

Also *simulation* or *pseudoexperiment*

<b>REAL EXPERIMENT</b>	<b>COMPUTER EXPERIMENT</b>
Record everything in a lab notebook	Record everything in a lab notebook
Choose method (device, assay)	Choose method (MD, MC, ...)
Build the experimental apparatus (from parts)	Download/buy/write a computer program (blocks of code)
Purchase chemicals, synthesise if not available	Get a force field, fit/calculate parameters if not available (e.g., partial charges)
Prepare the experiment	Prepare initial configurations, etc.
Perform the experiment, watch what's going on	Run the code, observe time development, control quantities, etc.
Analyse and calculate	Calculate mean values (with error estimates)
Clean the laboratory	Make backups, erase temporary files

MC and MD can be applied to similar systems.

## MD

- realistic models, complex molecules (bonds, angles...)
- condensed matter in general (fluids, solutions; biochemistry)
- kinetic quantities (diffusivity, viscosity...)
- better parallelization, more packages available

## MC

- simple qualitative models (lattice, hard-sphere-like)
- dilute systems
- critical phenomena
- fluid equilibria
- overcoming barriers is easier with MC
- fewer packages available

## Systematic errors:

- inaccurate molecular model (force field)
- neglected quantum effects, neglected many-body forces ...
- small sample (finite-size effects)
- insufficient time scale (long correlations, bottleneck problems)
- method problems: integration errors (too long timestep), inappropriate thermostat/barostat, not equilibrated enough, inaccurate treatment of Coulomb forces...

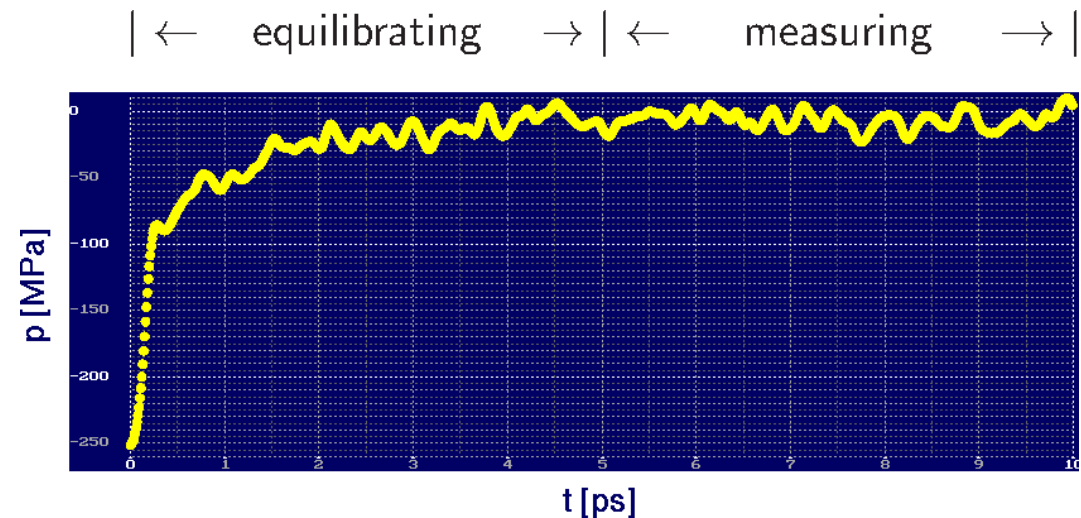
**Random (stochastic, statistical) errors** are essential in stochastic methods

