

Modely sebeskládajících DNA nanostruktur

Vypracoval: Jakub Klemsa

Školitel: Ing. Štěpán Starosta, Ph.D.

Fakulta jaderná a fyzikálně inženýrská
Matematická informatika

19. června 2014

1. Motivace
2. DNA vs. Chomského hierarchie
 - Regulární jazyky
 - Bezkontextové jazyky
 - Turingova univerzalita
3. Modely založené na Wangovo dláždění
 - Wangovo dláždění
 - aTAM
 - Studované složitosti
 - Výpočetní síla
 - Jiný důkaz TU
 - Meze studovaných složitostí
 - Důsledky
4. Návrh řešení NP problémů
 - Přizpůsobení modelu NP
 - Problém k -kliky
 - Počítačová simulace
 - Další vyřešené problémy

Motivace

There's plenty of room at the bottom – R. P. Feynman, 1959, [3]

- budoucnost počítačů na úrovni molekul

První experiment s DNA – L. Adleman, 1994, [1]

- využití sebeskládajících DNA molekul ke hledání Hamiltonovské cesty orientovaným grafem
- Hamiltonovská cesta projde každý vrchol grafu právě jednou, počáteční a cílový vrchol jsou označeny
- rozhodovací problém existence HC je NP-complete

Motivace

Výhody

- paralelizmus – ve zkumavce může být až 10^{18} „větších“ molekul

Nevýhody

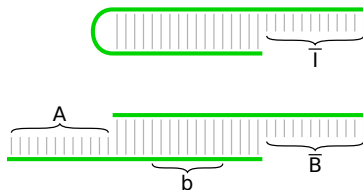
- 10^{18} není zdaleka neupočitatelné hrubou silou
- pravděpodobnostní povaha
- chybovost

Pole studia

- kinetika reakcí
- **abstraktní modely pohledem matematické informatiky**

Souvislost s Chomského hierarchií

Lineární vlákna \leftrightarrow regulární jazyky (E. Winfree [4])

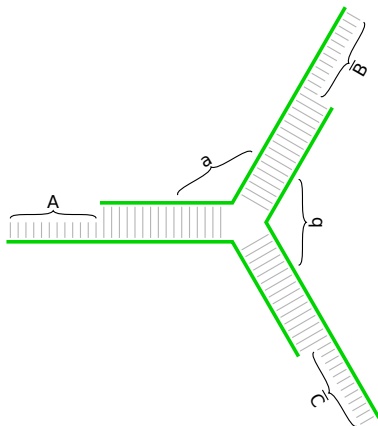


Obrázek: Vlákno pro iniciální symbol I a vlákno pro pravidlo $A \rightarrow bB$.

\bar{B} značí Watson-Crick komplementární sekvenci k B ($A \leftrightarrow T$, $C \leftrightarrow G$)

Souvislost s Chomského hierarchií

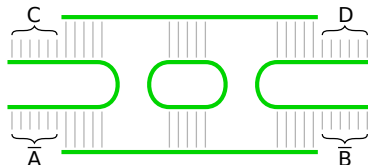
Stromové struktury \leftrightarrow bezkontextové jazyky (E. Winfree [4])



Obrázek: Struktura odpovídající pravidlu $A \rightarrow aBbC$.

Souvislost s Chomského hierarchií

Dvojkřížené molekuly \leftrightarrow Turingův stroj (TS) (E. Winfree [4])



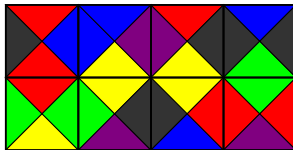
Obrázek: Dvojkřížená molekula.

Předpoklady budou upřesněny v abstraktním modelu.

