Cinvestav

Random methods to predict novel crystalline structures

Aldo Humberto Romero and Xavier Gonze

CINVESTAV Unidad Querétaro and UCL

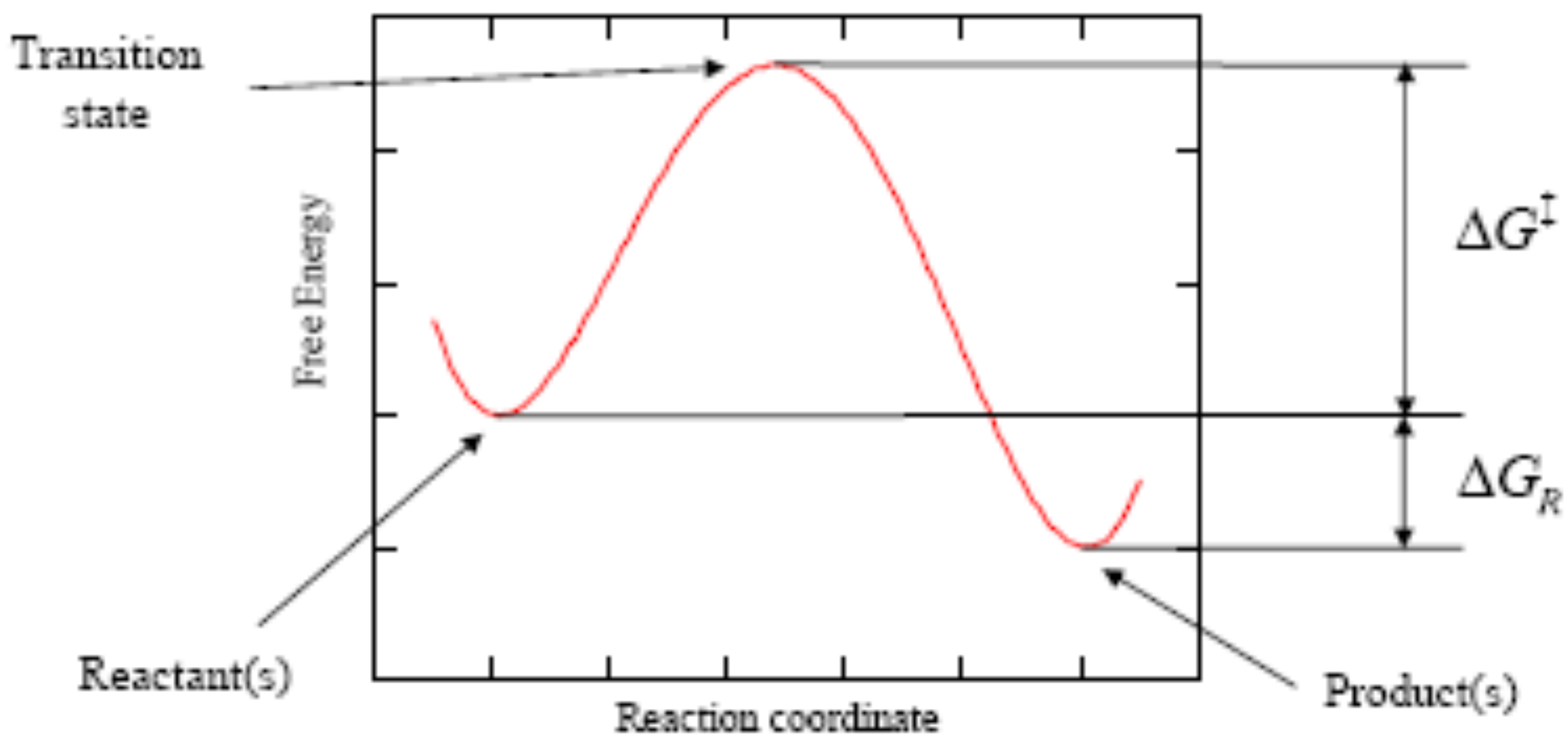
aromero@qro.cinvestav.mx

Today's Menu

- String method
  - ✳ Reminder: basics and implementation
  - ✳ Abinit variables, testing and uses
- Random Approach: Genetic Algorithm
  - ✳ Definition. Rules, fitness function, etc
  - ✳ Implementation in Abinit
- Generalizations for GA
- Conclusions and perspectives

String Method

- Method to Study rare events by finding a path between two given states. To find energy barrier and transition state.
- Given a potential, you can find a minimum energy path
- Processes that are more energetic than entropic



