Kinetic Monte Carlo Simulations

at Spatiotemporal Scales of Experiments

Karl-Heinz Heinig

Helmholtz-Zentrum Dresden-Rossendorf, Germany

Outline

1. **Motivation**
2. **Bit-coded, cellular automaton-based kMC code**
3. **Applications, need for further acceleration of simulations**
4. **MPP with GPUs of conservative (Kawasaki) Ising models**
5. **Tests and Applications**
6. **Outlook**

Supported by DFG, BMBF, DAAD

Coworkers:
Bartosz Liedke, Satoshi Numazawa, Jeffrey Kelling,
Lars Röntzsch, Torsten Müller, Andreas Kranz,
Matthias Strobel, Geza Odor,, Henrik Schulz

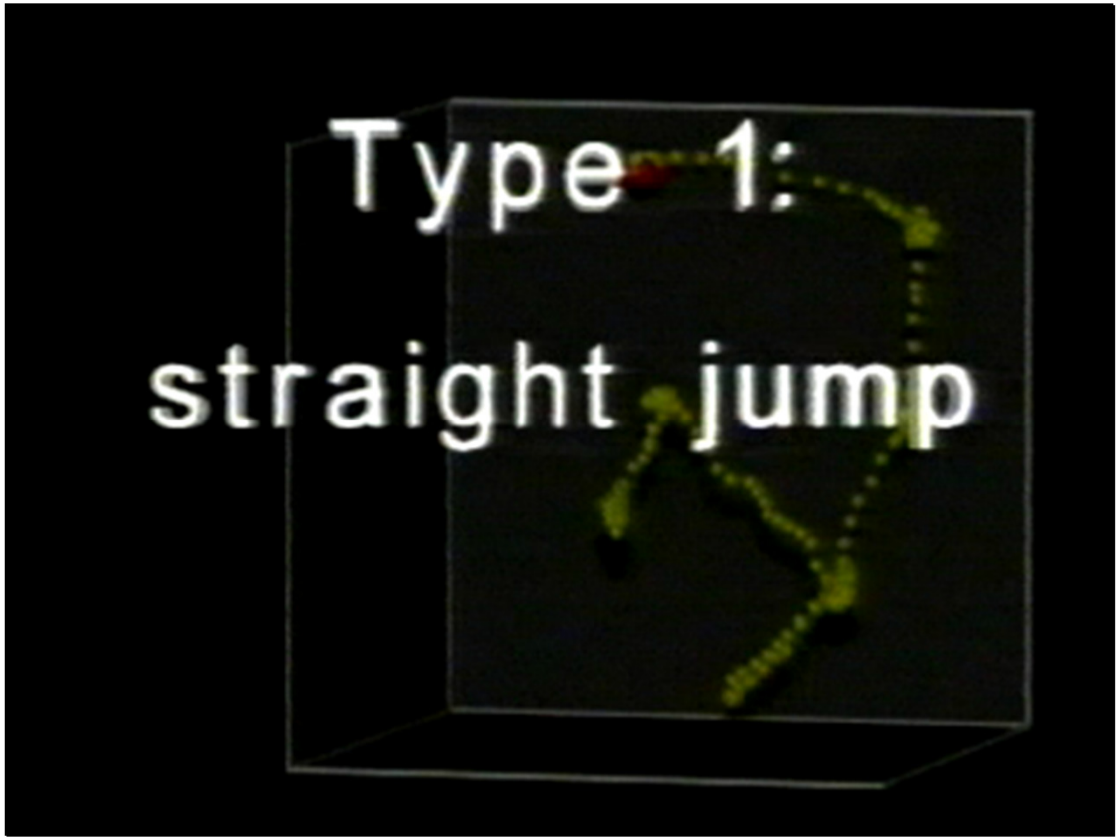
HZDR
HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

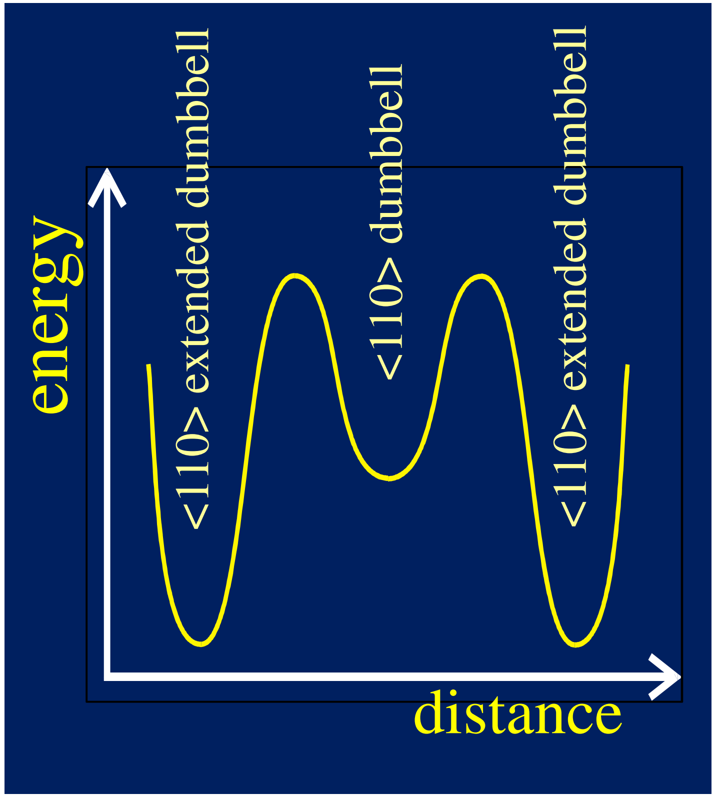
BEDMOD Workshop, March 26 - 29, 2012, Dresden

Motivation

Random walk of a Si self-interstitial in Si

Straight jump of a Si self-interstitial





MD, SW potential, 1200°C, ~ 10.000 atoms

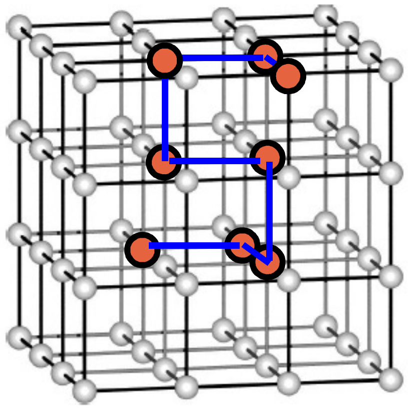
HZDR

Bit-coded, cellular automaton based kMC code

Steven Wolfram: kinetic lattice MC = statistic probabilistic cellular automaton

simulation cell

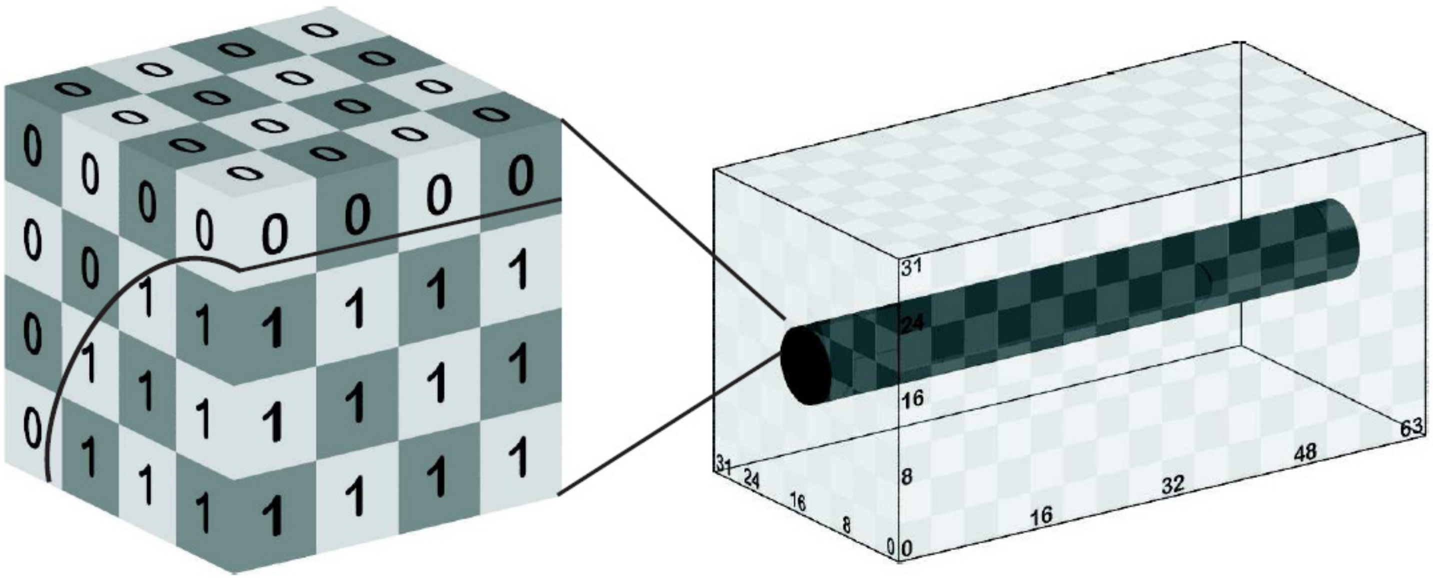
64 sites
= word of a 64 bit CPU



rigid cubic lattice

Bit-coded, cellular automaton based kMC code

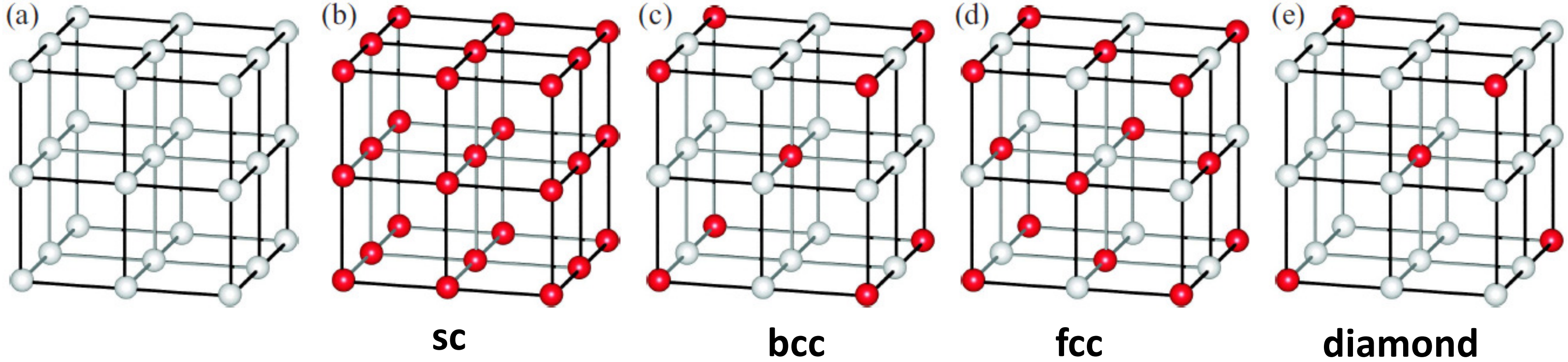
wire consisting of „phase 1“
embedded in „phase 0“ (e.g. vacuum)



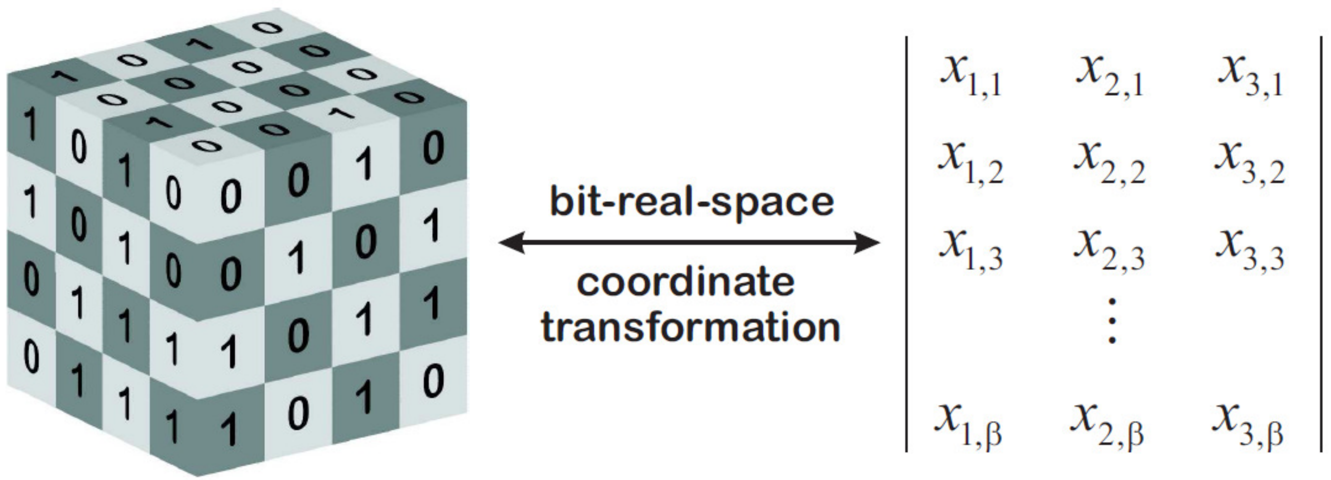
Bit-coded, cellular automaton based kMC code