Wangovo dláždění

Čtvercové dláždění roviny (její části), kde

- dlaždice mají na hranách barvu z konečné množiny barev (lepidel)
- máme-li fond, dlaždicemi je zakázáno rotovat nebo je překlápět
- soudedit smí pouze dlaždice se stejnou barvou na společné hraně



Obrázek: Wangovo dláždění.

aTAM

Rothemund, Winfree rozšířili definici

- každé lepidlo má přidružené přirozené číslo – síla lepidla
- existuje prázdné lepidlo se silou 0, které smí souedit se všemi
- dláždění se utváří
 - z iniciální dlaždice
 - po jedné dlaždici
 - součet právě připojených lepidel musí být větší nebo roven zadané hodnotě (tzv. *teplota*, ozn. τ)

Studované složitosti I

Biostep complexity $Bs(n)$

- počet laboratorních procedur
- jedna trvá až desítky minut
- za proveditelné budeme uvažovat pouze $Bs(n) \in O(1)$

Binding complexity $Bnd(n)$

- počet vazeb v koncovém dlážďení
- v nedeterministickém případě uvažujeme nejmenší přijímací
- v pravděpodobnostním případě uvažujeme střední hodnotu
- kvůli rostoucí psti chyby – proveditelné $Bnd(n)$ polynomiální

Studované složitosti II

Tile complexity $Ti(n)$

- počet různých dlaždic
- potřeba je syntetizovat – proveditelné $Ti(n)$ polynomiální

Glue complexity $Gl(n)$

- počet různých lepidel – sekvencí
- příliš dlouhé se mohou vázat chybně – proveditelné $Gl(n)$ polynomiální

Lemma

1. $Ti(n) \leq Gl^4(n)$,
2. $Gl(n) \leq 4 Ti(n)$.

Výpočetní síla

aTAM je Turingovsky univerzální (TU)

- ve 2D při teplotě $\tau = 2$, Winfree [4]
- ve 3D při teplotě $\tau = 1$, Cook [2]

Neví se ve 2D při teplotě $\tau = 1$

- existují modifikace aTAMu, které jsou TU

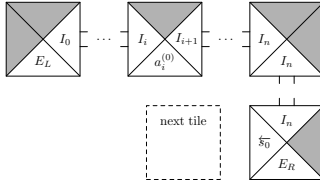
Důkaz TU ve 2D při teplotě $\tau = 2$ – převod na celulární automat

- neříká nic o spotřebě zdrojů

Důkaz TU ve 2D při $\tau = 2$

- přímočarý
- včetně odhadů studovaných složitostí v závislosti na čase a prostoru spotřebovaným simulovaným TS

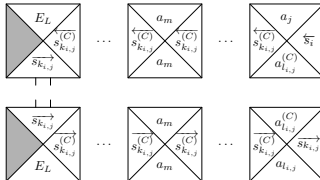
Assembly of input tape. E_L is left stop, $a_i^{(0)}$ is i -th symbol of input word $a_0^{(0)} \dots a_{n-1}^{(0)}$. Following assembly starts at a place denoted "next tile" while simulating a Turing machine reading $a_{n-1}^{(0)}$ and being in state s_0 . The arrow over s_0 means "comes from right".



Coming from right, being in state s_i , reading tape symbol a_j .
Transition function says: write $a_{i,j}$, switch to state $s_{k,i,j}$ and go left.



Situation is like before with only difference: go right.
Now the rest of the tape must be copied by special tiles which thus exist for all pairs $a_m, s_{k,i,j}$ of tape symbol and state, respectively.



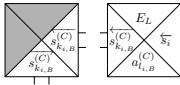
Coming from right, being in state s_i , reading left stop E_L which stands for blank symbol in terms of tape alphabet.

Transition function says: write $a_{l_{i,B}}$, switch to state $s_{k_{i,B}}$ and go left. Note that a glue of strength 2 must have been used.



Situation is like before with only difference: go right.

Note that tiles from previous figure are utilized to finish copying.



Coming from right, being in state s_i , reading blank symbol.

Transition function says: write $a_{l_{i,B}}$, switch to state $s_{k_{i,B}}$ and go left.



Situation is like before with only difference: go right.

Note that there already exist tiles which finish the turnaround.



Meze studovaných složitostí

Lemma

Studované složitosti v tomto systému jsou omezené jak následuje:

Biostep. $Bs(n) \in O(1)$.

Binding. $Bnd(n) \in O(s(n) \cdot t(n))$, kde $t(n)$ stojí za čas a $s(n)$ za prostor spotřebovaný simulovaným TS.