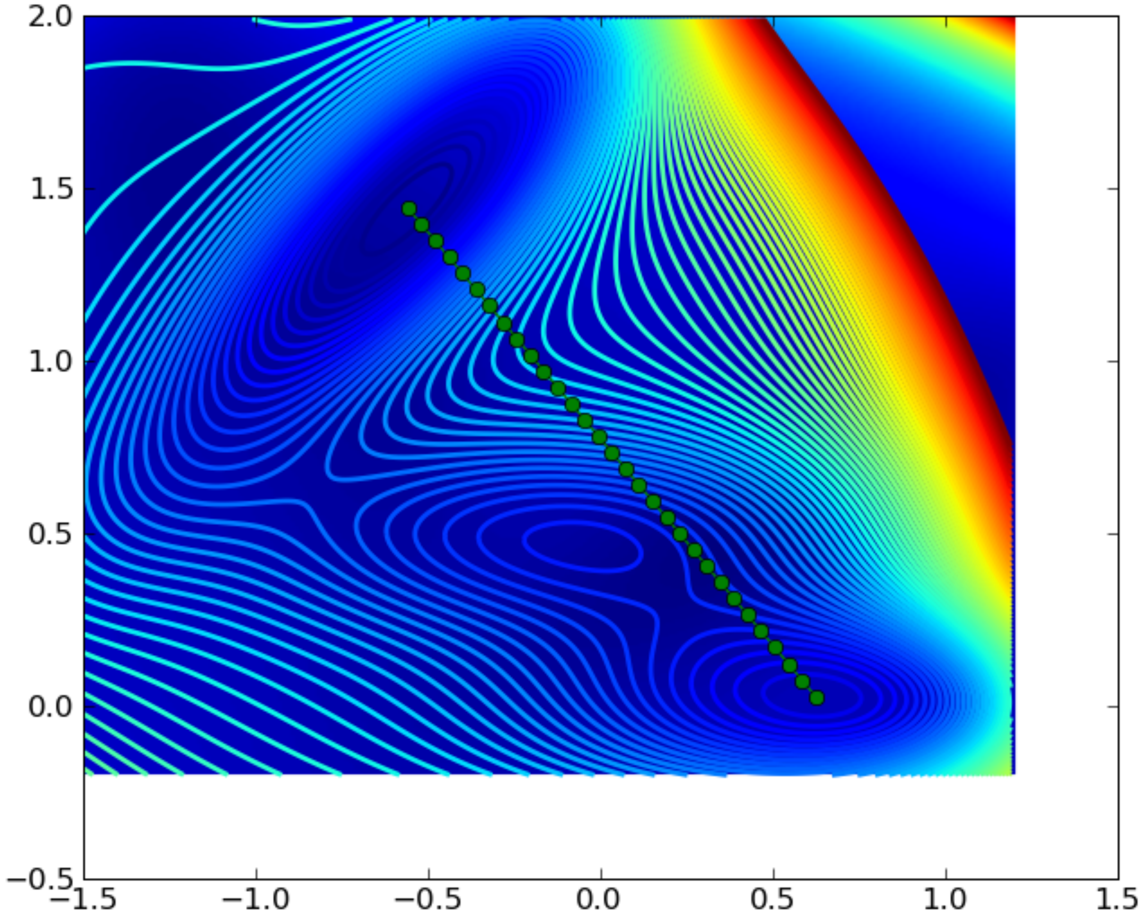
Finding reactions mechanisms

- Minimize a reaction path (a string) by evolving it according to the equation

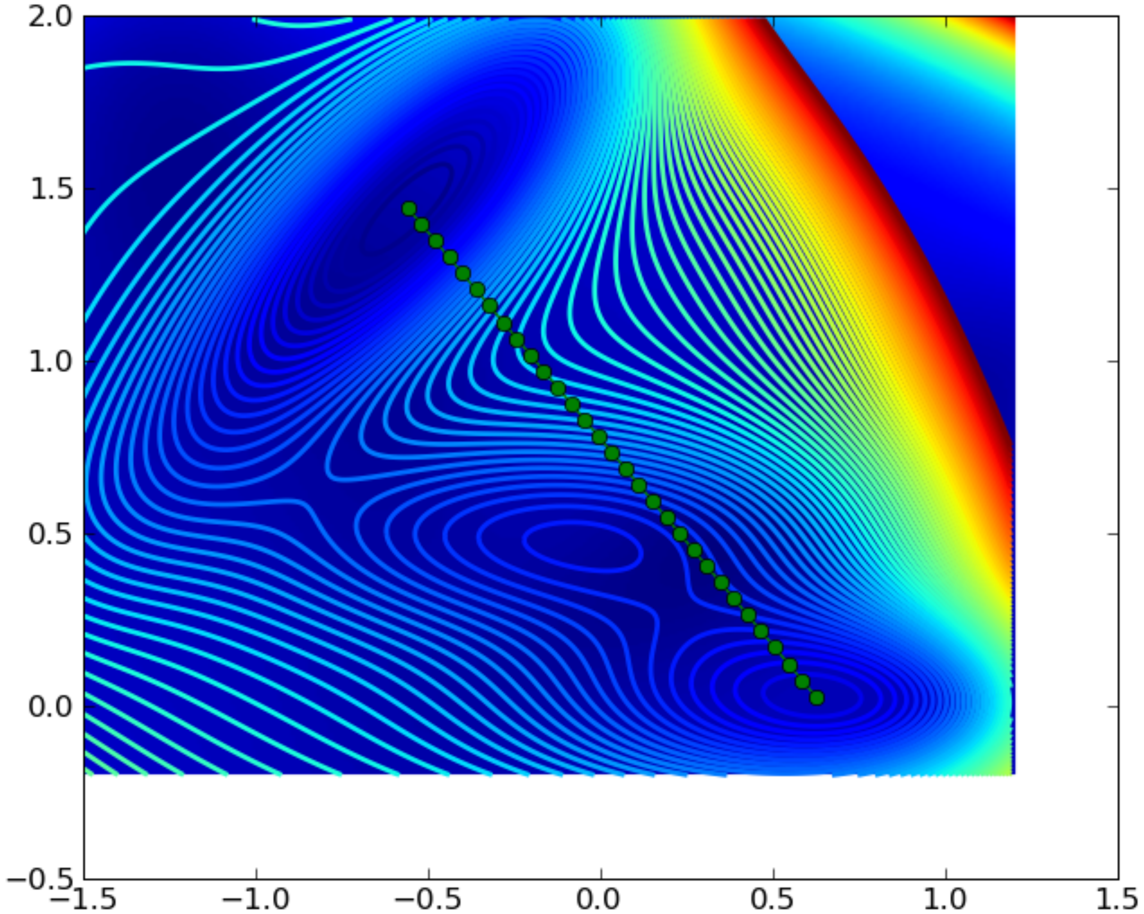
$$\frac{d\varphi_i}{dt} = -\nabla V(\varphi_i)$$

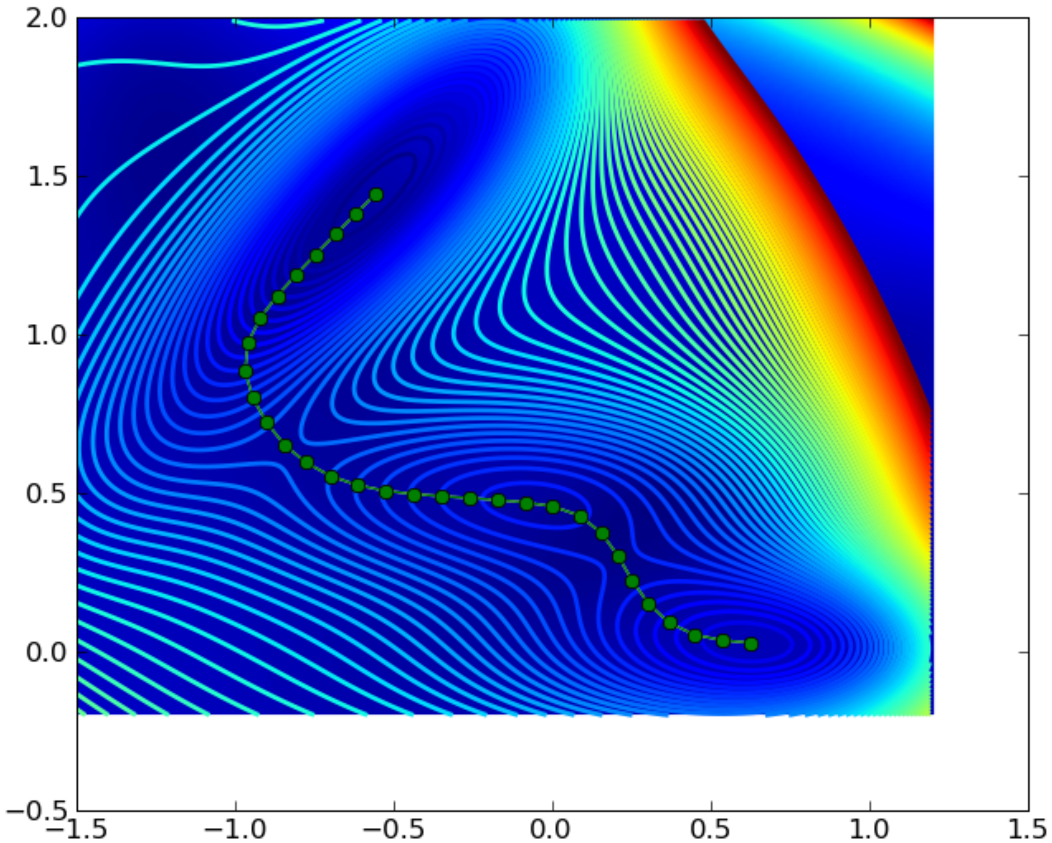
- To improve the resolution and decrease the changes of changing “too much”, the images positions are “reparametrized” along the path, such that Euclidean distance between them is almost constant.
- *NEB is a particular case of SM*

Weinan *et al.* J. Chem. Phys. 126, 164103 (2007)



2.0
1.5
1.0
0.5
0.0
−0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


2.0
1.5
1.0
0.5
0.0
−0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


2.0
1.5
1.0
0.5
0.0
−0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

Abinit instructions

- *ntimimage* 15
- *nimage* 10
- *imgmv* 2
- *chksymbreak* 0
- *tolimg* 1.d-5
- *dynimage* 0 8*1 0
- *xcart* 3*0d0 3*0.766911
- *xcart_lastimg* 3*0d0 2.233089 2*0.766911

Genetic Algorithm: survival of the fitness

- First proposed by Holland in "***Adaption in Natural and Artificial Systems***" (1992).
- Organisms reproduce themselves in offsprings which are similar but not equal, due to randomness (mutation) and sexual reproduction (crossing), which gives characteristics from their parents.
- Finally, the best adapted to the "environment" will survive to every generation, the population will become "better" and "better"
- This idea gives a robust search and optimal algorithm for complicated minima problems.
- A population is like a chrosomose, which is conformed by genes.

