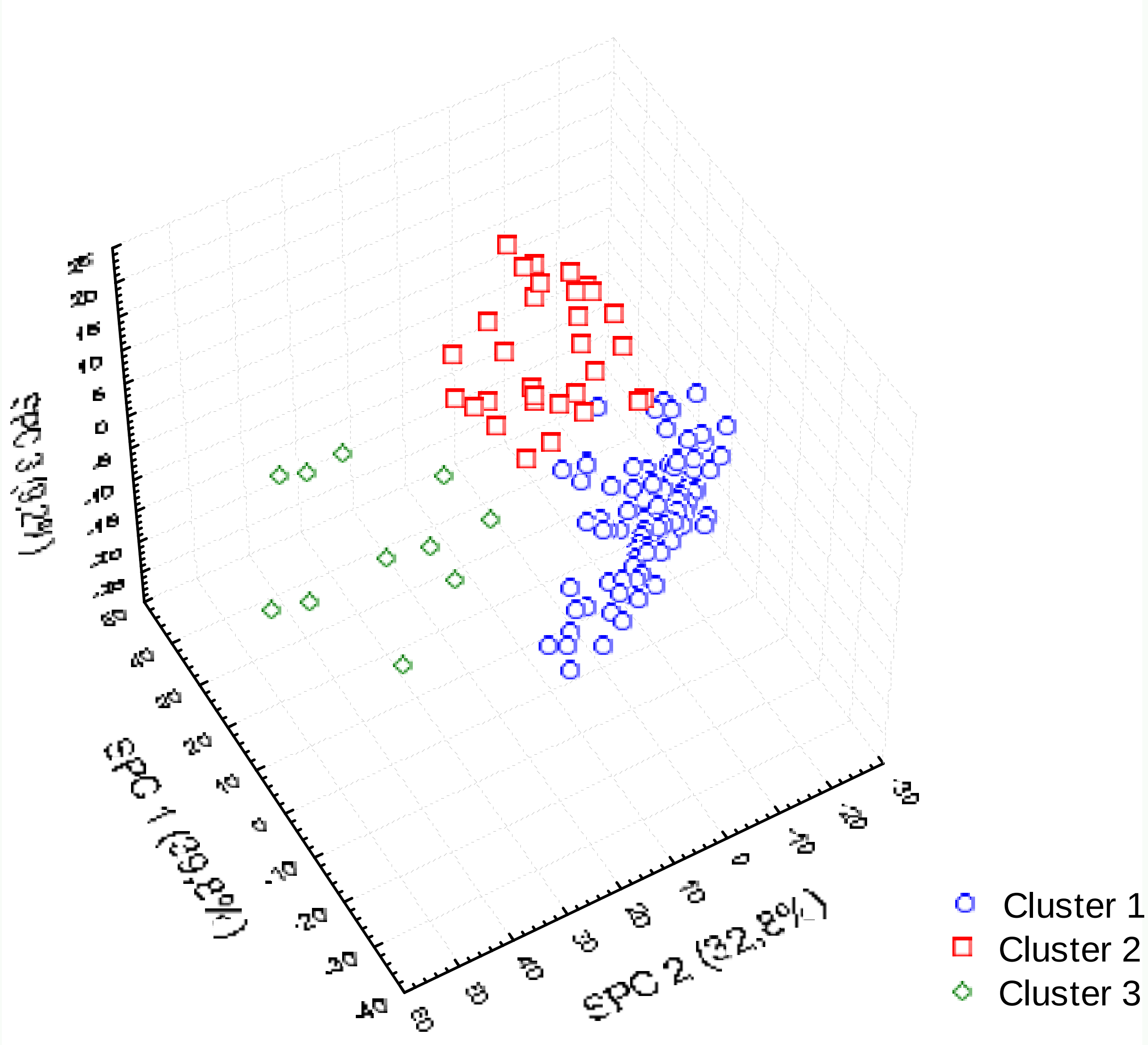
K-Means clustering



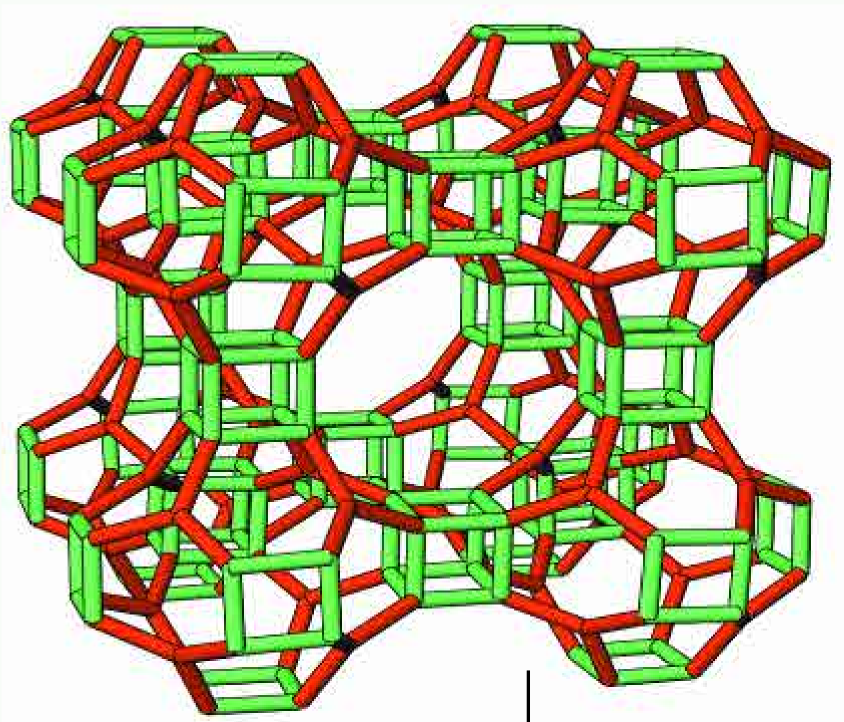


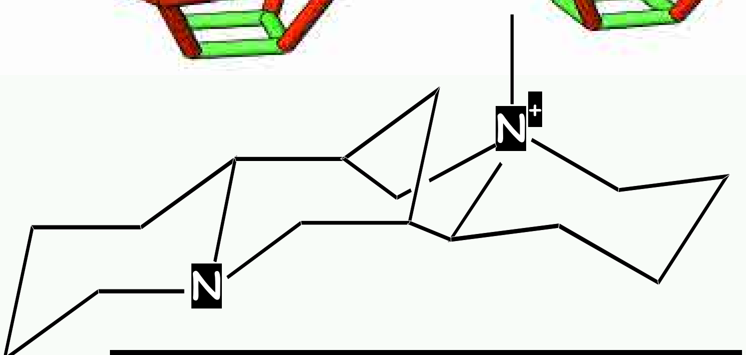
Cluster	Nr. Real phases	K-means clustering	Accuracy (%)
1. Amorph.	102	101	99.0
2. ITQ-21	29	26	89.7
3. Laminar	13	9	69.2

