Molekulový počítačový experiment

též „pseudoexperiment“

REÁLNÝ EXPERIMENT	**POČÍTAČOVÝ EXPERIMENT**
Vedení laboratorního deníku	Vedení laboratorního deníku
Zvol metodu (přístroj, protokol)	Zvol metody (MD, MC, . . .)
Stavba aparatury (z částí)	Stáhni/kup/napiš počítačový program, slož bloky kódu
Nakup chemikálie, syntetizuj, co není ke koupi	Stáhni silové pole, nafituj parametry, které nejsou dostupné
Připrav experiment	Připrav počáteční konfiguraci ap.
Proveď experiment, pozorně sleduj, co se děje	Spusť program, sleduj časovou závislost veličin vč. kontrolních
Analyzuj a počítej	Stanov střední hodnoty (s odhady chyb)
Ukliď laboratoř, popiš vzorky	Zapiš zálohy, vymaž nepotřebné soubory

MD nebo MC?

MC a MD se často dají použít na podobné systémy

MD

- realistické modely, složité molekuly (vazby, úhly. . .)
- kondenzovaná fáze obecně (tekutiny, roztoky; biochemie)
- kinetické veličiny (difuzivita, viskozita. . .)
- snazší paralelizace, existuje mnoho balíků

MC

- jednoduché kvalitativní modely (mřížkové, tuhé koule apod.)
- zředěné systémy
- kritické jevy
- fázové rovnováhy
- překonávání bariér, výměna molekul aj. triky jsou v MC snazší
- horší paralelizace, existuje málo balíků

Je to správně?

Systematické chyby:

- nepřesný molekulární model (silové pole)
- zanedbání kvantových jevů, nepárových sil (např. polarizovatelnosti)
- malý vzorek (*finite-size effects)*
- nedostatečná časová škála (dlouhé korelace, „hrdlo láhve“)
- nevhodná metodologie: chyby integrace (příliš dlouhý krok), nedostatečně zrovnovážněno, nepřesné Coulombovy síly . . .

Náhodné (statistické) chyby* jsou principiální pro stochastické metody

- časově korelované
- lze zmenšit delším výpočtem

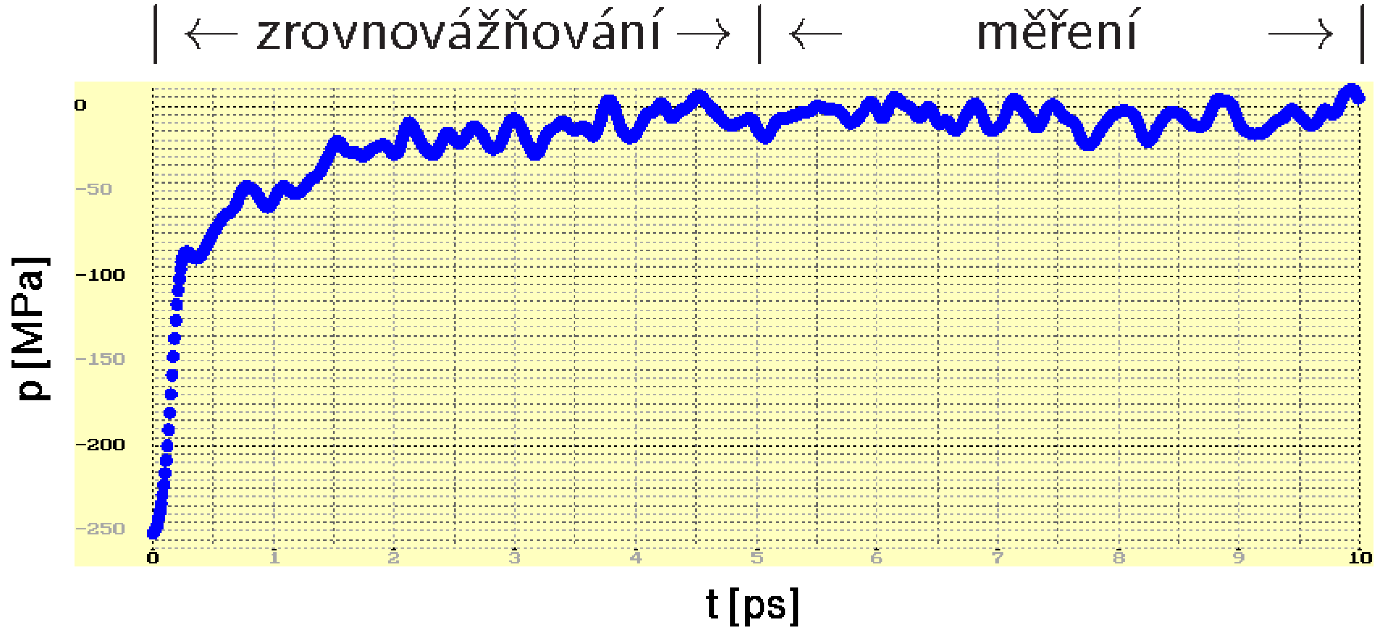
Nejistota* (v metrologii) zahrnuje kritické posouzení systematických i náhodných chyb*

