

Náročné simulace a zpracování dat ve vědě

Jiří Filipovič
fila@ics.muni.cz

Ústav výpočetní techniky, MU

21. října 2024

- simulace a modelování – co to je, proč nás zajímá
 - výpočetní náročnost – kdy potřebujeme zkušeného informatika
 - rekonstrukce dat z cryo-elektronové mikroskopie
 - simulace transportních procesů v proteinech
 - obecné úvahy o tom, co jsme se naučili
-
- společný příběh: co může informatik udělat pro léčbu COVID19?

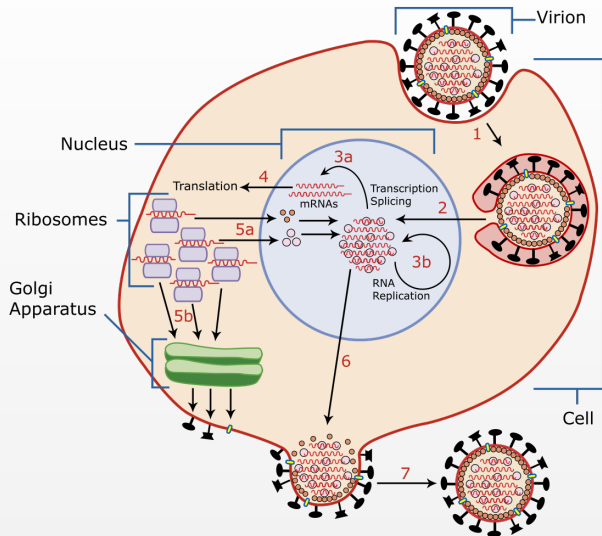
Virus

- cizopasník na pomezí mezi živým a neživým objektem
- „špatná zpráva v obálce”

Napadnutí buňky (zjednodušeně)

- navázání viru na buňku (její specifický receptor)
- penetrace do buňky (buňka virus „pozře”, nebo sfúzují membrány)
- replikace (buňka vyrobí kopie viru)
- uvolnění viru z buňky

Napadnutí buňky



Úvod

Osnova

O virech

Modelování a simulace

Náročné simulace

Šíření tepla

Paratelizace

cryo-EM

Motivace

Jak funguje

Práce pro informatika

COVID

Transport v proteinech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Úvod

Osnova

O virech

Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Jak bojovat s viry

- můžeme napadnout jeden z kroků v šíření viru (např. mu neumožnit navázat se na buňku)
- pro vývoj léčiv a vakcín musíme o viru něco vědět:
 - sekvence
 - struktura proteinů
 - mechanismus interakce s buňkou
- toto nelze jen tak pozorovat pod mikroskopem
 - ke všem krokům potřebujete počítač

Úvod

Osnova

O virech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

SARS-CoV-2

- objeven v roce 2019 ve Wu-chanu
- rychle se začal šířit do celého světa
- chceme tomu zabránit
- v této přednášce se dozvíte (nikoliv kompletní výčet) využití informačních technologií proti tomuto viru

Úvod

Osnova

O vírech

**Modelování a
simulace**

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Model

- fyzická či mentální náhrada reálného systému, vždy zjednodušená
 - hliněná maketa karoserie automobilu
 - idealizovaná představa voliče
 - matematická reprezentace molekuly
- napodobuje chování systému, o které se zajímáme

Modelování

- proces vytváření a zdokonalování modelu

Simulace

- proces, při kterém používáme model za účelem studia jeho vlastností
- ... a ideálně i vlastností reálného systému

Příklady využití simulací

Model interiéru automobilu

- testujeme pohodlí a ergonomii, automobil nemusí jezdit, prvky v interiéru nemusí plnit svou funkci
- pokud je něco špatně, snižuje náklady na vývoj a výrobu

Úvod

Osnova

O vírech

Modelování a
simulace

Náročná simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Příklady využití simulací

Model interiéru automobilu

- testujeme pohodlí a ergonomii, automobil nemusí jezdit, prvky v interiéru nemusí plnit svou funkci
- pokud je něco špatně, snižuje náklady na vývoj a výrobu

Simulace konfliktu se zákazníkem

- v bezpečném prostředí vytvoříme situaci, jejíž zvládnutí chceme natrénovat
- snižuje riziko nevhodné reakce v reálné situaci

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Příklady využití simulací

Model interiéru automobilu

- testujeme pohodlí a ergonomii, automobil nemusí jezdit, prvky v interiéru nemusí plnit svou funkci
- pokud je něco špatně, snižuje náklady na vývoj a výrobu

Simulace konfliktu se zákazníkem

- v bezpečném prostředí vytvoříme situaci, jejíž zvládnutí chceme natrénovat
- snižuje riziko nevhodné reakce v reálné situaci

Simulace protržení přehrady

- zjišťujeme, které oblasti budou zaplaveny (a jak rychle), jako podklad k evakuačnímu plánu
- zde je cena reálného experimentu nepřípustná

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Příklady využití simulací

Model interiéru automobilu

- testujeme pohodlí a ergonomii, automobil nemusí jezdit, prvky v interiéru nemusí plnit svou funkci
- pokud je něco špatně, snižuje náklady na vývoj a výrobu

Simulace konfliktu se zákazníkem

- v bezpečném prostředí vytvoříme situaci, jejíž zvládnutí chceme natrénovat
- snižuje riziko nevhodné reakce v reálné situaci

Simulace protržení přehrady

- zjišťujeme, které oblasti budou zaplaveny (a jak rychle), jako podklad k evakuačnímu plánu
- zde je cena reálného experimentu nepřípustná

Předpověď počasí

- bez simulace není předpověď možná, můžeme jen čekat

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Příklady využití simulací

Model interiéru automobilu

- testujeme pohodlí a ergonomii, automobil nemusí jezdit, prvky v interiéru nemusí plnit svou funkci
- pokud je něco špatně, snižuje náklady na vývoj a výrobu

Simulace konfliktu se zákazníkem

- v bezpečném prostředí vytvoříme situaci, jejíž zvládnutí chceme natrénovat
- snižuje riziko nevhodné reakce v reálné situaci

Simulace protržení přehrady

- zjišťujeme, které oblasti budou zaplaveny (a jak rychle), jako podklad k evakuačnímu plánu
- zde je cena reálného experimentu nepřípustná

Předpověď počasí

- bez simulace není předpověď možná, můžeme jen čekat

Modelování procesů v molekulách

- se současnými metodami nelze pozorovat
- simulace je jedinou cestou, jak tyto procesy přímo zkoumat

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteinech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Některé simulace se zcela obejdou bez počítačů

- model interiéru automobilu
- počítačová simulace je možná, ale může být méně vhodná (dražší, méně přesná), než fyzický model

Počítače můžou zvýšit přesnost

- předpověď počasí
- lidské schopnosti překonány numerickým modelem
- syntéza obrovského množství dat nerealizovatelná člověkem

Některé simulace prakticky neproveditelné bez počítače

- interakce molekul
- obrovské množství výpočtů i pro jednoduché modely

Zvládnutelné s tužkou a papírem, či tabulkovým procesorem

- „v březnu někdo přišel s matematickým modelem”
- malé množství dat, jednoduchý model

V určitém momentě začaly požadavky na model převyšovat lidské síly

- kosmické lety
- vývoj nukleární bomby

Komplikovanější modely snadno přesáhnou možnosti dnešních počítačů

- využívají se clustery/cloudy/superpočítače
- nestačí jen zapojit hodně procesorů, obvykle je za tím velký kus informatiky
- interdisciplinární oblast, kde musí spolupracovat doménoví experti s informatiky
- v přednášce se zaměříme především na tuto oblast

Šíření tepla v materiálu lze popsat pomocí parciální diferenciální rovnice

- analyticky prakticky neřešitelná pro komplikovanější systémy (nepravidelný tvar tělesa, kde se teplo šíří, nehomogenní materiál)
- lze aproximovat pomocí metody konečných diferencí – jednoduchá metoda, ale vyžaduje výpočetní výkon
- vizualizace viz <https://www.youtube.com/watch?v=TvLIIfSlLB0c>

Metoda konečných diferencí

- aproximuje derivace pomocí konečných diferencí
- prostorovou (popř. i časovou) doménu rozbijeme na konečně malé prvky, sousední hodnoty aproximují derivace

Co to teda znamená v případě výpočtu šíření tepla?