selection of the lattice structure

$$[\{(x \vee y \vee z) \wedge 1\} = 0]$$

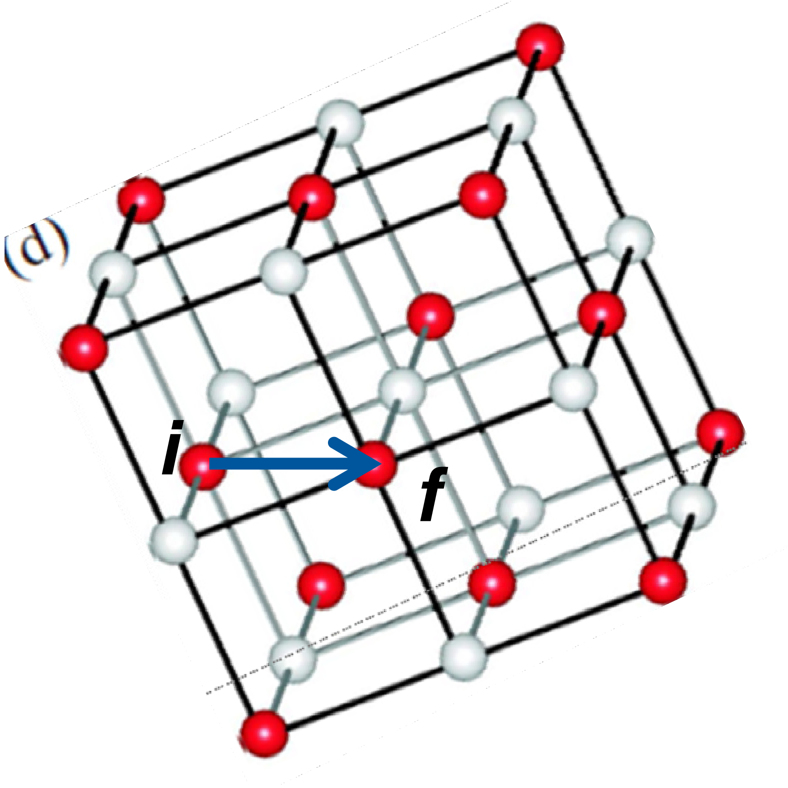


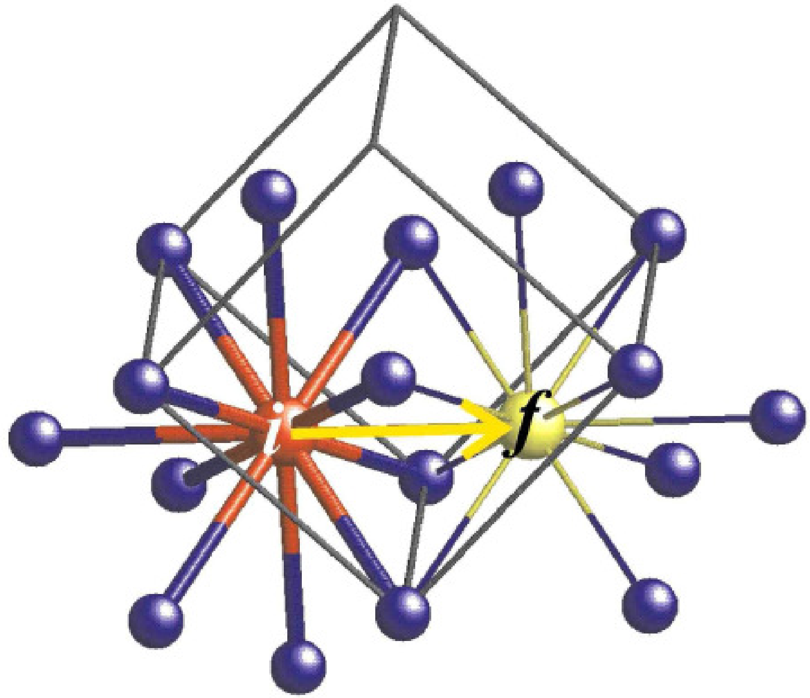
double book keeping (no NN-list needed)



HZDR

Bit-coded, cellular automaton based kMC code





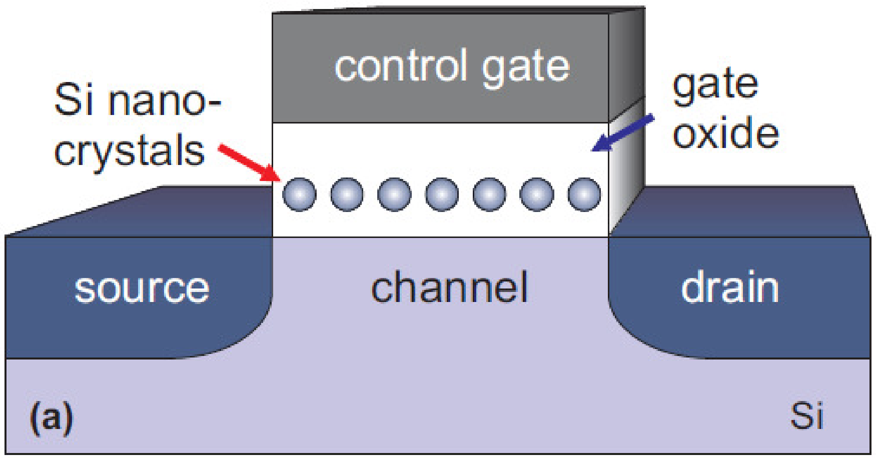
Jump-pair i→f has 18 NNs, i.e. there are 2^{18}~256000 configurations.

Totalistic CA like Ising model or RGL potential: only ~ 100 different jump configurations.

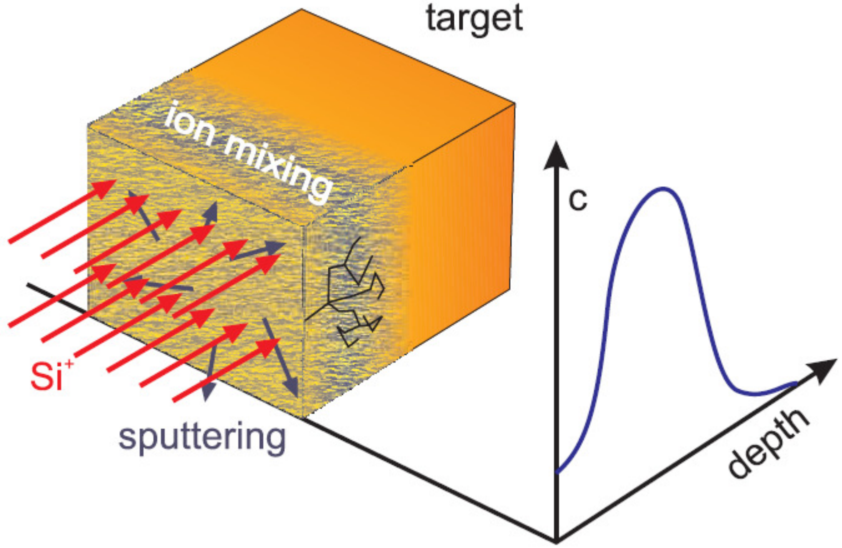
Look-up table, i.e. energies calculated once.

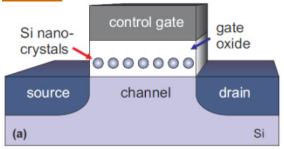
Application #1 (European FP6 project) : Process simulation of Nanodot FLASH Memory fabrication

non-volatile nanodot memory (FLASH)

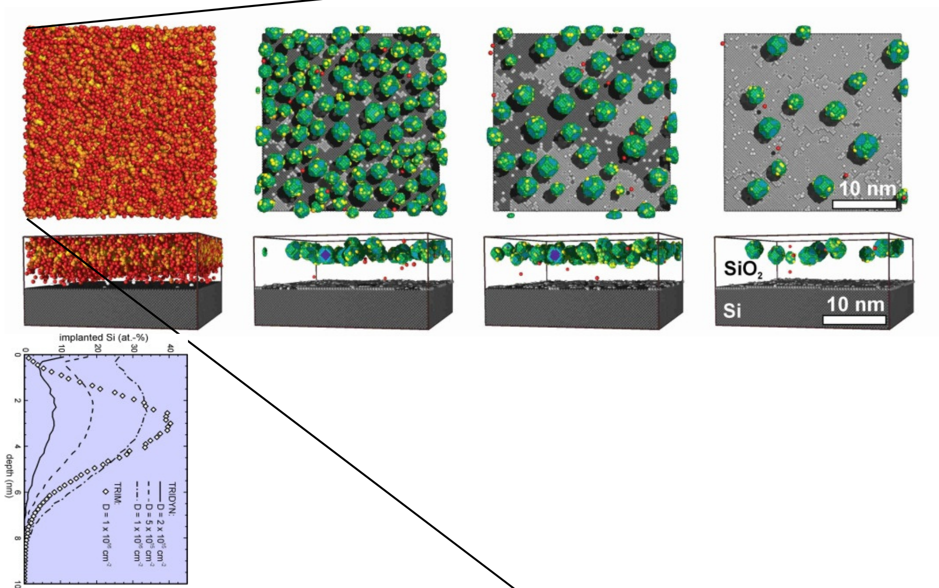


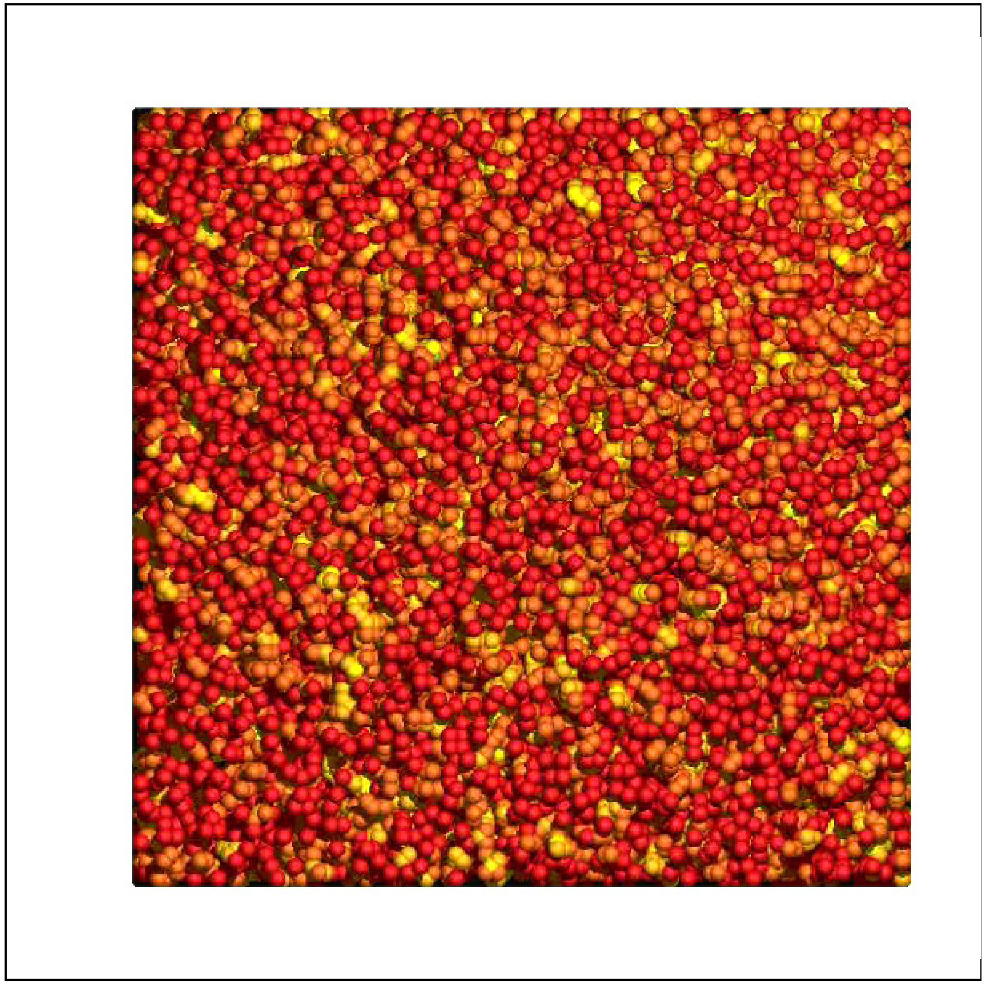
Si dot formation by Si^+ ion implantation into SiO_2 followed by phase separation





Application #1 (European FP6 project) : Process simulation of Nanodot FLASH Memory fabrication

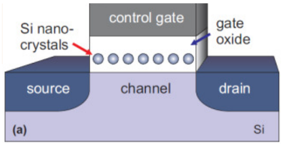




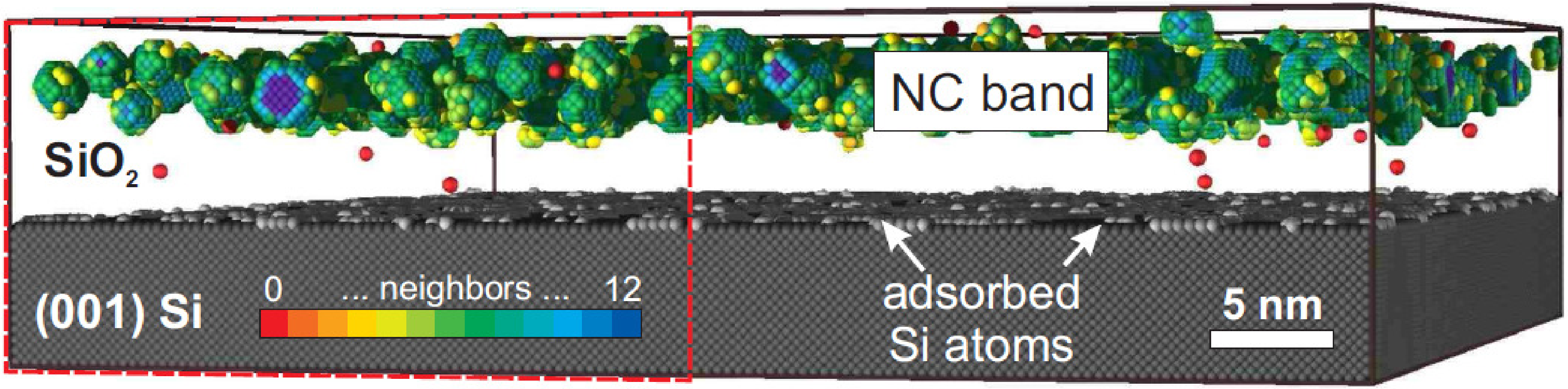
Structure:

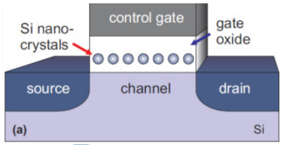
- NC δ-layer
- Denuded zone (self-adjusted tunneling oxide)

