Simulated annealing, effective but inefficient?

A case study for the 3D136 instance of the HP model of protein folding

Arnulf Möbius and Florian Günther

IFW Dresden, TU Chemnitz,
TU Dresden

CompPhys17, Leipzig, Germany, 30.11.2017

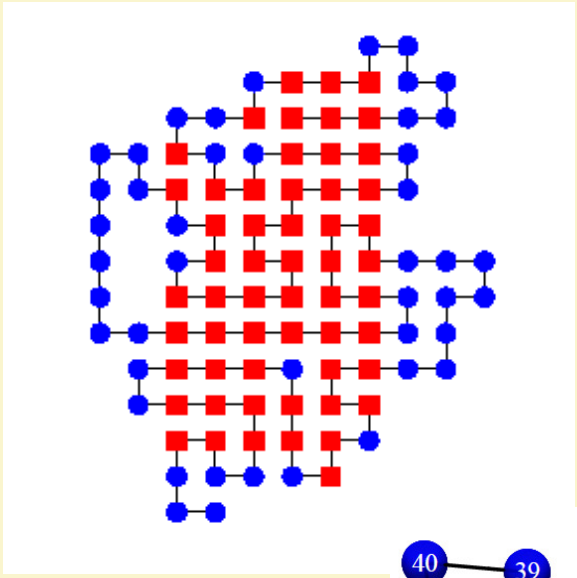
1. Background and motivation

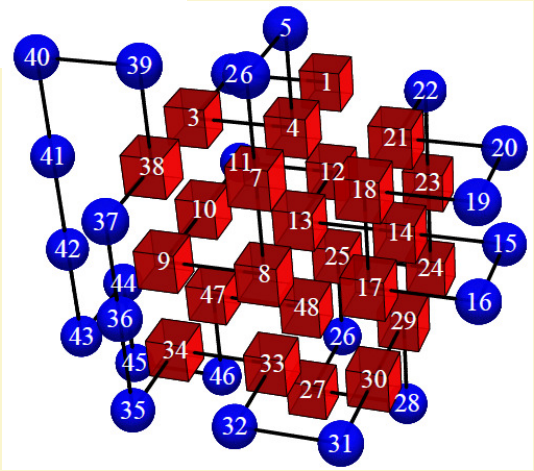
- *Structure optimization for models of protein folding via "partial distortion"-quench cycles*,
  F. Günther, A. Möbius, M. Schreiber, Telluride, 2013
- *Strukturoptimierung von Proteinmodellen mittels Zyklen aus lokaler Störung und nachfolgendem Quench,*
  Florian Günther, Masterthesis at TU Chemnitz, 2013,
- *Structure optimisation by thermal cycling for the hydrophobic-polar lattice model of protein folding,*
  F. Günther, A. Möbius, M. Schreiber,
  Eur. Phys. J. Special Topics **226** (2017) 639

- *Optimization by thermal cycling*, AM et al., PRL **79** (1997) 4297

Protein folding models studied by Florian

HP model





■ hydrophobic

● polar

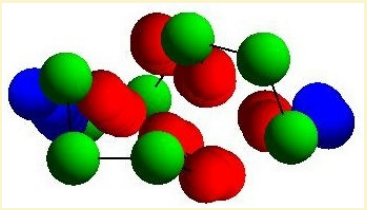
K.F. Lau, K. A. Dill,
Macromolecules **22** (1989) 3986

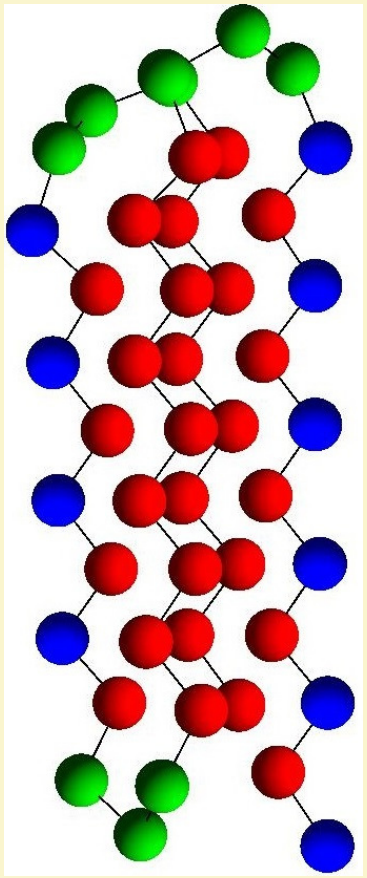
BLN model

● hydrophilic (L)

● neutral (N)

● hydrophobic (B)



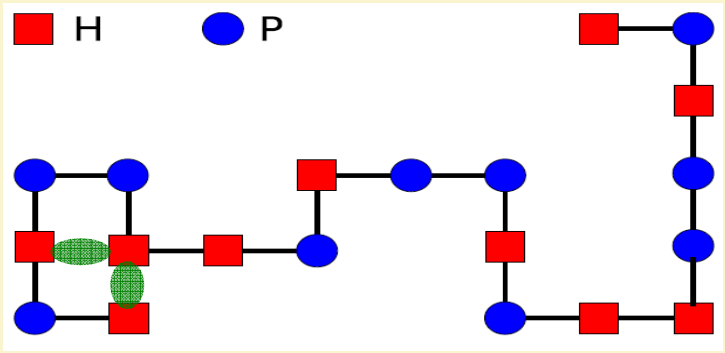


J.D Honeycutt, D. Thirumalai,
Biopolymers **32** (1992) 695

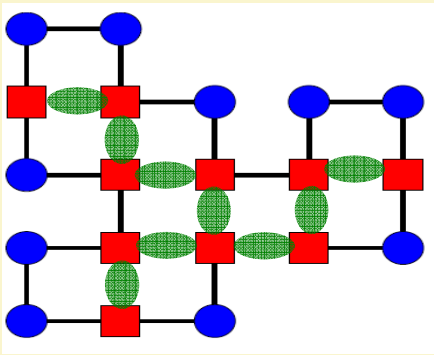
From now on focus on the hydrophobic-polar (HP) model[1]

- Only two kinds of monomers: hydrophobic (H) and polar (P)
- self-avoiding chain on regular lattice
- Energy = number of adjacent H-nodes, which are not adjacent in chain, times -1
- exact approach for not too large available[2]

a)



b)



Example: chain of twenty nodes which has 83,779,155 configurations:

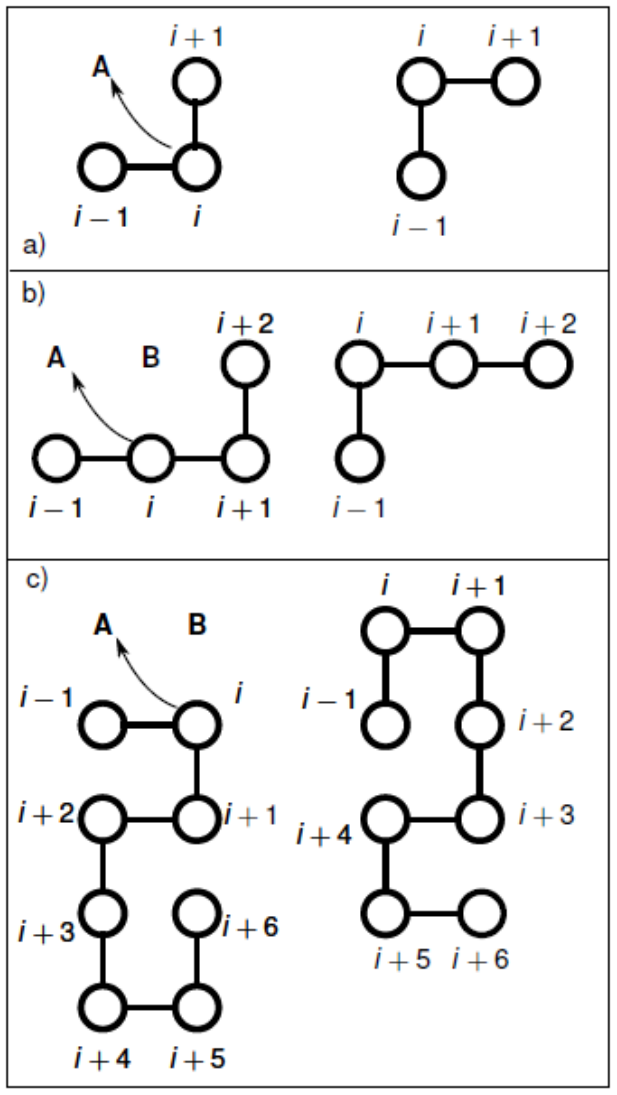
a) one of 12,473,446 configurations with energy -2;

b) one of 4 configurations with energy -9 (global minimum)

[1] K.F. Lau, K. A. Dill, Macromolecules **22** (1989) 3986

[2] R. Backofen , S. Will, Proc. of the 7th International Conference on Principle and Practice of Constraint Programming **2239** (2001) 494

Basic for our simulations: the pull move class



a) Start: shift node *i* to free location **A.**
If *i*+1 adjacent to **A**, move complete.

b) Otherwise: shift *i*+1 to **B**.
If *i*+2 is adjacent to **B**: move complete.

c) Otherwise: shift *i*+2 to position
previously held by *i*,
and so on until valid configuration
is reached.