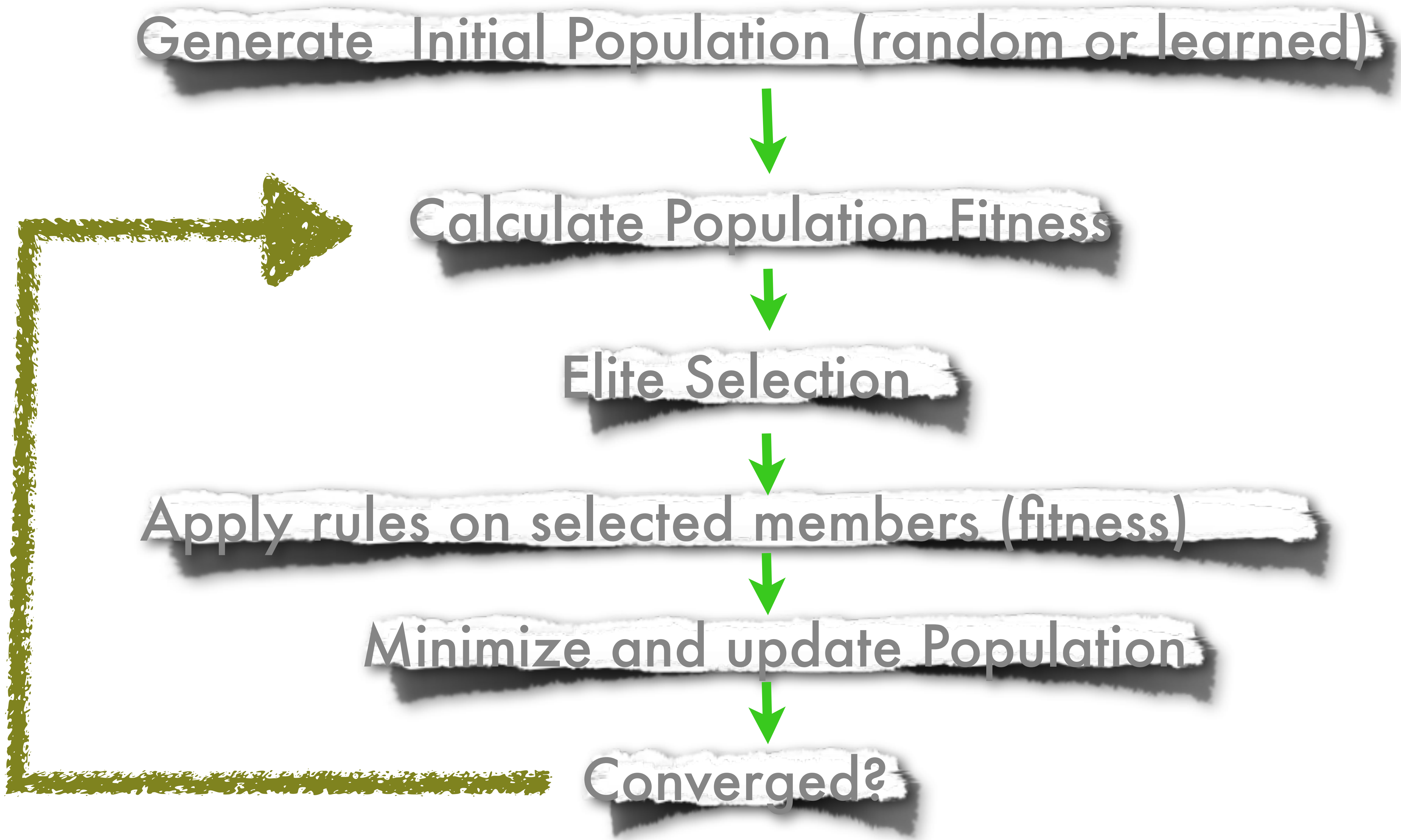
Basic parts

- Initial configurations. Usually random with no bias. Some input from “nice” configurations is welcome.
- Representation: binary or real
- Fitness function: function to be minimized and assigned to every population member.
- Selection. Stochastic based on Fitness.
- Rules: crossover, mutation, slicing, etc
- Stagnation: problems biasing the search
- Population number: Usually fixed but it can also vary.

Deaven and Ho PRL 75:288 (1995)

Genetic Algorithm: survival of the fitness

- A basic flow diagram is as follows:



Representation, initialization and fitness

- Every member of the population is represented in strings. Positions of atoms form a 3N vector.
- Every member is randomly initialized but also good initial configurations can help. (Lego pieces!).
- How "fit" is the criteria used for reproduction preventing early good structures to dominate.
- We use total energy (ideal enthalpy).
- Several Weights