Prostor, ve kterém simulujeme šíření tepla, rozbijeme pomocí pravidelné mřížky

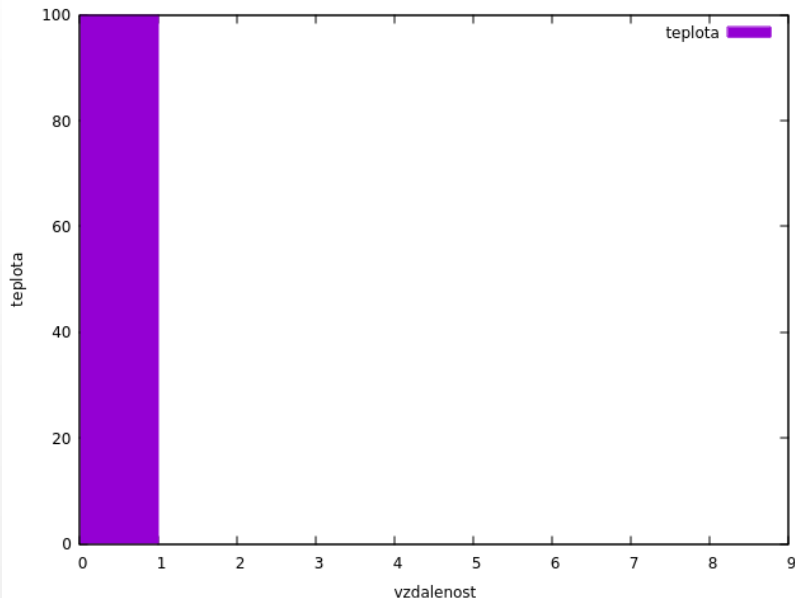
- v čase $t + 1$ nastavíme teplotu v každé buňce mřížky na základě její teploty a teploty sousedních buněk v čase t : pro jednorozměrný problém

$$u_j^{t+1} = (1 - r)u_j^t + ru_{j-1}^t + ru_{j+1}^t, r \leq 1/2$$

- typicky provádíme tak dlouho, dokud se teplota neustálí (změny v teplotě každé buňky jsou pod definované minimum)
- okraje mřížky představují okolní teplotu (mimo naši simulaci), v nejjednodušším případě nastavena konstantní teplota

Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



Úvod

Osnova
O virech
Modelování a simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro informatika
COVID

Transport v proteínech

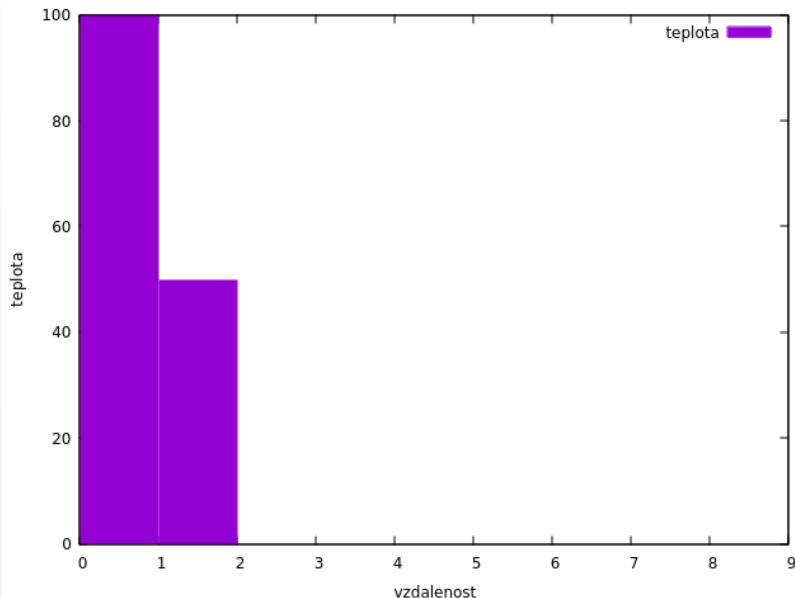
Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



Úvod

Osnova
O virech
Modelování a simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro informatika
COVID

Transport v proteínech

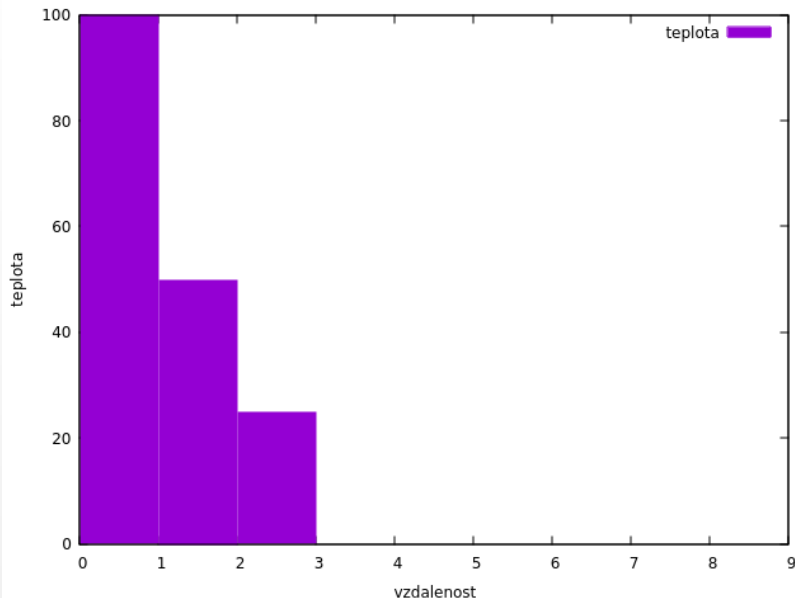
Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

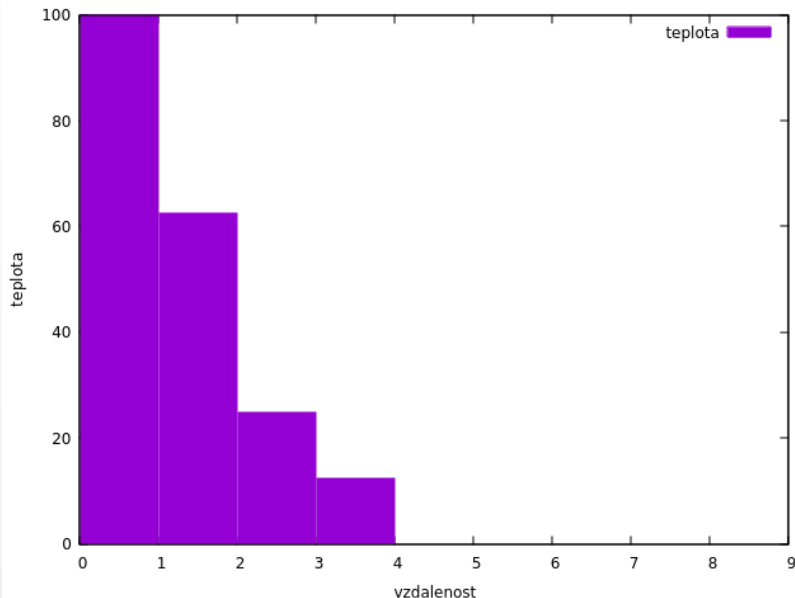
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



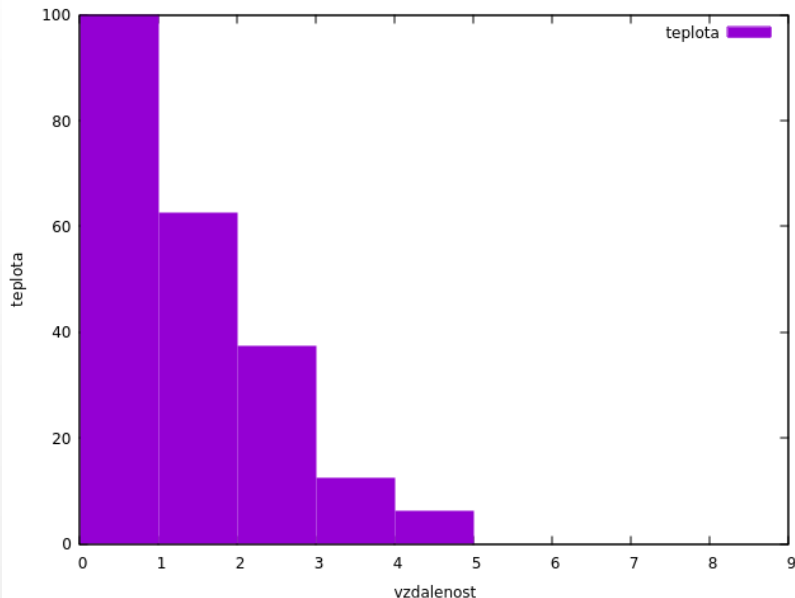
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



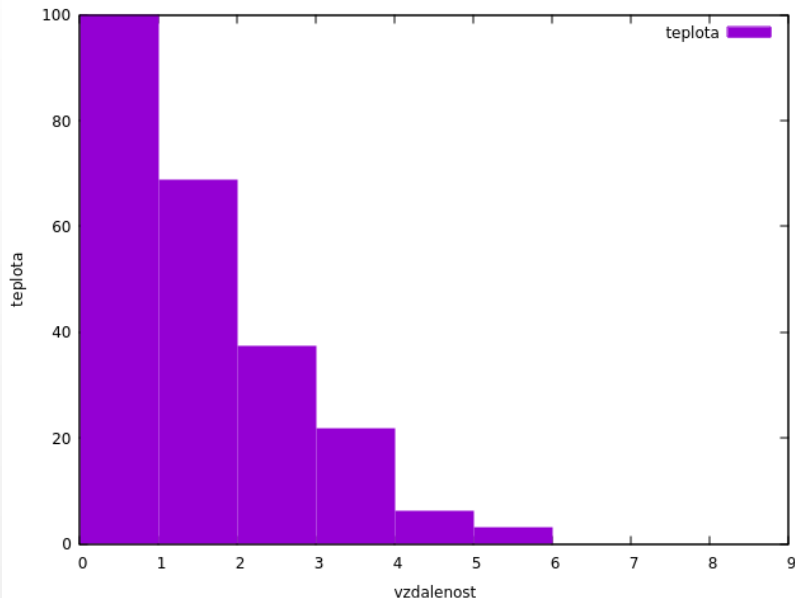
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