---

* Názvosloví kolísá podle oboru (matematická statistika, fyzika, metrologie, chemie)

Metodika simulace

- Start simulace (počáteční konfigurace):
  - experimentální struktura (biomolekuly)
  - krystal → kapalina, plyn → kapalina; Packmol
  - náhodná konfigurace (překryvy molekul = problém v MD)
    problém u špatně definovaných modelů (exp-6, TIP4P aj.)
  - možno předoptimalizovat k odstranění překryvů ap. (conjugované gradienty ap.)
  - mřížkové modely: krystal/chaos
  - MD: rychlosti = Maxwell–Boltzmann (stačí přibližně)
- Zrovnovážňování:
  sledovat časový (konvergenční) profil
- Měření veličin
  vč. odhadu statistických chyb (nejistot)



Výpočty

Příklad. Simulujeme kapku argonu v kubických periodických okrajových podmínkách. Mějme $N = 1000$ atomů za teploty $T = 85$ K. Vzdálenost mezi povrchy periodických obrazů kapky by měla být rovna průměru kapky. Vypočtěte velikost boxu v Å. Hustota kapalného argonu je $\rho = 1.4\,\mathrm{g\,cm^{-3}}$, molární hmotnost $M(\mathrm{Ar}) = 40$ g/mol.

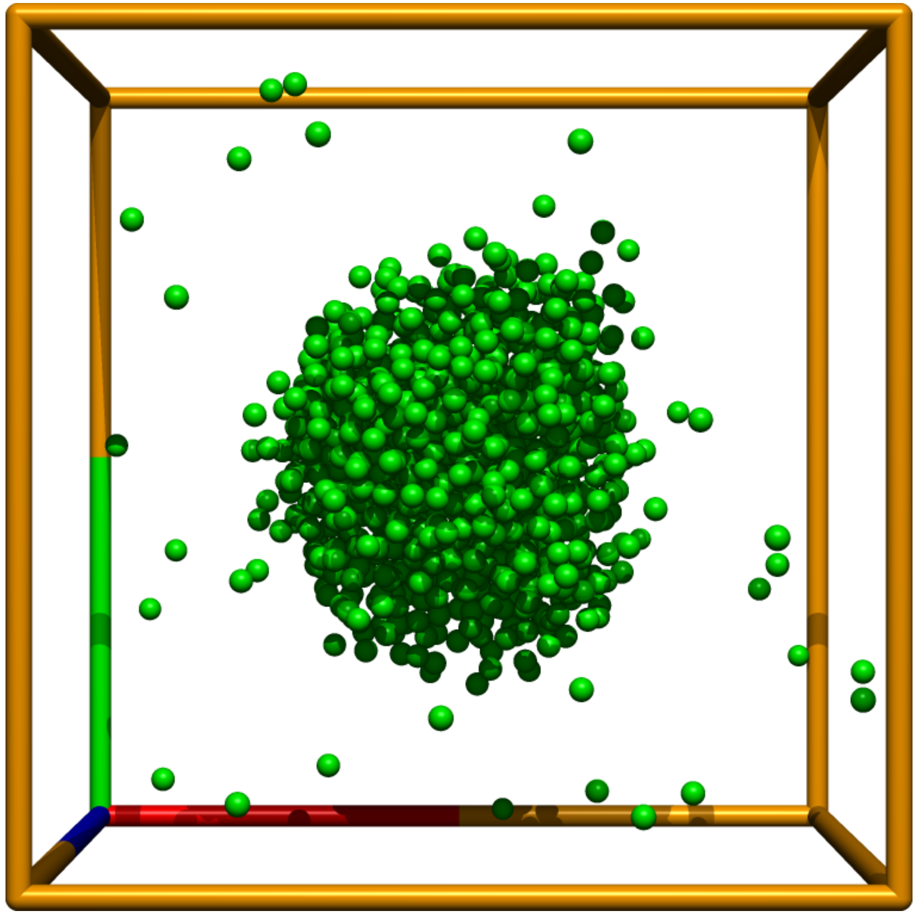
molární objem: $V_\mathrm{m} = M/\rho$

objem na 1 atom: $V_1 = V_\mathrm{m}/N_\mathrm{A}$

objem N atomů: $V = NV_1 = NM/\rho N_\mathrm{A}$
$= 1000 \cdot 0.040\,\mathrm{kg\,mol^{-1}}/(1400\,\mathrm{kg\,m^{-3}} \cdot 6.022\times10^{23}\,\mathrm{mol^{-1}})$