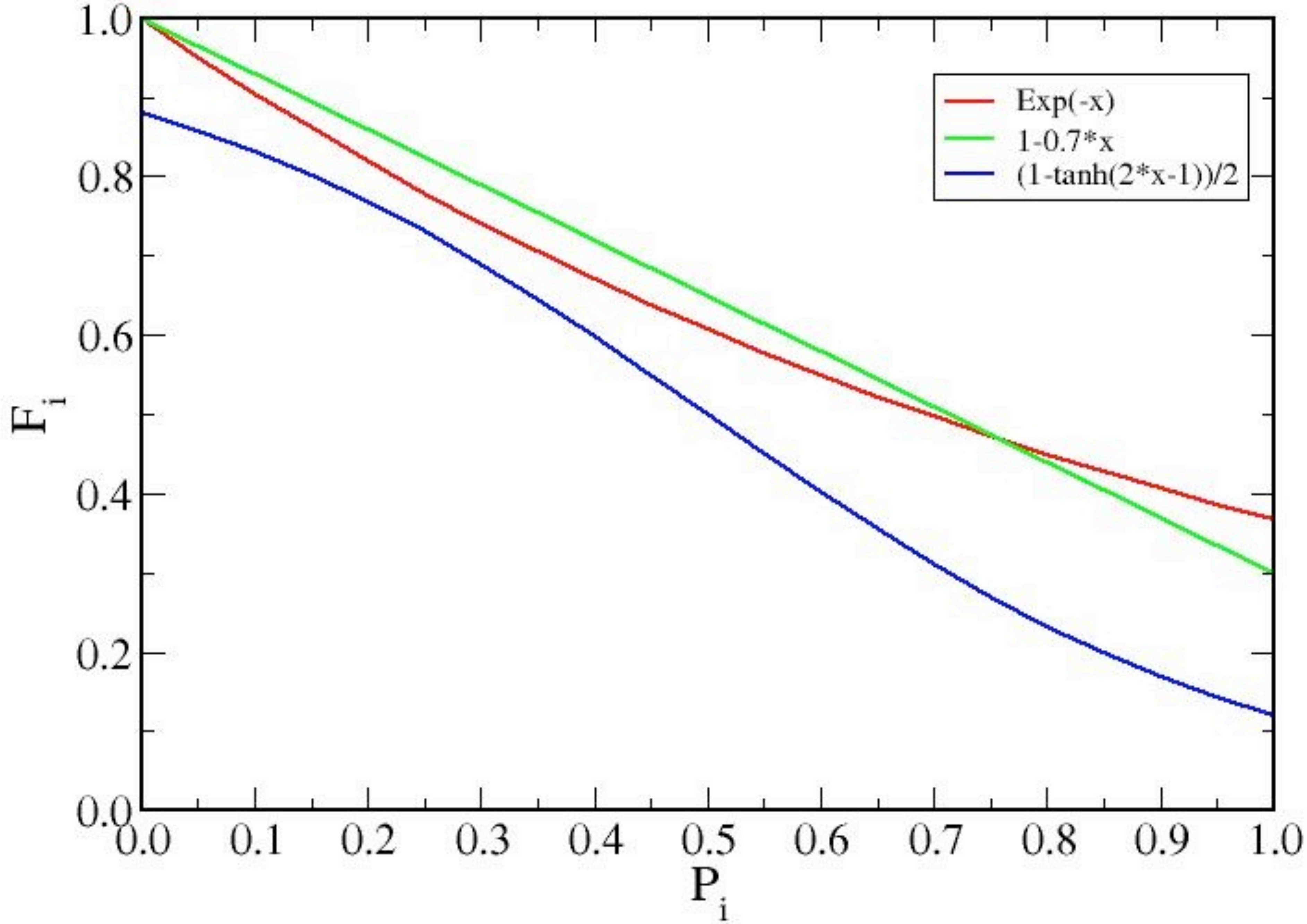
$$P_i = \frac{E_i - E_{Min}}{E_{Max} - E_{Min}}$$

$$F_i = Exp(-3P_i) \quad F_i = 1 - 0.7 * P_i \quad F_i = \frac{1 - tanh(2P_i - 1)}{2}$$

R.L. Johnston, Dalton Trans. page 4193 (2003)

Representation, initialization and fitness

- Every n ... s of atoms fo
- Every n configu
- How "fi ... d structur
- We use
- Several



$F_i = Ex$

R.L. Johnston, Dalton Trans. page 4193 (2003)

Selection

- Members of the population are selected for reproduction or mutation according to its fitness
- Prevent trapping of local minimum (high selectivity).
- Use of Elitism. The “super-population”.
- Use probability for selection $\phi_i = \frac{F_i}{\sum F_i}$ with a cumulative probability between i*th* and (i+1)*th*
- There are other selection methods like the tournament (two members are selected and the one with the lowest fitness is chosen)

Rules of GA

Two point crossover

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	4	9	8	7	6	5	10	11	12	13	14	15	16	17	18	19	20

Random Crossing

Parents

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
a	b	c	d	f	g	h	i	j	k	l	m	n	o	p	q	e	s	t	u

P=0.5

Children

a	2	3	d	5	6	h	8	j	10	11	m	13	14	p	16	17	s	19	20
1	b	c	4	f	g	7	i	9	k	l	12	n	o	15	q	e	s	t	u

N-point crossover

Parents

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
a	b	c	d	f	g	h	i	j	k	l	m	n	o	p	q	e	s	t	u

Children

1	2	3	4	5	6	7	8	9	10	11	12	n	o	p	q	e	s	t	u
a	b	c	d	f	g	h	i	j	k	l	m	13	14	15	16	17	18	19	20

Mutation (coordinates or cell parameters, volume)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	9	7	8	9	10	14	12	13	12	15	16	17	18	16	20

Arithmetic Average

Parents

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40

↓

Child

11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Geometric Average $\sqrt{a_i \cdot b_i}$