N. Lesh, M. Mitzenmacher, S. Whitesides, Proceedings of the 7th Annual International Conference on Research in Computational Molecular Biology (2003) 188.

What had we reached?

(1) Best-of-N idea applied to simulated annealing enables simple and effective parallelisation.

(2) Thermal cycling with appropriately chosen ensemble size is clearly faster than simulated annealing.

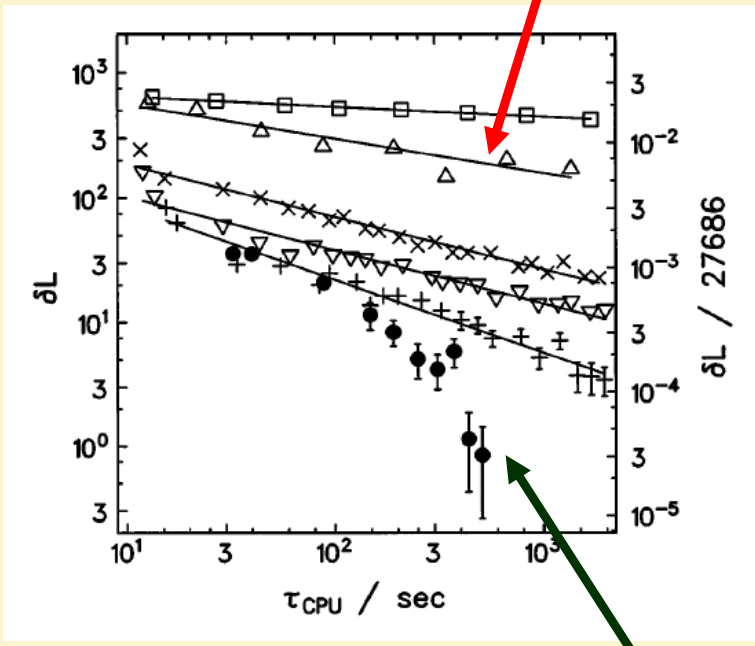
Name	TC, $n_{\mathrm{ens}} = 1$	TC, $n_{\mathrm{ens}} = 10$	TC, $n_{\mathrm{ens}} = 100$	SA	q_{acc}
2D64	0.9 (50%)	0.5 (50%)	1.4 (51%)	3.2 (53%)	6
2D85	10.7 (52%)	5.5 (51%)	6.6 (56%)	88.5 (55%)	16
2D100a	297 (54%)	122 (50%)	52.7 (54%)	226 (52%)	4
2D100b	>1200	169 (52%)	79.5 (50%)	1046 (58%)	13
3D48	4.8 (53%)	4.7 (52%)	7.9 (51%)	14.3 (55%)	3
3D58	561 (58%)	128 (58%)	64.9 (53%)	467 (50%)	7
3D64	>1500	354 (50%)	80.2 (56%)	1255 (61%)	16

(3) In these cases, ensemble thermal cycling seems to be superior even to the state-of-the-art competitor, the Wang-Landau approach.

=> Aims and challenges for the subsequent work

(A) Understanding, why advantage of thermal cycling (TC) over simulated annealing (SA) is far smaller in this case than for travelling salesman problem and for Coulomb glass

(B) Ground states for HP chains with more than 100 sites



TSP instance att532

Thus, to be done:

(a) Code 2.0 = improving technicalities

(b) Incorporation of second move class

(c) Utilization of branch-and-bound ideas

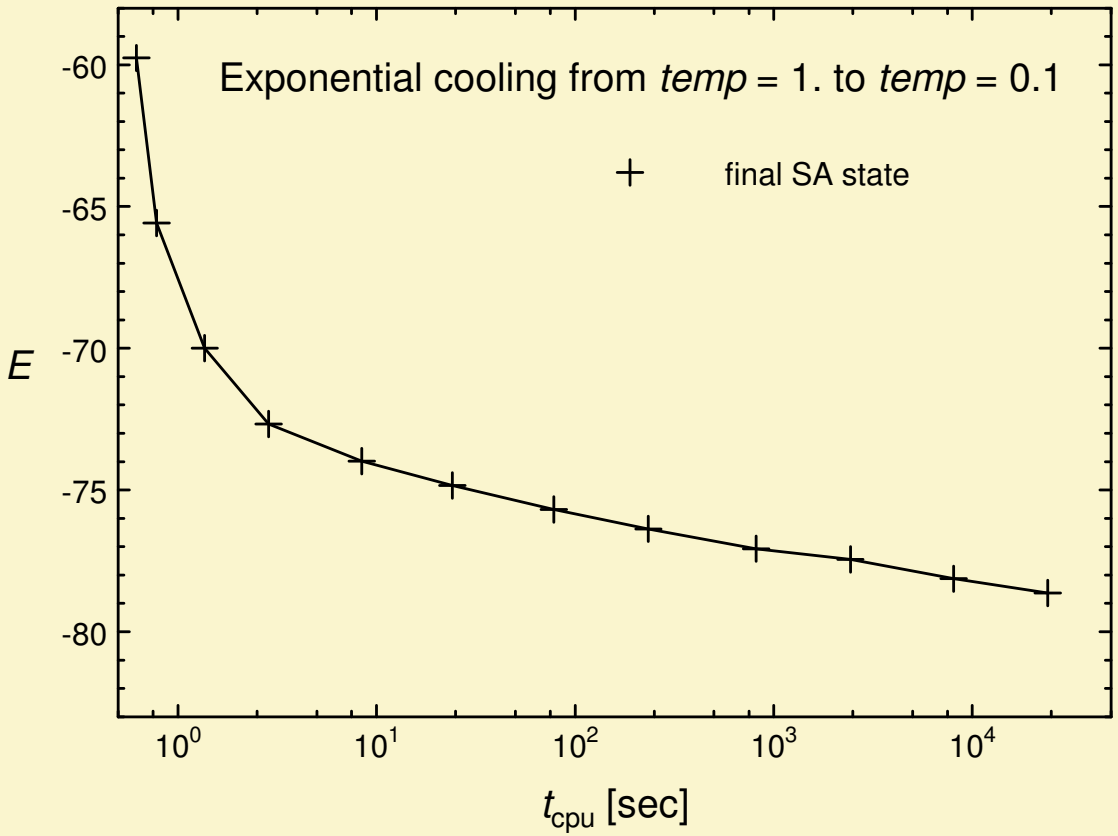
2. Test case: 3D136

HP instance under consideration here:

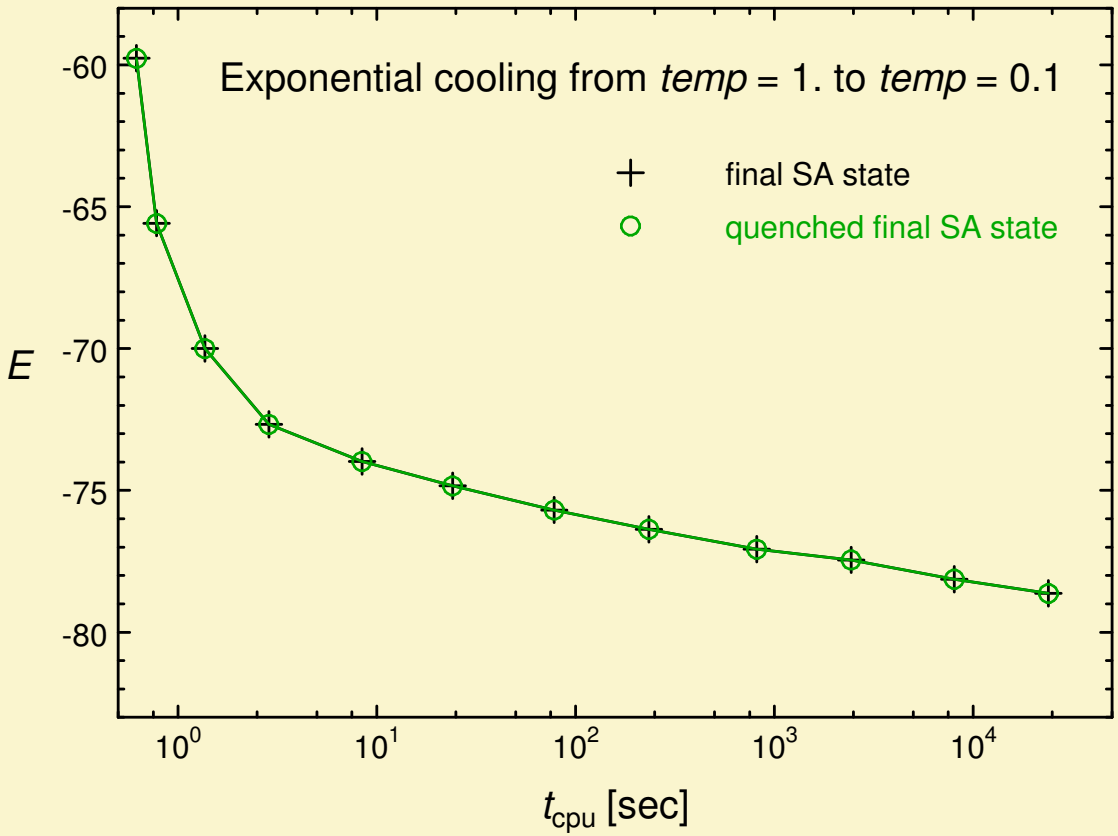
HPPPPPHPPPPHPHHPHHPPPPHPHHHPPPPHPHPHHHHPPPPP
PPPPPPHPPHPPPHPHHPPPHHPPHPPHPHPHPPPPPPPPHPPP
HHHHHHPPPHHPPHHHPPPHHPHHHHHPPPPPPPPPHPPPPHPH
PPPP

Really demanding for stochastic optimisation: Wüst and Landau needed more than 10 days to find a state with E = -83, the ground state energy reported by R. Backofen. His group had found it within 15 min by an exact approach! However, this state is highly degenerated.

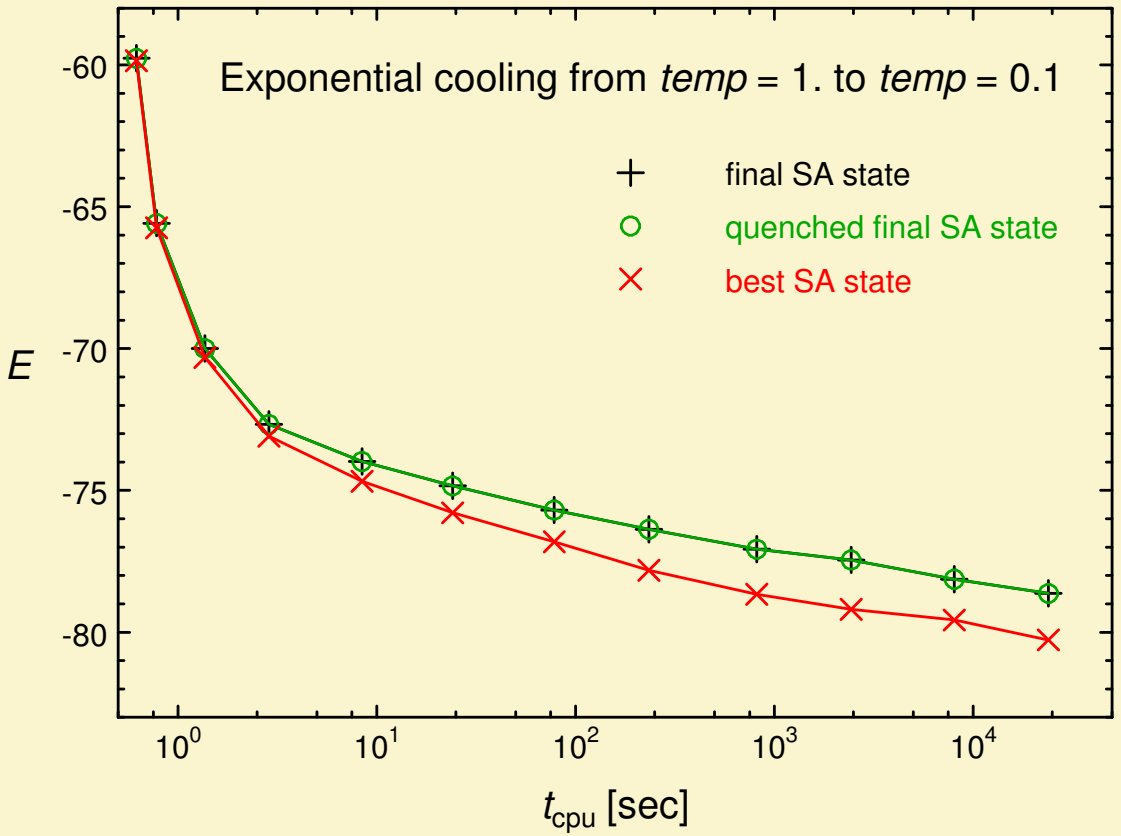
3. Some “surprises” in simulated annealing experiments



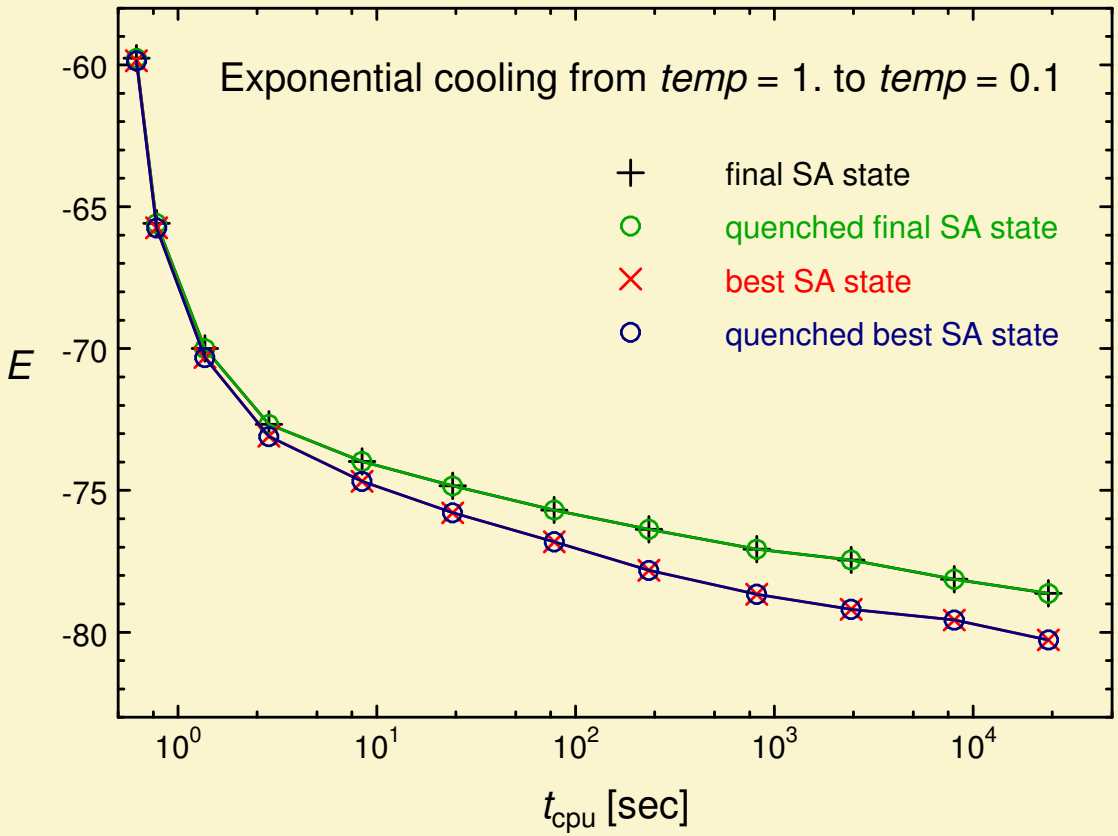
3. Some “surprises” in simulated annealing experiments