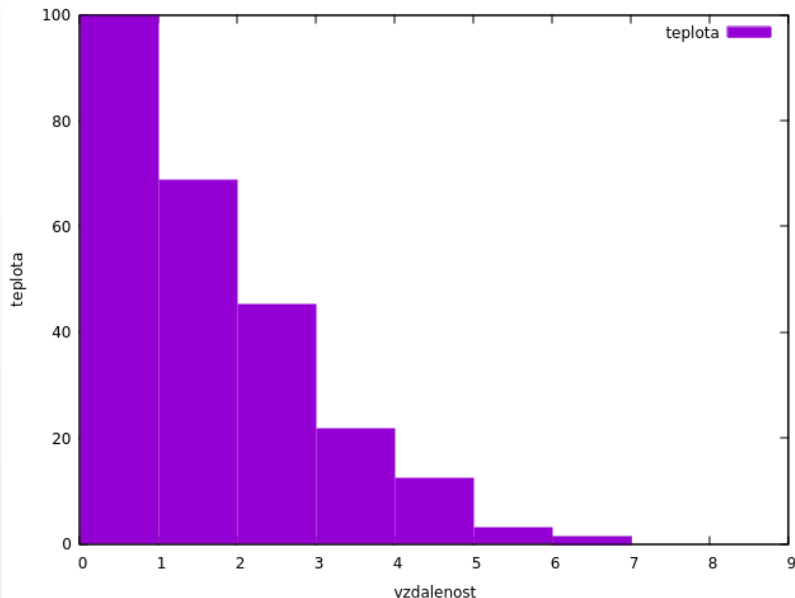
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



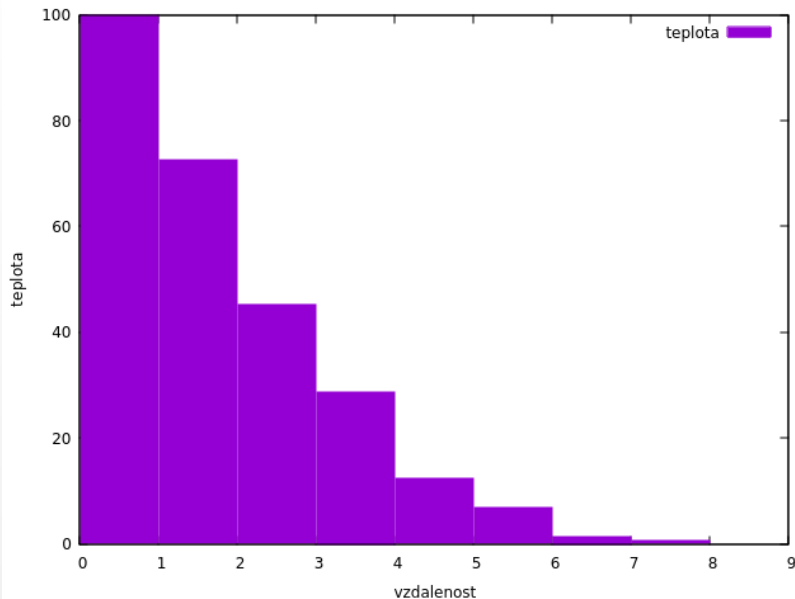
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



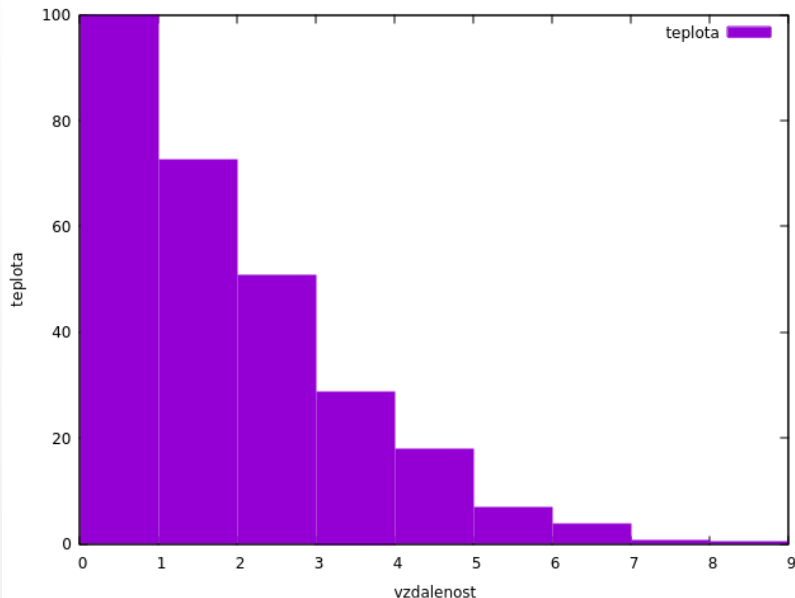
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



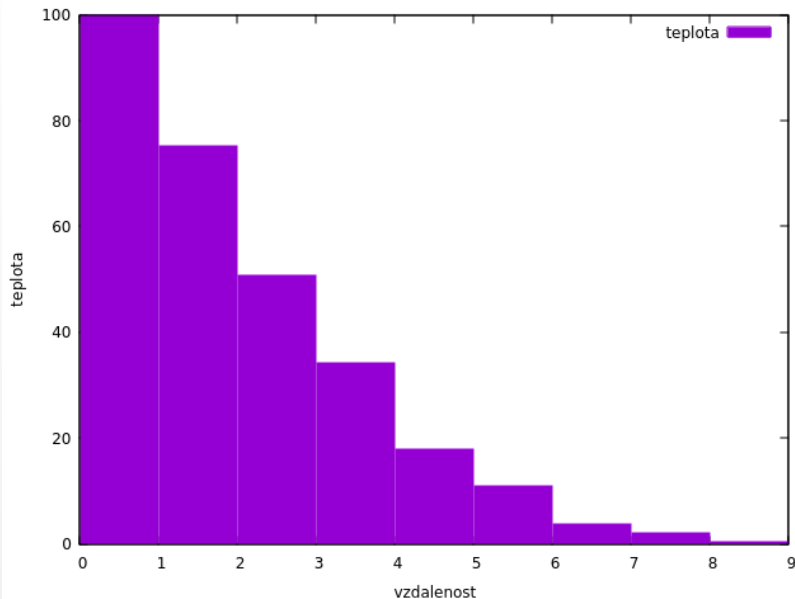
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



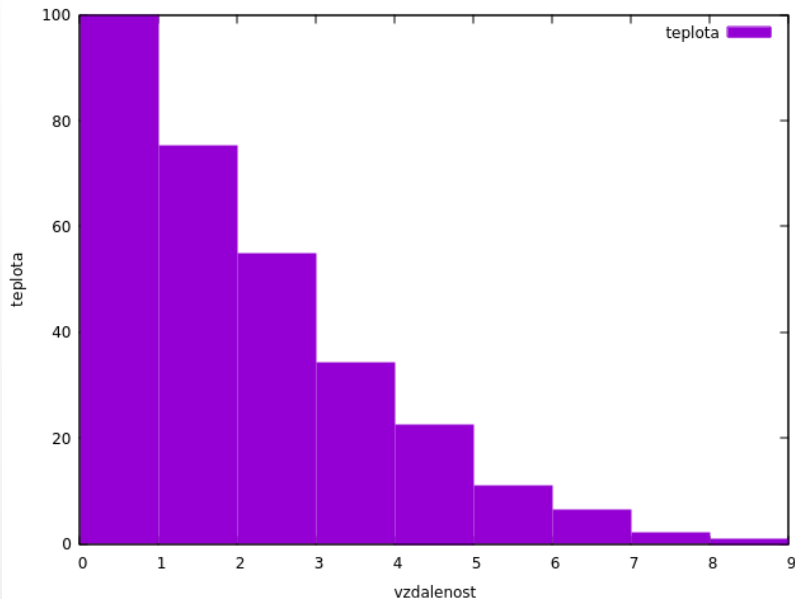
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



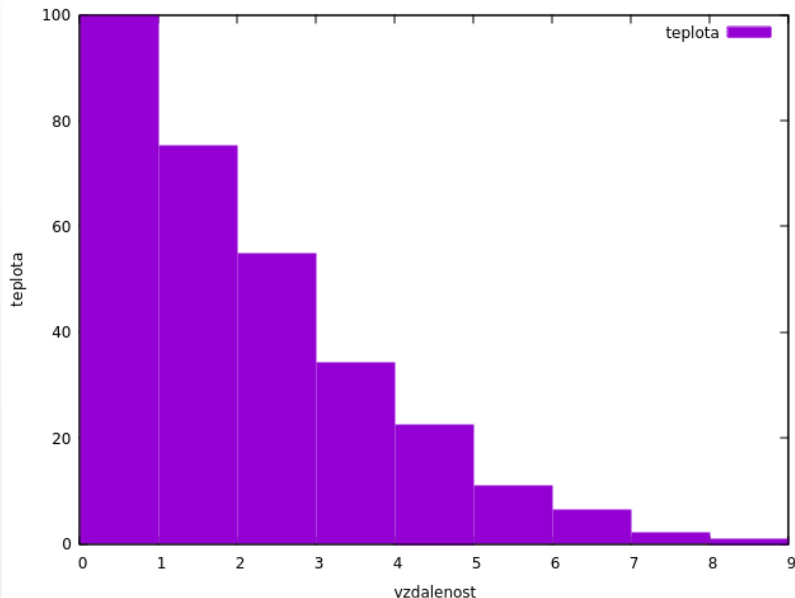
Příklad: šíření tepla v jednorozměrném prostoru