- time-correlated
- can be decreased by long calculations

**Uncertainty** (in metrology) includes critical assessment of both the systematic and random errors

**Warning:** there is no generally adopted terminology

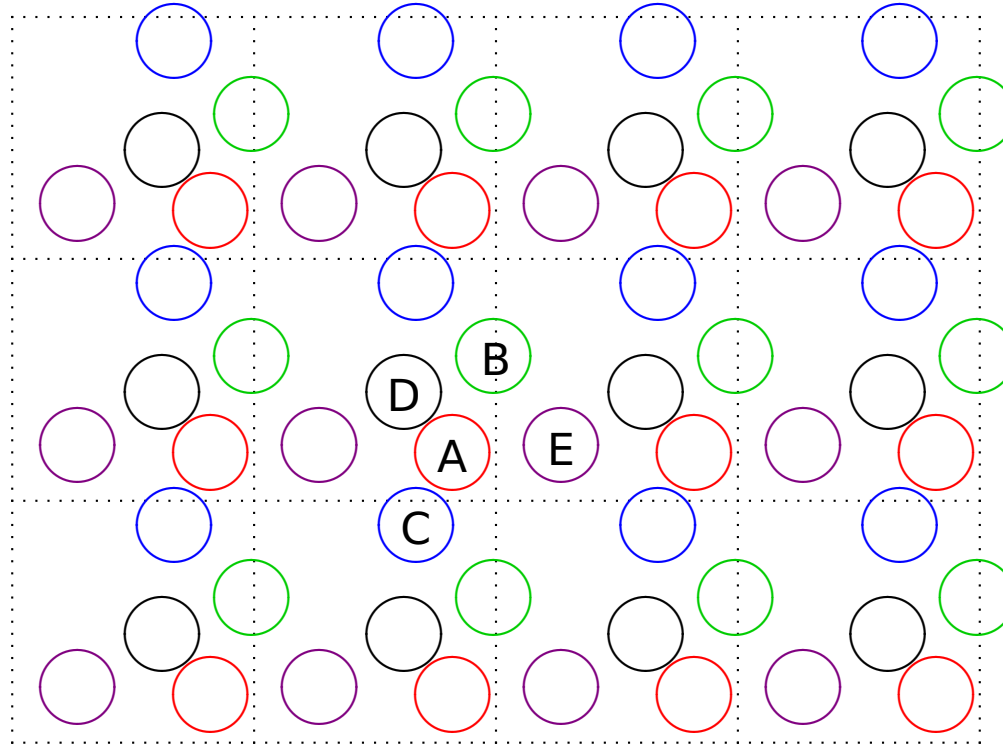
- Start (initial configuration):
  - experimental structure (biomolecules)
  - crystal  $\rightarrow$  liquid (melt), gas  $\rightarrow$  liquid (shrink); Packmol
  - random configuration (overlaps of molecules = problem in MD)  
problem for “ill-defined” models (TIP4P etc.)
  - lattice models: crystal/chaos
  - MD: velocities = Maxwell–Boltzmann (approximation enough)
- Equilibration  $\rightarrow$  watch graphically  
(convergence/time profile)
- Measuring the quantities of  
interest incl. estimates of errors



# Boundary conditions

● free (vacuum) – droplet, protein in vacuum ...

● periodic (cyclic, torroidal)



● walls (hard, soft, smoothed, made of atoms), pores, slab, ...

# Periodic boundary conditions: MD

+ 6/16  
s06/2

```
REAL L edge size of the cubic simulation box (cell)
VECTOR r1, r2 where vector r = (r.x,r.y,r.z)
           both vectors must lie in the basic box
VECTOR dr := r2 - r1 difference of vectors
           (ignoring the boundary conditions)
```

```
IF      dr.x < -L/2 THEN dr.x := dr.x + L
ELSE IF dr.x >  L/2 THEN dr.x := dr.x - L
```

```
IF      dr.y < -L/2 THEN dr.y := dr.y + L
ELSE IF dr.y >  L/2 THEN dr.y := dr.y - L
```

```
IF      dr.z < -L/2 THEN dr.z := dr.z + L
ELSE IF dr.z >  L/2 THEN dr.z := dr.z - L
```

*Vector dr now goes from r1 to the nearest image of r2*

*Squared distance to the nearest image:*

```
REAL rr := dr.x**2 + dr.y**2 + dr.z**2
```

# Periodic boundary conditions: MC

+ 7/16  
s06/2

In MC, usually the vector  $\vec{r}_{12} = r_2 - r_1$  is not needed, the distance is enough

REAL L *edge size of the cubic simulation box (cell)*

VECTOR r1, r2 *where vector r = (r.x,r.y,r.z)*

*both vectors must lie in the basic box*

VECTOR dr := r2 - r1 *difference of vectors*

*(ignoring the boundary conditions)*

```
REAL rr := (L/2 - abs(L/2-abs(dr.x)))**2  
          + (L/2 - abs(L/2-abs(dr.y)))**2  
          + (L/2 - abs(L/2-abs(dr.z)))**2
```