Principal Component Analysis

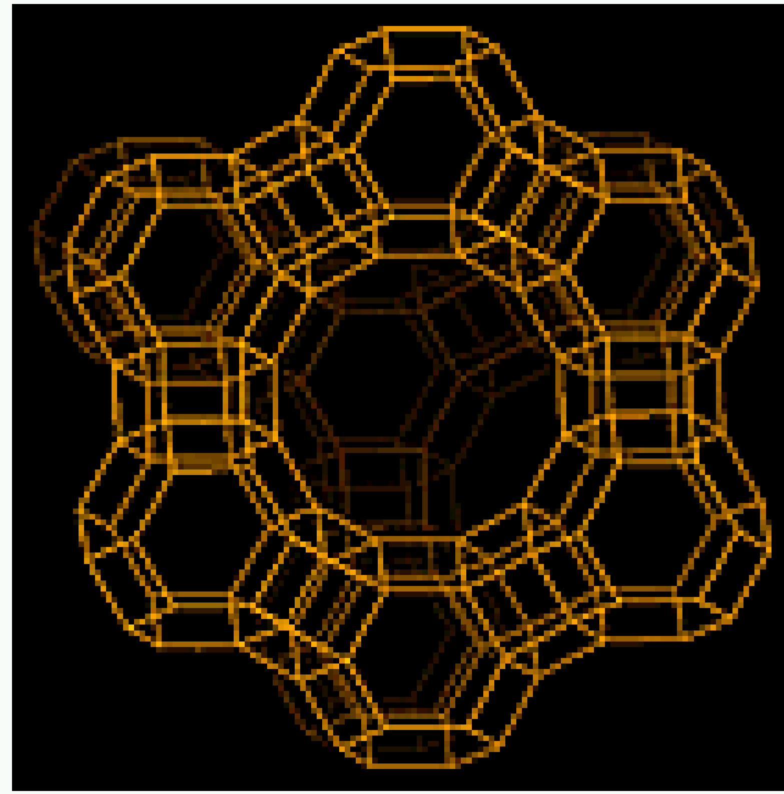


STRUCTURE OF ZEOLITE ITQ-21



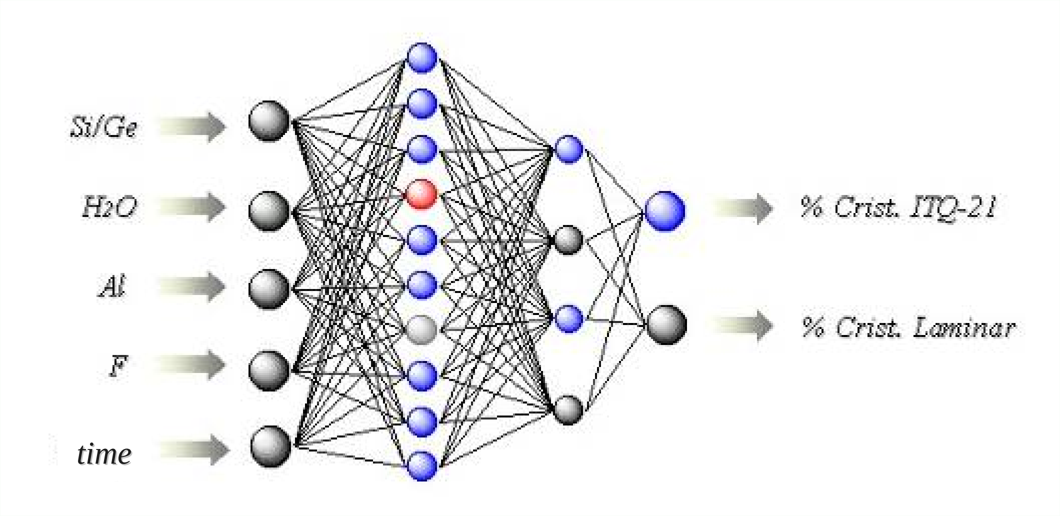


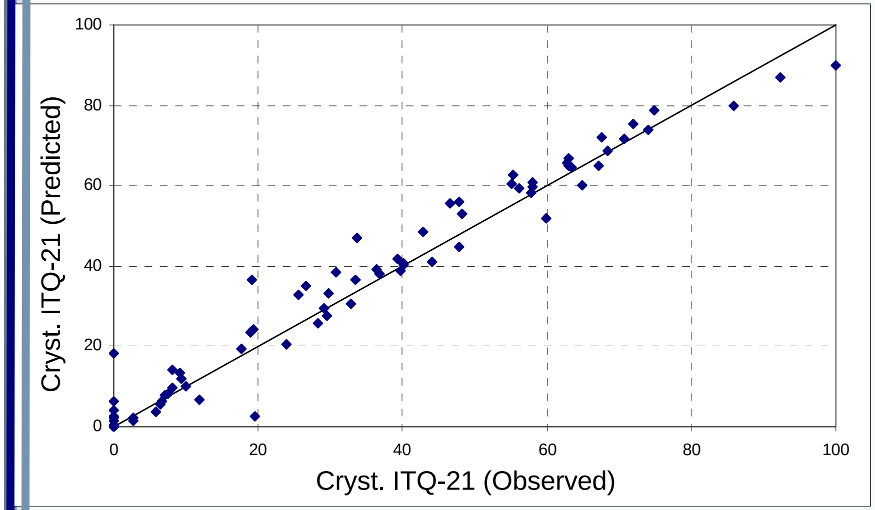
N-methylsparteinium

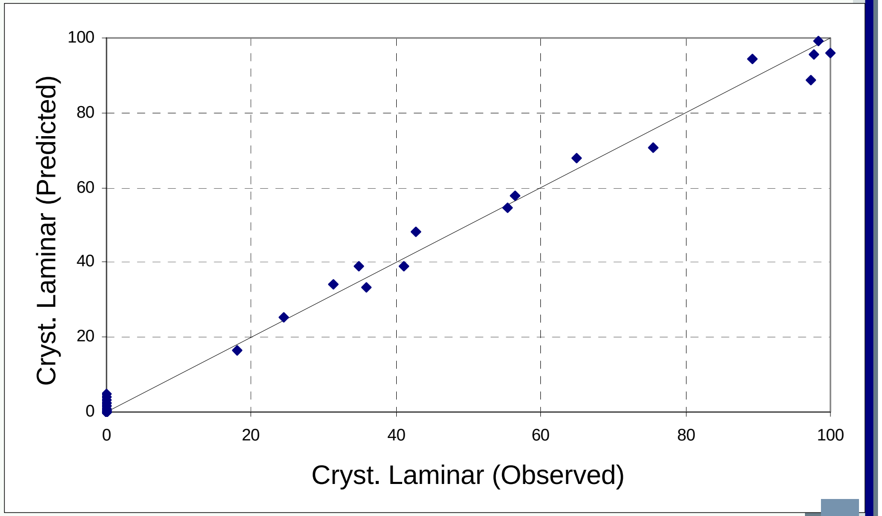


A. Corma, et al NATURE (2000)

Neural Networks







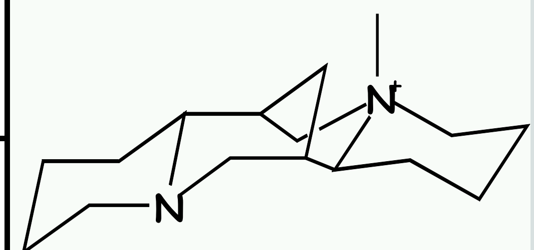
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

CSIC UNIVERSIDAD POLITECNICA DE VALENCIA

Factorial Experimental Design (144)

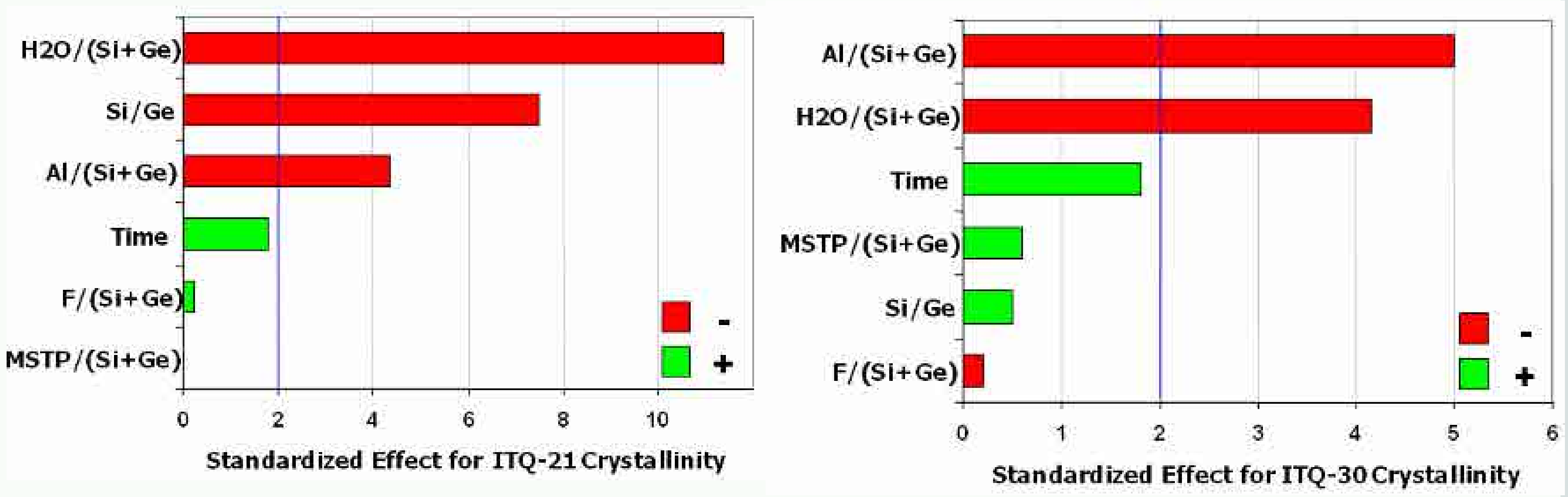
Time	Si/Ge	Al/(Si+Ge)	(MSPT & F)/(Si+Ge) = 0.25, H2O/(Si+Ge) = 2	(MSPT & F)/(Si+Ge) = 0.25, H2O/(Si+Ge) = 5	(MSPT & F)/(Si+Ge) = 0.25, H2O/(Si+Ge) = 10	(MSPT & F)/(Si+Ge) = 0.5, H2O/(Si+Ge) = 2	(MSPT & F)/(Si+Ge) = 0.5, H2O/(Si+Ge) = 5	(MSPT & F)/(Si+Ge) = 0.5, H2O/(Si+Ge) = 10
1 day	15	0.02	ITQ-21	ITQ-21	Amorphous	ITQ-30	ITQ-21	Amorphous
1 day	15	0.04	ITQ-21	Amorphous	Amorphous	ITQ-21	ITQ-21	Amorphous
1 day	15	0.067	ITQ-21	Amorphous	Amorphous	ITQ-21	ITQ-21	Amorphous
1 day	20	0.02	ITQ-21	ITQ-21	Amorphous	ITQ-30	ITQ-21	Amorphous
1 day	20	0.04	ITQ-21	Amorphous	Amorphous	ITQ-21	ITQ-21	Amorphous
1 day	20	0.067	Growth ITQ21	Amorphous	Amorphous	ITQ-21	Amorphous	Amorphous
1 day	25	0.02	ITQ-21	ITQ-21	Amorphous	ITQ-30	ITQ-21	Amorphous
1 day	25	0.04	ITQ-21	Amorphous	Amorphous	ITQ-21	Amorphous	Amorphous
1 day	25	0.067	Growth ITQ21	Amorphous	Amorphous	ITQ-21	Amorphous	Amorphous
1 day	50	0.02	Growth ITQ21	Amorphous	Amorphous	ITQ-30	Amorphous	Amorphous
1 day	50	0.04	Growth ITQ21	Amorphous	Amorphous	Growth ITQ21	Amorphous	Amorphous
1 day	50	0.067	Amorphous	Amorphous	Amorphous	Amorphous	Amorphous	Amorphous
5 days	15	0.02	ITQ-21	ITQ-21	Amorphous	ITQ-30	ITQ-30	Amorphous
5 days	15	0.04	ITQ-21	Amorphous	Amorphous	ITQ-21	ITQ-21	Amorphous
5 days	15	0.067	ITQ-21	Amorphous	Amorphous	ITQ-21	Growth ITQ21	Amorphous
5 days	20	0.02	ITQ-30	ITQ-21	Amorphous	ITQ-30	ITQ-30	Amorphous
5 days	20	0.04	ITQ-21	Amorphous	Amorphous	ITQ-21	ITQ-21	Amorphous
5 days	20	0.067	Growth ITQ21	Amorphous	Amorphous	ITQ-21	Growth ITQ21	Amorphous
5 days	25	0.02	ITQ-30	Amorphous	Amorphous	ITQ-30	ITQ-30	Amorphous
5 days	25	0.04	ITQ-21	Amorphous	Amorphous	ITQ-30	Growth ITQ21	Amorphous
5 days	25	0.067	Growth ITQ21	Amorphous	Amorphous	ITQ-21	Amorphous	Amorphous
5 days	50	0.02	ITQ-30	Amorphous	Amorphous	ITQ-30	ITQ-30	Amorphous
5 days	50	0.04	Growth ITQ21	Amorphous	Amorphous	Growth ITQ21	Amorphous	Amorphous
5 days	50	0.067	Amorphous	Amorphous	Amorphous	Amorphous	Amorphous	Amorphous