Prvních 10 iterací



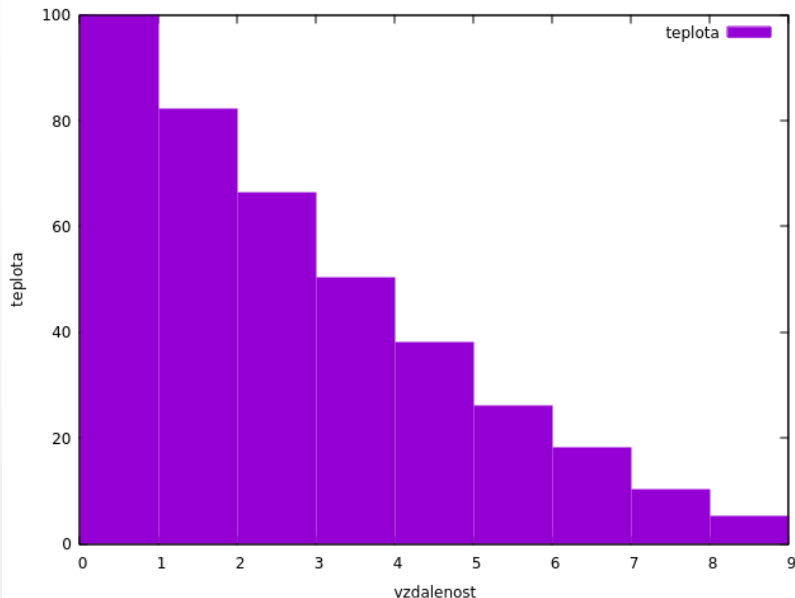
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



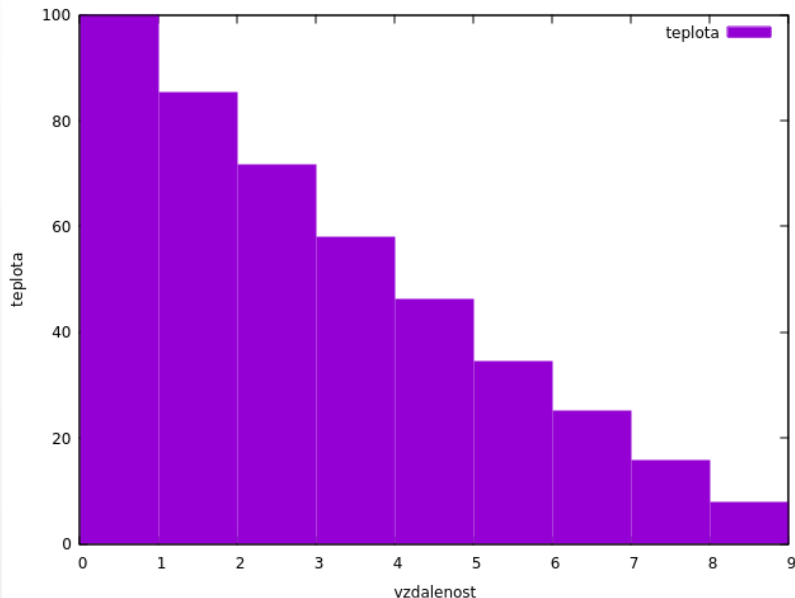
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



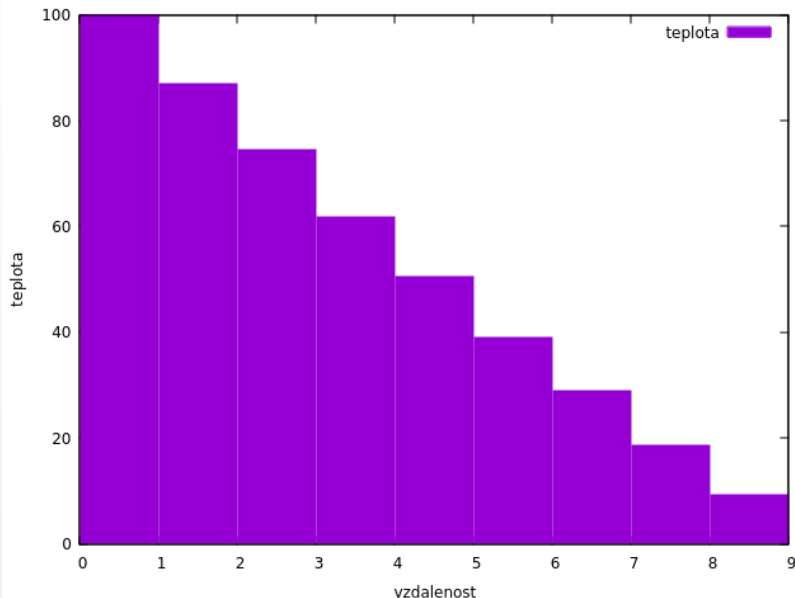
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



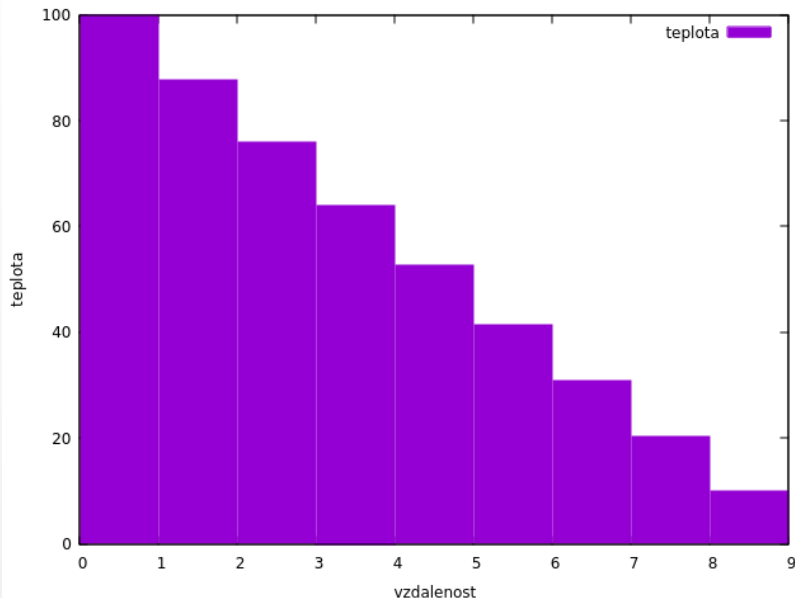
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



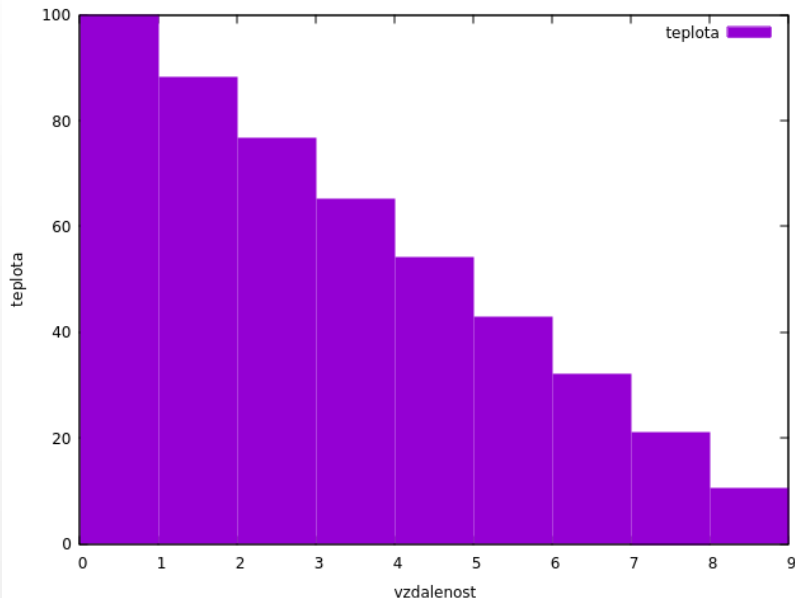
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



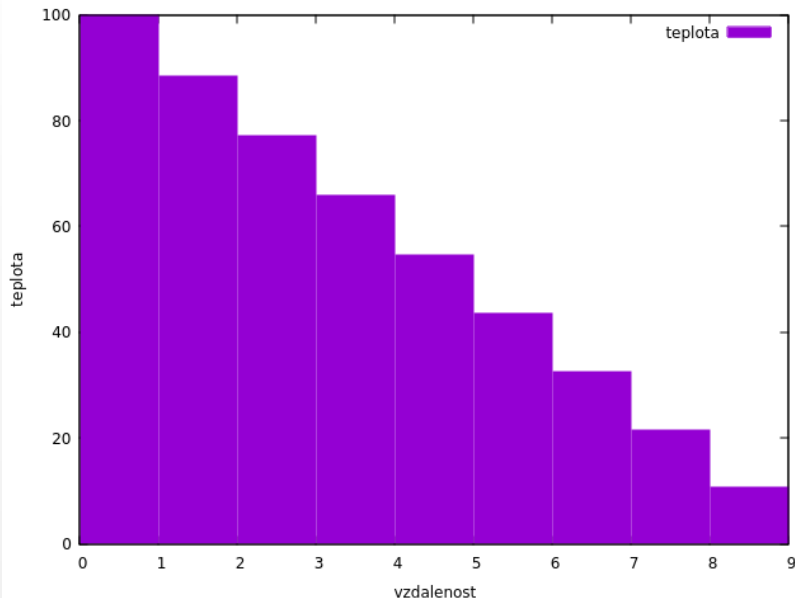
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