$= 4.744\times10^{-27}\,\mathrm{m^3}$

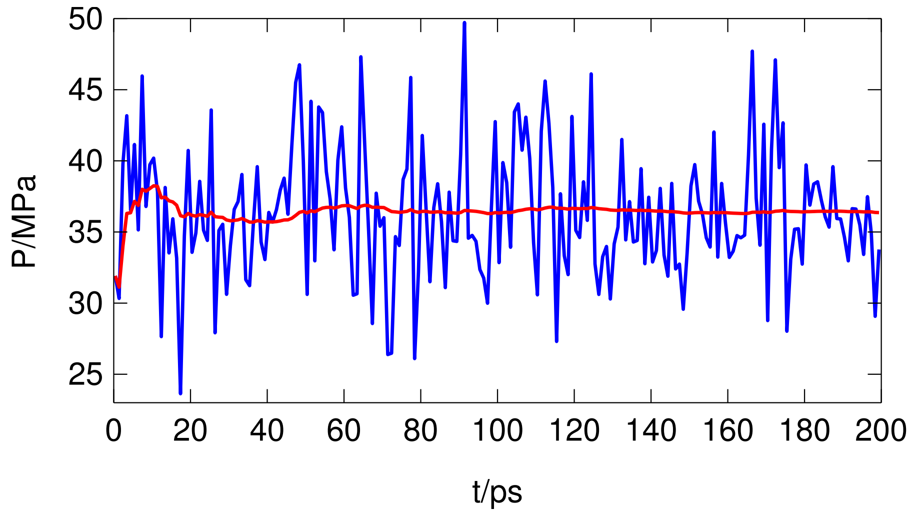
poloměr kapky: $\frac{4}{3}\pi R^3 = V \Rightarrow R = 2.24\times10^{-9}$ m

velikost boxu: $\underline{L = 90\ \text{Å}}$



Měření

Trajektorie = posloupnost konfigurací (MD: v čase)



Konvergenční profil:

- vývoj veličiny (časový profil, —)
  problémy jsou líp vidět
- kumulativní (*running average*, —)
  lze odhadnout nepřesnost

Typ zpracování:

- střední hodnoty (← ergodická hypotéza)
- méně často fluktuace

Typ veličiny:

- mechanické (teplota, tlak, vnitřní energie, parametry uspořádání. . .)
- entropické (S, F, μ,. . .)
- strukturní (korelační funkce, počet sousedů, analýza klastrů. . .)
- pomocné či kontrolní veličiny (parametry uspořádání, integrály pohybu v MD)

Náhodné chyby

veličina = (odhad střední hodnoty) ± (odhad chyby)

Aritmetický průměr = příklad **statistiky**; též „statistický funkcionál“, v metrologii „měřicí funkce“

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

Standardní chyba = standardní odchylka statistiky, obv. se označuje σ

$$\sigma(\overline{X}) = \sqrt{\left\langle \left(\overline{X} - \langle X \rangle\right)^2 \right\rangle}$$

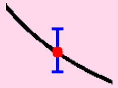
Pro **nekorelované** (nezávislé) X_i a velké m má $\overline{X}$ Gaussovo normální rozdělení.