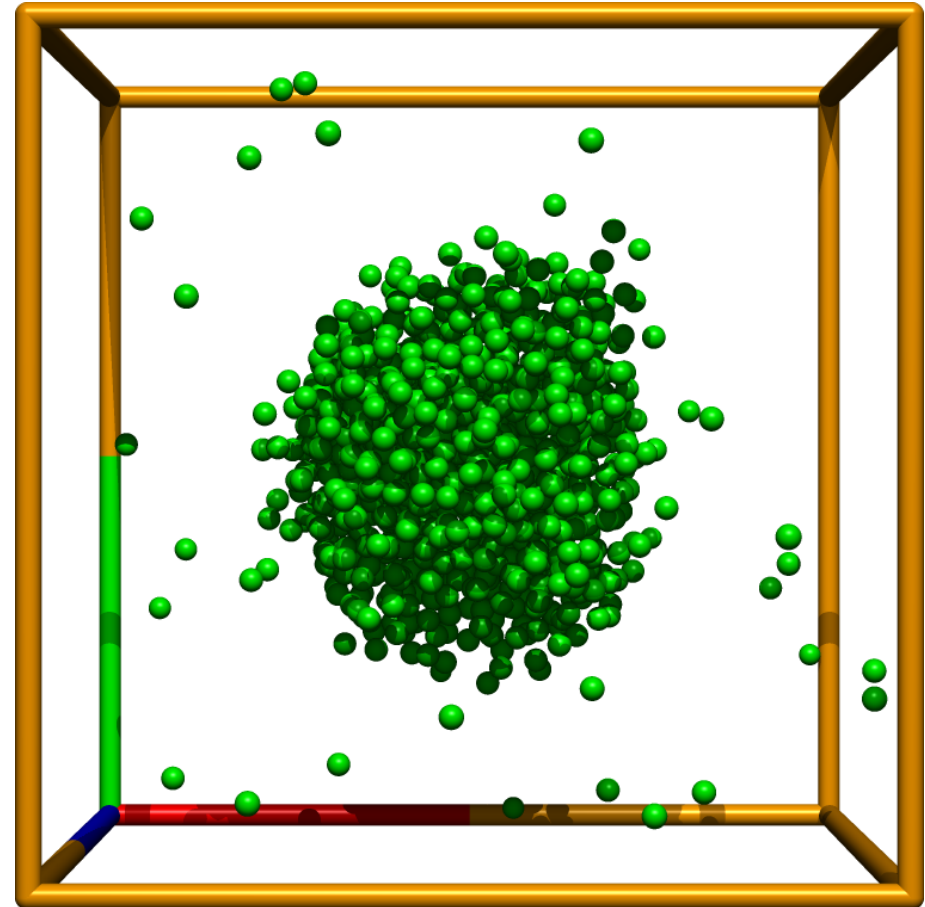
# Calculations

**Example.** We simulate an argon droplet in a periodic cubic simulation cell. Let us have 1000 atoms and temperature 85 K. The distance between surfaces of periodic images of droplets should be equal to the droplet diameter. Calculate the size of the box in Å. Argon density is  $1.4 \text{ g cm}^{-3}$ .

8/06

```
N[0]=1000
x=90          ! [Å]
L[0]=x L[1]=x L[2]=x
dt.prt=1 dt.plb=1 ! [ps]
LJcutoff=11    ! [Å]
cutoff=LJcutoff
;

T=85
h=0.005 noint=200 ! po 1 ps
thermostat="Berendsen" tau.T=1
x=1 ! [K*k_B]
center.K[0]=x center.K[1]=x center.K[2]=x
init="crystal" no=10 ! 10 ps
;
thermostat="Andersen"
;
thermostat="Berendsen"
no=80
center.K[0]=0 center.K[1]=0 center.K[2]=0
;
```