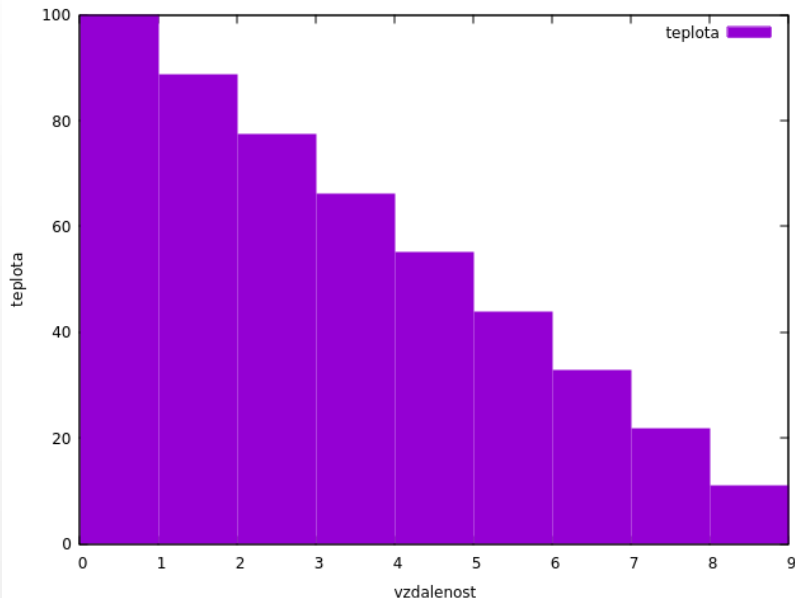
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



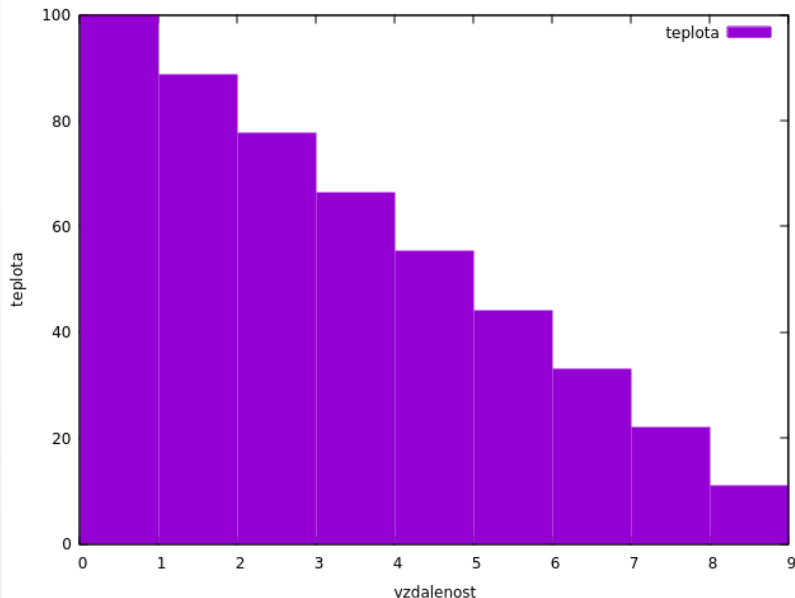
Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



Příklad: šíření tepla v jednorozměrném prostoru

Pro zrychlení skákejme po 10 iteracích...



Simulujeme šíření tepla v bloku motoru o rozměrech $1m \times 1m \times 1m$, v rozlišení 1mm

- potřebujeme celkem 1 000 000 000 buněk
- nejméně 3 000 iterací je zapotřebí jen k tomu, abychom přenesli nějaké teplo z jednoho rohu do druhého
- řádově alespoň miliardy aktualizací teploty v buňkách, jednotky GB paměti

Dokážeme být obecní?

- nechť n značí počet buněk (rozlišení) v každé dimenzi a i počet iterací výpočtu
- pro jednoduchost uvažujme stejný počet buněk v každém rozměru
- pro třírozměrný objekt vyžaduje $\mathcal{O}(n^3 i)$ výpočtů
- důsledek: zdvojnásobení rozlišení vede při zachování počtu iterací k osminásobné výpočetní náročnosti

Úvod

Osnova

O vírech

Modelování a
simulace

Náročná simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Předpokládejme, že na našem počítači zvládneme zpracovat miliardu buněk za sekundu. Zamysleme se, jak velké problémy zvládneme vyřešit.

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná
simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v
proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Předpokládejme, že na našem počítači zvládneme zpracovat miliardu buněk za sekundu. Zamysleme se, jak velké problémy zvládneme vyřešit.

n	i	čas
1 000	10 000	2 hodiny 47 minut

Předpokládejme, že na našem počítači zvládneme zpracovat miliardu buněk za sekundu. Zamysleme se, jak velké problémy zvládneme vyřešit.

n	i	čas
1 000	10 000	2 hodiny 47 minut
2 000	20 000	44 hodin 27 minut

Předpokládejme, že na našem počítači zvládneme zpracovat miliardu buněk za sekundu. Zamysleme se, jak velké problémy zvládneme vyřešit.

n	i	čas
1 000	10 000	2 hodiny 47 minut
2 000	20 000	44 hodin 27 minut
10 000	100 000	3 roky 62 dní

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Obdobné výpočty provádíme i v předpovědi počasí

- potřebujeme relativně jemné rozlišení pro zachycení malých, ale signifikantních jevů
- mřížka přes celou planetu je fakt velká
- extrémní množství výpočtů
- problém se vstupními daty: pro predikci budoucnosti potřebujeme rozumně přesně znát současnost
- na příkladě výše je zřejmé, že nemůžeme úplně snadno zvyšovat rozlišení

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Když nám počítač nestačí

- pořídíme dva počítače (či dva tisíce počítačů, dva miliony počítačů...)

Jak bychom řešili výpočet šíření tepla na více počítačích?

- zkuste si představit, že řešíte dvourozměrný problém „ručně“ na čtverečkováném papíru

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Metoda konečných diferencí se paralelizuje snadno

- počítače zpracovávají každý svou část prostoru, musí si vyměňovat hranice
- ne všechny výpočty se paralelizují tak snadno

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Uvažujme problém seřazení posloupnosti čísel

- zkuste navádět přednášejícího, jak to udělat
- a následně problém zobecnit

Má více možných řešení, zde si představíme jednoduché (ne nejrychlejší)

- mějmě vstupní posloupnost čísel velikosti n a výstupní posloupnost (na začátku prázdnou)
- opakujme $n \times$
 - odstraň nejvyšší (nejnižší) číslo se vstupní posloupnosti
 - vlož jej na konec výstupní posloupnosti
- celkově n^2 kroků (lze lépe, ale jako ilustrace stačí)

Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Jak zrychlit takovéto řazení na více počítačích?

- zkuste navádět dva přednášející, jak to udělat :-)
- hledejte způsob, který by fungoval pro hodně dlouhé posloupnosti a velký počet přednášejících

Jak zrychlit takovéto řazení na více počítačích?

- zkuste navádět dva přednášející, jak to udělat :-)
- hledejte způsob, který by fungoval pro hodně dlouhé posloupnosti a velký počet přednášejících

Řešení

- rozdělíme vstup na dvě části, každou seřadíme zvlášť
- tyto seřazené části spojíme tak, aby byl výsledek opět seřazený (do výsledné posloupnosti přiřazujeme vždy vyšší (nižší) prvek z obou struktur)
- bonusová otázka – kolik to stojí operací? nenarazili jsme náhodou na rychlejší algoritmus?

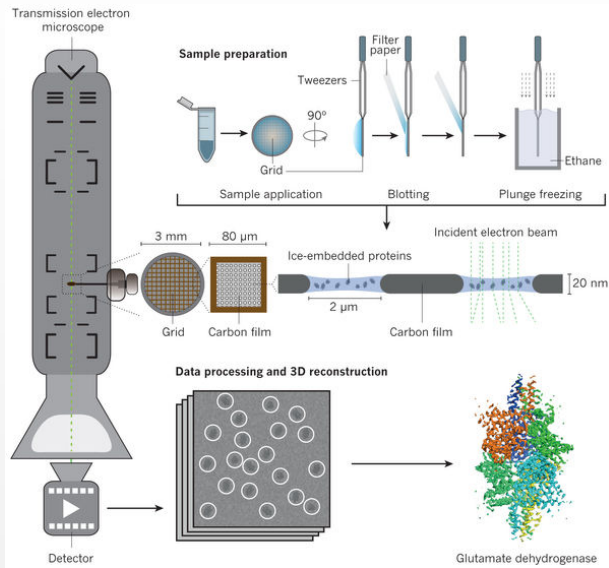
Pro porozumění mnoha biologických procesů je zapotřebí znát molekulární strukturu látek, které se na procesech podílí

- zobrazení molekul lze realizovat pomocí krystalografie, NMR a elektronové mikroskopie
- jednotlivé metody mají své omezení