APL 81 (2002) 3049
APL 85 (2004) 2373



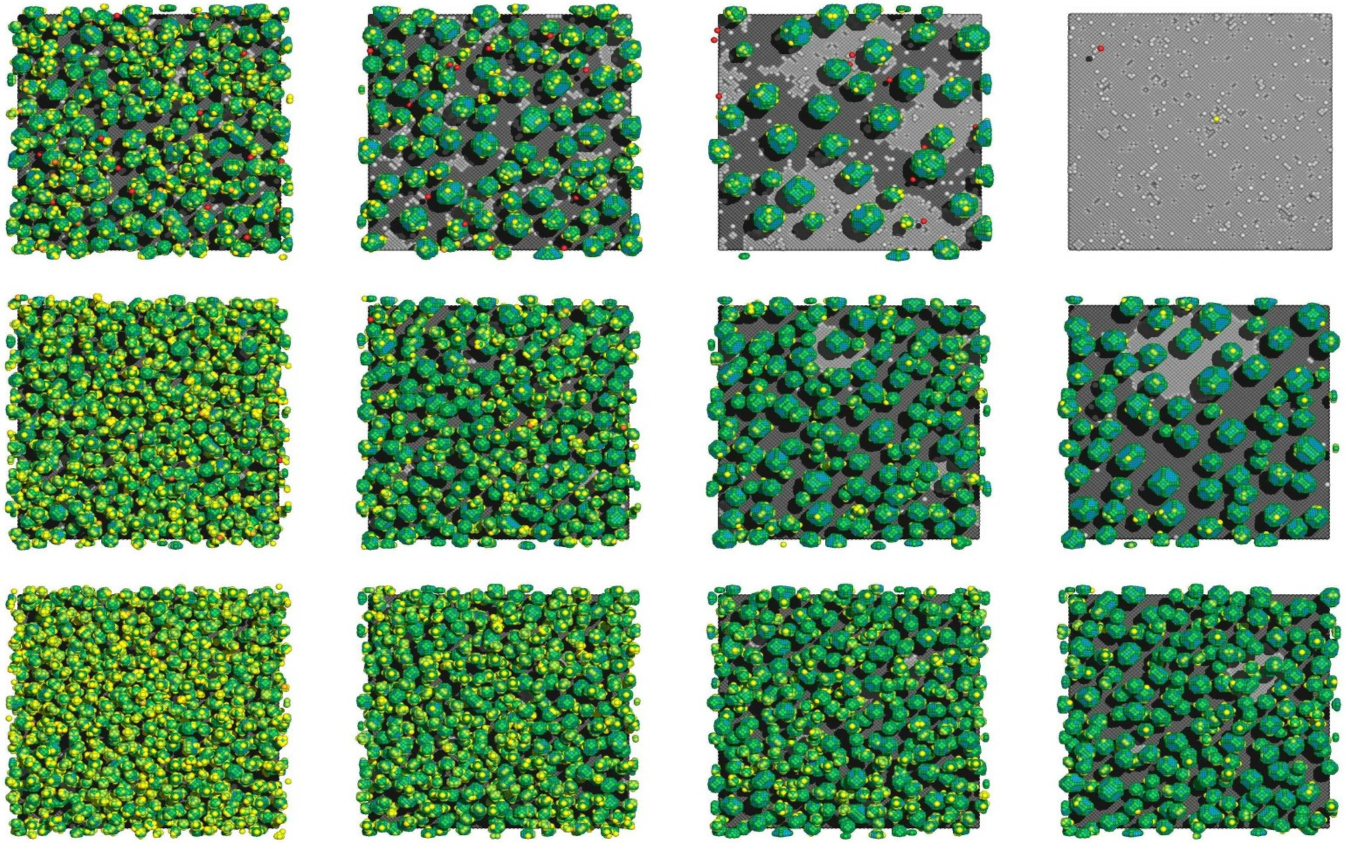
Application #1 (European FP6 project) :
Process simulation of Nanodot FLASH Memory fabrication

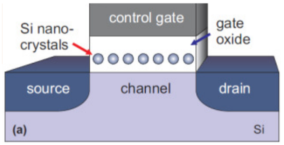




Application #1 (European FP6 project) : Process simulation of Nanodot FLASH Memory fabrication

Control of NC size by annealing conditions

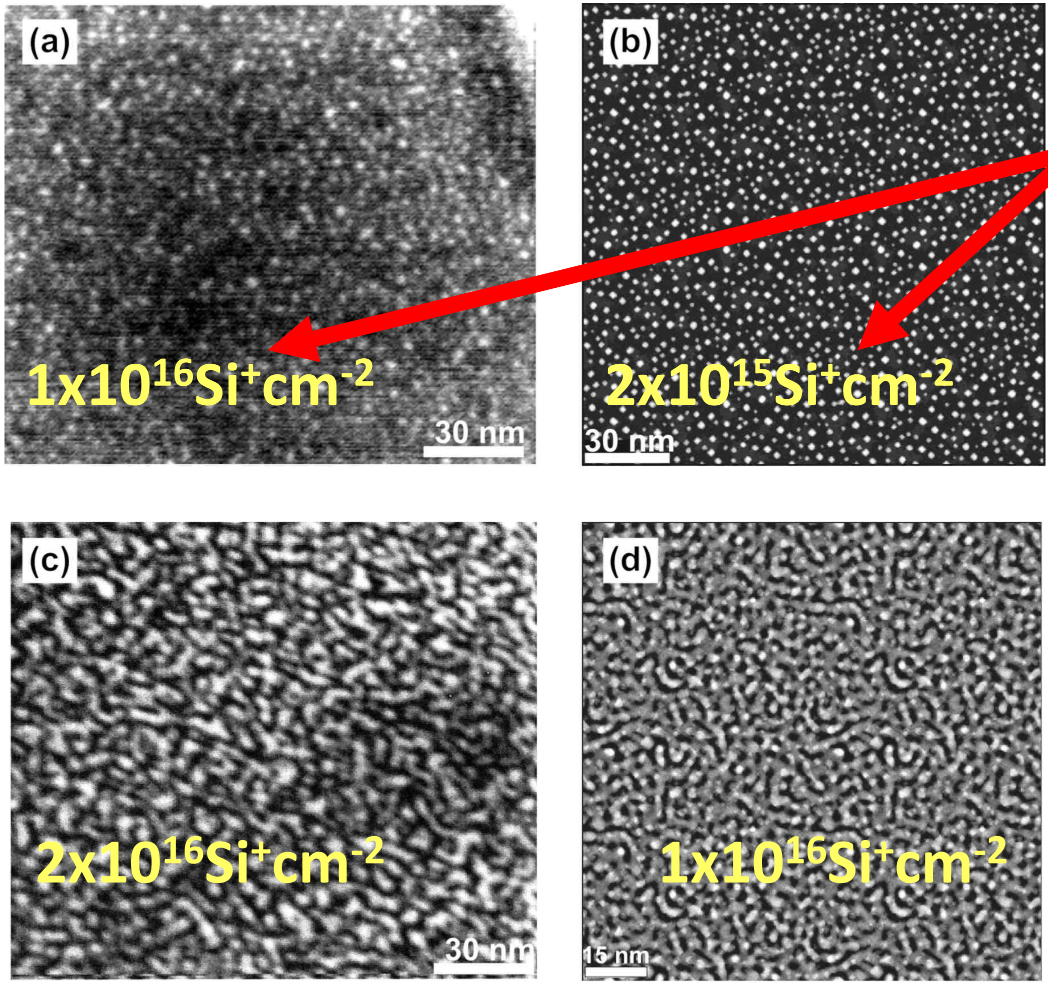




Application #1 (European FP6 project) :
Energy-Filtered TEM (EFTEM) vs kMC

1 keV Si^+ into 8nm SiO_2

EFSTEM | kinetic MC



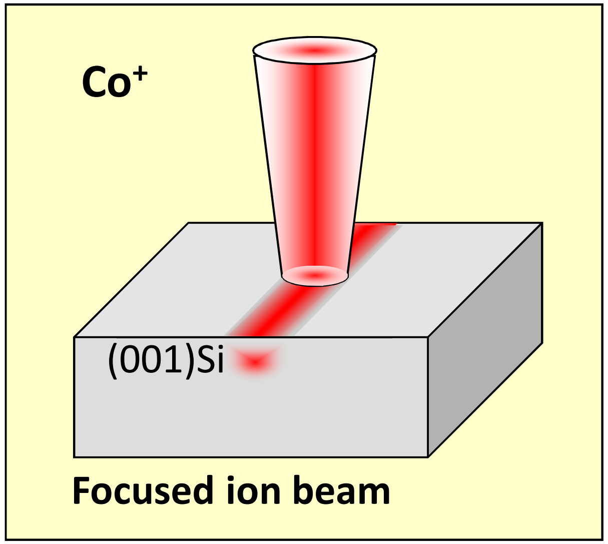
Good agreement ? No !
Exp. & theory differ by a factor 5 !!

Theory predictive,
Exp. „wrong“ !!! (humidity)

kMC-based process simulations could identify a parasitic oxidation of ion implanted silicon

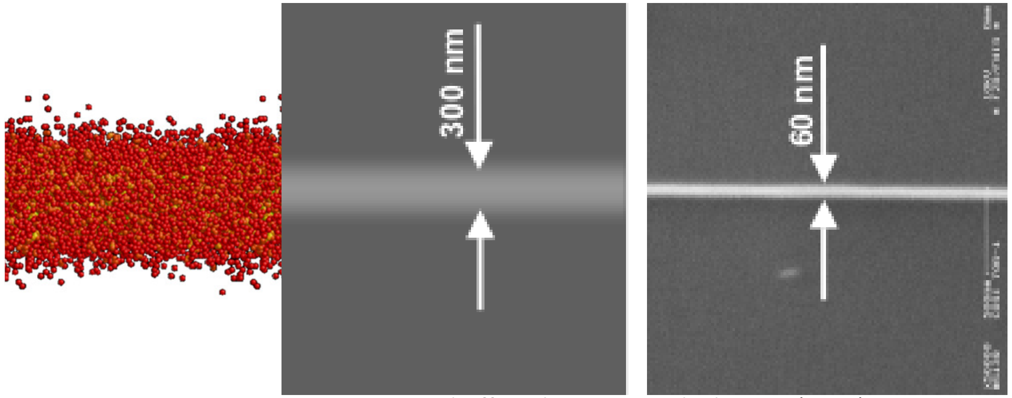
APL 85 (2004) 2373

Application #2 (DFG priority programm): Synthesis of functional nanowires



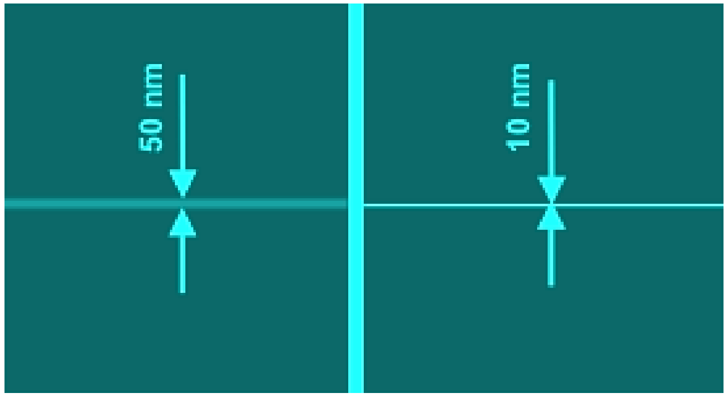
FIB: $10^{17} Co^{2+} cm^{-2}$ @70keV & 415 °C into (001)Si along <110>

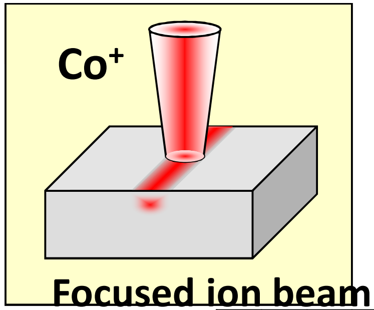
Annealing in N_2 for 60'@ 600 °C and for 30'@ 1000 °C



S.Hausmann, L.Bischoff et al., Jpn. J. Appl. Phys. 38 (1999) 7148
L.Bischoff, K.H.Heinig et al., Nucl. Instr. Meth. B112 (1996) 201

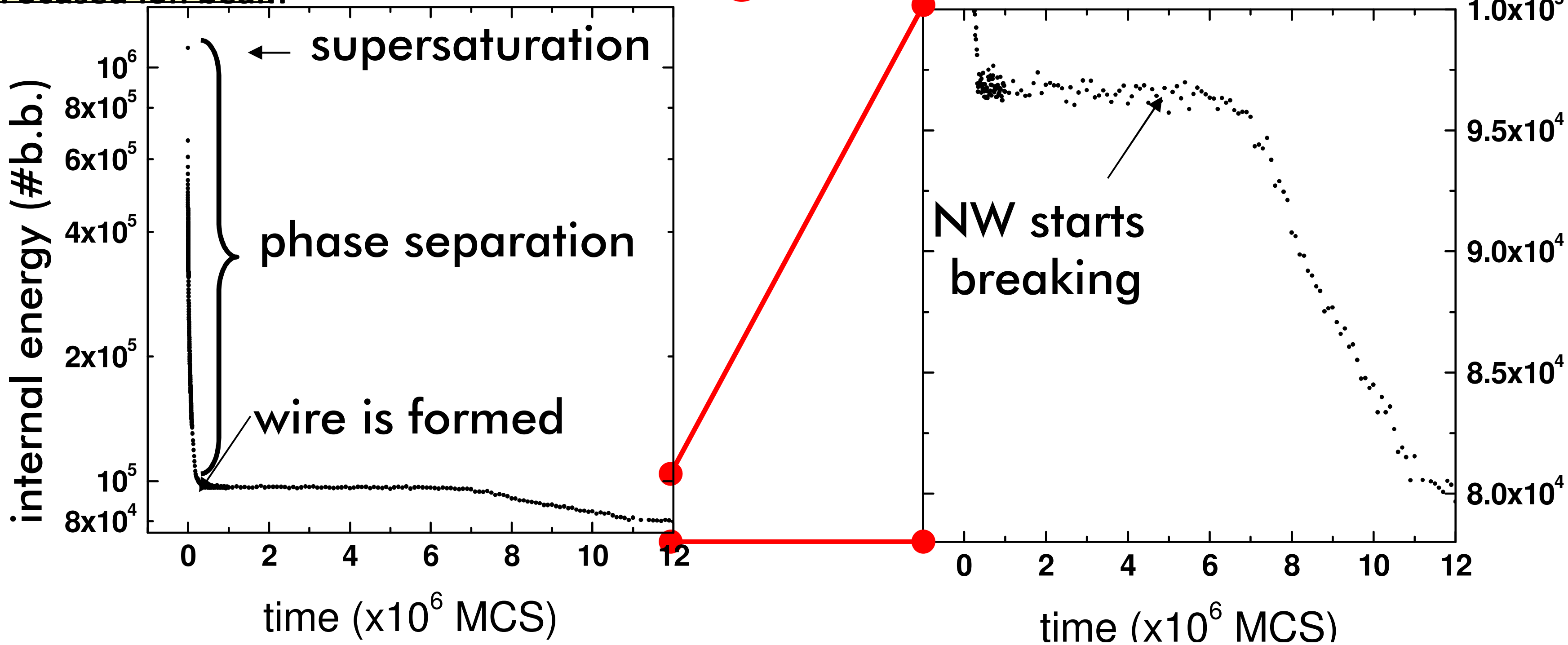
- Implantation of Co lines of 50 nm width
- Expected $CoSi_2$ nanowire after annealing: ~10 nm