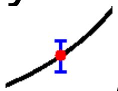
Odhad standardní chyby aritmetického průměru z nekorelovaných dat:

$$\sigma(\overline{X}) \approx \sqrt{\frac{\sum_{i=1}^{m} \Delta X_i^2}{m(m-1)}}, \qquad \text{kde } \Delta X_i = X_i - \overline{X}$$

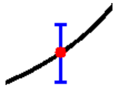
Zvyky a zlozvyky

Jak udávají nejistotu měřených hodnot různé obory:

v příp. nesymetrie $Q = 1.23^{+0.67}_{-0.45}$ 

- **Fyzika:** $Q = 123.4 \pm 0.5 \equiv 123.4(5) \equiv 123.4_5$  ,
  kde 0.5 = $\sigma(Q)$ = odhadnutá směrodatná/standardní chyba/nejistota statistiky Q (např. $Q = \overline{X}$) počítané z výběru (*sample*),
  také:standardní/směrodatná odchylka (rozumí se aritmetického průměru či jiné statistiky)
  nepřesně jen:(odhadnutá) chyba/nejistota, standardní/směrodatná odchylka

  V případě normálního rozdělení s pravděpodobností 68 % platí $\langle Q \rangle \in 123.4 \pm 0.5$

- **Biologie, ekonomie, inženýrství, politologie, farmakologie:** $Q = 123.4 \pm 1.0$ 
  $\pm 1.0 = \pm 2\sigma(Q)$ = interval spolehlivosti (*confidence interval*) na hladině (spolehlivosti) 95 %
  nepřesně jen: ±1.0 = interval spolehlivosti, 1.0 = chyba/nejistota, . . .

  V případě normálního rozdělení s pravděpodobností 95 % platí $\langle Q \rangle \in 123.4 \pm 1.0$

- **Chemie**: často ignorováno; pokud udáno, tak nikdo neví, jaká je hladina spolehlivosti ☹

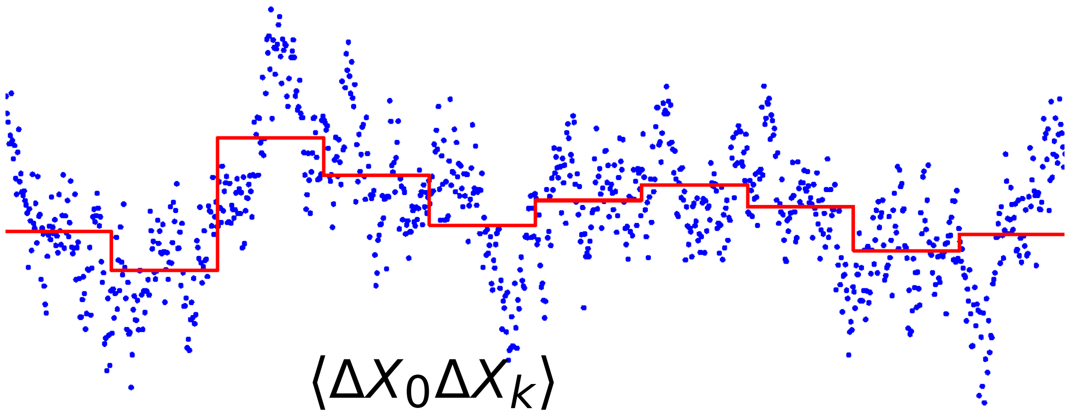
- **„Fyzikální jistota“** začíná na $\pm 5\sigma(Q)$ (hladina spolehlivosti 0.999 999 43)

Vždy nutno udat typ chyby/nejistoty resp. hladinu spolehlivosti

α = hladina významnosti (*significance level*), často 5 %
$1 - \alpha$ = hladina spolehlivosti (*confidence level*), často 95 %