Legend: Amorphous | Growth ITQ21 | ITQ-21 | ITQ-30

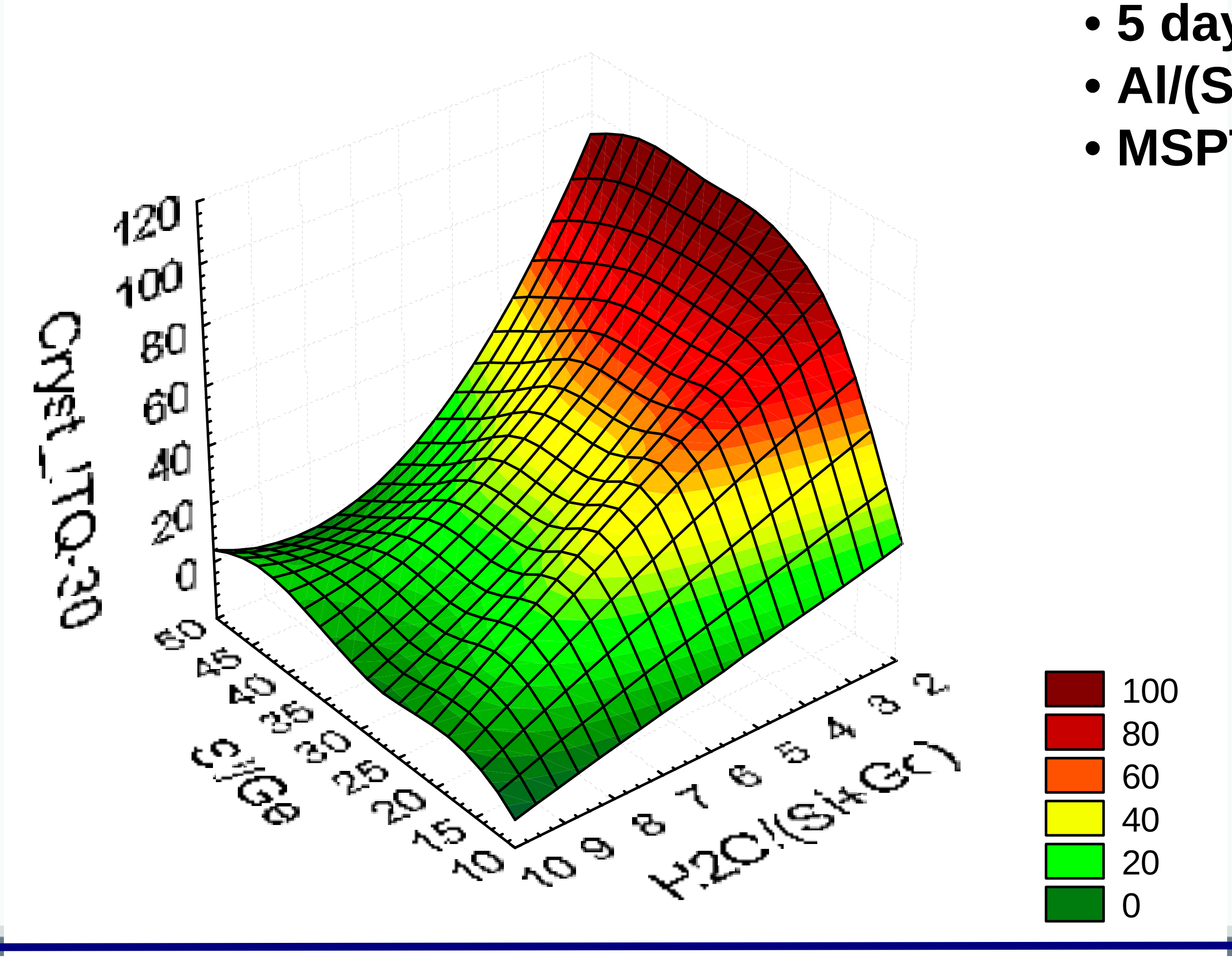


Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Pareto analysis showing the relative influence of each synthesis variable over crystallinity of the different phases



ITQ-30 crystallinity

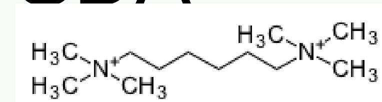


Synthesis conditions:

- **5 days**
- **Al/(Si+Ge) = 0.02**
- **MSPTF/(Si+Ge) = 0.5**

Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

HT phase diagram using hexamethonium as SDA

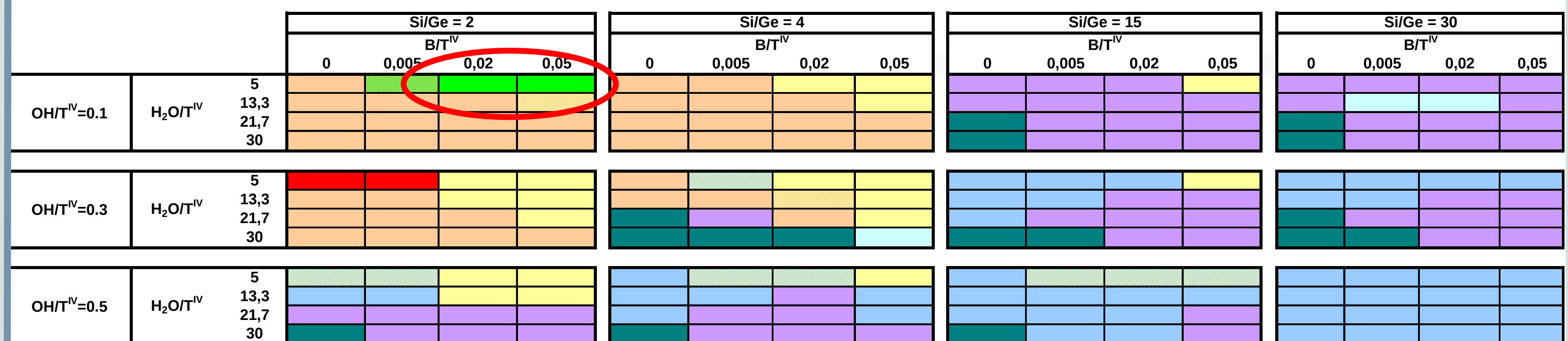


Factorial Design: $4^3 \times 3^1$

Variables	Levels	Level 1	Level 2	Level 3	Level 4
Si/Ge	4	2	4	15	30
OH/(Si+Ge)	3	0.1	0.3	0.5	
B/(Si+Ge)	4	0	0.005	0.020	0.050
H_2O/(Si+Ge)	4	5	13.3	21.7	30

→ **192 experiments**

SDA/(Si+Ge)=0.25 T=175ºC t=14 days



Phases

ITQ22	LAM
SSZ31	EU1
ITQ24	Amorp

Mixtures

- LAMELLAR-ITQ24
- ITQ22-ITQ24-ITQ33
- ITQ22-ITQ24
- ITQ24-ITQ17
- ITQ33-ITQ24

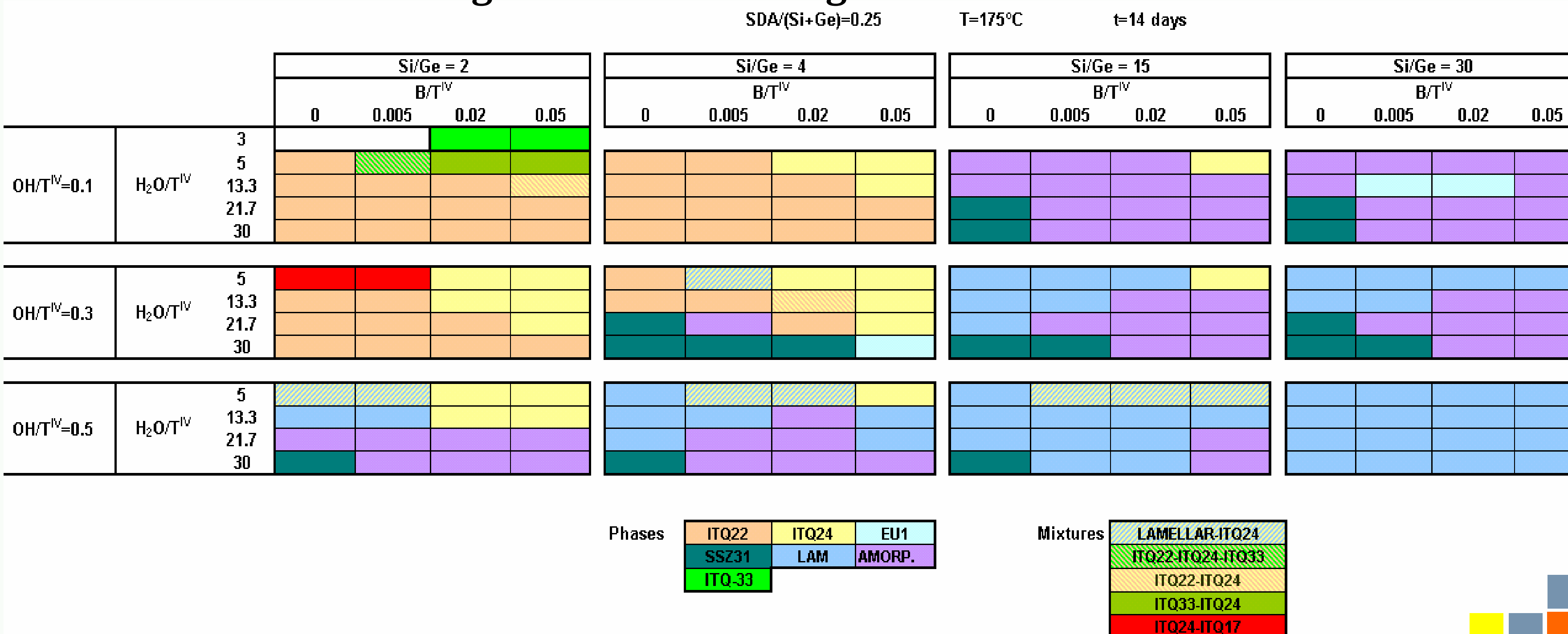
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