Parents

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40

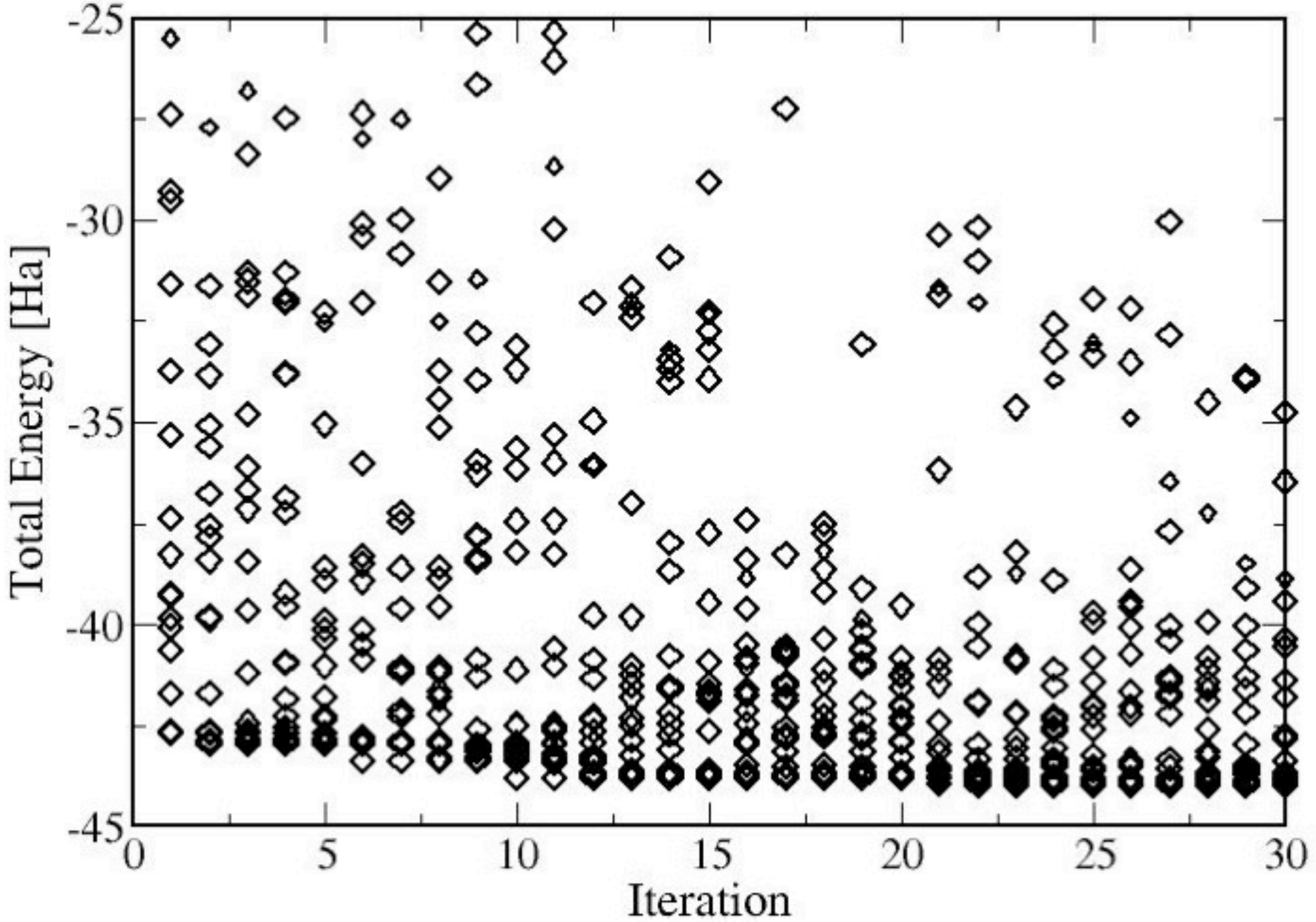
↓

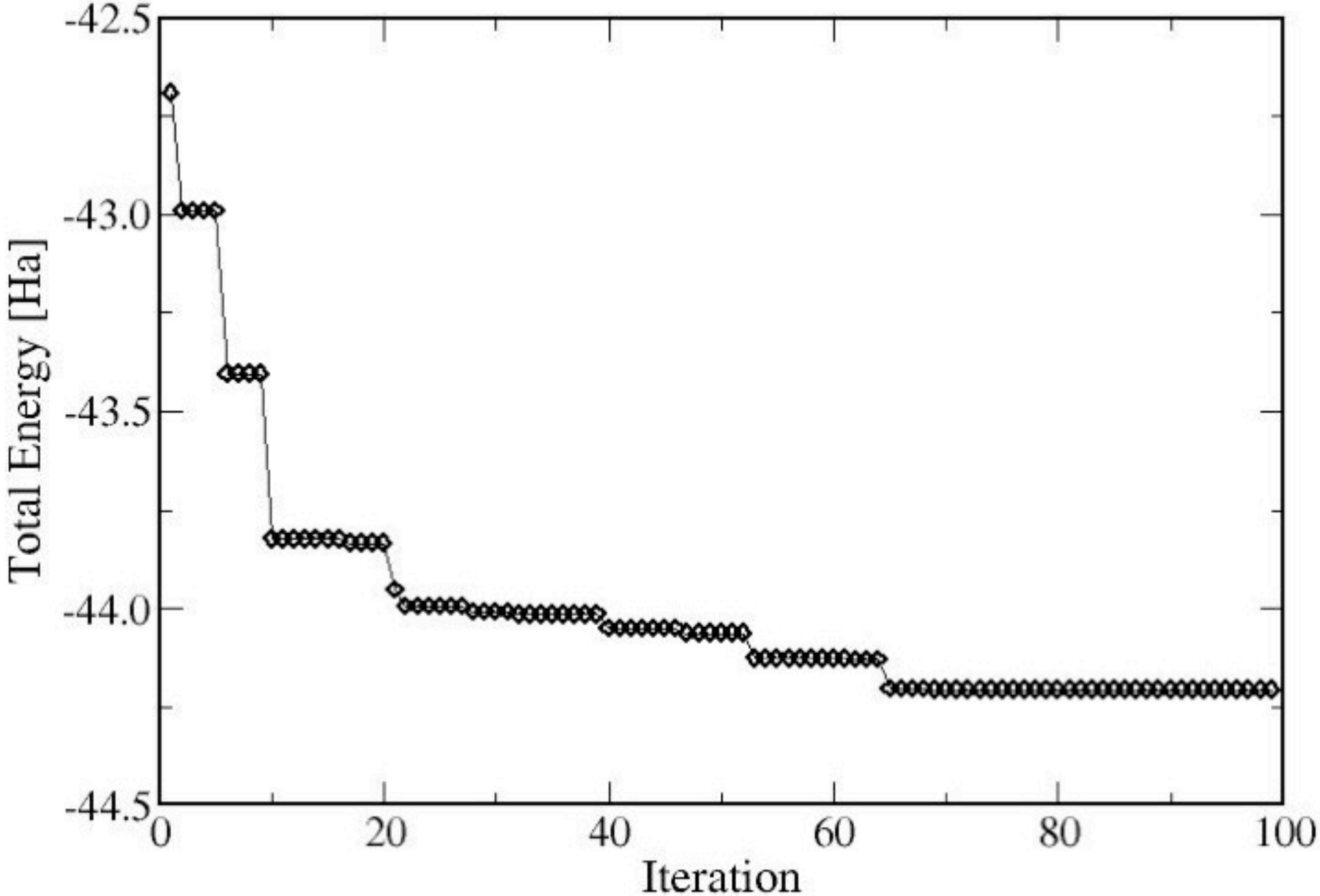
4.58	6.63	8.31	9.79	11.18	12.49	13.75	14.96	16.15	17.32	18.47	19.60	20.71	21.81	22.91	24.00	25.08	26.15	27.22	28.28

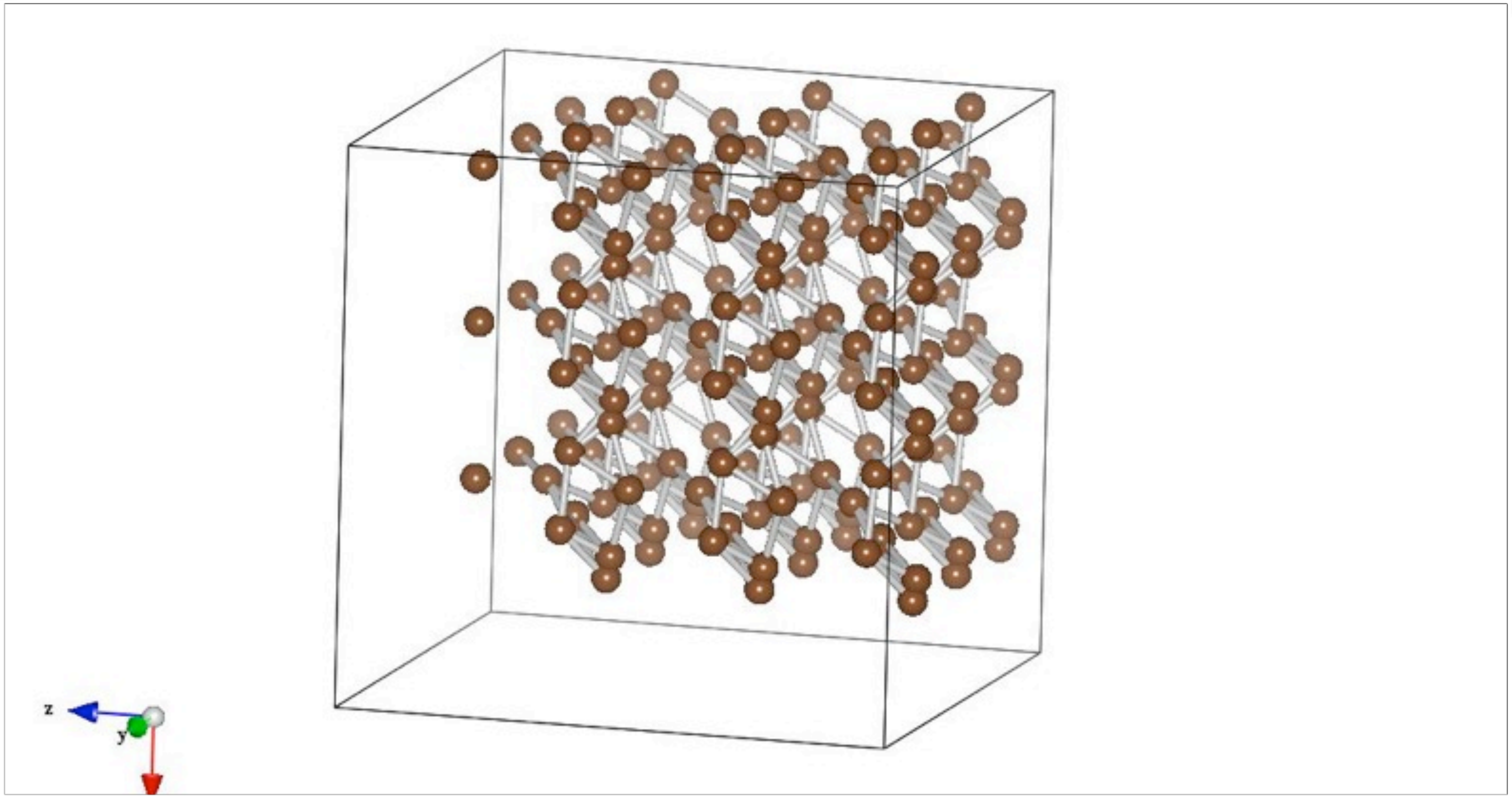
Additional issues: changes and constraints