3. Some “surprises” in simulated annealing experiments

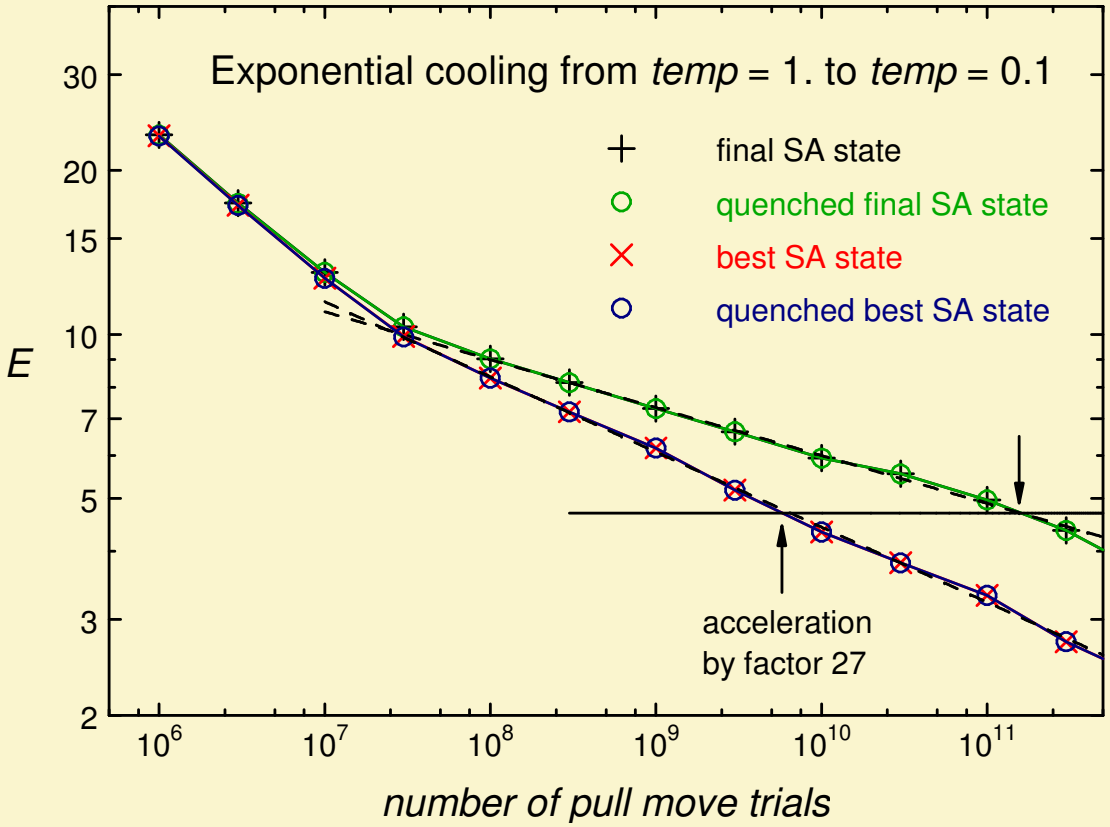


3. Some “surprises” in simulated annealing experiments



- Florian’s focus to the best state hit in the turn of SA rather than to the final one seems very effective, but demands only a few additional lines of code!
- For the parameters chosen, the subsequent quench has no effect.

Logarithmic presentation yields better insight

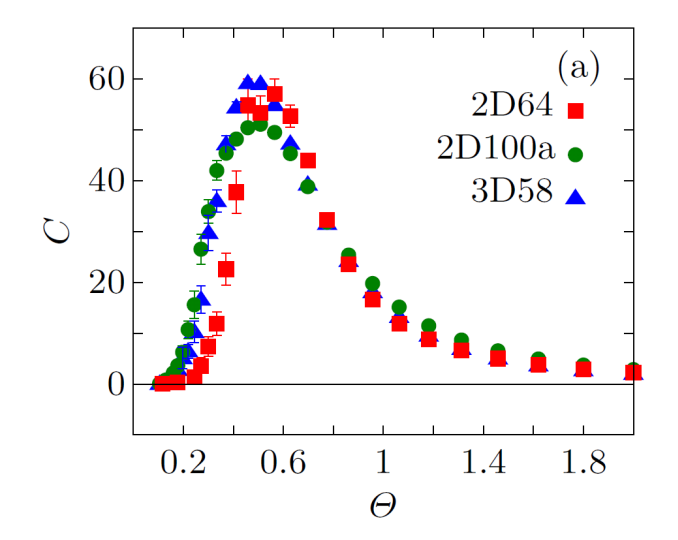


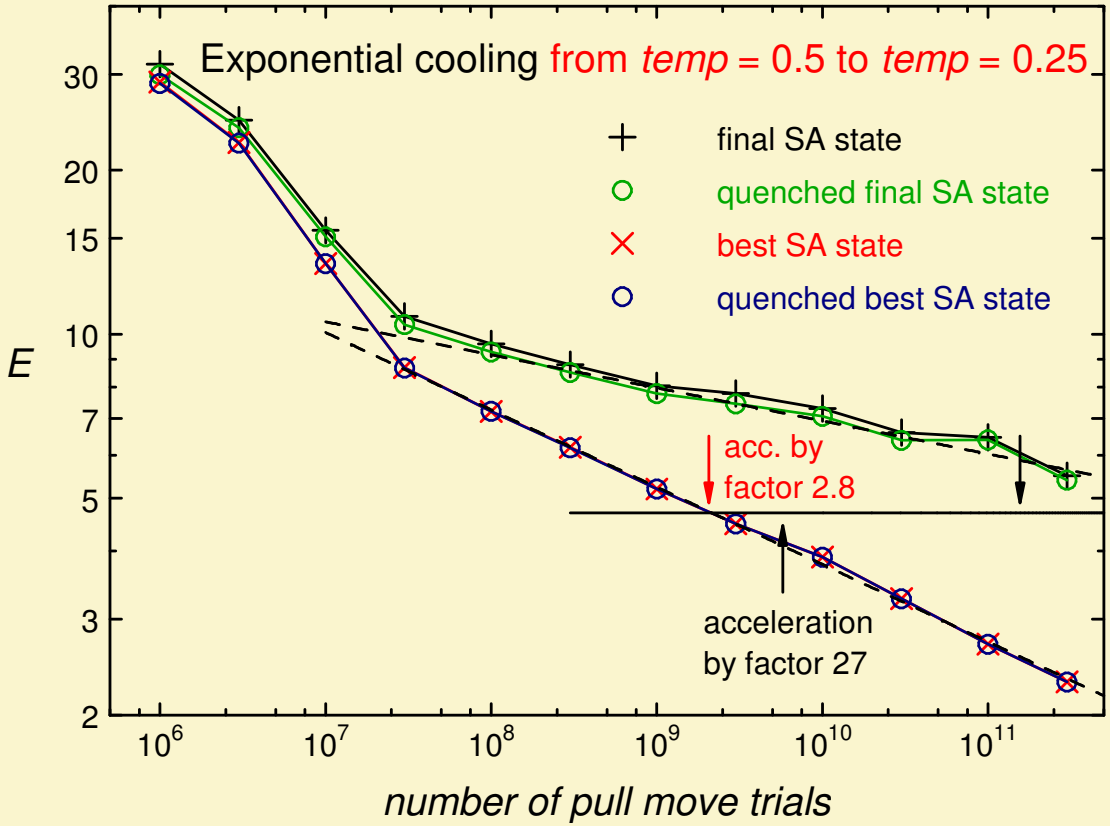
Convergence is not only better, but performance relation is even **steeper**