ITQ-33 Case

- 8 different phases
- Many mixtures
- One new phase to be found
- Germanium is shifting the peaks
- Great variations in peaks intensity due to composition diversity

Diverse structures:
EU-1, ITQ-17, ITQ-22, ITQ-33
ITQ-24, SSZ-31, a lamellar phase

Phase diagram obtained using *hexamethonium* as SDA



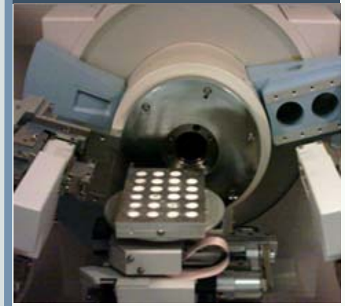
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

A Challenging Task

SYNTHESIS

- Great number of samples
- Impurities

Throughtput



- Data quality loss

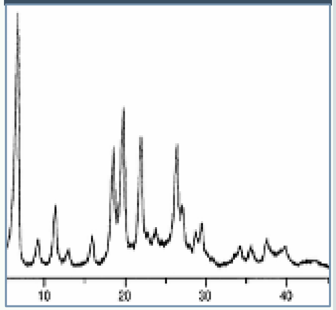
Additional

- Time constraint
- Fully automatic solution
- Minimize user interaction
- Integration of chemical knowledge (potential phases)

For one specific structure, large intensity & angles changes due to:

- Level of crystallinity
- Crystallite size
- Chemical composition

Powder Diffrac.

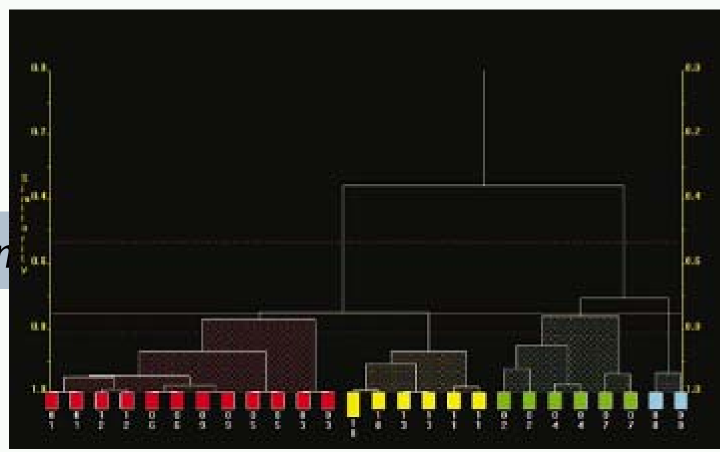


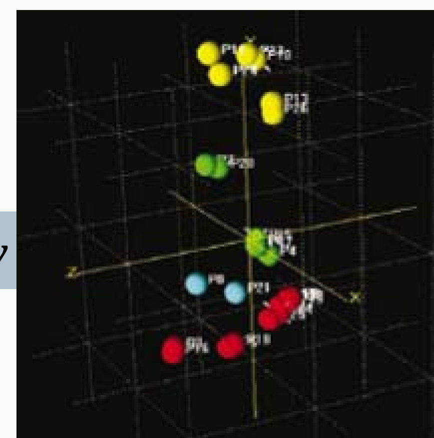
For mixtures this becomes a challenge !!

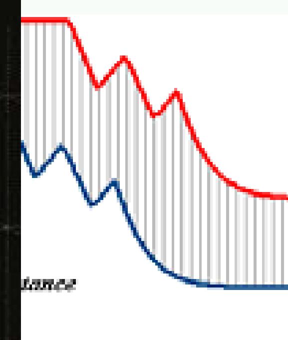
Existing Solutions

Approaches	Pattern	Methods	Associated techniques	Advantages	**Drawbacks**
Peak search and indexing	Reduced (Stick)	d-spacing and Intensity	• Hanawalt (8 strongest peaks) • Fink	- Low storage constraint - Speed - Ease of DB building	- Peak determination (overlap, shoulders) - Number of peaks to consider - Weak peaks are discarded Strong sample preparation and hardware dependence
Full profile	Full	Similarity-based on Euclidian distance	• Statistics [i] • PCA, MDS [i] • Clustering [i]	- Full use of info.	- No commercial database - Decision for pre-processing

The criterion … ***ey***







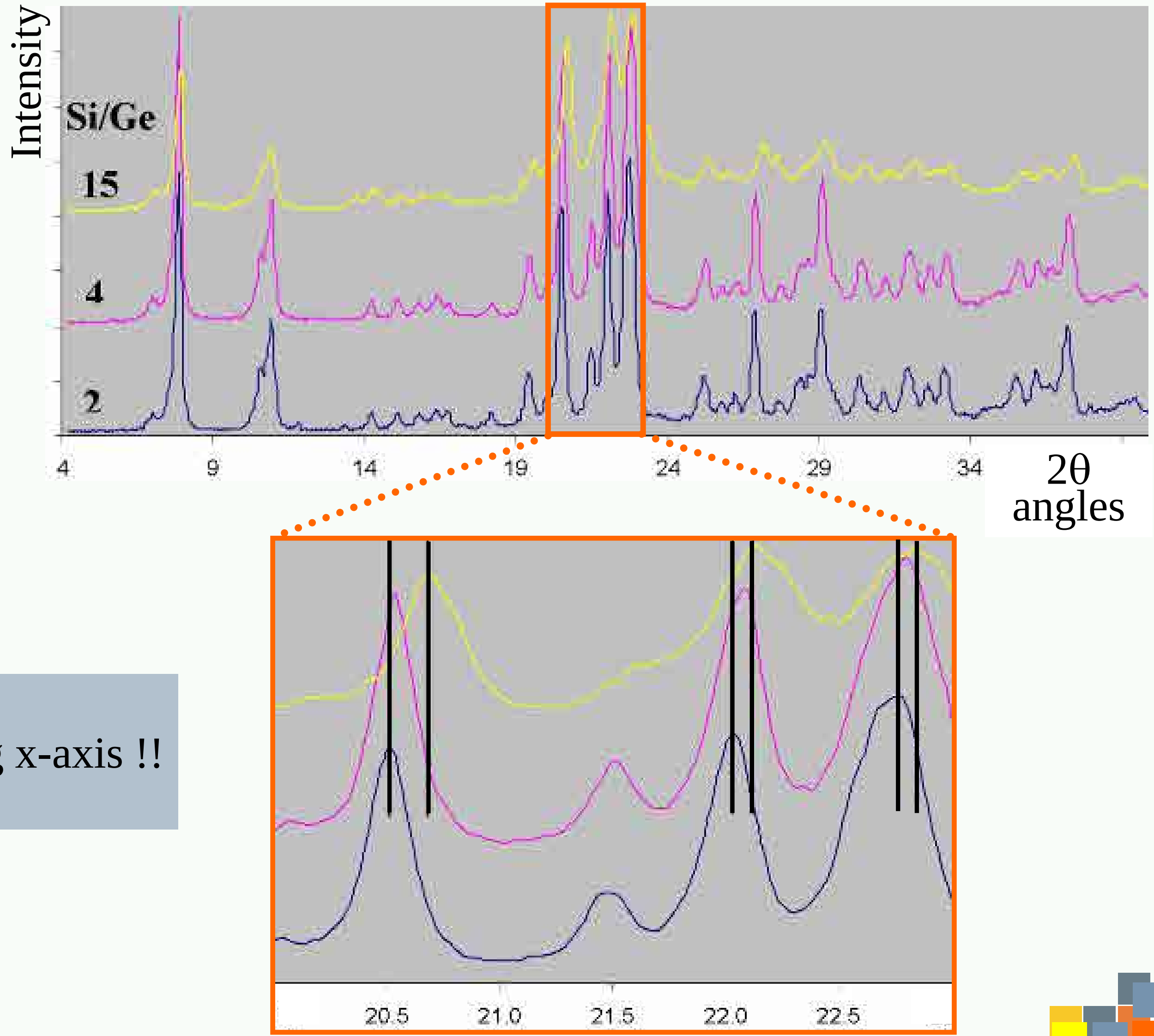
$$Eucl(Q,C) = \sum_{i=1}^{m} |q_i - c_i|$$

[i] C. J. Gilmore, *et al. J. Appl. Cryst*. **2004**, 37, 231-242.
A. Corma, *et al. Chem. Mater.*, **2006**, *18*, 3287-3296