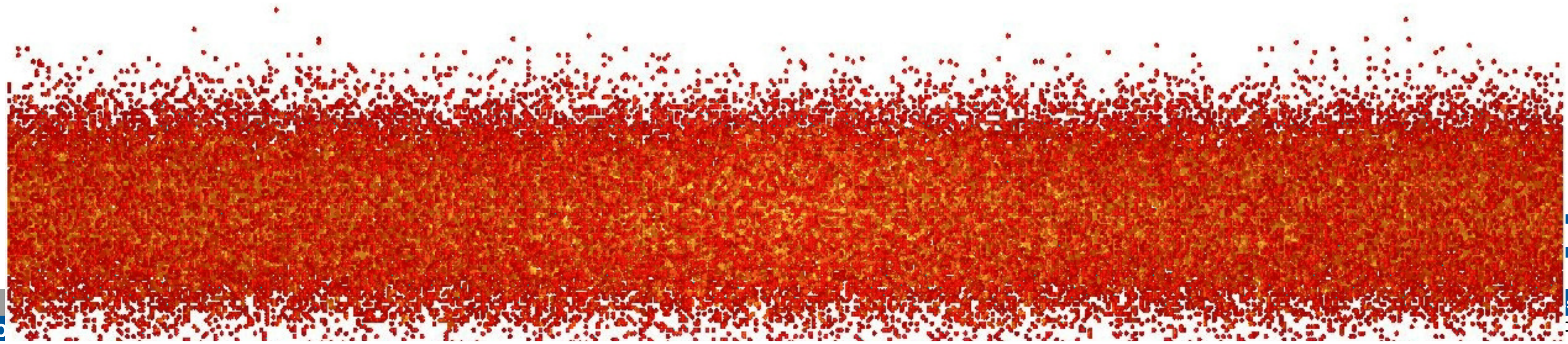




Application #2 (DFG priority programm): Synthesis of functional nanowires

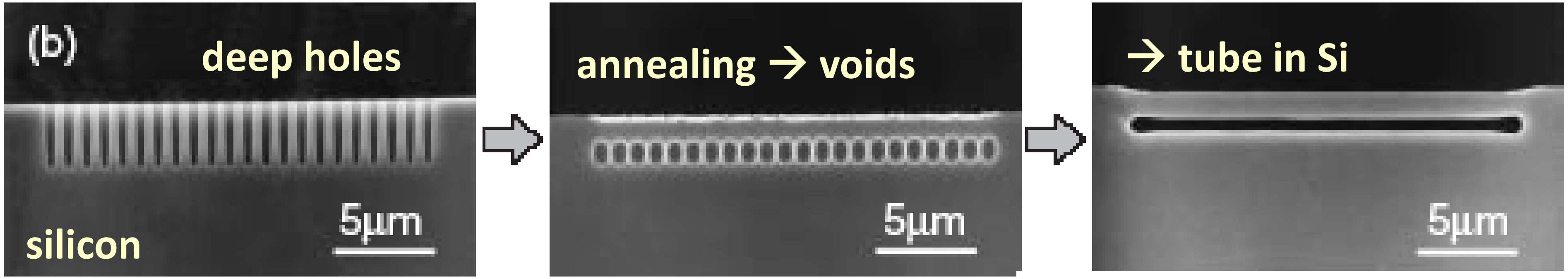
Short vs. Long Term Relaxation



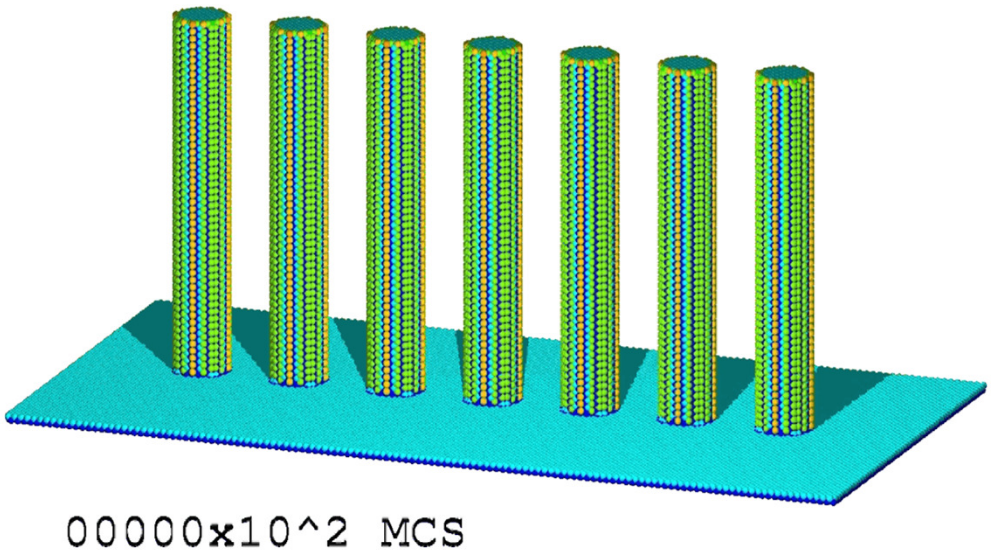


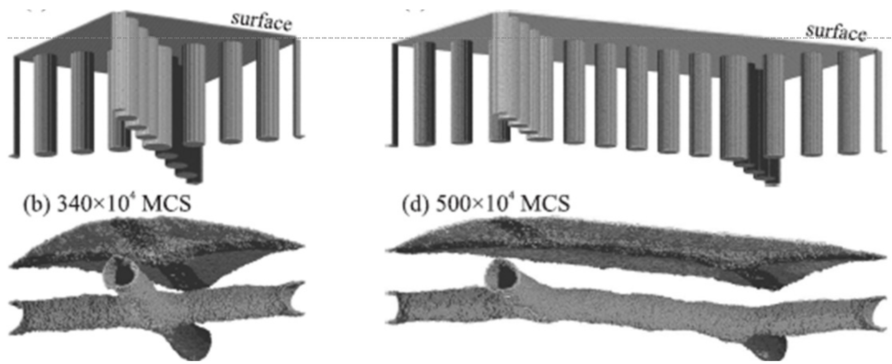
Application #2 (DFG priority programm): Synthesis of functional nanowires

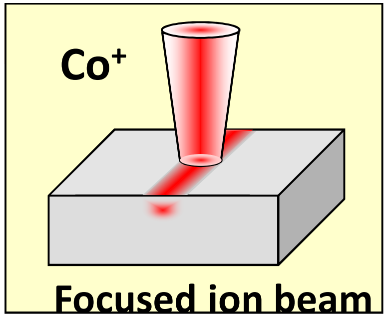
Many interesting applications of nanocapillarity (tube = inverse wire)



Japanese Journal of Applied Physics
Vol. 43, No. 1, 2004, pp. 12–18

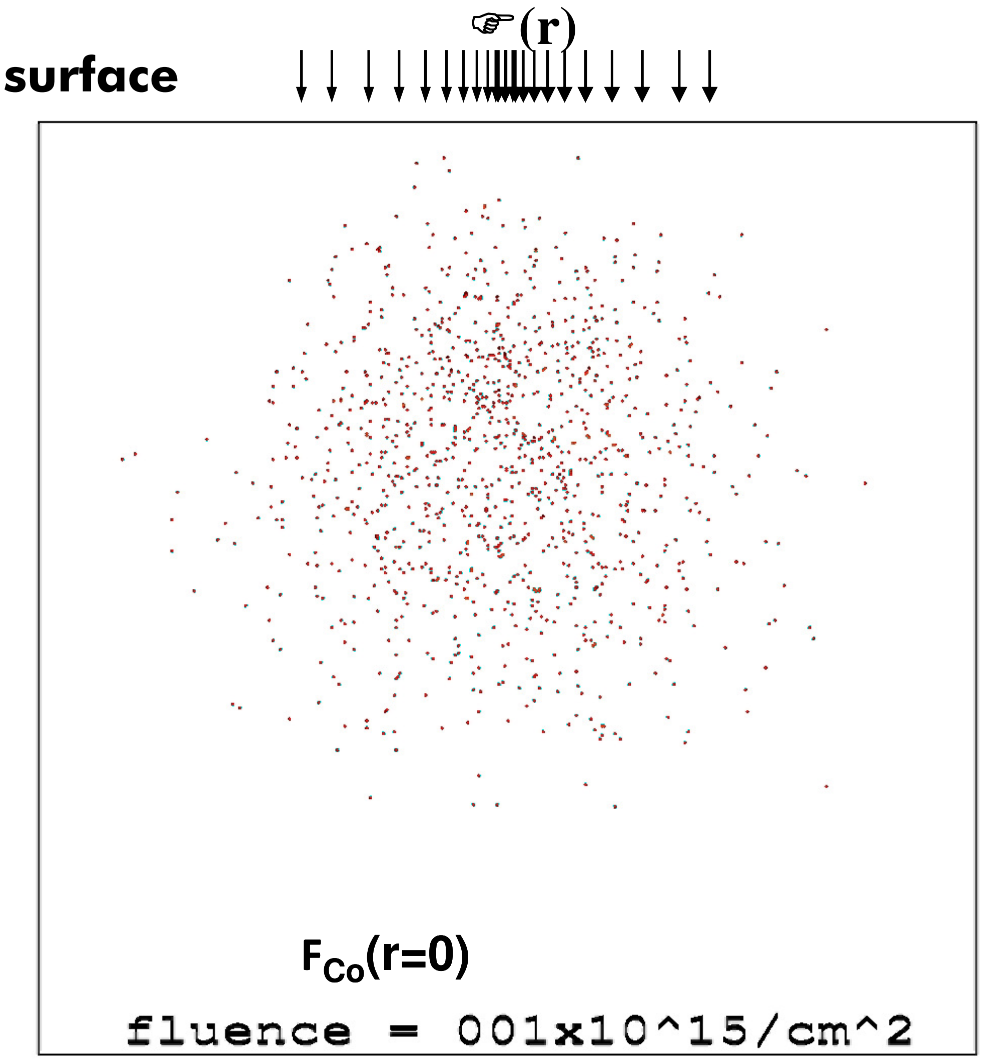




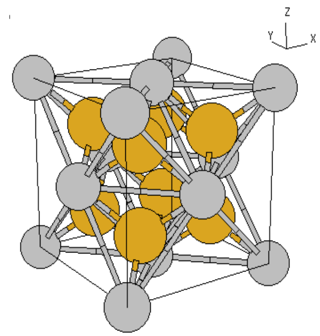


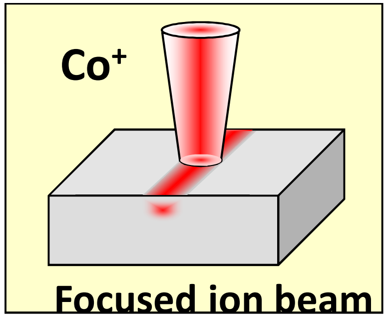
Application #2 (DFG priority programm): Synthesis of functional nanowires

simulation of implantation
of a realistic FIB Co profile



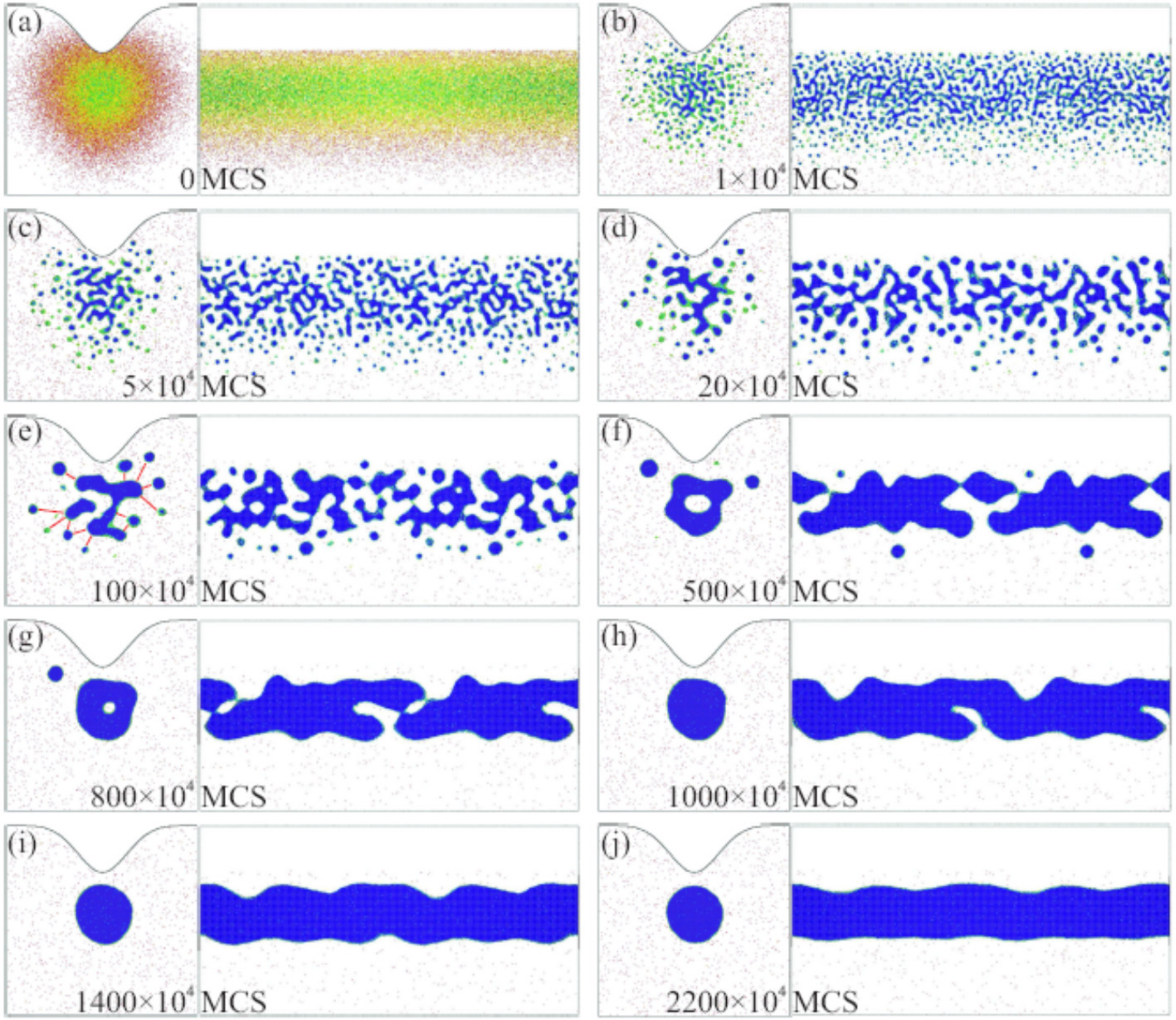
- (100)Si substrate
- Gaussian Co beam (FWHM 50nm)
- E_{Co} = 60 keV
- color → coordination
- a_{Si}~a_{CoSi2}





Application #2 (DFG priority programm): Synthesis of functional nanowires

simulation of annealing
of a realistic FIB Co profile



simulation parameters:

Gaussian beam profile,
Width = 50 nm
E_{Co} = 60 keV
$F_{Co}(r=0)$ = 8×10^{16} ions/cm^2
Cell = (512 x 512 x 512) a^3
$\#_{atoms}$ = 4,743,197

almost 2 years CPU time !

decay takes 100x longer !