Here, difference to ground state energy is given, E_0 = -83 from R. Backofen,

www.ima.umn.edu/materials/2007-2008/W1.14-18.08/4814/HP-benchmark.pdf

Acceleration by modifying the temperature interval bounds

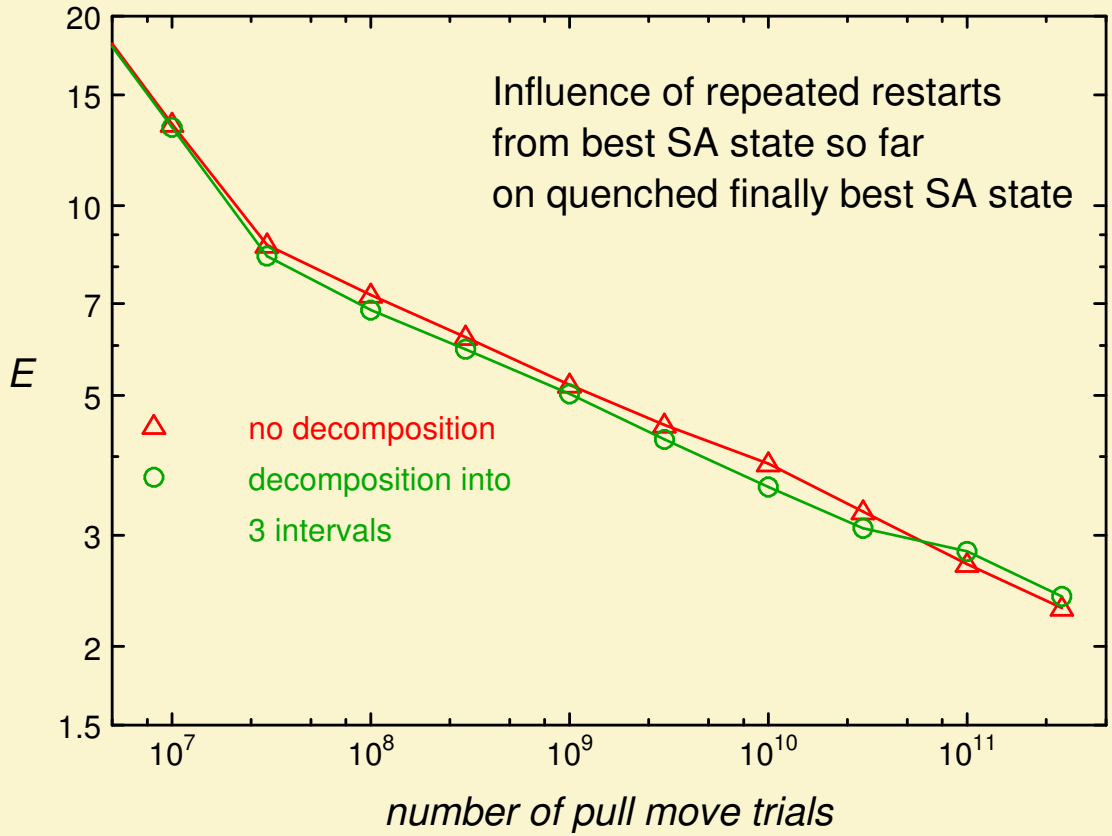




Note: Optimal choice of bounds depends on the aim!

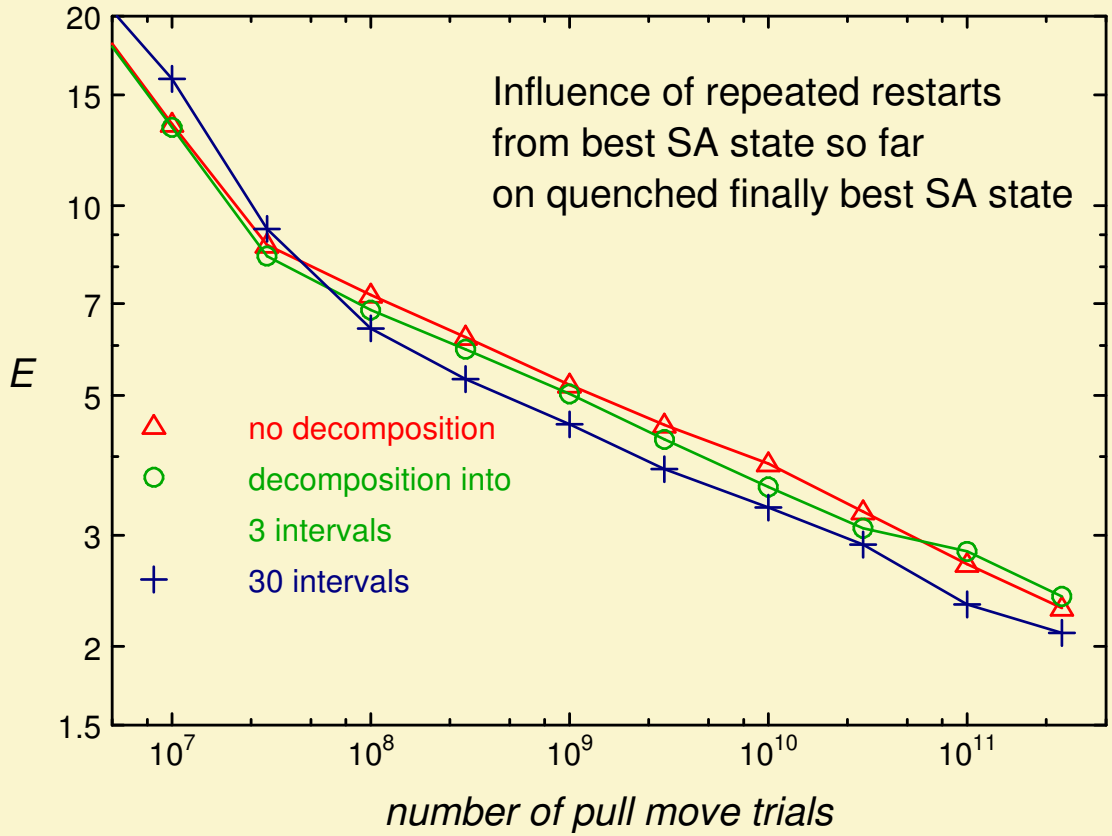
Taking non-equilibrium characteristics of SA better into account

The success of focusing on the best state so far rather than on the final one is a non-equilibrium feature. Already in early stages of the cooling, making use of previous experience could hold the system on the “right track”.



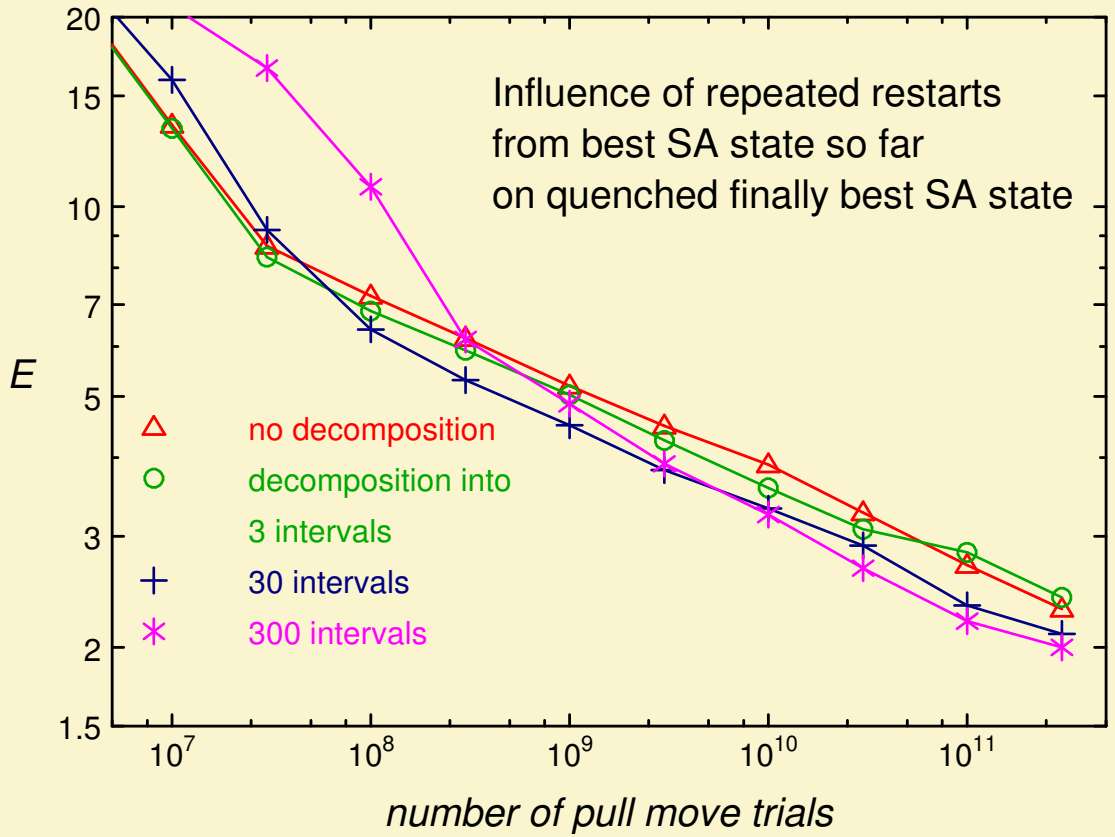
Taking non-equilibrium characteristics of SA better into account

The success of focusing on the best state so far rather than on the final one is a non-equilibrium feature. Already in early stages of the cooling, making use of previous experience could hold the system on the “right” track.