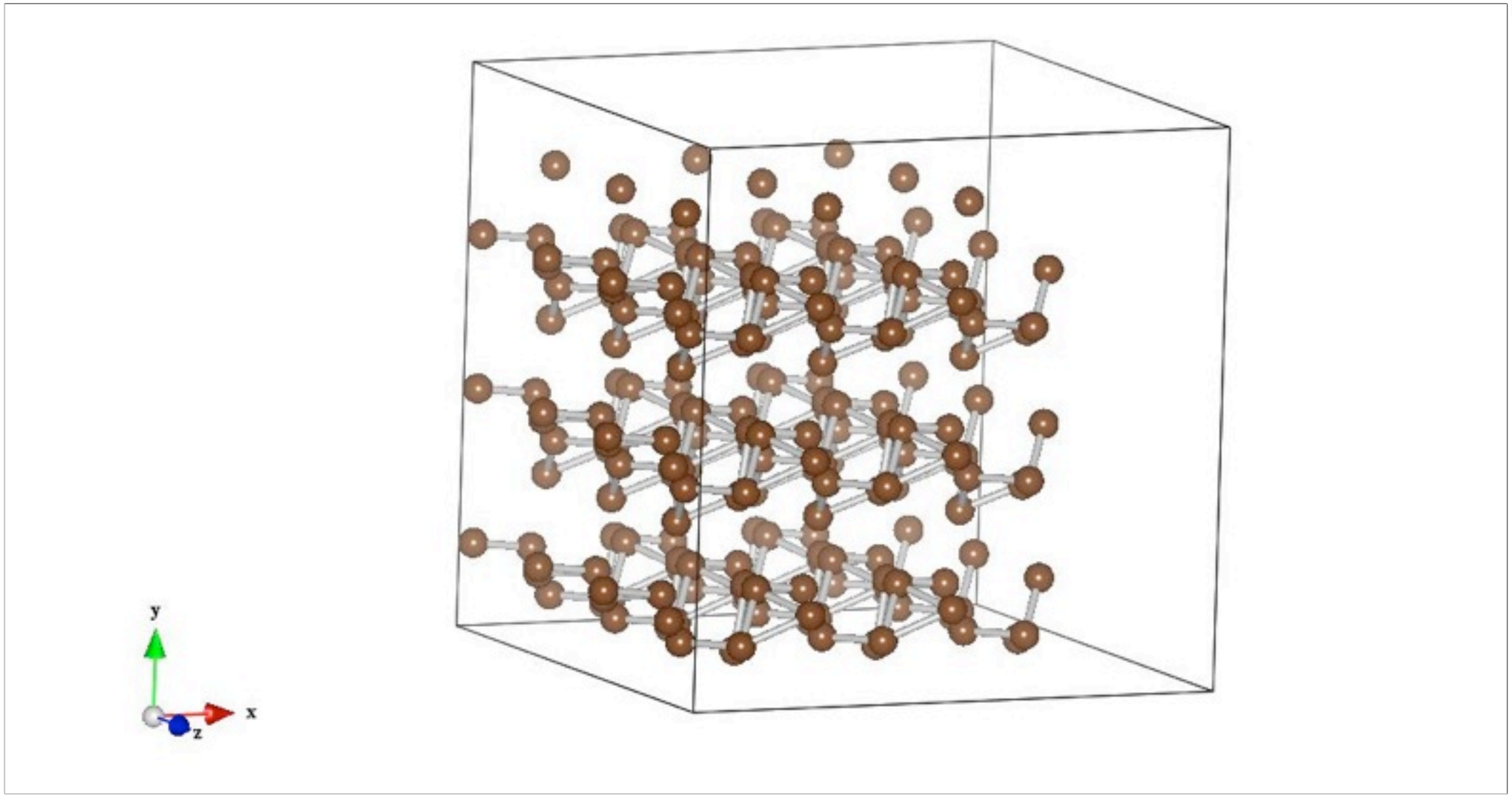
- Random change in cell volume
- Check atomic distances after a new child has been created (close atoms less than a given distance or same position)
- Check configurational distance.
  - ✓ Find distance from origin
  - ✓ rearrange distances from large to small (d1< d2< d3< ..<dn)
  - ✓ Find distance between two configurations