Requirements

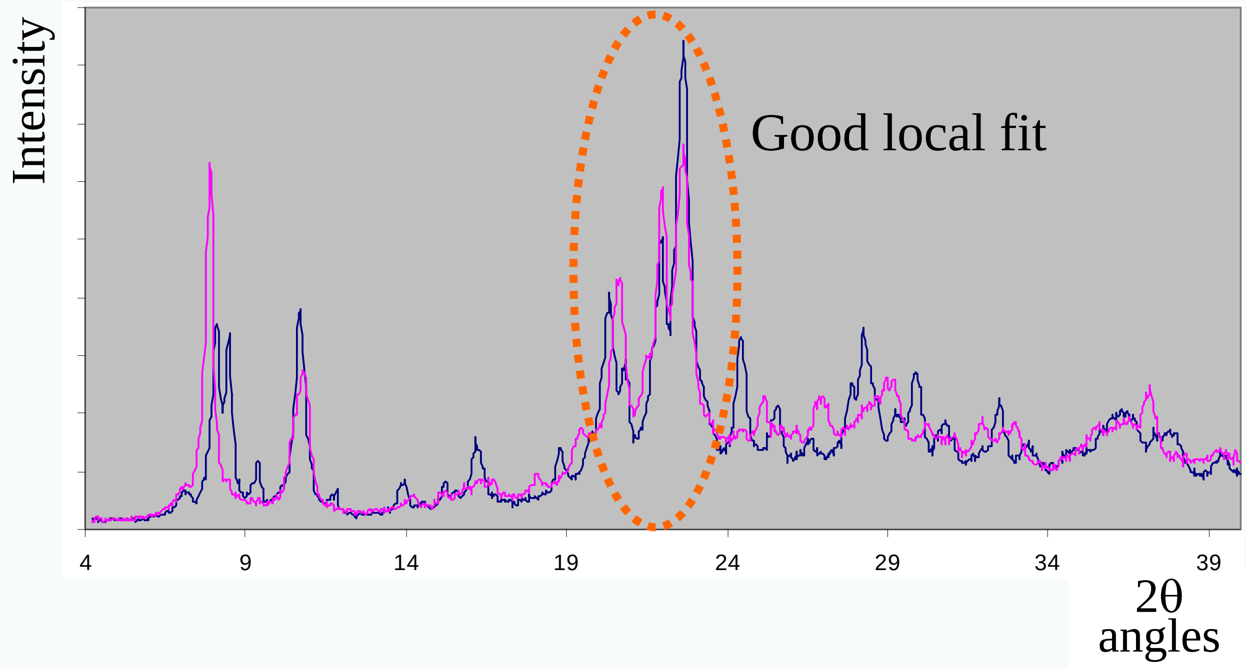
"2θ-shifting"

<u>ITQ-24 zeolite</u>
obtained with Hexamethonium as SDA,
Si/Ge is varied



Not constant along x-axis !!

Requirements



"high peaks"

<u>ITQ-22 zeolite</u>
is also obtained with Hexamethonium

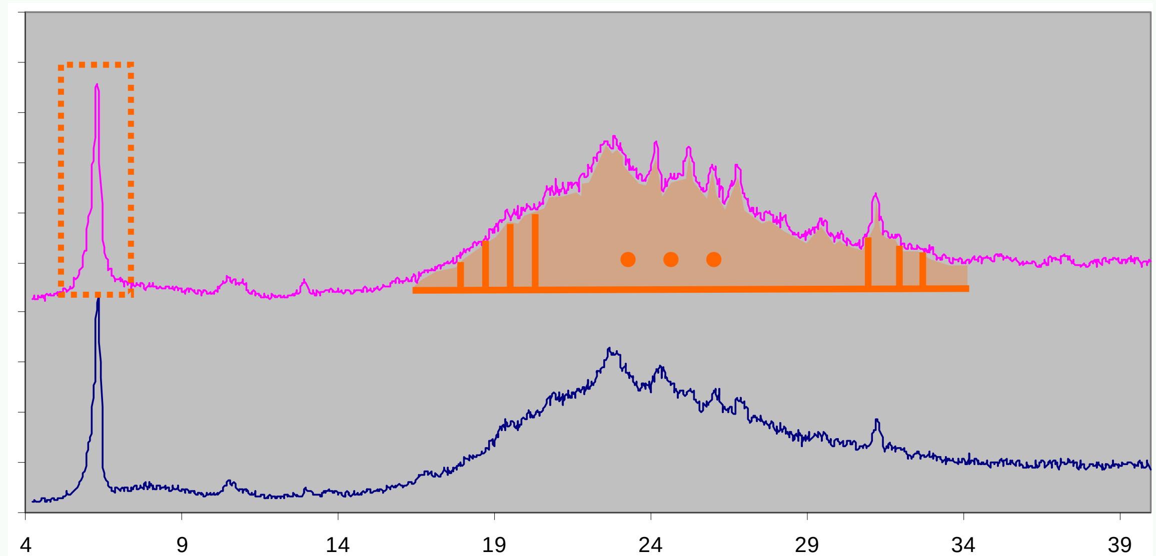
Correctly size the influence of the highest peaks

When traditional approach (PolySnap2©) is used:

ITQ-22 (---, blue) is predicted as ITQ-24 (---, pink)

Requirement

"growing phases"
& "amorphous"
A <u>lamellar</u> is also obtained with Hexamethonium

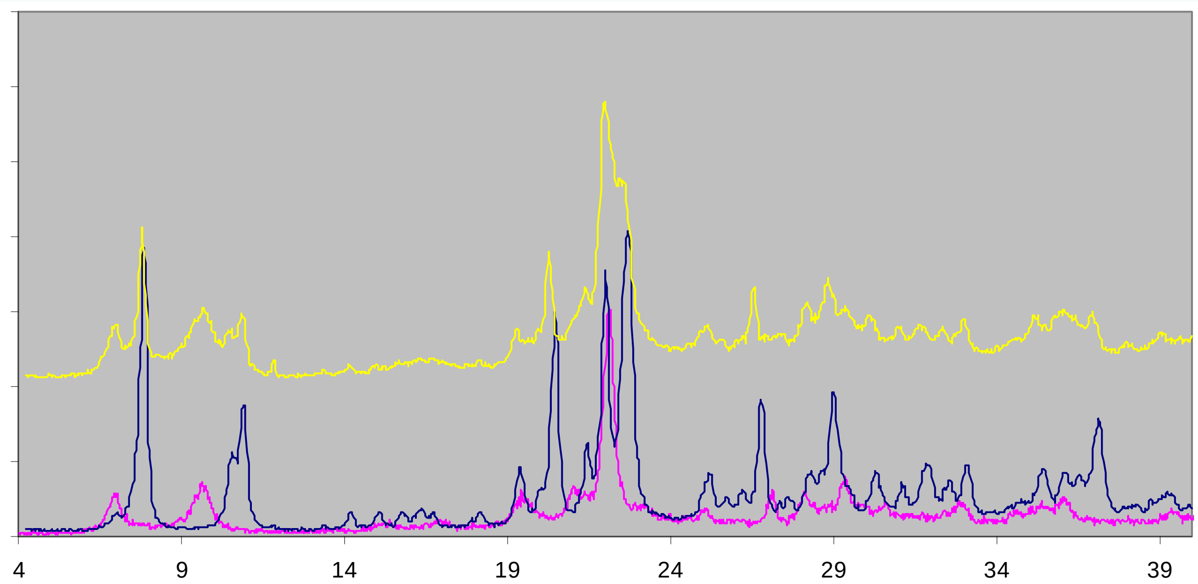


The Lamellar phase is growing,
being the traditional approach prediction "amorphous"

Correctly deal with amorphous

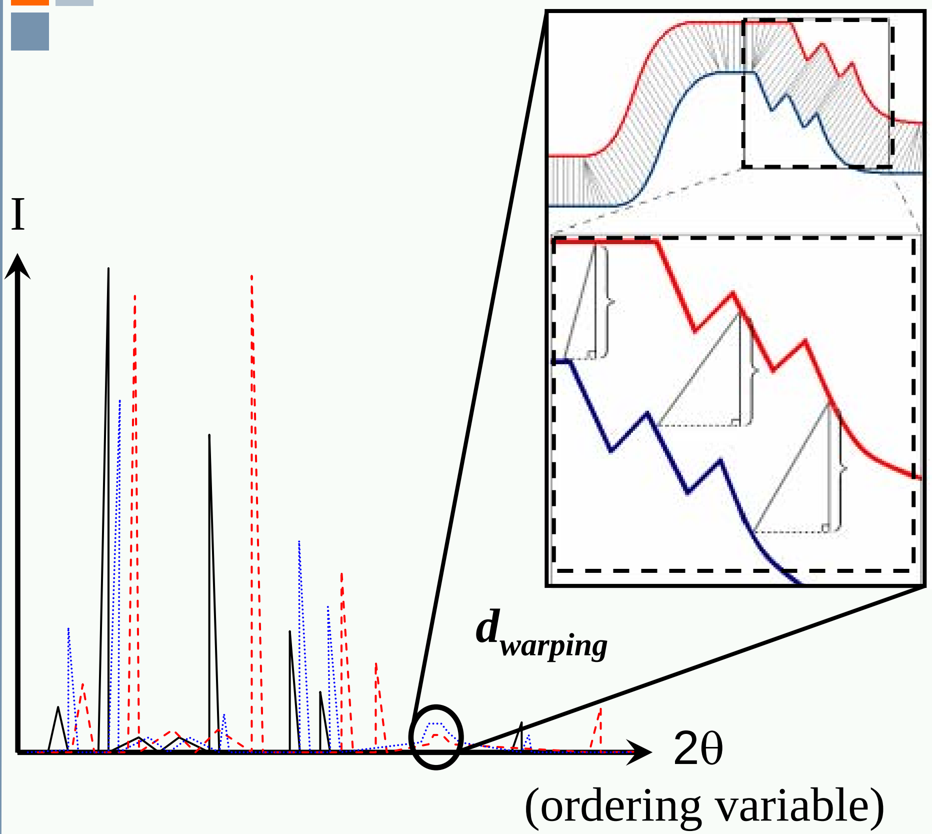
Requirement

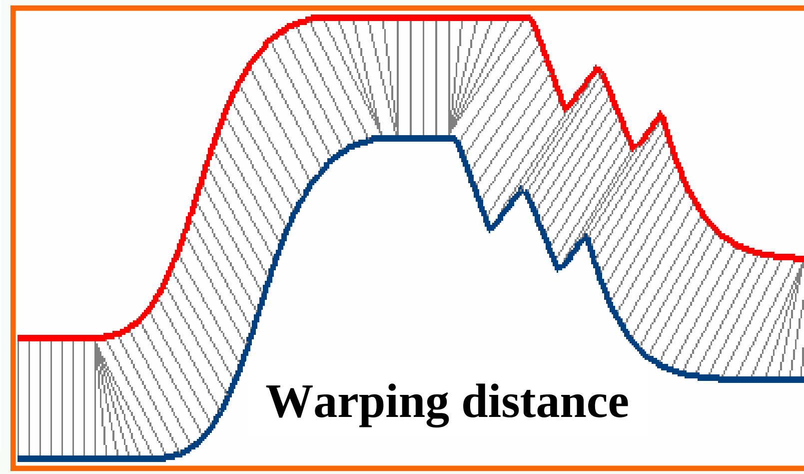
"Mixtures"
are also obtained with Hexamethonium