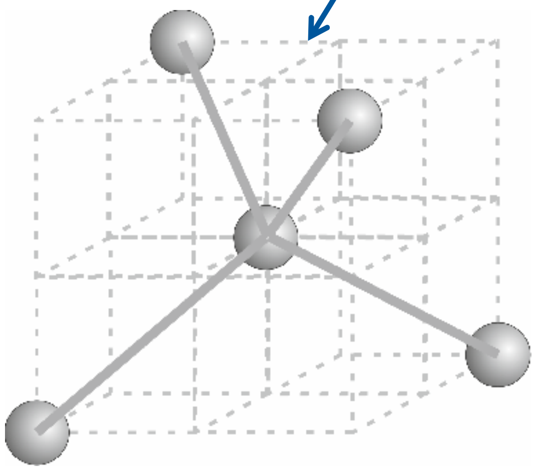
Beyond standard lattice kMC ?

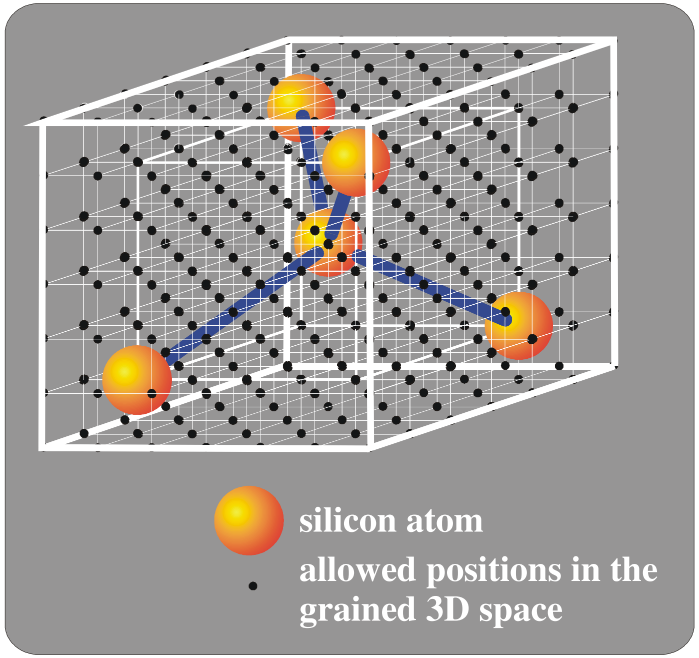
a) Massively Parallel Programming

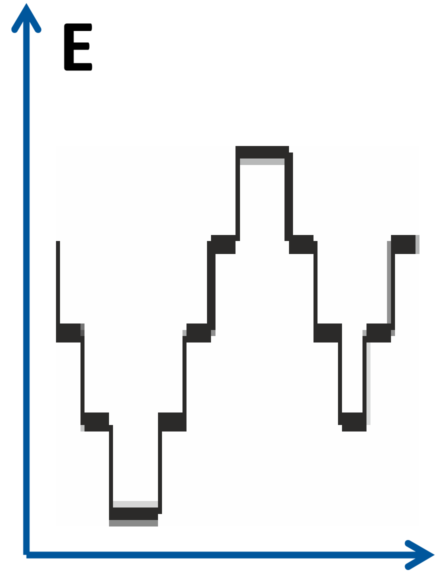
b) Exploitation of the Cellular Automaton Concept

Fine-grained lattice kMC for silicon

$$\left(\left[\{(x \vee y \vee z) \wedge 1\} = 0\right] \vee \left[\{(x \wedge y \wedge z) \wedge 1\} = 1\right]\right) \wedge \left[\{(x \oplus y \oplus z) \wedge 10\} = 0\right]$$





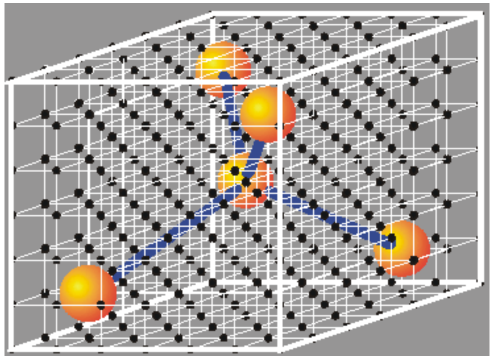


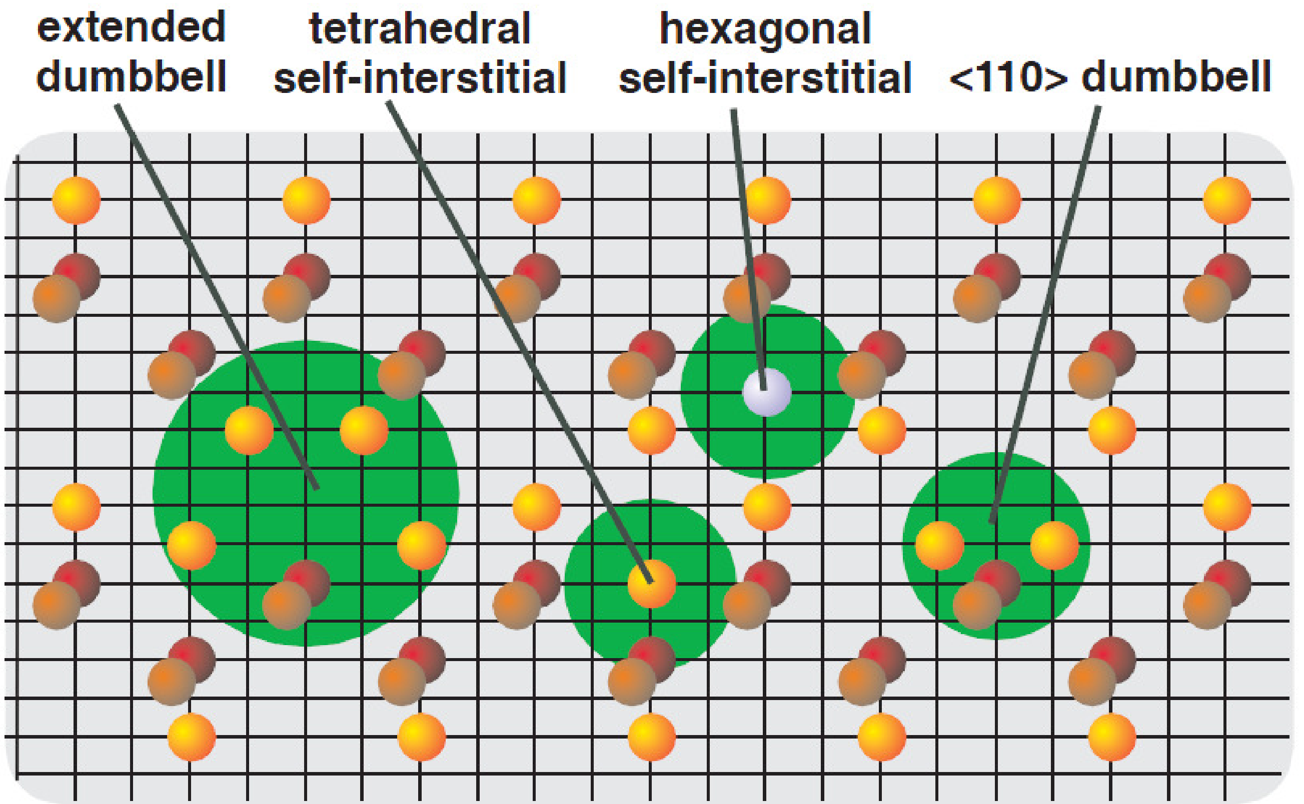
- “standard” lattice kMC
- 8 atoms per 64 sites
- jumps to lattice site

- “fine grained” lattice kMC
- 1 atom per 64 sites
- Jumps on fine lattice

Fine-grained lattice kMC for silicon

3 atomic layers along a <110> direction

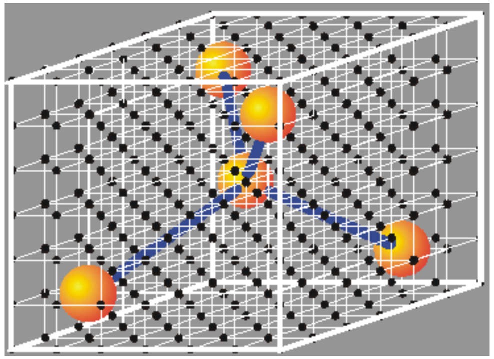




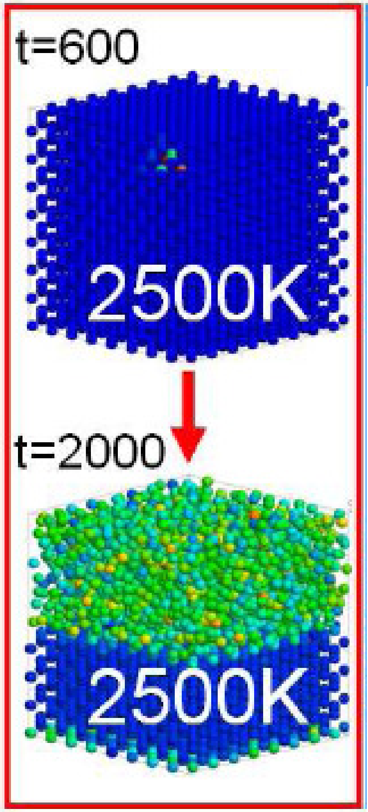
- atoms of the (110) lattice plane having self-interstitials
- (110) lattice plane above this defective plane
- (110) lattice plane below this defective plane
- hexagonal selfinterstitial between two (110) lattice planes

Location of point defects is similar to MD results (even relaxation)

Fine-grained lattice kMC for silicon

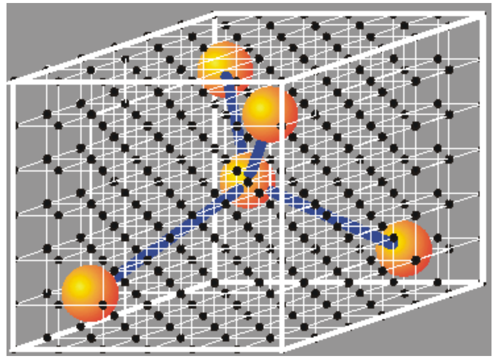


Amorphization of Si and
Solid Phase Epitaxial (SPE) regrowth

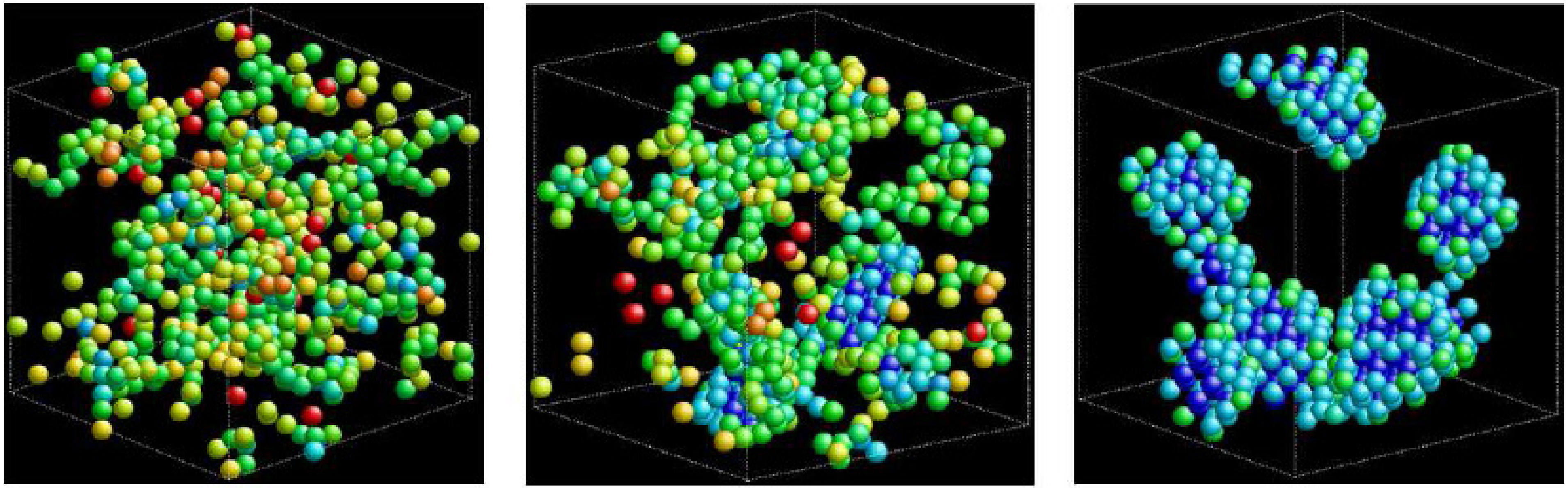


t=12000

Fine-grained lattice kMC for silicon



Nanocrystal nucleation and growth from dissolved silicon



t=100 t=1000 t=10000

Massive Parallel Programing (MPP) of kMC

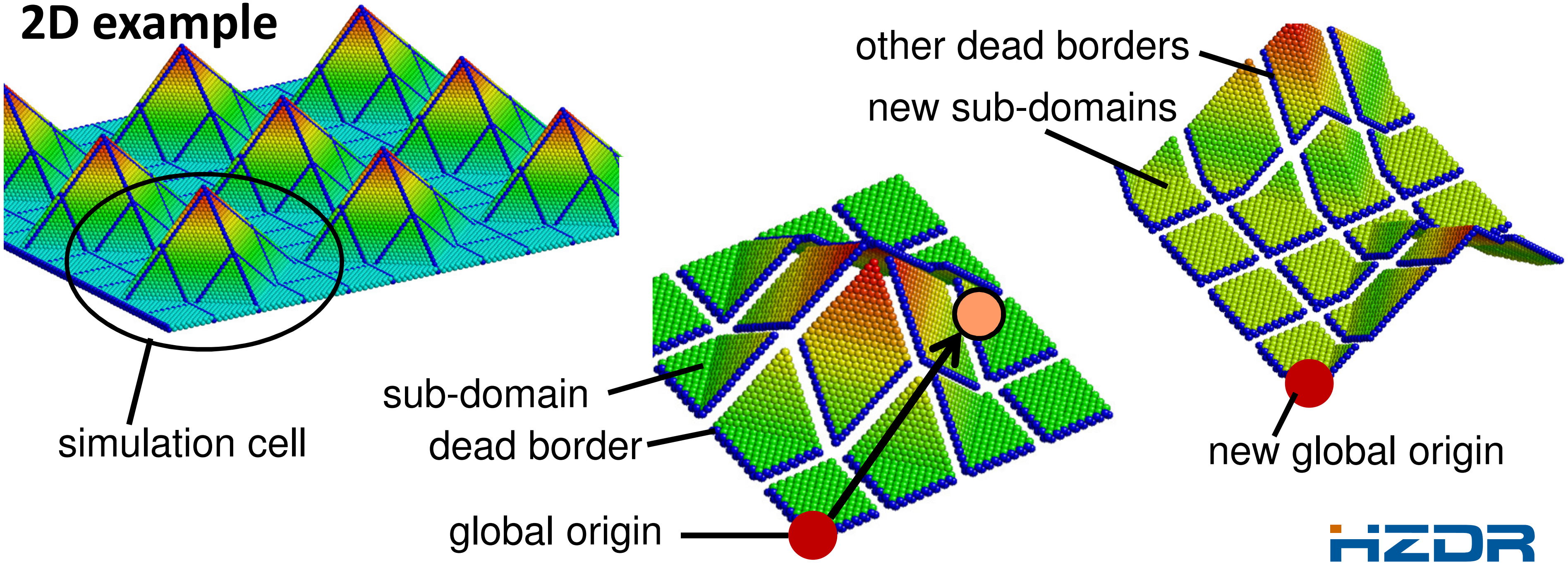
a) with CPU cores (multi-core CPUs, LINUX clusters,…)

b) with graphic cards (NVIDIAs Tesla and Fermi card, ATI,…)