Analýza časové řady a odhad náhodných chyb

Problém: korelace

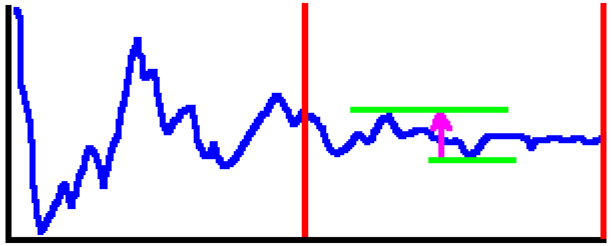


- bloková metoda: $\overline{X}_j = \frac{1}{B}\sum_{i=1}^{B} X_{i+(j-1)B}$
- analýza korelací ⇒

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

MC: c_k monotonně klesá [přesně: $c_k = \sum_{\lambda\neq 1} c_\lambda \lambda^k$, $\lambda \in (-1,1)$]
MD: $c_k \to c(t)$ (časová autokorelační funkce): tlumené oscilace

- lépe = kombinace:
mírně zblokovat a pak $\tau \approx c_1$



- running average (≈ 10 bloků):

$$\sigma_X \approx 0.6[\max_{2.\ \text{půlka}}(X) - \min_{2.\ \text{půlka}}(X)]$$

nebo pro jistotu (vzorec je nepřesný):

$$\text{“chyba”} \approx \max_{2.\ \text{půlka}}(X) - \min_{2.\ \text{půlka}}(X)$$

pak $\langle X\rangle \in (\overline{X} - \text{“chyba”}, \overline{X} + \text{“chyba”})$ s pravděp. ≈ 85% (pro dlouhé časové řady)

Cvičení/ukázka

- Vygenerujte korelovaná náhodná data (proces 1. řádu)

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

kde $u = u_{[0,1)}$ nebo u_{Gauss} apod. a $|q| < 1$.

- Vypočtěte aritmetický průměr i chybu různými metodami.

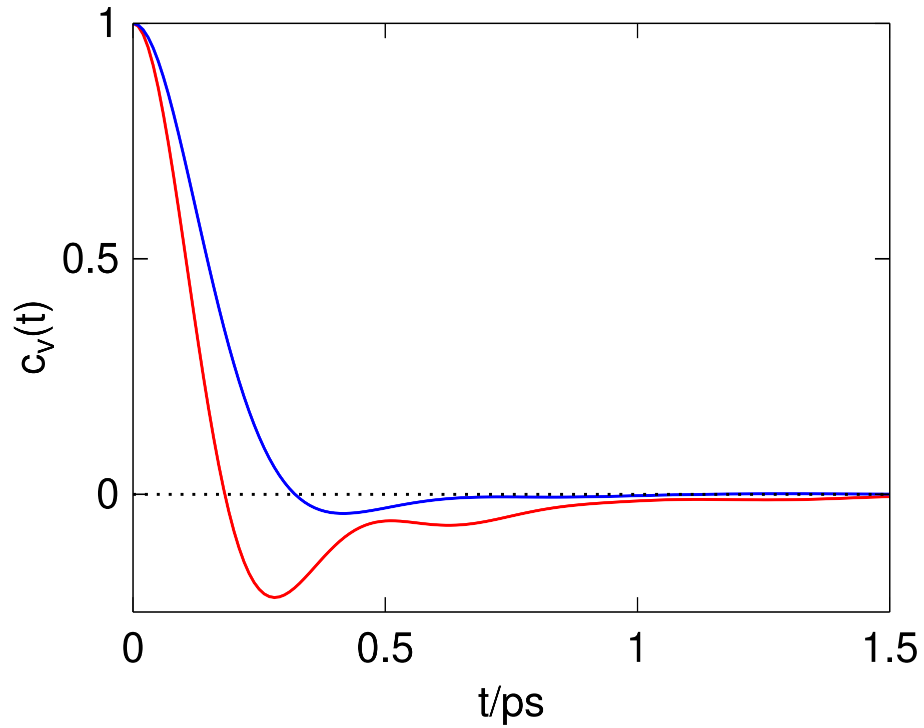
Pozn.: analytický výsledek je

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

kde variance (fluktuace) je $\mathrm{Var}\,X = \langle (X-\overline{X})^2\rangle$