Pure ITQ-17 (---, yellow) is predicted,
being a mixture between ITQ-17 (---, pink)
and ITQ-24 (---, blue)

Distance and Similarity Criterion



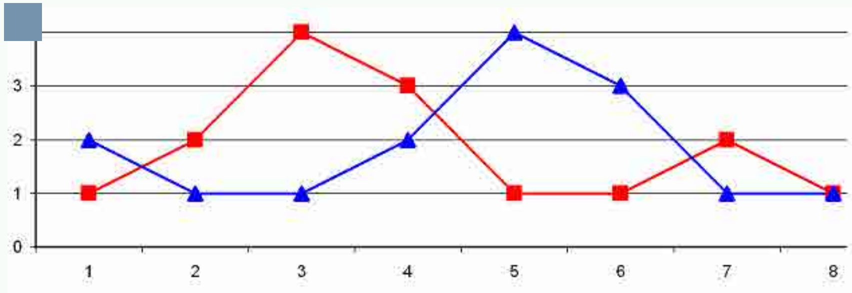


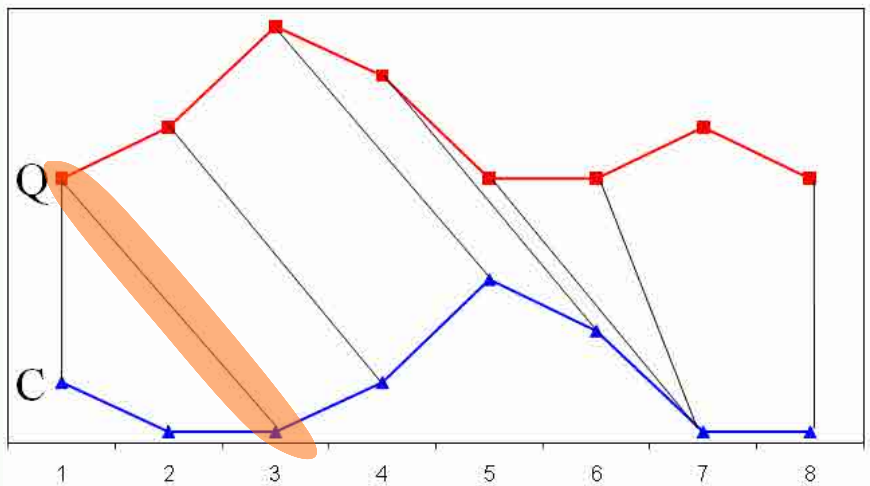
A recursive definition:
Matching in (i) depends of (i-1)

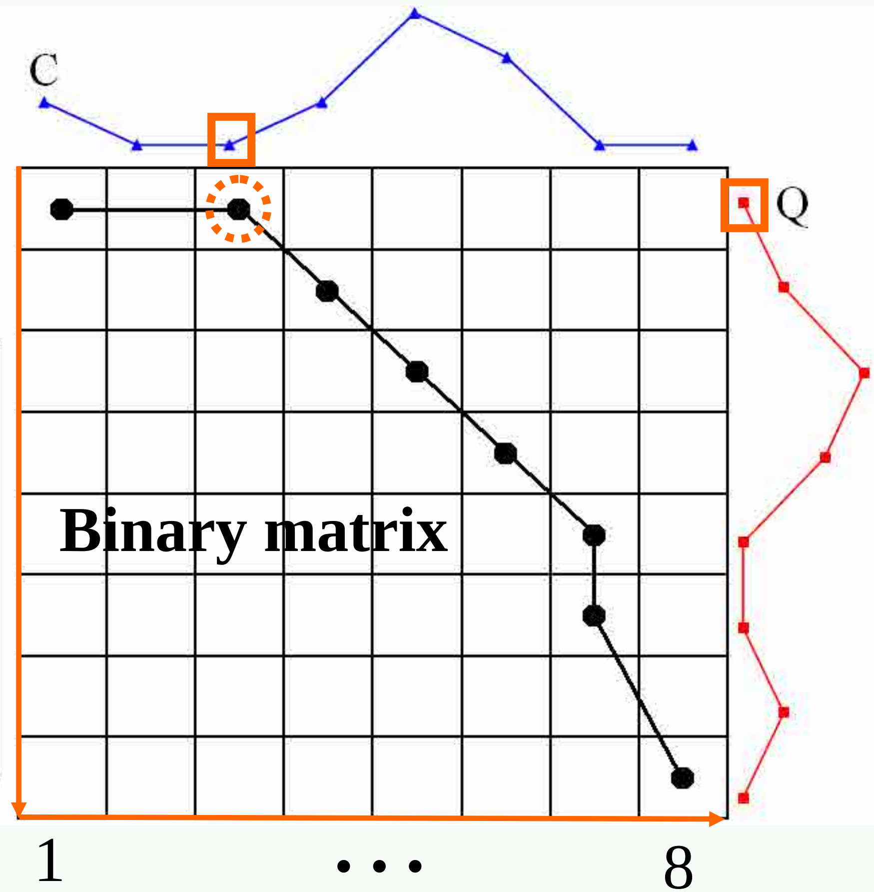
$$\gamma(q_i, c_j) = \left| q_i - c_j \right| + \min\left\{ \gamma(q_{i-1}, c_j), \gamma(q_i, c_{j-1}), \gamma(q_{i-1}, c_{j-1}) \right\}$$

$$DTW(Q, C) = \gamma(q_m, c_n)$$

Warping Path







We look for the **<u>BEST</u>** WP minimizes the distance

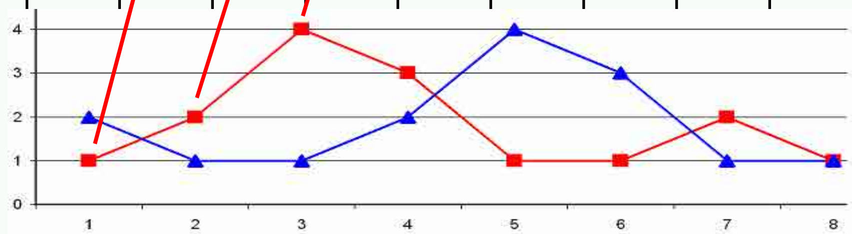
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Best Warping Path

$$X_{i,j} = |Q_i - C_j| + \mathrm{Min}\{|Q_{i-1} - C_{j-1}| \,;\, |Q_{i-1} - C_j| \,;\, |Q_i - C_{j-1}|\}$$

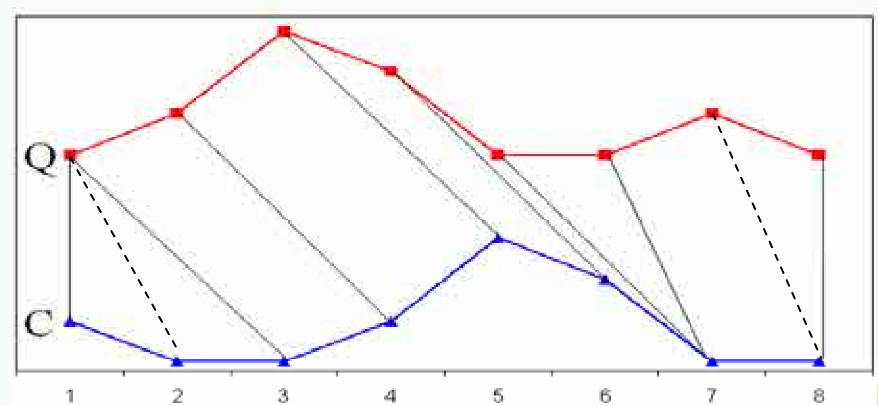
	1	2	3	4	5	6	7	8
1	1	1	1					
2	1	2	2	1				
3	3	4	5	3	1			
4		5	6	4	2	1		
5			5	5	5	3	1	
6				6	8	5	1	1
7					8	6	2	2
8						8	2	2

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
Q	1	2	4	3	1	1	2	1
C	2	1	1	2	4	3	1	1



Search_Best_Path(X)

	1	2	3	4	5	6	7	8
1	**1**	**1**	**1**					
2	1	2	2	**1**				
3	3	4	5	3	**1**			
4		5	6	4	2	**1**		
5			5	5	5	3	**1**	
6				6	8	5	**1**	1
7					8	6	2	**2**
8						8	2	**2**



Weighting

The proposed *Adaptable Time Warping* (ATW) is designed such as 1-norm, DTW, and **<u>more generally all warping and non-warping distances</u>** are particular cases of ATW.