Zobrazení částic v přirozeném prostředí

- pro spoustu biologicky relevantních částic (proteiny, viry) je přirozené prostředí voda
- pokud bychom je studovali mimo vodu, jejich struktura se zhroutí
- pokud bychom je studovali v tekuté vodě, budou se hýbat
- cryo-elektronová mikroskopie zobrazuje částice v tenké vrstvě ledu

Cryo-elektronová mikroskopie



Simulace a zpracování dat

J. Filipovič

Úvod

Osnova
O virech
Modelování a simulace

Náročné simulace

Šíření tepla
Paratelizace

cryo-EM

Motivace

Jak funguje
Práce pro informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Dašší úvahy

Vlastnosti modelů
AI

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paratelizace

cryo-EM

Motivace

Jak funguje

Práce pro

informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Jedním z potenciálních terčů terapie je spike-glykoprotein, kterým se virus váže na lidskou buňku (ACE2 receptor)

- spike protein můžeme zablokovat, aby se na ACE2 receptor nevázal
- můžeme naučit imunitní systém spike protein rozeznávat (a likvidovat)
- zajímá nás tedy, jak tento protein vypadá a jak se chová
- využijeme cryo-elektronovou mikroskopii

Velké množství exemplářů studovaného vzorku zmrazíme v tenké vrstvě ledu a vložíme do elektronového mikroskopu

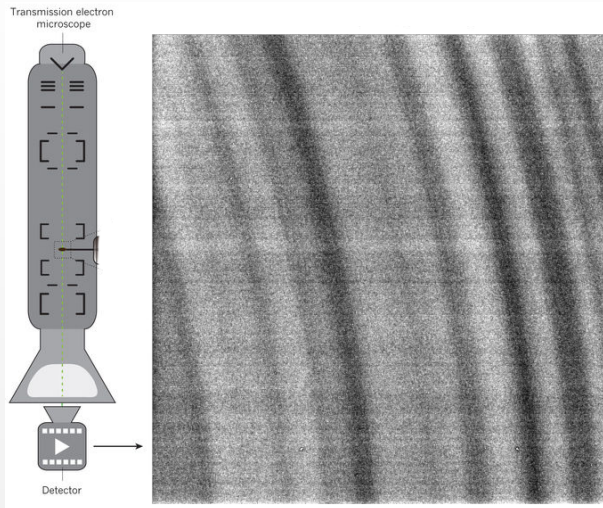
- získáme velmi zašuměný obraz (pouhým okem nedokážeme rozlišit jednotlivé částice)
- chceme získat 3D strukturu vzorku

Proč je to komplikované

- vidíme atomy vzorku obklopeného atomy tvořícími vodu
- vysoký podíl přirozeného šumu (elektronový paprsek je velmi slabý, jinak by nám zničil vzorek)
- nedokonalá data (mikroskop se třese, optika není dokonalá a má vady)
- studované částice jsou v neznámé orientaci, mohou být kontaminovány

Cryo-elektronová mikroskopie

Hrubá data, která získáme z mikroskopu



Úvod

Osnova
O vírech
Modelování a simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro informatika
COVID

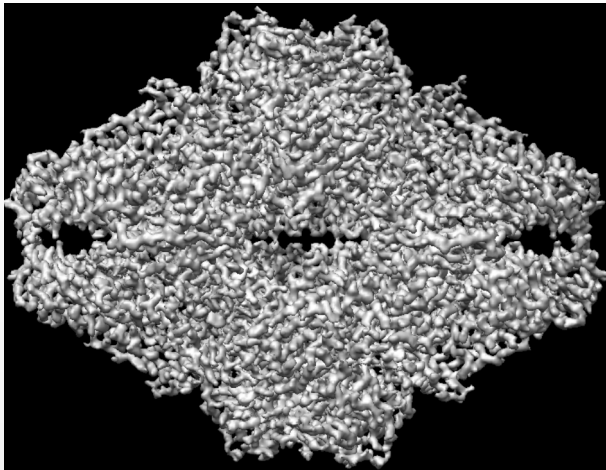
Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Výstup, který očekáváme



Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteinech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Jak získáme smysluplná data z šumu?

- šum je náhodný, vzorek stále stejný
- pokud sečteme velké množství zašuměných obrázků, šum se potlačí a získáme signál

Analogie: roj včel

- představte si, že mezi vámi a přednášejícím létá hustý roj včel
- pokud uděláte jednu fotku, vidíte v podstatě jen včely, velmi malé procento obrazu obsahuje kousky přednášejícího
- opravdu velké množství fotek bude dohromady obsahovat celého přednášejícího, který bude vždy stejný, zatímco včely budou vždy na lehce jiné pozici
- pokud fotky spojíte, začne být přednášející zřetelný

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Opravdu je to tak jednoduché?

- v mikroskopu se nedíváme s různým šumem na jednu částici, ale na mnoho jejich kopií
- každá kopie může být jinak otočená
- v analogii s rojem včel: představte si, že se přednášející při každé fotce otočí do jiné pozice či poodejde

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

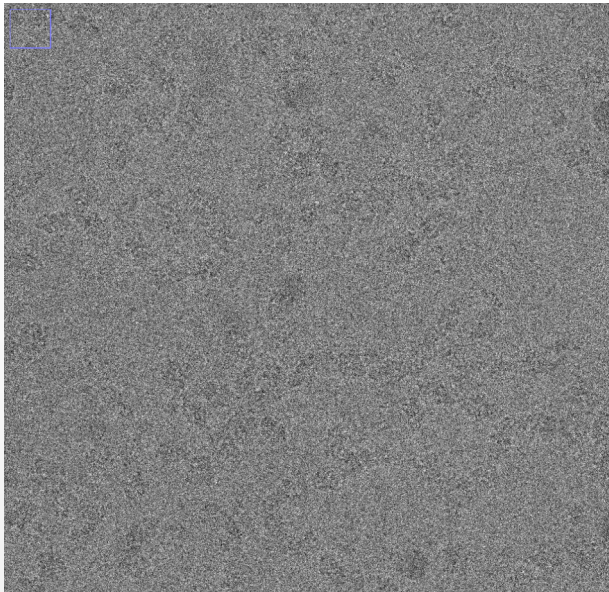
Další úvahy

Vlastnosti modelů
AI

Hlavní kroky v cryo-EM

- vytvoření movies: sloučení více fotek stejného kusu ledu (potlačení šumu daného slabým proudem elektronů)
- vybrání a kategorizace částic: vyřazení kontaminace, nalezení částic zachycených ze stejného úhlu
- 3D rekonstrukce: vytvoření 3D objektu z jednotlivých projekcí částice

Vytváření movies



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

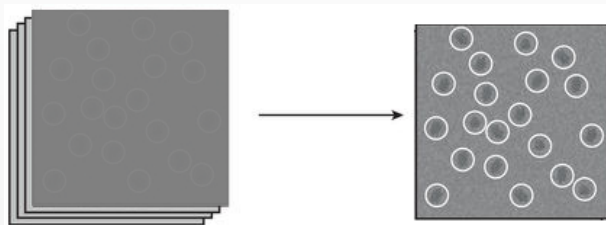
Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI



Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

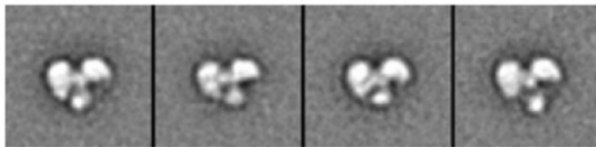
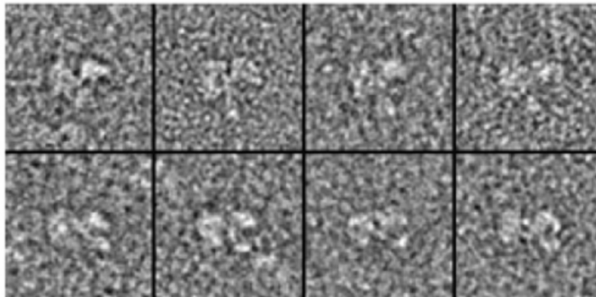
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Vytvoření 2D projekcí



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

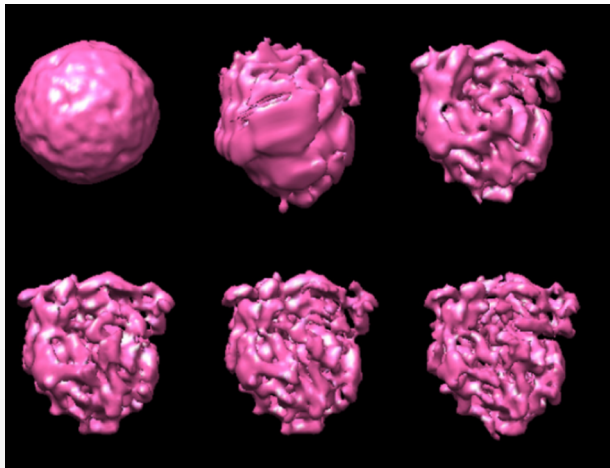
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

3D rekonstrukce



Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paratelizace

cryo-EM

Motivace

Jak funguje

**Práce pro
informatika**

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Máme velké množství dat, model, a hledáme jeho parametry

- vytvoření movies: model deformace v mikroskopu (parametry jsou posun a vzdouvání obrazu)
- vybrání a kategorizace částic: metrika podobnosti jednotlivých částic (parametry jsou translace a rotace)
- 3D rekonstrukce: 3D model projekcí částic (parametry jsou úhly částic)

Hlavní výpočetně náročné části zpracování obrazu

- vytvoření movies: sloučení více fotek stejného kusu ledu (potlačení šumu daného slabým proudem elektronů)
- vybrání a kategorizace částic: vyřazení kontaminace, nalezení částic zachycených ze stejného úhlu
- 3D rekonstrukce: vytvoření 3D objektu z jednotlivých projekcí částice

Práce s velkým množstvím dat

- z mikroskopu dostaneme TB
- identifikujeme stovky tisíc částic
- iterativní rozdělování částic do tříd, zpřesňování 3D modelu atp.