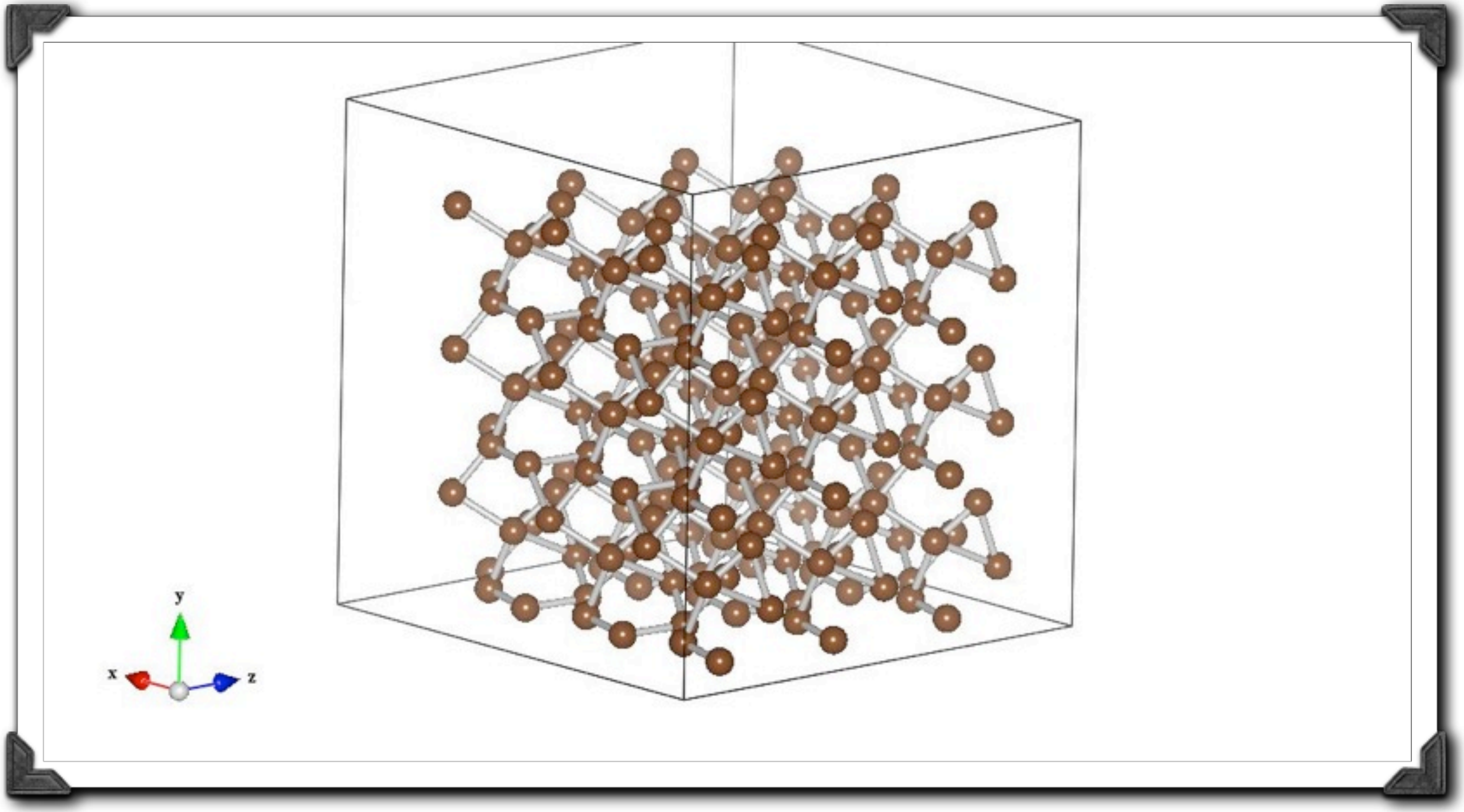
$$\sum_{i=1}^{N} |d_i^{\alpha} - d_i^{\beta}|$$