$\gamma_A(A'_i, B'_j, \pi_{ij})$ is the recursive function that computes the distance between A'_i and B'_j

$$\Pi = \begin{pmatrix} \pi_{11} & \cdots & \pi_{1t} \\ \vdots & \ddots & \vdots \\ \pi_{t1} & \cdots & \pi_{tt} \end{pmatrix} \quad \pi_{ij} \in \mathbb{R}^+, \quad \forall i, j \in [1, t]$$

$$ATW(A, B, \Pi) = \gamma_A(A'_t, B'_t, \pi_{tt})$$

$$with\ \gamma_A(A'_i, B'_j, \pi_{ij}) = \begin{cases} \infty \quad if \quad \pi_{ij} = \infty \\ |a_i - b_j| \times \pi_{ij} + \begin{cases} 0, if \quad i = j = 1 \\ \gamma_A(A'_i, B'_{j-1}, \pi_{i(j-1)}), if \quad i = 1\ and \quad j > 1 \\ \gamma_A(A'_{i-1}, B'_j, \pi_{(i-1)j}), if \quad i > 1\ and \quad j = 1 \\ \min \begin{Bmatrix} \gamma_A(A'_{i-1}, B'_j, \pi_{(i-1)j}) \\ \gamma_A(A'_i, B'_{j-1}, \pi_{i(j-1)}) \\ \gamma_A(A'_{i-1}, B'_{j-1}, \pi_{(i-1)(j-1)}) \end{Bmatrix}, else \end{cases} \end{cases}$$

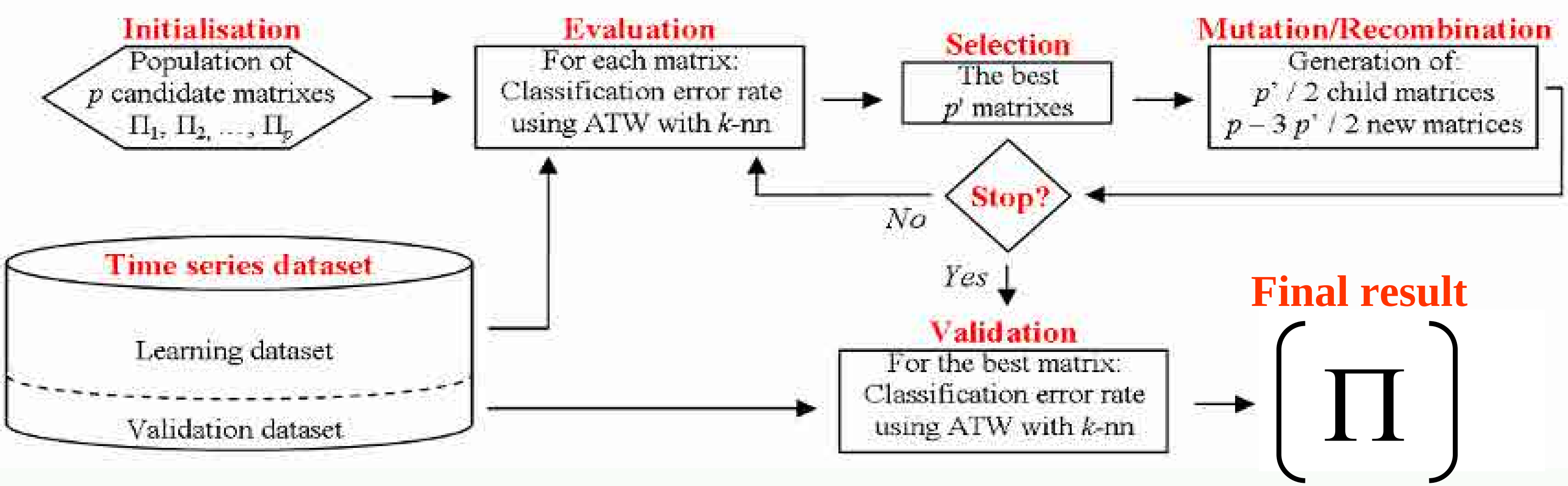
Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

Heuristics, GAs, & Machine Learning

- Even considering the reduced case where each π_{ij} can **only take *v* different values**

The resulting total complexity of the exhaustive search is $O\left(v^{t^2}.t^2\left(n^2-n\right)/2\right)$

Such an optimization process is **<u>unreachable</u>** even for small **v** values, few (**n**) and short (**t**) series. This implies the use of a heuristic which aims at quickly finding a solution, i.e a locally optimal matrix Π.



Optimized.

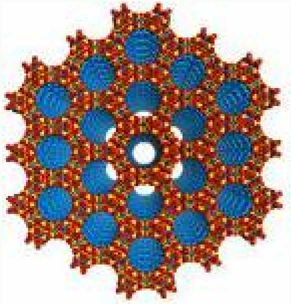
→Phases specificities are captured

ITQ-33 Case

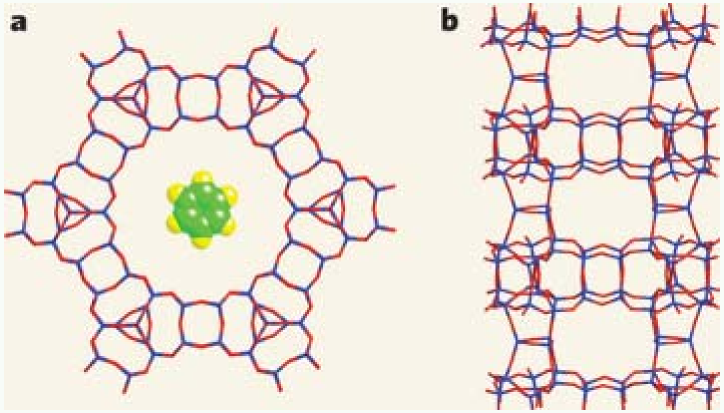
192 diffractograms with a parallelized XRD to follow the formation of the different phases.

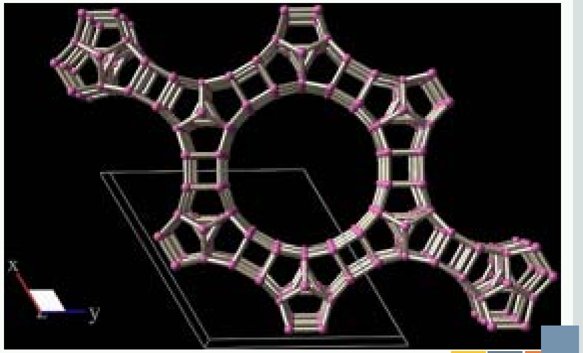
Strategy for extracting automatically each crystallographic phase from powder diffraction data

We use the new kind of "distance" that adapts itself for providing optimum discrimination



A very unique zeolite structure extra-large 18MR connected with medium 10MR pores[i]





Dr. Laurent Allan, BAUMES (baumesl@itq.upv.es)

[i] A. Corma, *et al. Nature*. **2006**, 44

Results: 92% Error Less!!

PolySnap 61

ATW 5

ATW

PolySnap©

No false negative

Predicted \ Real	Amorphous	ITQ-22	ITQ-24	EU-1	SSZ-31	Lamellar	Lamellar + 24	24 + 33	17 + 24	22 + 24	22 + 24 + 33	
Amorphous	55 / 55	2	3	2	3	28	2					55 / 95
ITQ-22		36 / 36	2							1	1	36 / 40
ITQ-24			17 / 8				2					17 / 10
EU-1				3 / 1								3 / 1
SSZ-31					15 / 13							15 / 13
Lamellar						46 / 18	1					46 / 19
Lamellar + 24							8 / 0					8 / 0
24 + 33								2 / 0				2 / 0
17 + 24									2 / 0			2 / 0
22 + 24		**2**								2 / 0		4 / 0
22 + 24 + 33											1 / 0	1 / 0
Other mixtures			2 / 6		**1**		3	2	2	1		3 / 13
	55	38	19	3	16	46	8	2	2	2	1	187/192=97 131/192=68

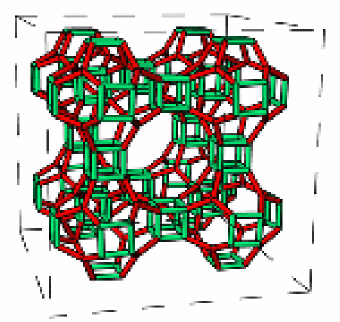
Quantitative analysis is carried out using all the measured data points with **singular value decomposition** (SVD) as the tool of matrix inversion to ensure computational stability. The percentage composition of the sample is determined. $M = U\Sigma V^*$

- No background removal
- No peak extraction
- No smoothing (wavelets)
- No interpolation
- No region exclusion
- No pre-defined shift

- No user interaction
- Fast, effective
- Original and new
- No false negative
- New phase detection

The difficulty is to find which phase is present!!

20 benchmarks
3 other real cases



Baumes, et al. (2008), Cryst Eng Comm.

Baumes et al. (2009) Chem A Eur J.

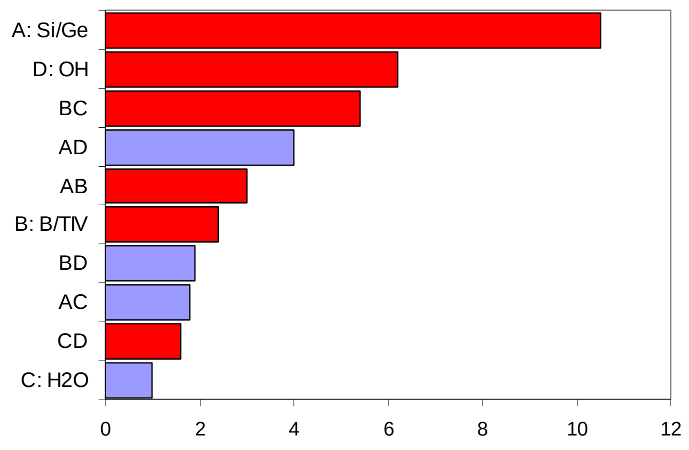
&

Reliability and Robustness

The proposed *Adaptable Time Warping* (ATW) is designed such as 1-norm, DTW, and **more generally all warping and non-warping distances** are particular cases of ATW.