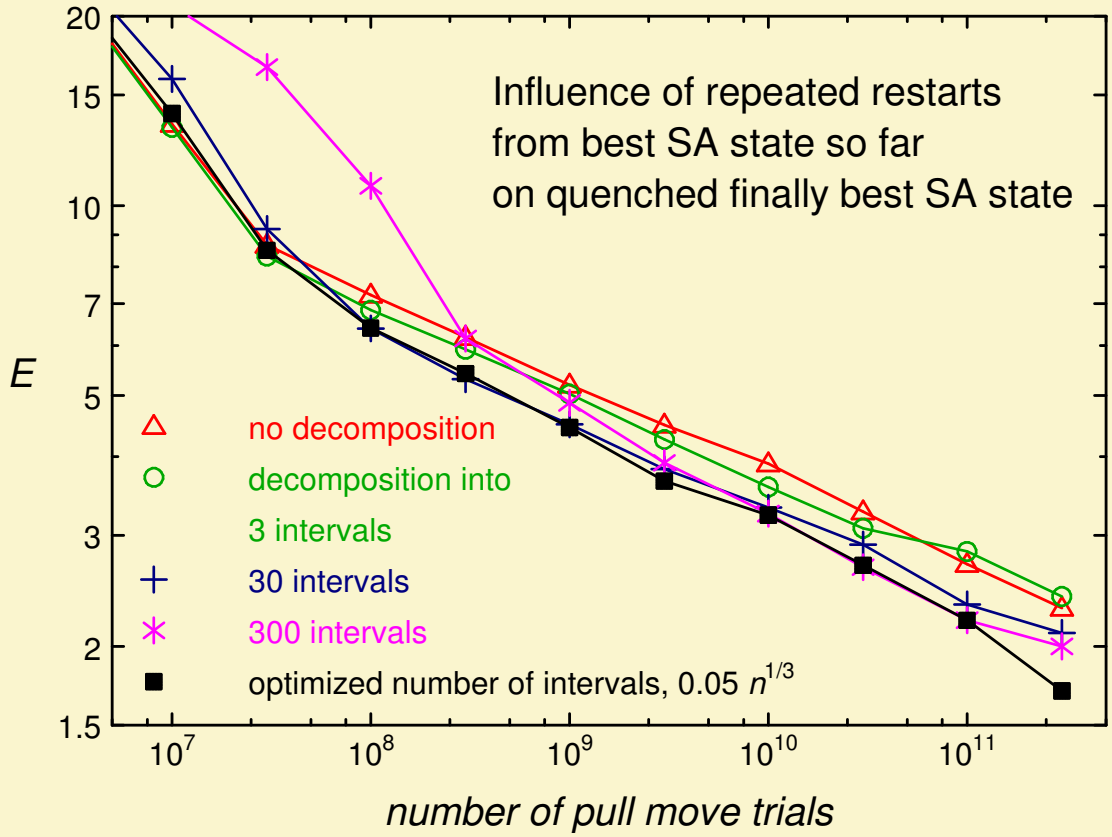


Taking non-equilibrium characteristics of SA better into account

The success of focusing on the best state so far rather than on the final one is a non-equilibrium feature. Already in early stages of the cooling, making use of previous experience could hold the system on the right track.

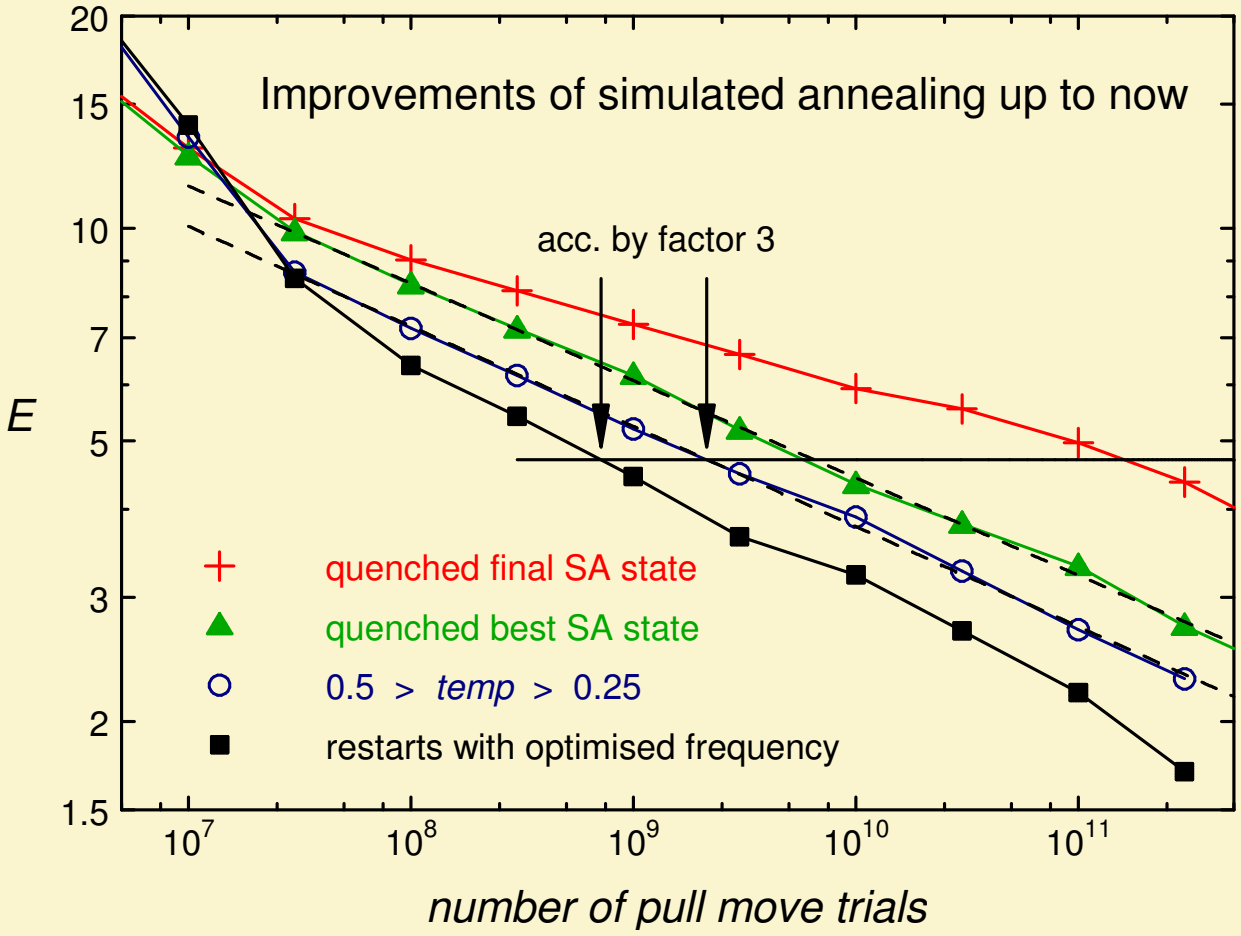


It is tempting, to optimize the number of restarts:

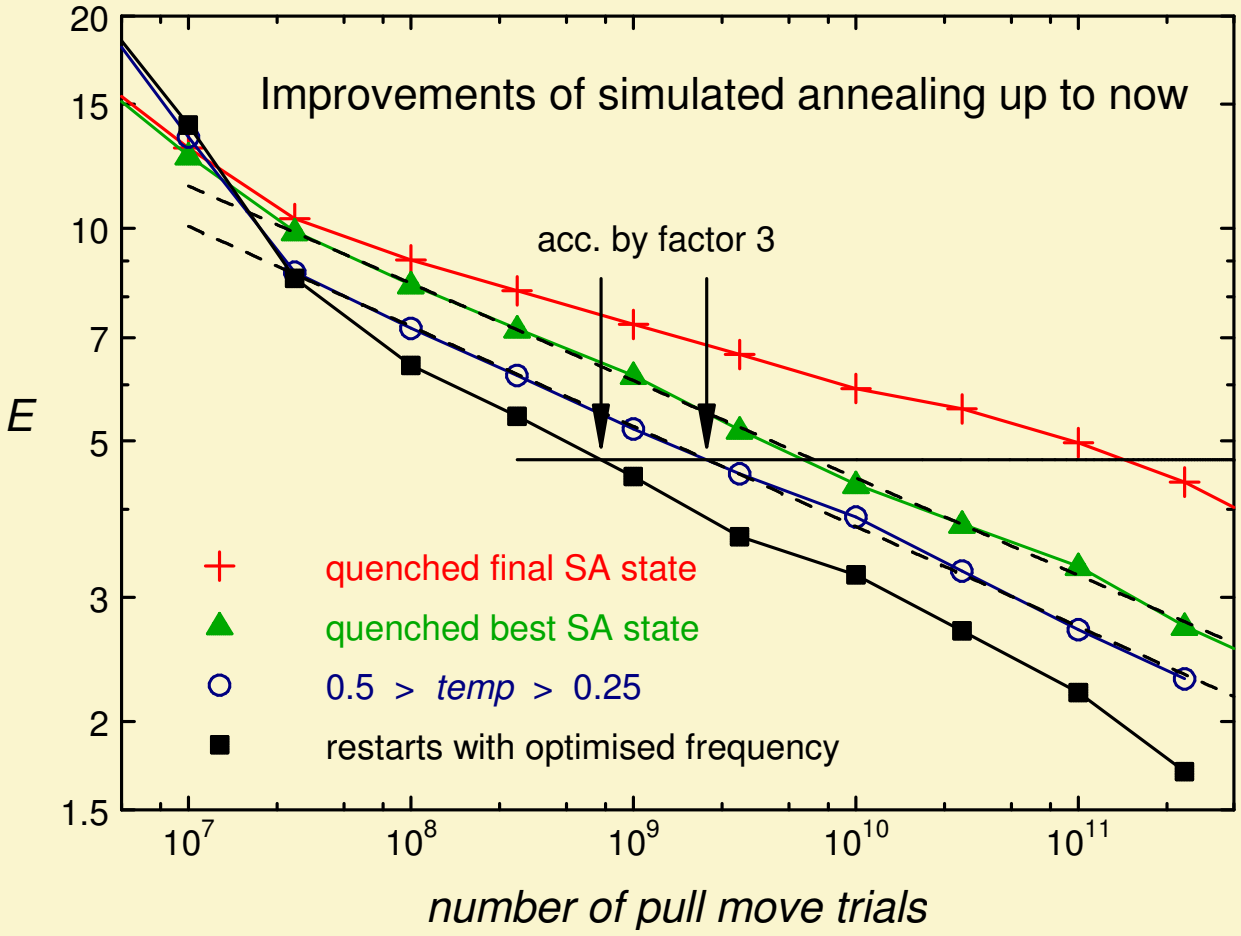


This effect is not new. Quotation from en.wikipedia.org: *"Sometimes it is better to move back to a solution that was significantly better rather than always moving from the current state. This process is called restarting…"*
However, a corresponding remark is missing e.g. in de.wikipedia.org.

Quantification of the success of restarts:



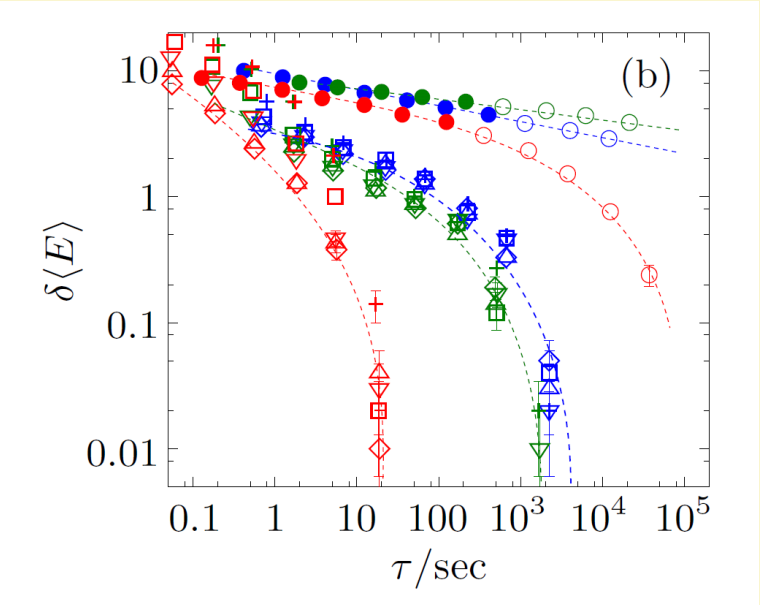
Quantification of the success of restarts:



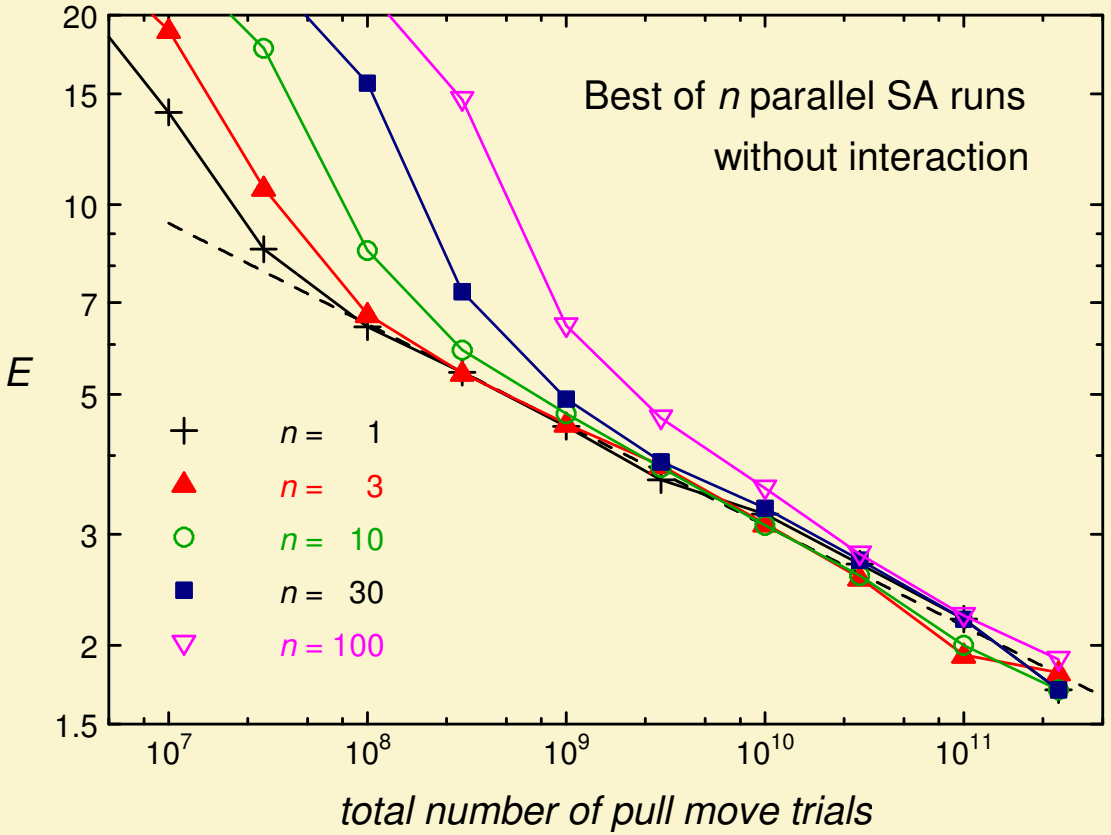
And what now?

Parallelisation may help. Let's have a closer look:

Instead of performing one "careful" SA run, we now make use of several "more superficial" runs with the same total effort, and consider the best of their quenched finally best states as final result of the collective simulation.