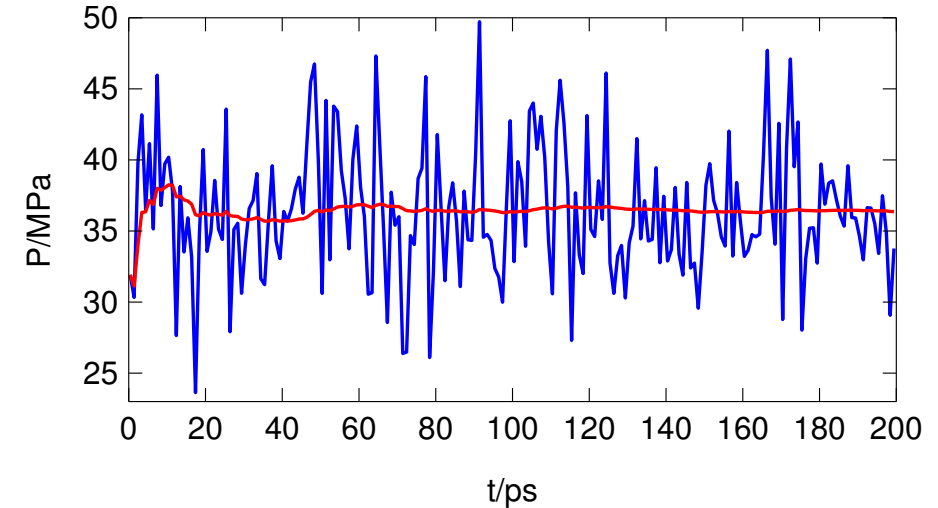




Trajectory = sequence of configurations (MD: in time)

## Convergence profile:

- time development of a quantity (time profile, —) problems better seen
- cumulative (running average, —) can estimate the inaccuracy



## Type of statistical treatment:

- averaged values ( $\leftarrow$  ergodic hypothesis)
- less often fluctuations

## Type of quantity:

- mechanical (temperature, pressure, internal energy, order parameters...)
- entropic ( $S, F, \mu, \dots$ )
- structure (correlation functions, number of neighbors, analysis of clusters...)
- auxiliary or control quantities (order parameters, integrals of motion in MD)

quantity = (estimate of the mean value)  $\pm$  (estimate of the error)

Arithmetic average (example of a **statistic**\*):

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i$$

**Standard error** = standard deviation of the statistic, usually denoted as  $\sigma$

$$\sigma_X = \sqrt{\langle (\bar{X} - \langle X \rangle)^2 \rangle}$$

For **uncorrelated** (independent)  $X_i$  and large  $m$ ,  $\bar{X}$  has Gaussian distribution.

●  $\langle X \rangle \in (\bar{X} - \sigma_X, \bar{X} + \sigma_X)$  with probability  $\approx 68\%$

●  $\langle X \rangle \in (\bar{X} - 2\sigma_X, \bar{X} + 2\sigma_X)$  with probability  $\approx 95\%$

The estimate of the standard error of uncorrelated data:

$$\sigma_X^{\text{estim}} = \sqrt{\frac{\sum_{i=1}^m \Delta X_i^2}{m(m-1)}}, \quad \text{where } \Delta X_i = X_i - \bar{X}$$

\*in metrology called the *measurement function*

**Physics:**  $\sigma_X^{\text{physics}} = \sigma_X \rightarrow \sigma_X^{\text{estim}}$  = estimated standard error/uncertainty; loosely (estimated) error/uncertainty, standard deviation (= of the average or other statistic). Often denoted  $\sigma$ .

Common notation:  $123.4 \pm 0.5 \equiv 123.4(5)$

Custom certainty level =  $5\sigma_X$  (confidence level 0.999 999 43)

**Biology, economy, engineering:** level of confidence 95 % (data are with the probability of 95 % in the interval given). In case of a Gaussian distribution:

$$\sigma_X^{\text{biology}} \approx 2 \sigma_X^{\text{physics}}$$

**Chemistry:** mostly ignored, if given, nobody knows whether  $\sigma_X^{\text{chemistry}} = \sigma_X$  or  $2\sigma_X$

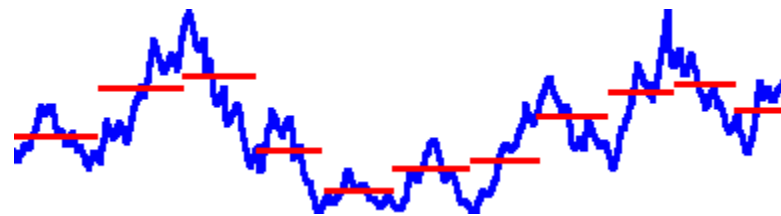
The type of error/uncertainty must be specified