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje

**Práce pro
informatika**
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Zrychlování výpočtu

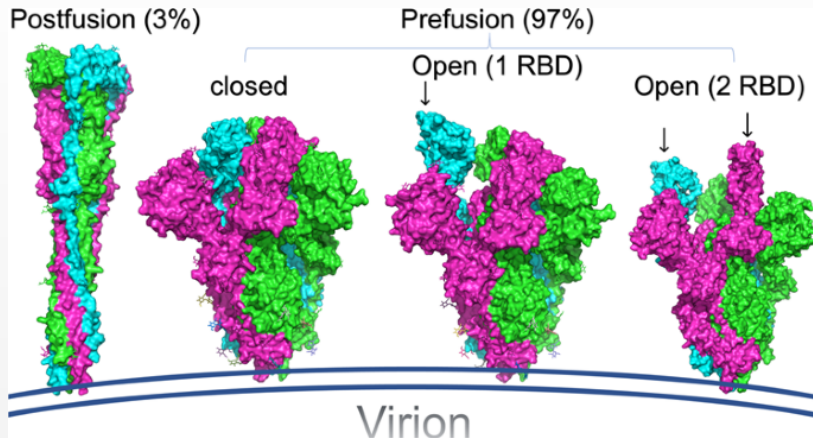
- matematicky odlišné metody
- chytřejší algoritmy
- paralelizace a GPU akcelerace

Proč potřebujeme rychlost?

- menší rekonstrukce zvládnutelné na desktopu či malém clusteru
- proces je částečně interaktivní
- rozdíl, jestli vidíte změnu parametru za 30 minut, nebo druhý den

Zpět k COVIDu

První struktura spike proteinu byla zjištěna pomocí cryo-EM



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paratelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika

COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Ismail, A.M., Elfiky, A.A. SARS-CoV-2 spike behavior in situ: a Cryo-EM images for a better understanding of the COVID-19 pandemic. Sig Transduct Target Ther 5, 252 (2020).

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Co nám struktura prozrazuje?

- trimer se na konci rozevívá (získal přezdívku Demogorgon), na ACE2 se váže v otevřeném stavu (viz <https://www.youtube.com/watch?v=ieF7ER1wvT0&t=1s>)
- díky znalosti struktury vazebné domény lze hledat léčiva, co ji zablokují
- spike protein se vyskytuje i v postfúzním tvaru (možná ochrana proti imunitní reakci)

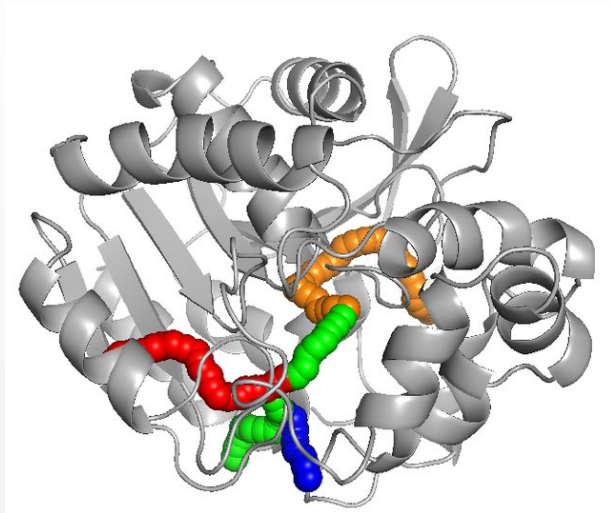
Proteiny jsou velké biomolekuly

- stavební prvky živých organizmů
- biologická funkce založená na interakci s malými molekulami (ligandy) či jinými proteiny

Transport ligandů

- ligand může upravit funkci proteinu, či tuto funkci zablokovat
- v případě enzymů naopak protein (enzym) katalyzuje přeměnu ligandu
- studium reakcí protein-ligand důležité mimo jiné při vývoji léčiv (inhibice funkce nežádoucího proteinu)
- u části proteinů musí ligand projít tunelem, než dojde k interakci – nezajímá nás jen výsledná poloha ligandu, ale i cesta proteinem

Cesty v proteinu



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Chceme najít cestu z venkovního prostředí na specifické místo v proteinu

- působení různých chemických sil
- ligand i protein jsou flexibilní tělesa, můžou se měnit

Možné přístupy

- molekulový docking: rychlý, ignoruje cestu, hledá nejlepší polohu v cílovém místě
- geometrický: rychlý, ignoruje chemické síly
- molekulová dynamika: simuluje systém protein-ligand v čase, výpočetně náročný
- hybridní: jednodušší výpočetní model, ale stále zahrnuje síly

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Zaveďme si analogii

- představte si, že máte hospodu a chcete najít vhodný typ zákazníků, kteří do ní budou chodit
- zákazníkům se musí líbit na místě (musí jim chutnat nabízené pivo)
- zákazníkům se musí chtít do hospody chodit (jeden zákazník bude preferovat hospodu v centru města, druhý v lese, třetí na kopci)

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Molekulový docking

- víme, jak se bude zákazníkům líbit v hospodě, ale nevíme, jestli jsou spokojení s cestou

Geometrický přístup

- víme, zda zákazník projde dveřmi

Molekulová dynamika

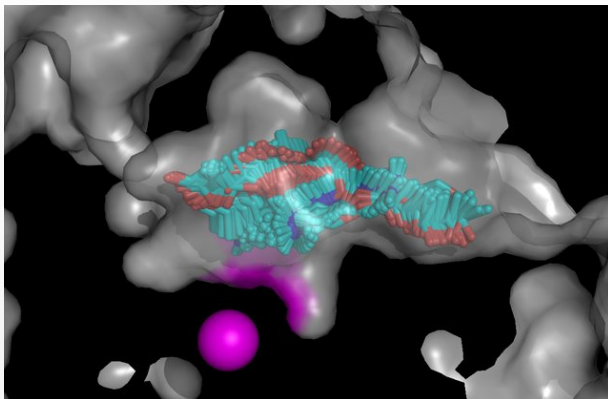
- kompletně simuluje zákazníky, včetně spánku, cesty do práce atp.

Na ÚVT se ve spolupráci s PŘF zabýváme vývojem hybridní metody, implementované v software CaverDock

- molekulový docking upravený tak, aby dokoval postupně podél tunelu až do místa, které nás zajímá
- ligand „chytne“ za jeden atom a táhneme jej přes tunel (v každém kroku hledáme nejlepší ne příliš vzdálenou polohu)
- oproti molekulové dynamice nižší náročnost: simulujeme jen cestu, která nás zajímá, oproti geometrické metodě známe chemické síly

V naší analogii

- vezmeme zákazníka do hospody, po cestě a na místě měříme, jak je spokojený



<https://www.fi.muni.cz/~xfilipov/caverdock/linb-wt-p1-wiew1.mp4>

Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Jak velké množství voleb cesty pro ligand máme?

- v každém kroku se můžeme posunout ve třech dimenzích, rotovat dle tří os a ohýbat vazby v ligandu: celkově n dimenzí, typicky $n > 10$
- abychom došli z venkovního prostředí na místo, musíme udělat m kroků, kde m je typicky několik desítek
- celkově bez detekce dualit $\mathcal{O}(n^m)$ možných cest

Můžeme zkontrolovat všechny cesty?

- řekněme, že máme 10 dimenzí (a pro jednoduchost v každé můžeme provést jen dva možné posuny), 50 kroků a zvládneme spočítat pozici a energii pro 1 000 000 kroků za sekundu
- při jednoduché implementaci (bez sjednocování duplicit) bychom potřebovali více než 10^{36} let (odhadované stáří vesmíru je 10^{10} let)
- exponenciální algoritmus: 10 kroků zvládneme za 3 hodiny
- zde nás nespasí rychlejší počítač, potřebujeme chytřejší algoritmus

Úvod

Osnova

O vírech

Modelování a
simulace

Náročná simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteinech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Budeme akceptovat jen některé polohy ligandu

- takové, které jsou v energetickém minimu a zároveň dostatečně blízko předešlé polohy
- namísto hrubého prohledávání se jedná o matematickou optimalizaci
- při pohybu tunelem vycházíme z nejlepší dosud známé cesty, ostatní ignorujeme
- pokud se dostaneme do pozice s příliš vysokou energií, zkusíme najít jinou polohu (bez požadavku na blízkost) a couvat

Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

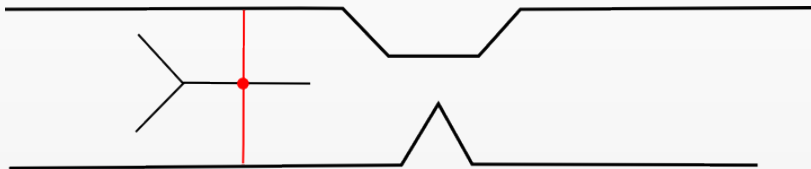
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Můžeme detekovat, že ligand neprojde