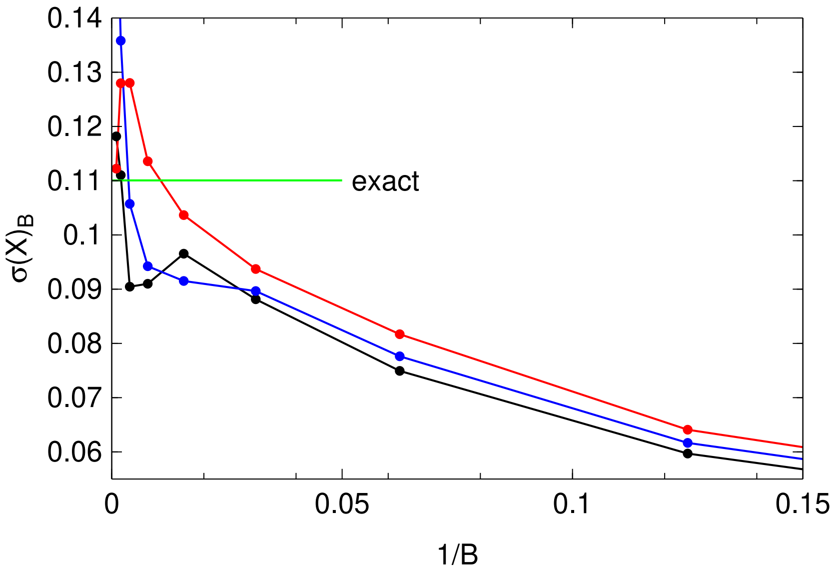
Časová autokorelační funkce

Rychlostní časová autokorelační funkce (*velocity–velocity autocorrelation function*) kapalného argonu:
— 150 K, 1344 kg m^{-3},
— 120 K, 1680 kg m^{-3}.
Výsledky trajektorie dlouhé 100 ps pro 216 Lennard--Jonesových částic



Typické chování (MC i MD):

- tekutina: $\lim_{t\to\infty} c(t) = \text{konst}\, t^{-3/2}$ (hydrodynamic tail)
  bloková metoda: $\sigma(X)_B$ je lineární funkcí $1/\sqrt{B}$ pro $B \to \infty$
- přeskoky mezi stavy (vysoká bariéra): $c(t) \propto \lambda^t$ (λ těsně pod 1)
  bloková metoda: $\sigma(X)_B$ je lineární funkcí $1/B$ pro $B \to \infty$



Analýza chyb – sčítání a odčítání

Součet nezávislých měření: dvojmoci standardních odchylek jsou aditivní

Příklad. Počítáme $I = \int_0^1 f(x)\mathrm{d}x$ přibližně Simpsonovým vzorcem,

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

Pro $f(x)$ máme následující data se standardními chybami:

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

Vypočtěte I včetně odhadu chyb.

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

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

$$\underline{I = 1.565(24)}$$

Analýza chyb – násobení a dělení

Dvojmoci relativních chyb jsou aditivní

Příklad. Vypočtěte 3.46(7)/0.934(13).

podíl: $3.46/0.934 = 3.704$

rel. chyba $= \sqrt{\left(\frac{0.07}{3.46}\right)^2 + \left(\frac{0.013}{0.934}\right)^2} = 0.0246$

abs. chyba $= 3.704 \times 0.0246 = 0.091$

$\underline{3.46(7)/0.934(13) = 3.70(9)}$ nebo zaokrouhleno nahoru: 3.70(10)

Analýza chyb

Chyba funkce f proměnné s chybou je (linearizovaný vztah, tj. pro malé σ):

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

$$\ln(x \pm \sigma_x) = \ln x \pm \frac{\sigma_x}{x}, \qquad \exp(x \pm \sigma_x) = \exp x \pm \sigma_x \exp x, \qquad \frac{1}{x \pm \sigma_x} = \frac{1}{x} \pm \frac{\sigma_x}{|x|^2}$$

Příklad. Vypočtěte aktivitu H^+ z pH = 2.125(5).

aktivita:

$$a_{H^+} = 10^{-2.125} = \exp(-2.125 \times \ln 10) = 0.00750$$

metoda 1:

$$\sigma = 0.005 \times \ln 10 \times a = 0.000086$$

metoda 2:

$$\sigma = |10^{-2.125} - 10^{-2.125-0.005}| = 0.000087$$

aktivita s odhadem nejistoty:

$$\underline{a_{H^+} = 0.00750(9)}$$