Massive Parallel Programing (MPP) of kMC with CPU cores

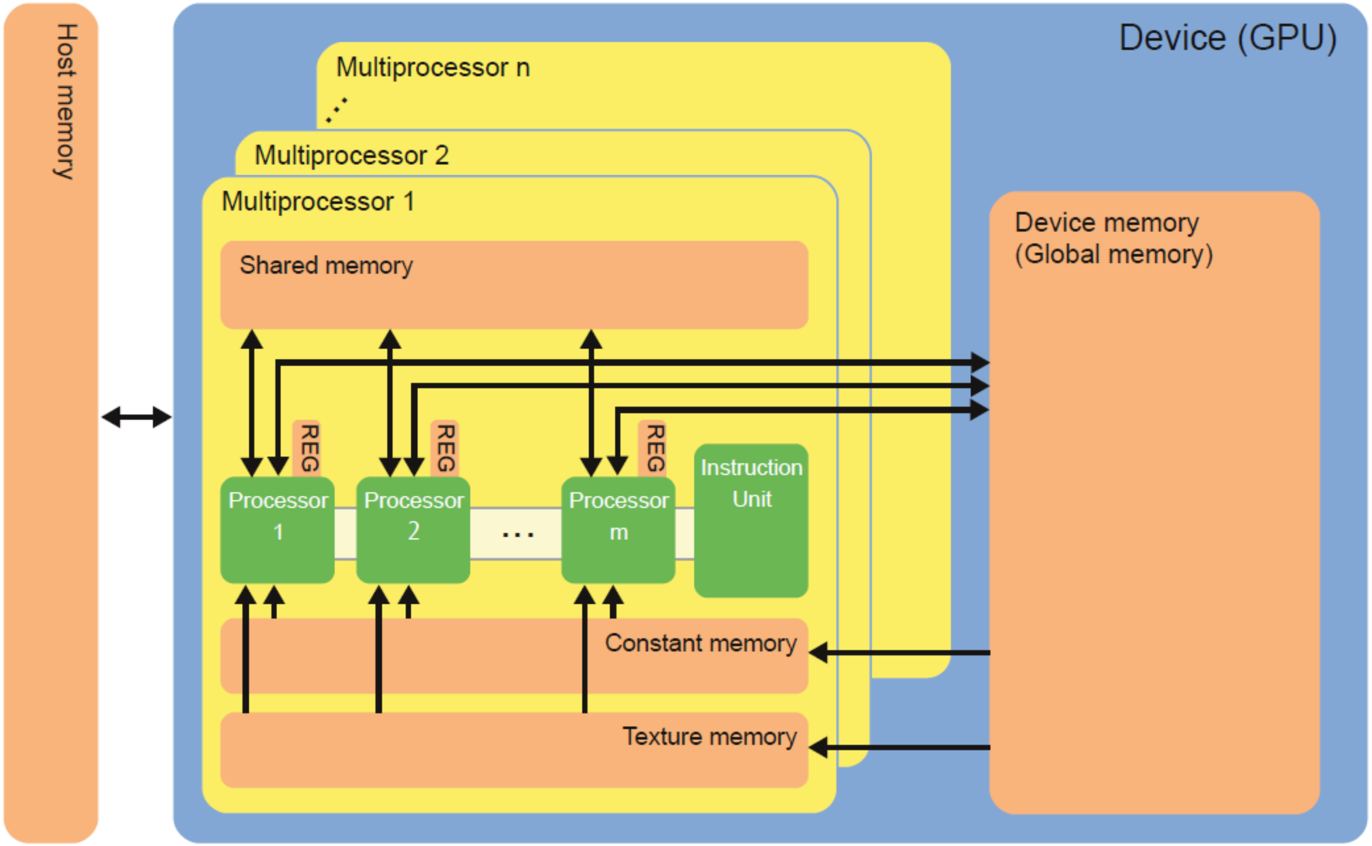
<u>"Dead borders" concept</u>:

- Break the simulation cell (periodic BCs) into sub-domains
- Run kMC simulations of the sub-domains in parallel
- Avoid "talking" between sub-domains by "dead borders" (small statistical error)
- After a "short" time, shift the global origin randomly → other dead borders



Massive Parallel Programing (MPP) of kMC with graphic cards (GPUs)

Schematic representation of the architecture of current GPUs.



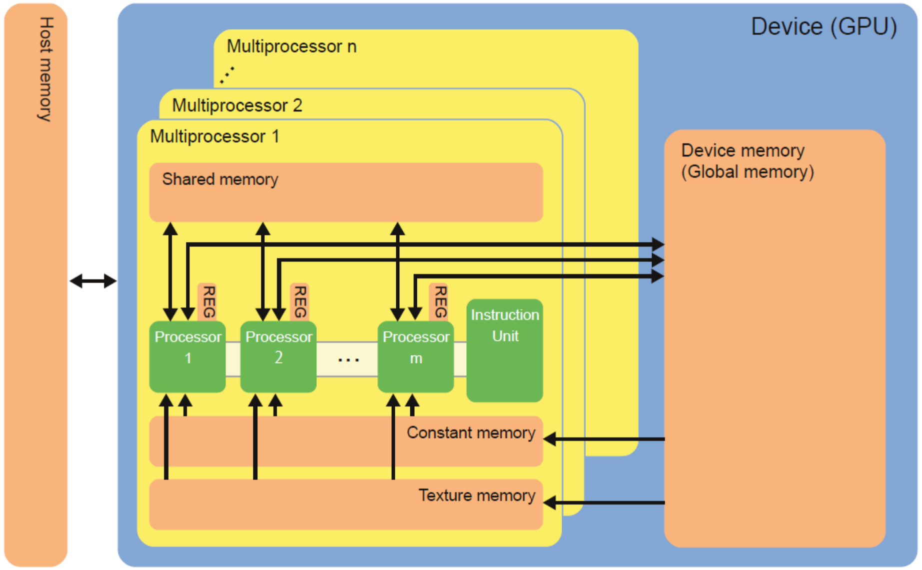
Tesla:
- 30 multiprocessors
- 8 processors each

→ 240 cores

Fermi:
- 14-16 multiprocessors
- 32 processors each

→ 512 cores

Weigel, Int. J. of Mod. Phys. C (2011)



Registers: each multiprocessor is equipped with several thousand registers with local, zero-latency access;

Shared memory: processors of a multiprocessor have of on chip, small latency shared memory;

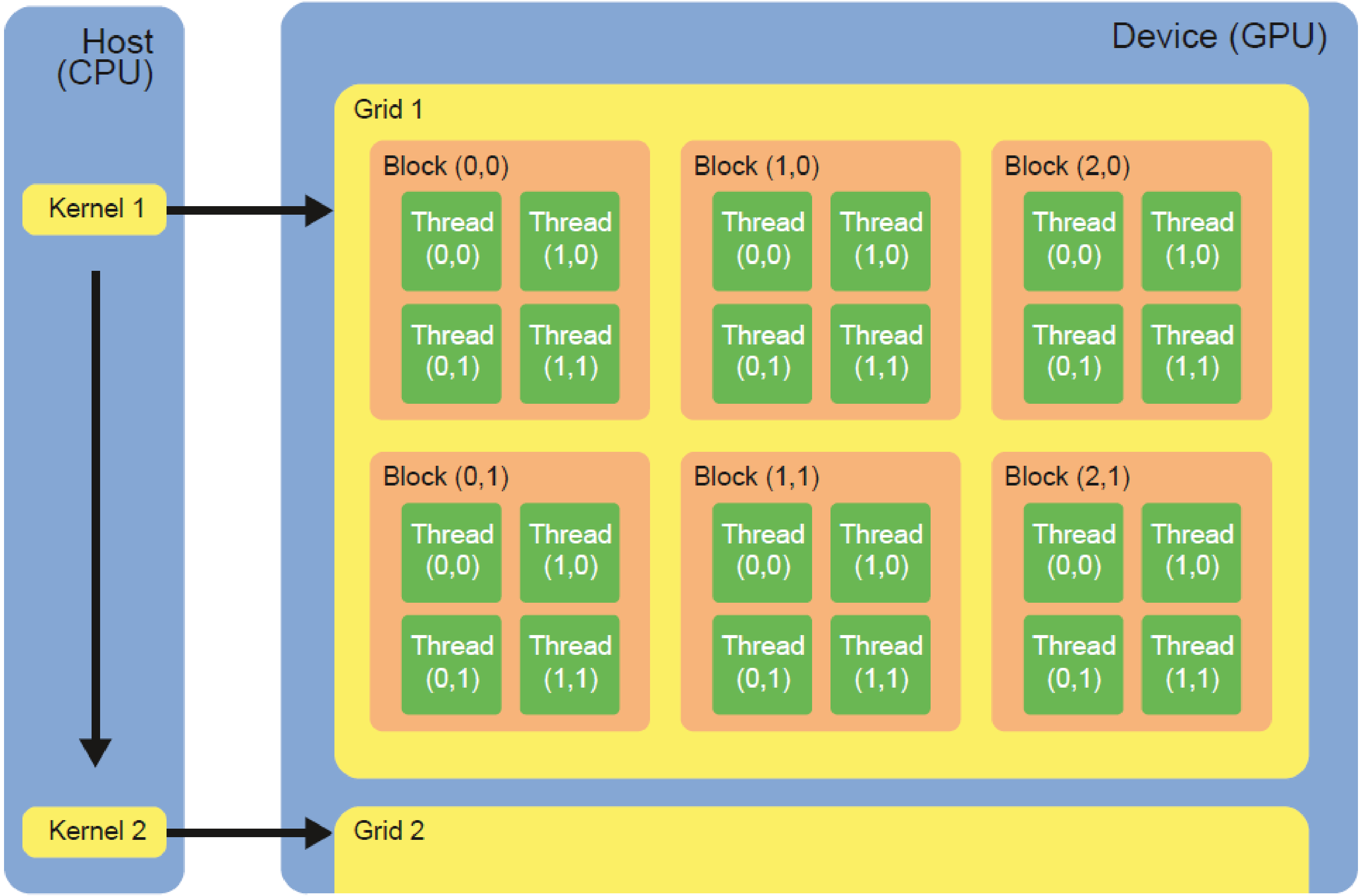
L1 and L2 caches: 16/48 kB L1 cache and 768 kB L2 cache;

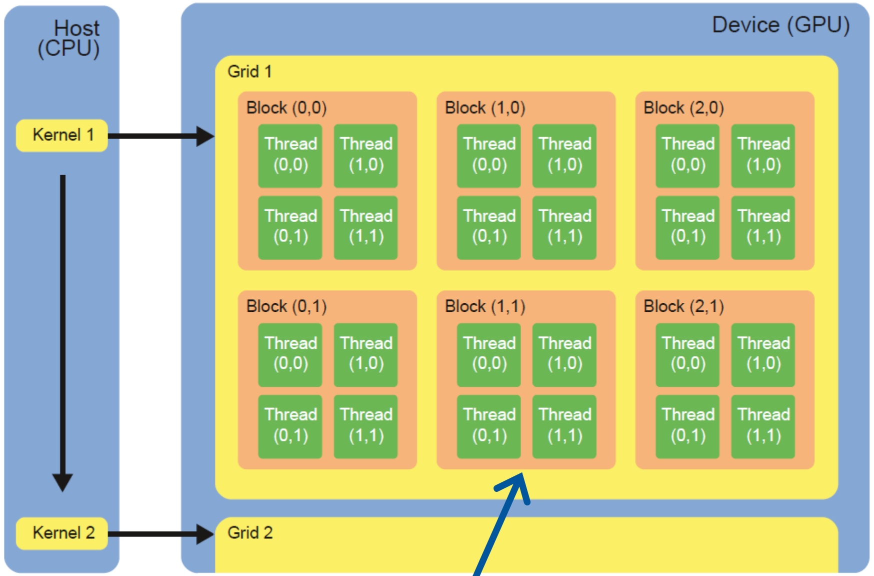
Global memory: large amount (currently up to 6 GB) of memory on separate DRAM chips with access from every thread on each multiprocessor with a latency of several hundred clock cycles;

Constant and texture memory: read-only memories of the same speed as global memory, but cached;

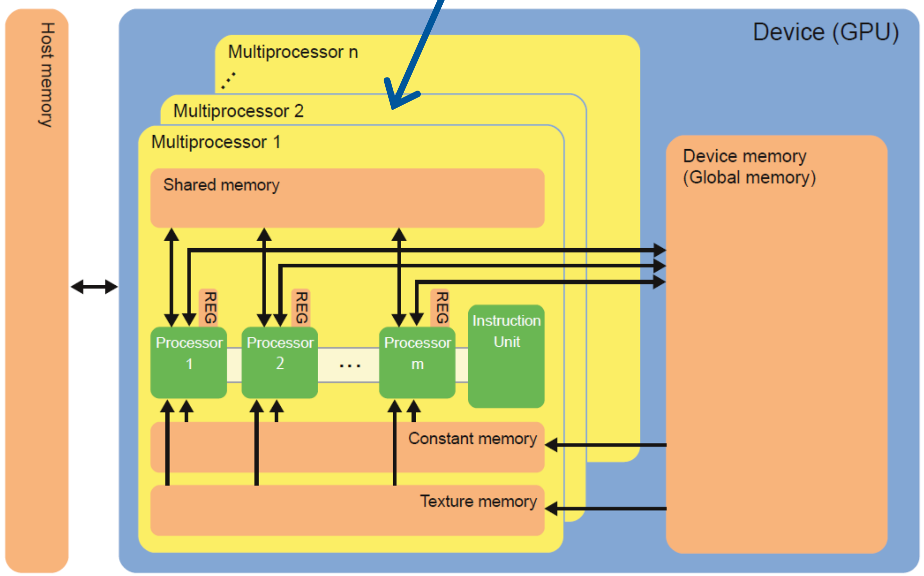
Host memory: cannot be accessed from inside GPU functions, relatively slow transfers.