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

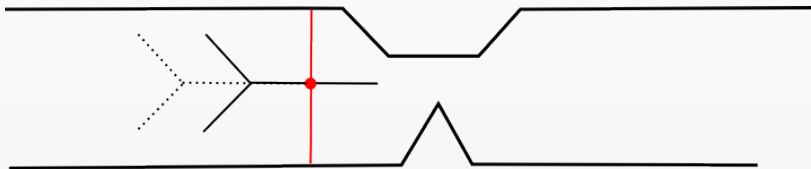
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Můžeme detekovat, že ligand neprojde



Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

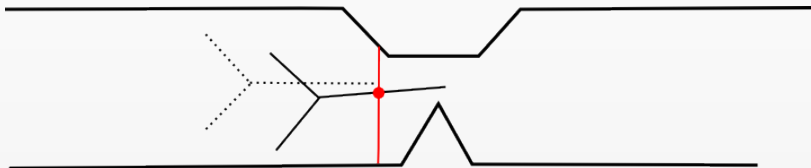
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Můžeme detekovat, že ligand neprojde



Úvod

Osnova
O virech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

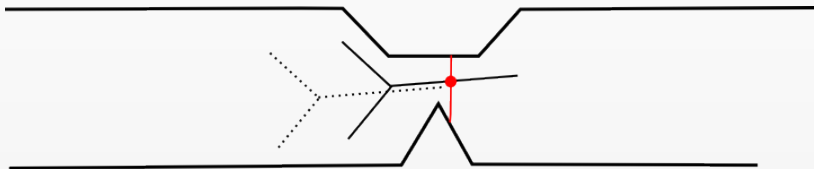
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Můžeme detekovat, že ligand neprojde



Úvod

Osnova
O vířech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

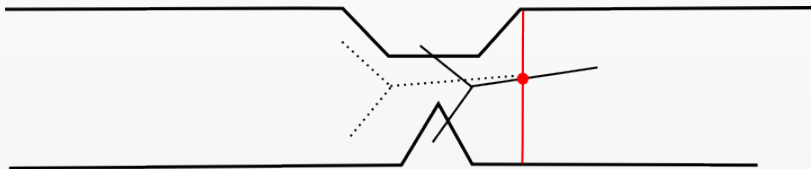
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Můžeme detekovat, že ligand neprojde



Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

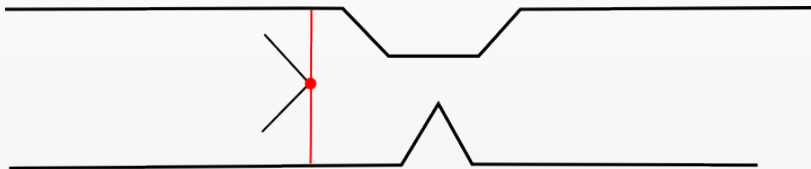
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Úvod

Osnova
O vířech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

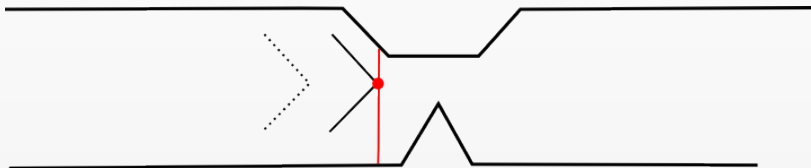
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Úvod

Osnova
O vířech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

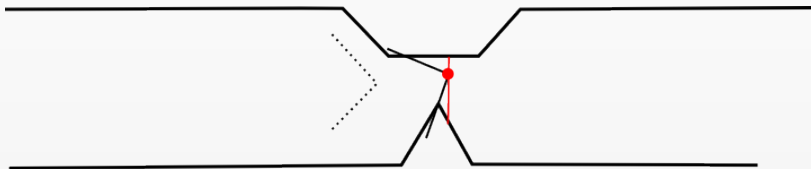
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

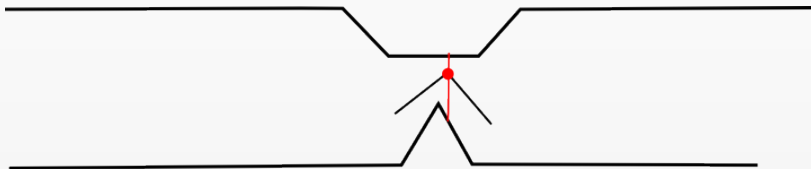
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Úvod

Osnova
O vířech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

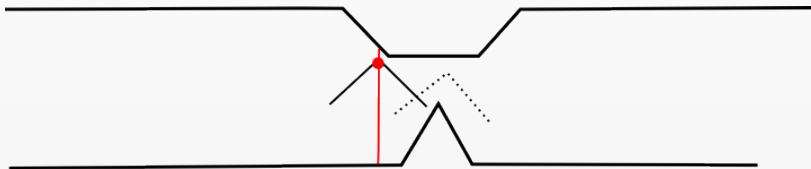
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Úvod

Osnova
O vírech
Modelování a
simulace

Náročné simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

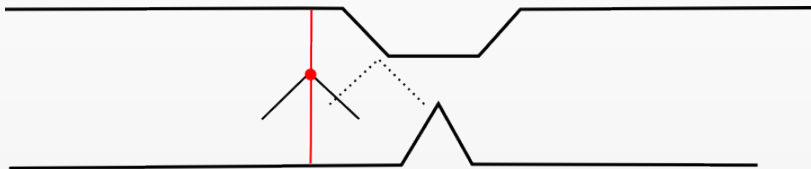
Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

V některých případech lze nalezenou překážku překonat změnou pozice ligandu



Kolik výpočetních kroků udělá chytřejší algoritmus?

- tunel délky m , v každém kroku se můžeme vrátet až na začátek
- jsme omezeni na $\mathcal{O}(m^2)$ spuštění matematické optimalizace

Srovnání s prohledáváním

- předpokládejme, že matematická optimalizace je $100000\times$ pomalejší než prostý výpočet energie
- pro příklad s 10 dimenzemi a 50 kroky v tunelu: nejhůře 250 sekund (naivní algoritmus 10^{36} let)

Útok na spike protein koronaviru

- aby došlo na navázání na lidský ACE2 receptor, musí se vazebné domény otevřít
- pokud bychom našli léčivo, co zapadne do spike proteinu a podrží vazebné domény u sebe, virus nedokáže buňky napadat (zalepíme Demogorgonovi pusu)
- chceme nalézt takovou molekulu, která nasedne do vnitřní dutiny spike proteinu a bude v ní silně držet
- navrhli jsme několik schválených léčiv, které by mohly spike protein blokovat¹

¹G. Pinto et al. Screening of world approved drugs against highly dynamical spike glycoprotein of SARS-CoV-2 using CaverDock and machine learning. Computational and Structural Biotechnology Journal, Vol. 19, 2021.

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteinech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Statistický vs. výpočetní model

- výsledky vs. procesy
- nalezení vlastností vs. predikce chování
- účinnost léčiva vs. simulace interakce léčiva s relevantními proteiny
- rekonstrukce cryo-EM dat vs. transportní procesy v proteinech
- obě metody trpí na garbage in – garbage out

Stabilita výpočetního modelu

Golfový míček ve stabilní a nestabilní pozici



Stabilní systémy

- cena-poptávka
- populace kořisti a dravců

Nestabilní systémy

- epidemie
- známost celebrit

Úvod

Osnova

O vírech

Modelování a
simulace

Náročné
simulace

Šíření tepla

Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v
proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Úvod

Osnova
O vírech
Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace
Jak funguje
Práce pro
informatika
COVID

Transport v proteínech

Motivace
Výpočetní model
COVID

Další úvahy

Vlastnosti modelů
AI

Co to je umělá inteligence?

- intuitivní (vágní) definice: uměle vytvořený systém, který dokáže provádět nějakou kognitivní funkci typickou pro inteligentní bytost (člověka)
- přesnější definice: uměle vytvořený systém, který se dokáže v daném prostředí chovat tak, že maximalizuje své šance dosáhnout stanoveného úkolu

Strojové učení

- schopnost systému zlepšovat sebe sama bez explicitních instrukcí od tvůrce systému
- podmnožina umělé inteligence

Úvod

Osnova

O vírech

Modelování a
simulace

Náročná simulace

Šíření tepla
Paralelizace

cryo-EM

Motivace

Jak funguje

Práce pro
informatika

COVID

Transport v proteínech

Motivace

Výpočetní model

COVID

Další úvahy

Vlastnosti modelů

AI

Aniž bychom to zmiňovali, AI se vyskytuje i v dnes probraných případech

- CaverDock: hledání cesty pro ligand v proteinu patří do kategorie umělé inteligence, nikoliv však strojového učení (algoritmus nedokáže vylepšovat sám sebe)
- cryo-EM: jistá forma umělé inteligence je parametrizace modelu, navíc při kategorizaci částic lze využít neuronové sítě: např. pro naučení, jak zhruba vypadá hledaná částice z určitého úhlu, následně taková síť rozpoznává částice ze stejné kategorie
- strojové učení vytváří statistický model: nemusí rozumět procesům stojícím za chováním modelu, ale dokáže model zlepšit na základě pozorování výsledků procesů