# Analysis of time series and error estimation

**Problem:** correlations (example:  $x_{i+1} = rx_i + u_{\text{Gauss}}$ )

- block method:  $\bar{X}_j = \frac{1}{B} \sum_{i=1}^B X_{i+(j-1)B}$

- analysis of correlations  $\Rightarrow$

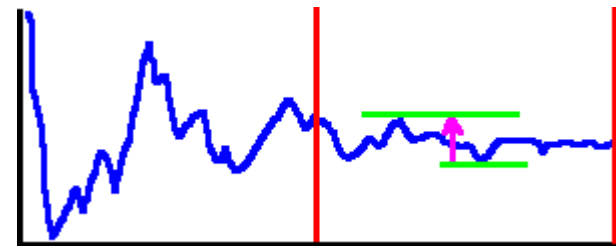


$$\sigma_X = \sqrt{\frac{\sum_{i=1}^m \Delta X_i^2}{m(m-1)} (1 + 2\tau)} \quad \tau = \sum_{k=1}^{\infty} c_k \quad c_k = \frac{\langle \Delta X_0 \Delta X_k \rangle}{\langle (\Delta X)^2 \rangle}$$

MC:  $c_k$  is monotonously decreasing [ex.:  $c_k = \sum_{\lambda \neq 1} c_\lambda \lambda^k, \lambda \in (-1, 1)$ ]

MD:  $c_k \rightarrow c(t)$  (time autocorrelation function): damped oscillations

- even better = both approaches combined:  
first to block a bit, then  $\tau \approx c_1$



- from running average (roughly  $\approx 10$  blocks):

$$\sigma_X^{\text{estim}} \approx 0.6 [\max_{2\text{nd half}}(X) - \min_{2\text{nd half}}(X)]$$

or to be on the safe side (this formula is approximate):

$$\text{err}_X \approx \max_{2\text{nd half}}(X) - \min_{2\text{nd half}}(X)$$

$\Rightarrow \langle X \rangle \in (\bar{X} - \text{err}_X, \bar{X} + \text{err}_X)$  with probability  $\approx 85\%$  (for long enough time series)

## Exercise/Example

- Generate random correlated data (1st order process):

$$X_{k+1} = qX_k + u$$

where  $u = u_{[0,1)}$  or  $u_{\text{Gauss}}$  etc., and  $|q| < 1$ .

- Calculate the arithmetic average incl. error by different methods

Note: it is known analytically,

$$\sigma_X = \sqrt{\frac{1+q}{1-q}} \sqrt{\frac{\text{Var} X}{m}} = \frac{1}{1-q} \sqrt{\frac{\text{Var} u}{m}}$$

where the variance, or fluctuation, is defined by  $\text{Var} X = \langle (X - \bar{X})^2 \rangle$