Total Energy [Ha]
-25
-30
-35
-40
-45
0
5
10
15
20
25
30
Iteration


-42.5
-43.0
-43.5
-44.0
-44.5
Total Energy [Ha]
0
20
40
60
80
100
Iteration

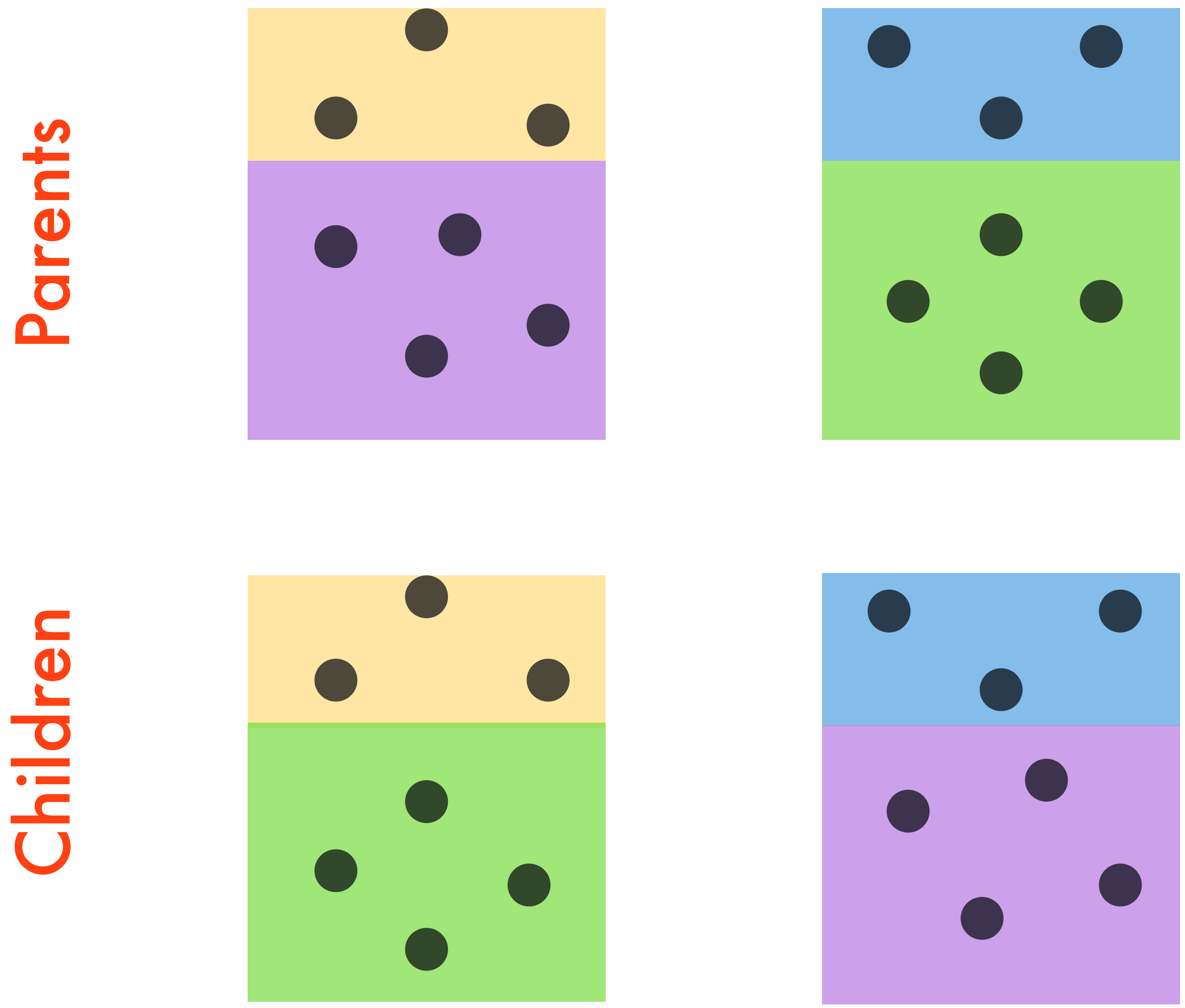

z
y

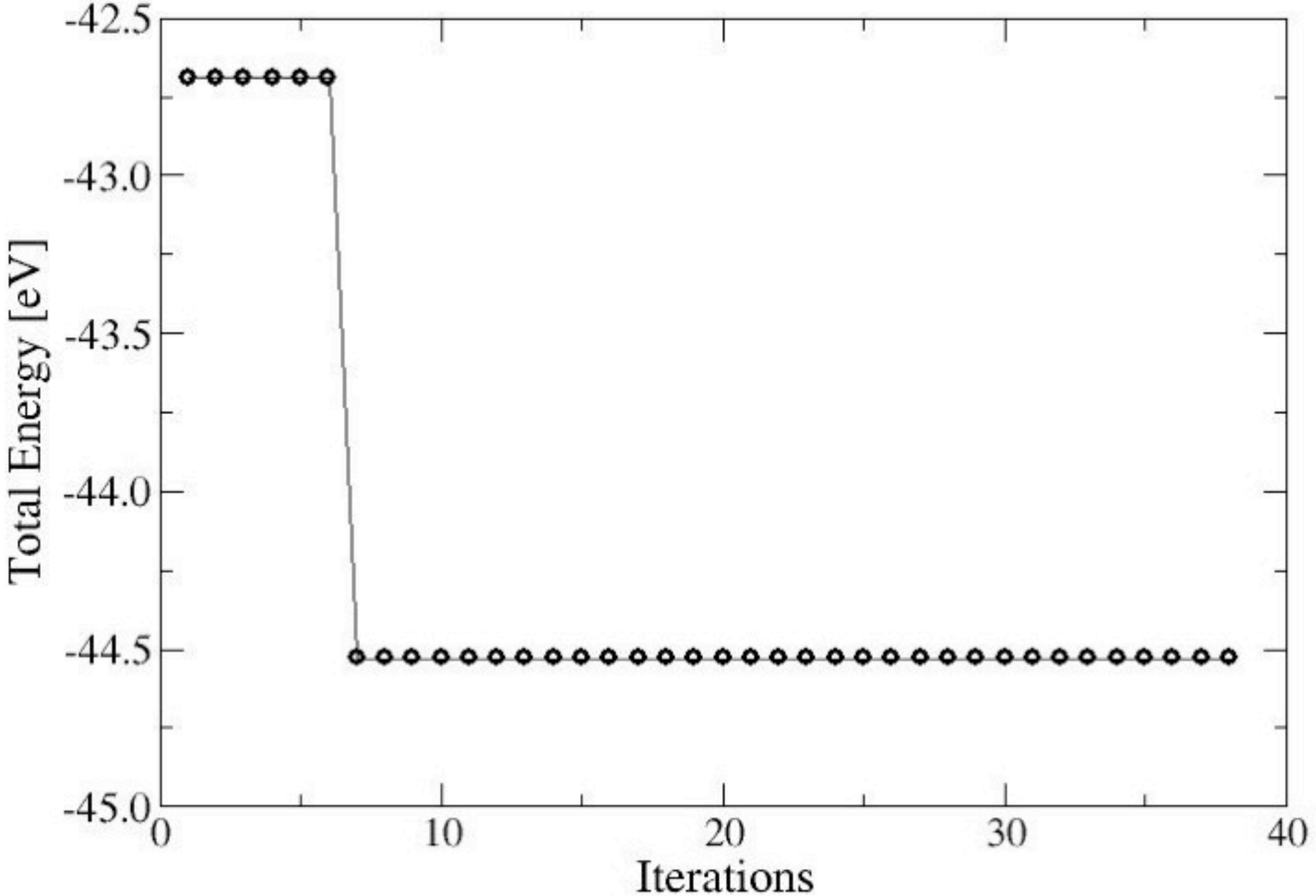

y
x
z

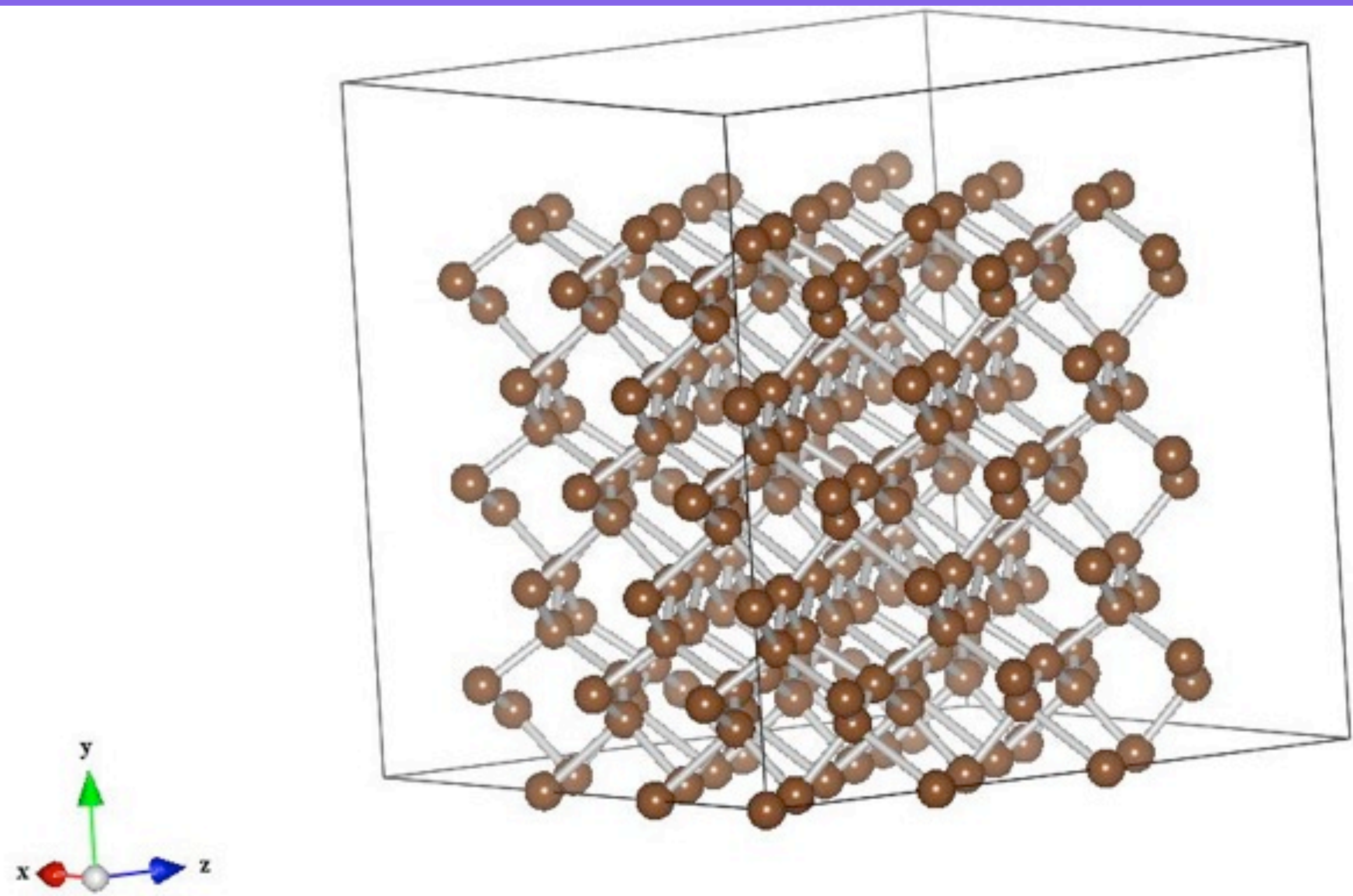

y
x
z

Still.. not enough

- Slicing method (in the unitary cell)




-42.5
-43.0
-43.5
-44.0
-44.5
-45.0
0
10
20
30
40
Total Energy [eV]
Iterations


y
x
z

Input example GA

ntimimage 100
optcell 0
dilatmx 1.3
nimage 20
chksymbreak 0
random_atpos 2
imgmov 4

acell 3*6.0d0
natom 8
nstep 100
ntypat 1
typat 8*1
znucl 6
kptopt 1
ngkpt 3 3 3

Future and status (Marc's bar!)

- Include cell deformation. (optcell 2+ dilatmx) Almost!
- Use the Enthalpy. Almost!
- Parallelization. Almost!.
- Lego approach. Typical structures. Maybe from g(r)
- Generalization for several species (some rules needs to be modified)
- Include different fitness functions
- Etching technique. N is not constant. (Landman)
- Improve searching: Bee algorithm or graph theory