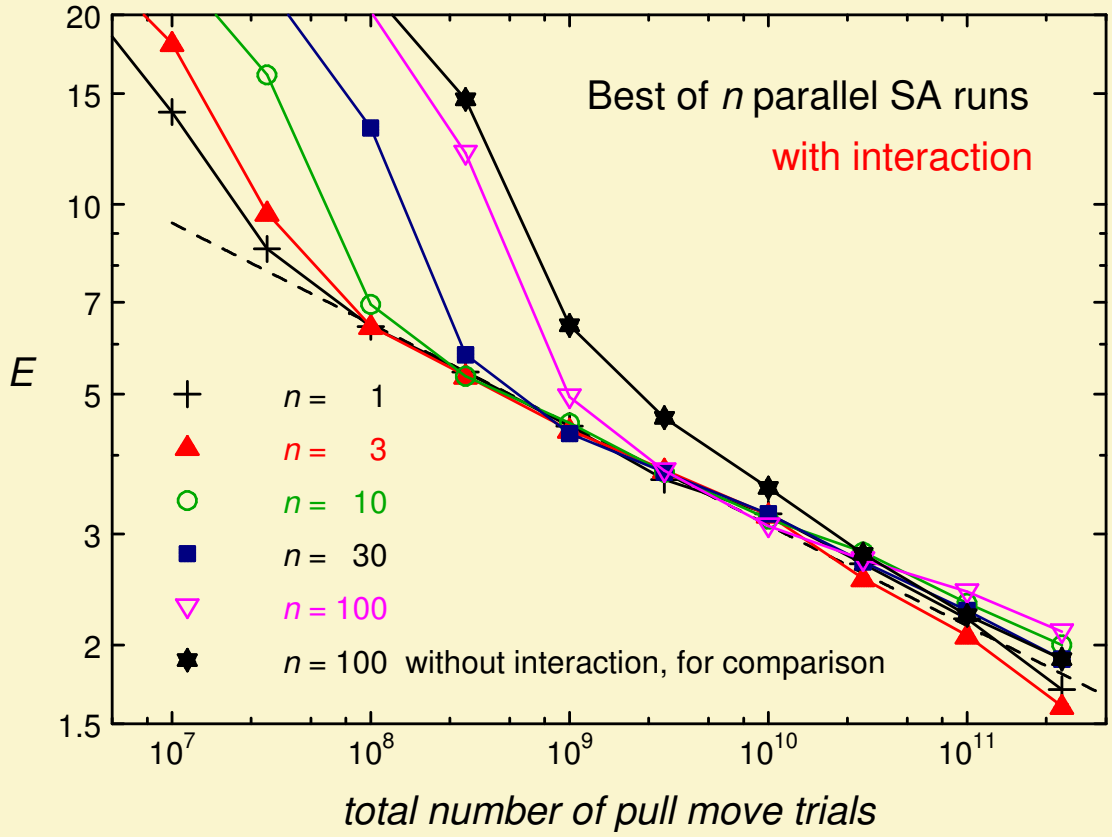
F. Günther, A. M., and M. Schreiber, EPJST **226** (2017) 639



Motivated by B.A. Huberman, R.M. Lukose, T. Hogg, Science **275** (1997) 51

How does an interaction of the threads during the restarts act?

To answer this question, we substitute part of the walkers which have reached the lowest states in the previous interval for the worse ones. But each state is duplicated maximum once in this interaction process.



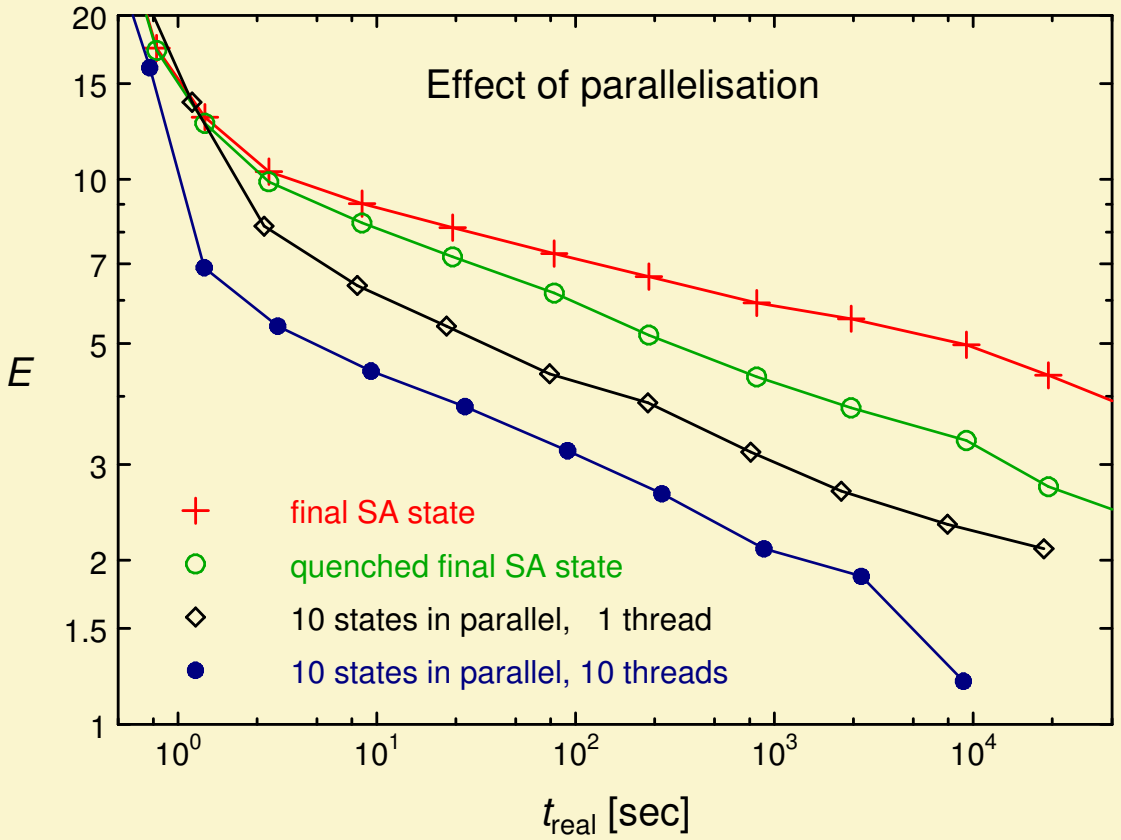
Result: Interaction only stabilizes the simulation in the left region of the plot, but seems to have no influence in the right area.

Parallelisation by means of OpenMP works nicely

Problems overcome:

- Each thread needs an independent random number generator. The own one, which is based on non-integer parts of logarithms of prime numbers, was appropriately adapted.
- For our FORTRAN code, segmentation faults could be cured by allocating the large arrays instead of defining them by the DIMENSION command.
- In the first attempts the parallelised code was rather inefficient. This could be traced back to memory access conflicts. Adding unused space in the RAM between the variables accessed by different threads solved that problem.

Comparison of the standard approach and our parallelised tuned version



Our simulations produced a large number of states with E = -82 and E = -83.

Post-relaxation using dressed bridge moves could not improve performance.

Further acceleration by means of using more cores seems possible.

4. Conclusions and challenges for future work

Take home, in particular for teaching simulated annealing to students: SA is considerably better than often thought. It can be accelerated by orders of magnitude by taking care of its non-equilibrium nature as follows.

First, do not forget to focus on the best state hit in simulated annealing instead of on the final one of this simulation.

Second, optimization of the interval bounds by means of preliminary fast runs can be very useful.

Third, restarts help a lot. However, their optimum frequency depends on the length of the SA run to be performed. Empirical rule may be tried.

Fourth, parallelisation with a medium number of threads may work fine, but there are clear efficiency losses for large numbers of threads.

Five, interaction in the restart stage may stabilize the simulations.

The unpleasant part of the take home message:

3D136 seems to exhibit a huge degree of ground state degeneracy …

Thus, this instance of the HP model seems to be “not really biological”;

moreover, “common wisdom” cannot be exploited.

Directions of our future work:

- Try other schedules.
- Apply the considered ideas to other models, for example TSP.
- Compete with thermal cycling once more, also concerning parallelisation.

The unpleasant part of the take home message:

3D136 seems to exhibit a huge degree of ground state degeneracy …

Thus, this instance of the HP model seems to be “not really biological”;

moreover, “common wisdom” cannot be exploited.

Directions of our future work:

- Try other schedules.
- Apply the considered ideas to other models, for example TSP.
- Compete with thermal cycling once more, also concerning parallelisation.

Many thanks to Wolfhard Janke and his team

for organising CompPhys17!