# Time autocorrelation function

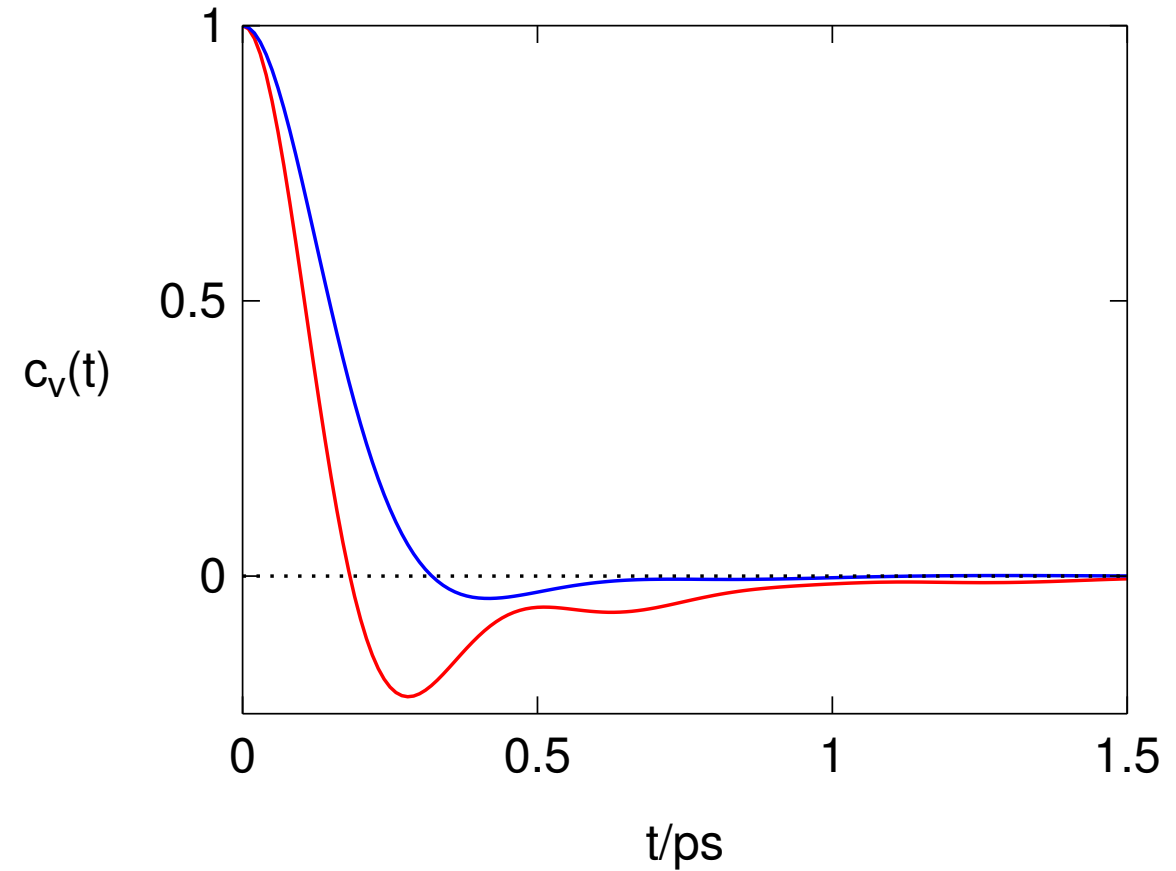
+ 14/16  
s06/2

Velocity-velocity autocorrelation function of liquid argon:

— 150 K,  $1344 \text{ kg m}^{-3}$ ,

— 120 K,  $1680 \text{ kg m}^{-3}$ .

Results from a 100 ps trajectory for 216 Lennard-Jones particles



Typical behavior (MC + MD):

- fluid:  $\lim_{t \rightarrow \infty} c(t) = \text{konst } t^{-3/2}$  (hydrodynamic tail)
- jumps between states:  $c(t) \propto \lambda^t$  ( $\lambda$  just below 1)

Sum of independent measurements: squares of standard deviations are additive

**Example.** We calculate  $I = \int_0^1 f(x)dx$  approximately by the Simpson formula,

$$I = \int_0^1 f(x)dx \approx \frac{1}{6} [f(0) + 4f(0.5) + f(1)]$$

For  $f(x)$  we have the following data with standard errors:

$x$	0	0.5	1
$f(x)$	1.34(5)	1.57(3)	1.77(6)

Calculate  $I$  including the error estimate.

$$I = \frac{1}{6} [1.34 + 4 \times 1.57 + 1.77] = 1.565$$

$$\sigma(I)^2 = (0.05/6)^2 + (0.03 \times 4/6)^2 + (0.06/6)^2 = 0.000569 \Rightarrow \sigma(I) = 0.024$$

$$I = 1.565(24)$$

For division and multiplication, the same holds true for the relative errors

**Example.** Calculate  $3.46(7)/0.934(13)$ .

Error of function  $f$  of a variable with error is (linearized; i.e., for small  $\sigma$ ):

$$f(x \pm \sigma_x) = f(x) \pm f'(x)\sigma_x$$

$$\ln(x \pm \sigma_x) = \ln x \pm \frac{\sigma_x}{x}, \quad \exp(x \pm \sigma_x) = \exp x \pm \sigma_x \exp x,$$

$$\frac{1}{x \pm \sigma_x} = \frac{1}{x} \pm \frac{\sigma_x}{|x|^2}$$

**Example.** Calculate the activity of  $H^+$  from  $pH = 2.125(5)$ .

$$a(H^+) = 0.00750(9)$$