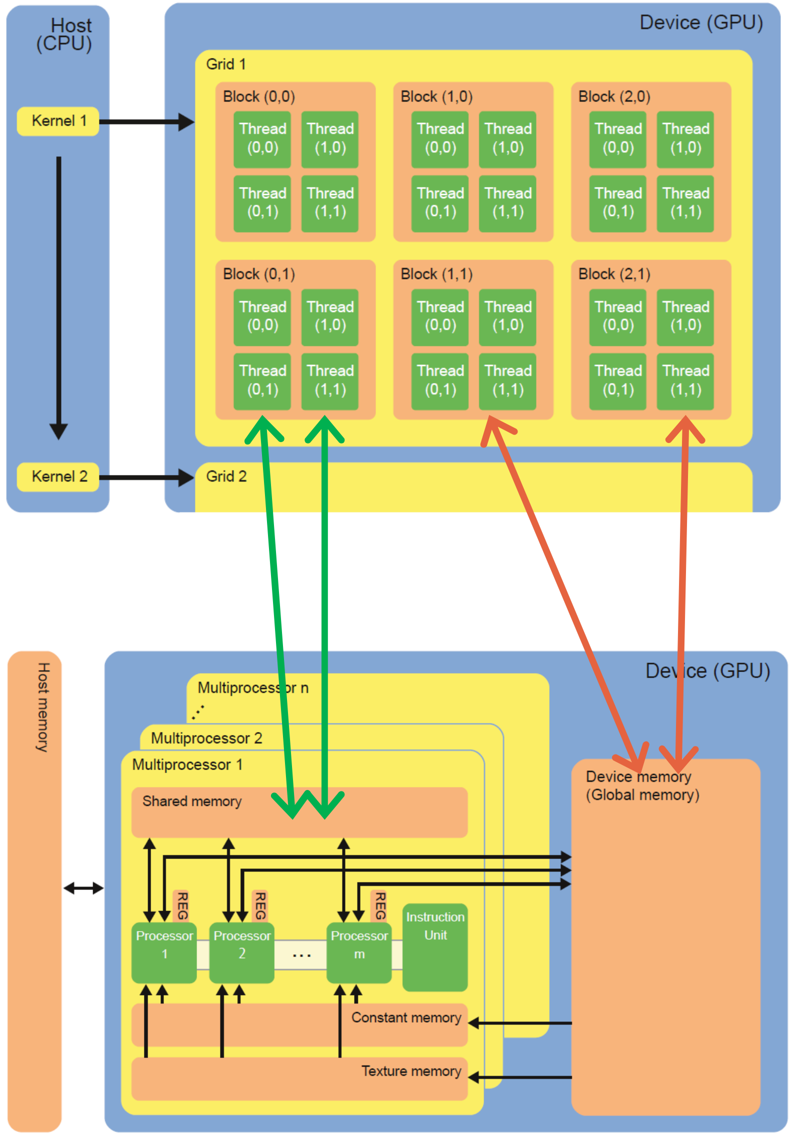
Parallel execution of a GPU program ("kernel") in a grid of thread blocks. Threads within a block work synchronously on the same data set. Different blocks are scheduled for execution independent of each other.





Each block of threads can be thought of as being executed on a single multiprocessor unit





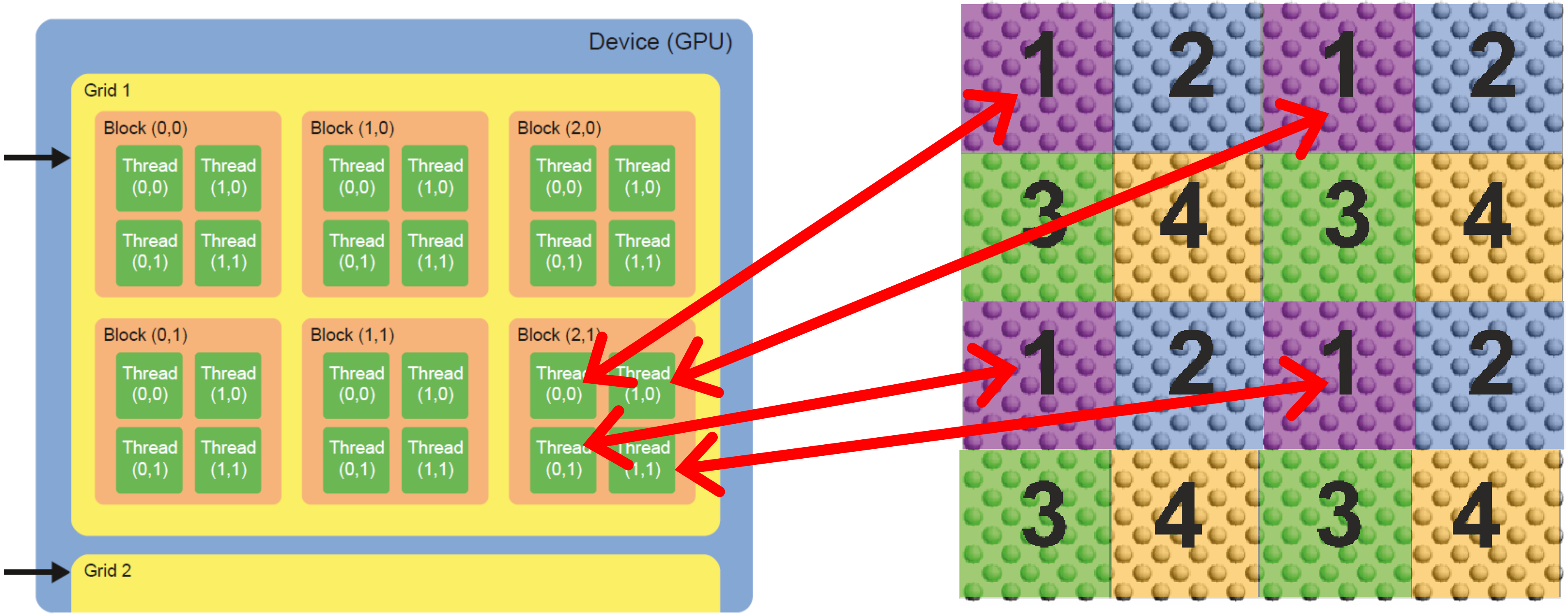
Block:

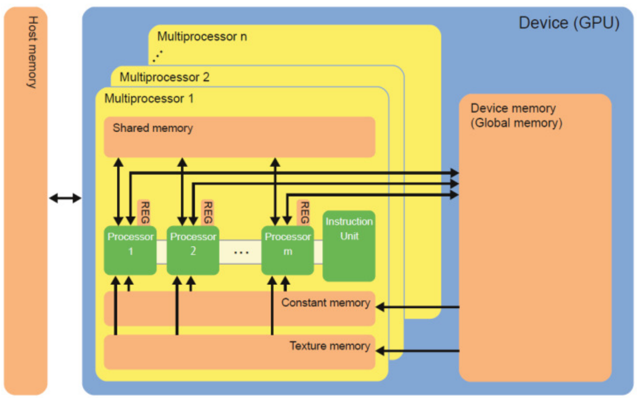
- Threads (max 512 for Tesla; 1024 for Fermi)
- Its threads access the same bank of shared memory concurrently
- Each thread should execute (exactly) the same instructions

Grid:

- The blocks of a grid (up to 65536 x 65536) are scheduled independently of each other and can only communicate via global memory accesses

hierarchical level 1: parallelisation of threads





e.g. thread cells #1 treated parallel (no overlap), then cells #2, …

Example:

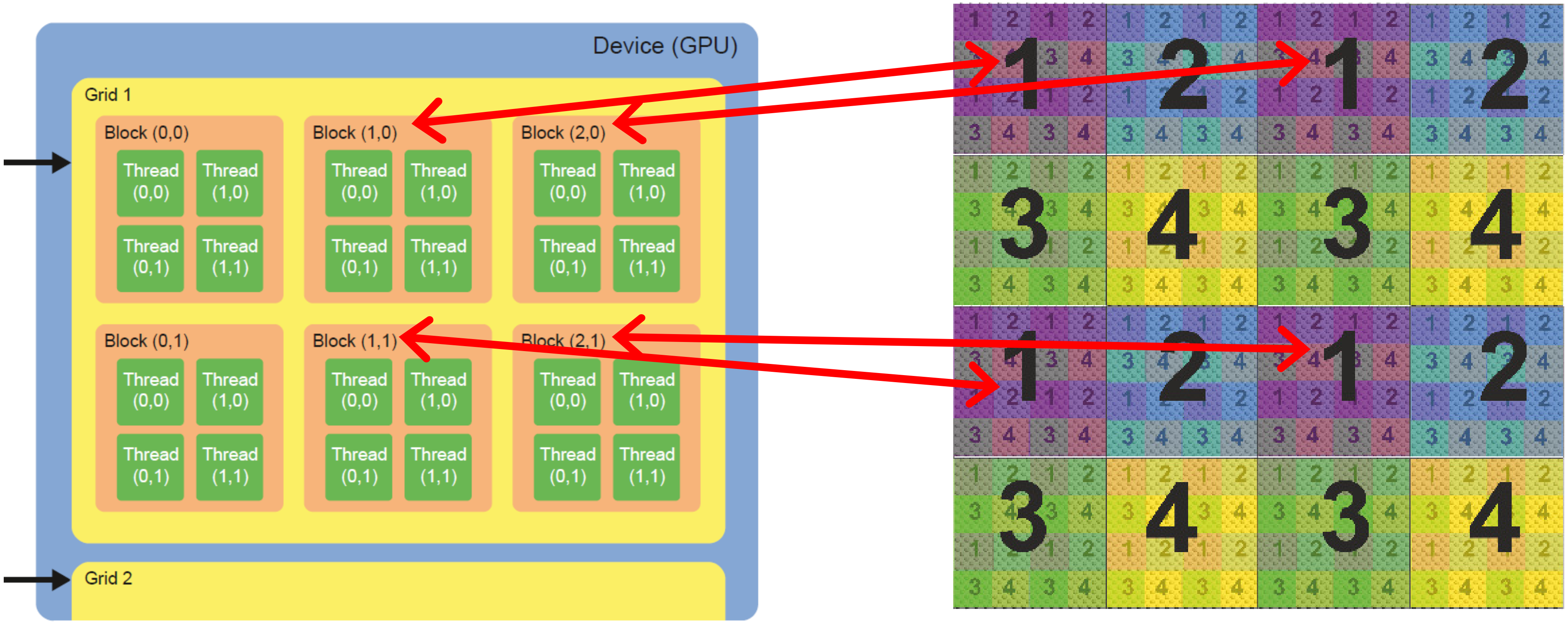
512 Threads/block

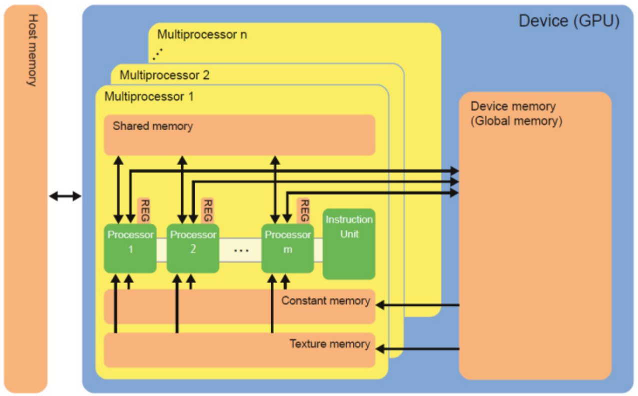
32 sites/ thread cell

→ 16000 sites/block

HZDR

hierarchical level 2: parallelisation of blocks





e.g. thread blocks #1 treated parallel (no overlap), then blocks #2, …

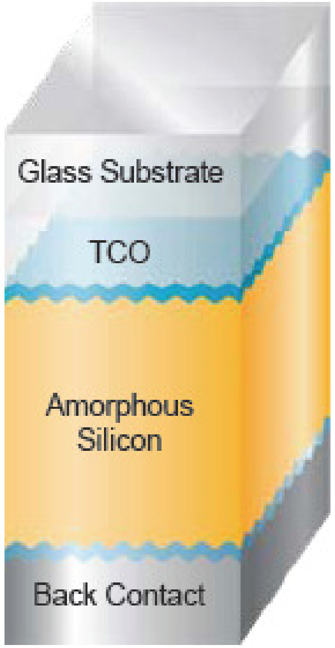
Example:

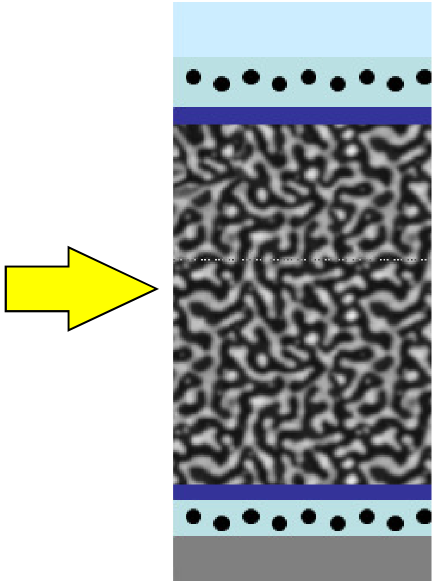
16000 sites/block

64 blocks/grid

→ 1 mill. sites

Application #3 (int. BMBF project): Synthesis Si/SiO_2 nanocomposite (sponge) for photovoltaics