It gives you performances at least as good as the other methods !!!!

Statistical variable influence for getting pure ITQ-33 using Pareto Analysis



FD: Framework density
(T atoms per 1000 Å^3)

Variables:

A= Si/Ge

B= B/(Si+Ge)

C= H_2O/(Si+Ge)

D= OH/(Si+Ge)

Most important statistical variables

- Si/Ge
- OH/(Si+Ge)
- Combination of trivalent and water content

Second Dataset of experiments

			OH = 0,1			
			B/(Si+Ge)			
			0,01	0,02	0,05	0,07
Si/Ge=1	H2O/(Si+Ge)	5				
Si/Ge=2	H2O/(Si+Ge)	2				
		5				

			Al/(Si+Ge)		
			0,01	0,02	0,05
Si/Ge=2	H2O/(Si+Ge)	2			
		5			

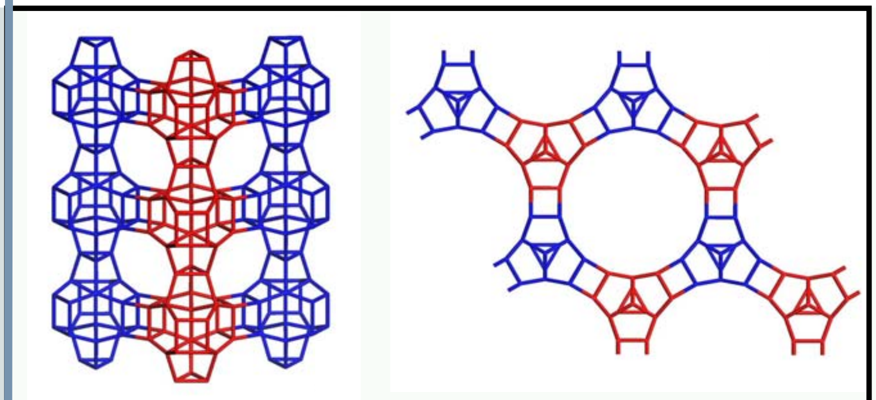
	Amorphous		ITQ-33-ITQ-24
	ITQ-33-amorp		ITQ-33-GeO2

3rd Dataset in F media to obtain pure ITQ-33

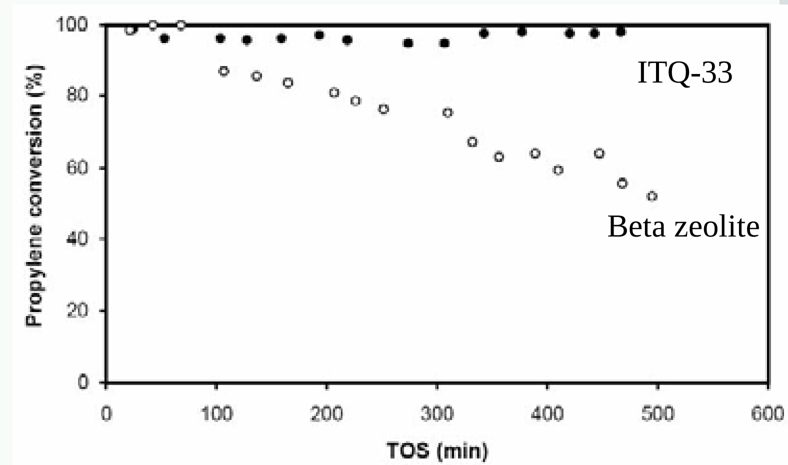
		$H_2O/T^{IV}=3$ $SDA/T^{IV}=0.25$								
		$F^-/T^{IV}=0.1$			$F^-/T^{IV}=0.3$			$F^-/T^{IV}=0.5$		
		$T^{III}/T^{IV}=0$	$B/T^{IV}=0.05$	$Al/T^{IV}=0.05$	$T^{III}/T^{IV}=0$	$B/T^{IV}=0.05$	$Al/T^{IV}=0.05$	$T^{III}/T^{IV}=0$	$B/T^{IV}=0.05$	$Al/T^{IV}=0.05$
Si/Ge	2	ITQ-17	ITQ-17-ITQ-24	IM-10	ITQ-17	ITQ-17-ITQ-24	ITQ-33	ITQ-17	ITQ-17-ITQ-24	ITQ-17
	5	ITQ-17	ITQ-17-ITQ-24	IM-10	ITQ-17	ITQ-24	EU1	ITQ-17	ITQ-24	ITQ-17
	15	ITQ-13	Amorphous	Amorphous	ITQ-13	ITQ-24	EU1	ITQ-17	ITQ-24	EU1
	INF	Amorphous	Amorphous	Amorphous	ITQ-13	Amorphous	EU1	Amorphous	Amorphous	EU1

Legend
Amorphous
ITQ-17
ITQ-24
ITQ-17-ITQ-24
ITQ-13
IM-10
EU1
ITQ-33

36 experiments

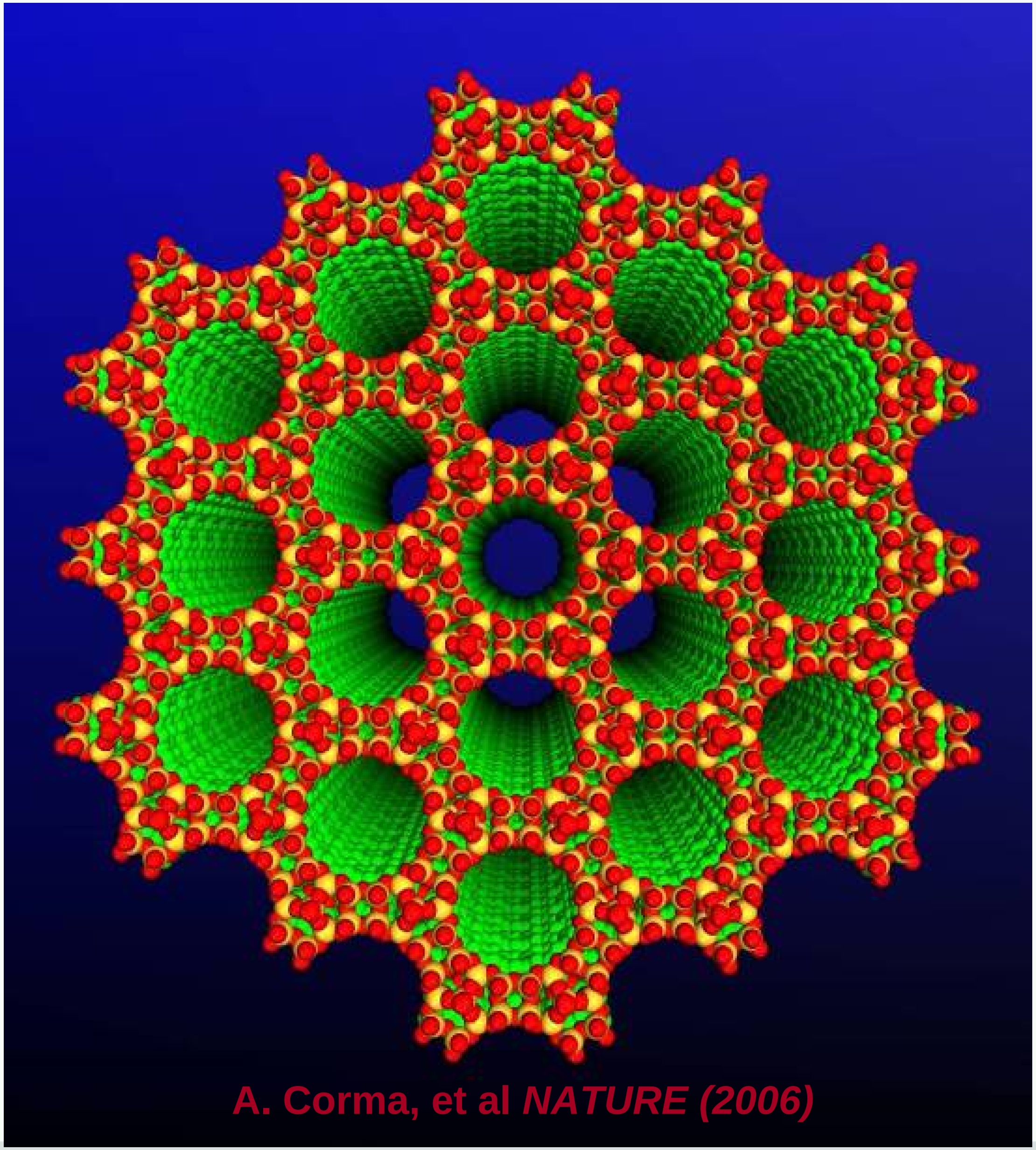


ITQ-33: Interconnected channels of 18- and 10-MR



Alkylation of benzene with propylene

Corma, et al. (2006), *Nature, 443, 842*


A. Corma, et al NATURE (2006)

Thanks!

- Avelino Corma
  Santiago JIMENEZ
  Jose Serra
  Manolo Moliner
  Pedro Serna
  Mª-jose Diaz

From UNIVERSIDAD POLITÉCNICA DE VALENCIA CSIC

Funding

- 6th European Framework Program for R&D **TopCombi**
- **Spanish government**: project MAT 2006-14274-C02-01, grant FPU AP2003-4635
- **European funding Project I3P**

TOPCOMBI