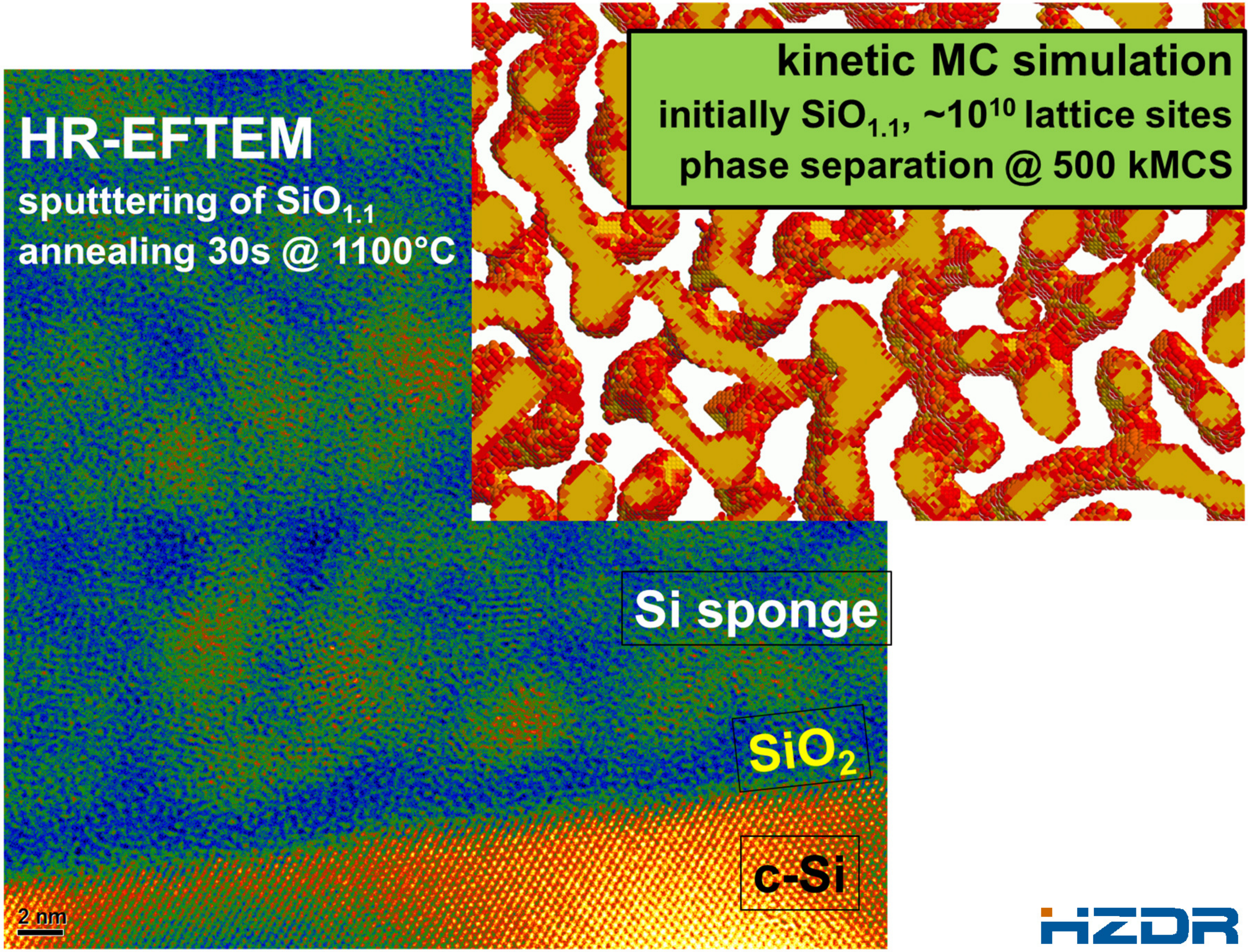


Substitution of a-Si:H by Si nanowire network or nanosponge embedded in SiO_2

Fabrication:
- SiO deposition
- heating (laser)
- phase separation
- $2SiO \rightarrow Si + SiO_2$

Process optimisation:
- Which time/temperature is needed @ laser annealing?
- How can the structure size tuned (band gap engineering)?
- Phase separation at interfaces?
- …

Morphology

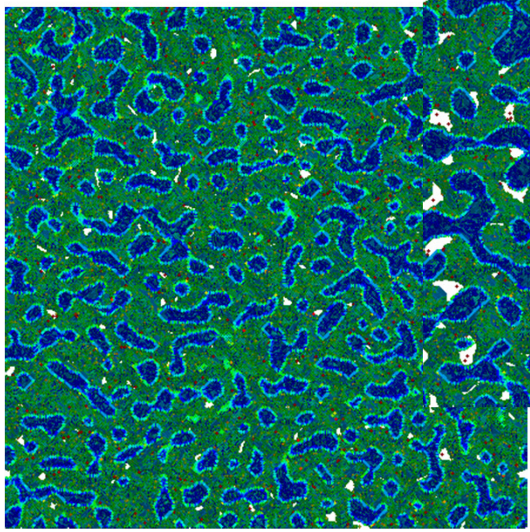


kinetic MC simulation
initially $SiO_{1.1}$, ~10^{10} lattice sites
phase separation @ 500 kMCS
HR-EFTEM
sputttering of $SiO_{1.1}$
annealing 30s @ 1100°C
Si sponge
SiO_2
c-Si
2 nm

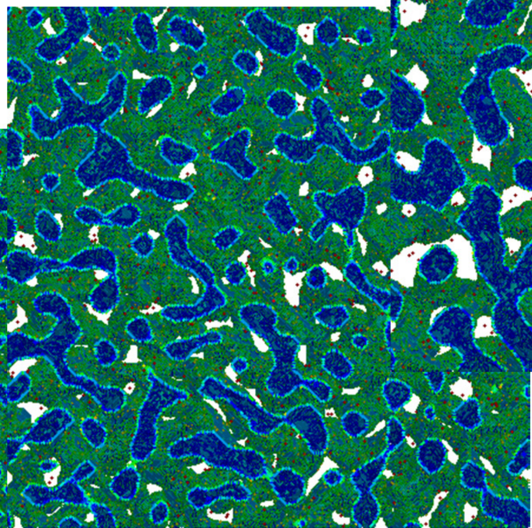
Application #3 (int. BMBF project): Synthesis Si/SiO_2 nanocomposite (sponge) for photovoltaics

system size 512^3 = 128.000.000 lattice sites
initial state: 32,5% Si sites; 67,5 % SiO_2 sites
eff. temperature ε = 1.5

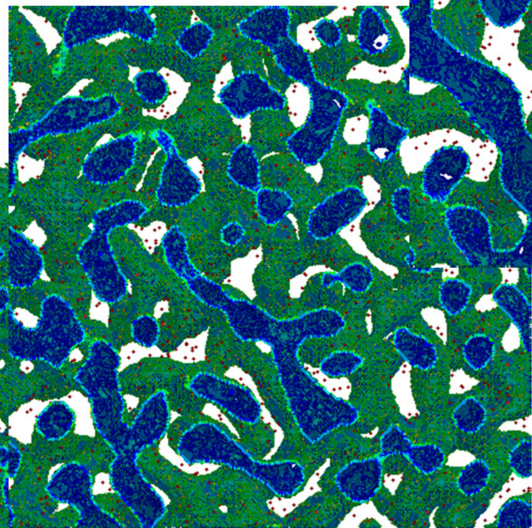
0.1 MMCS



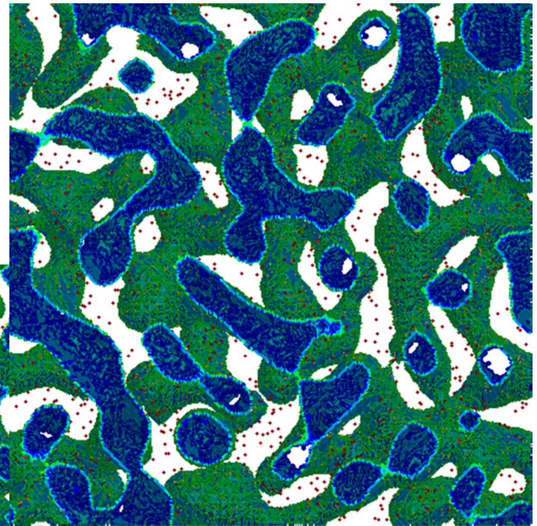
0.6 MMCS



2 MMCS

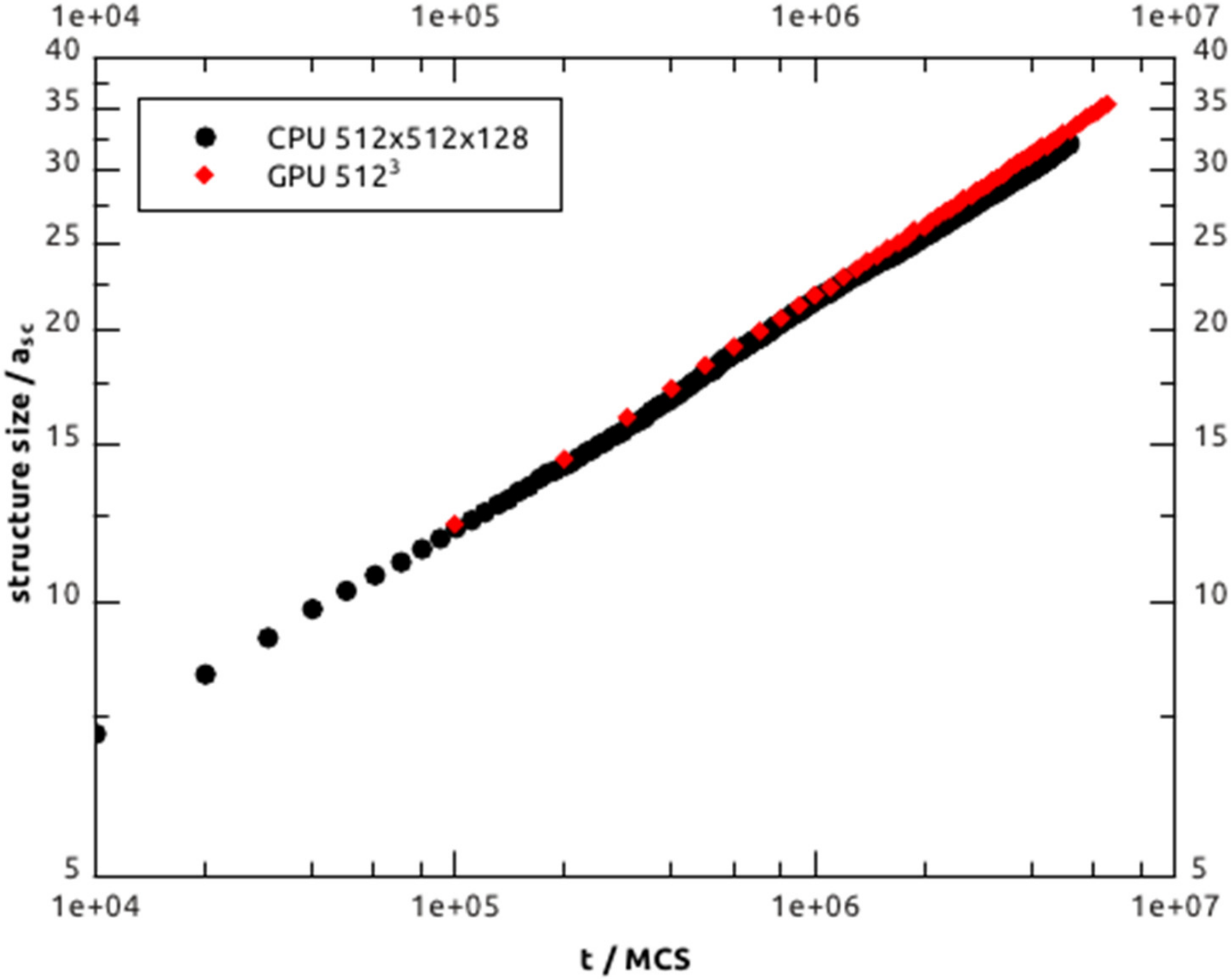


7.5 MMCS



Application #3 (int. BMBF project): Synthesis Si/SiO_2 nanocomposite (sponge) for photovoltaics

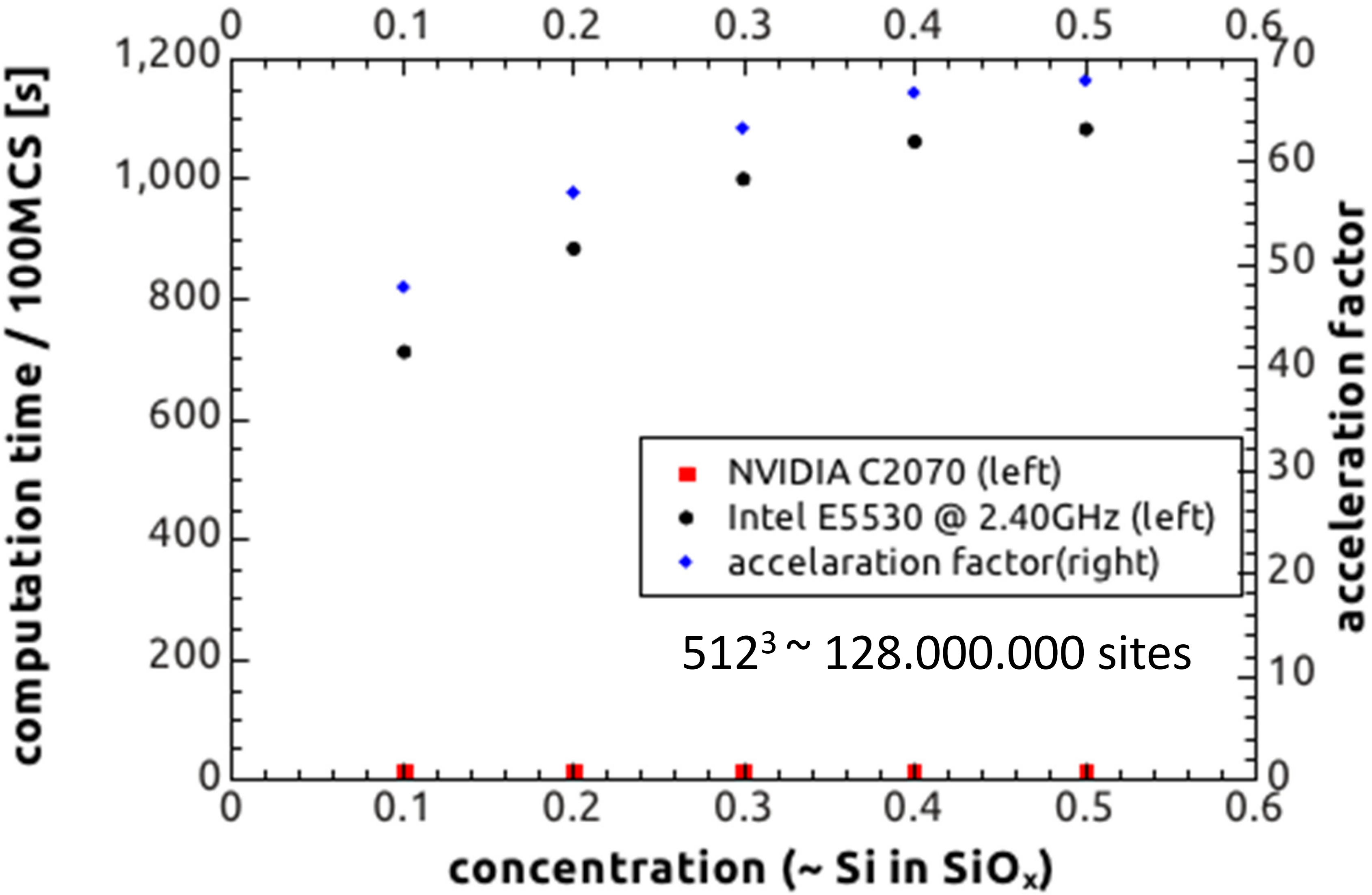
Scaling of the structure size = coarsening during phase separation: comparison CPU vs GPU



CPU results by Bartosz Liedke

HZDR

Comparison E5530 @ 2.4 GHz vs NVIDIA C2070: ~70x



512^3 ~ 128.000.000 sites

Summary

- CA-based, bit-coded lattice kMC is now an established method for process development/optimization in nanotechnologies.
- Using the CA concept and a fine grained lattice, a method in-between MD and kMC can be established.
- Using Massively Parallel Programming MPP on GPUs, kMC can be accelerated by about 2 orders of magnitude.