Tile. $Ti(n) \in O(n)$.

Glue. $Gl(n) \in O(n)$.

Zachování tříd složitosti

Důsledek

Ve 2D při $\tau = 2$ se za použití příslušného modelu (determ., nedeterm. nebo pravd'.) zachovávají složitostní třídy odolné k polynomiálnímu zpomalení pro všechny studované složitosti (tj. P, ZPP, RP, BPP, NP, ...).

Navíc, biostep complexity zůstane v $O(1)$.

Proveditelnost BPP ve 2D při $\tau = 2$

BPP se považuje za proveditelnou na TS, z předchozího důsledku (zachování tříd a $Bs(n) \in O(1)$) pak plyne:

Důsledek

BPP je proveditelná ve 2D při $\tau = 2$.

Poznámka

$P, ZPP, RP, \text{co-RP} \subseteq BPP$.

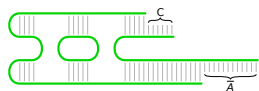
Přizpůsobení aTAMu

Odvozen z Winfreeho ukázky řešení problému Hamiltonovské cesty – avšak srozumitelnější

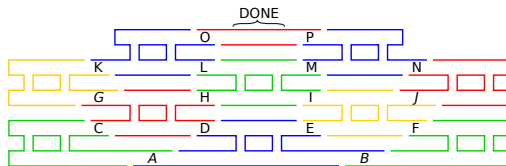
- $\tau = 2$
- 5 dalších typů dlaždic včetně daných sil lepidel
- pevně nastavená počáteční t_0

Poznámka

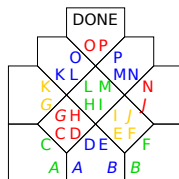
Tento model lze snadno simulovat klasickým aTAMem.



(a) Rohová molekula



(b) Schéma sebeskladu

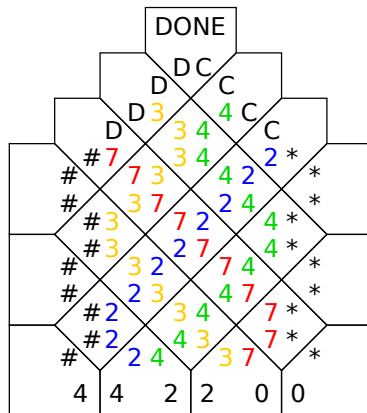
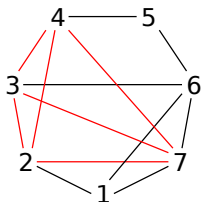


(c) Model

Obrázek: Evoluce modelu od molekul k dlaždicím. Zde lepidla A, B, G a J mají sílu 2, všechna ostatní mají sílu 1.

Problém k -kliky

NP-complete problém, $Bnd \sim 5/4k^2$, $Ti \sim 2k^2e + 3kn$, $Gl \sim kn$.



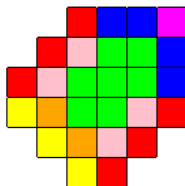
Obrázek: Nalezení k -kliky. Pořadí barev je dáno jejich vlnovou délkou.

Simulace v xgrow

xgrow je open-source simulátor jak kinetických tak abstraktních modelů.

- skriptem k zadanému grafu generuji potřebné dlaždice
- se zapnutím kinetiky není jednoduché dosáhnout bezchybného dláždění

```
flake 1 (32 by 32, seed 1 @ (8,2))
105932 events, 27 tiles, 0 mismatches
([DX] = 0.827988 uM, T = 41.315 C, 5-mer s.e.)
Gmc=17.0 Gse= 8.6 k=1000000 T= 2.0
t =      316.722 sec; G =      -19.016
105932 events (52979a,52953d,0h,0f), 27 tiles total
```



- BOX/bonds
- TILE/err
- RUN/pause
- export [ONE]
- sample
- next/big/prev
- cool heat
- NO fission
- FIXED/wander
- clean/fill/Rx
- restart
- quit

left: identify
middle:puncture
right: Gmc Gse
EW '98-'04

Další vyřešené problémy

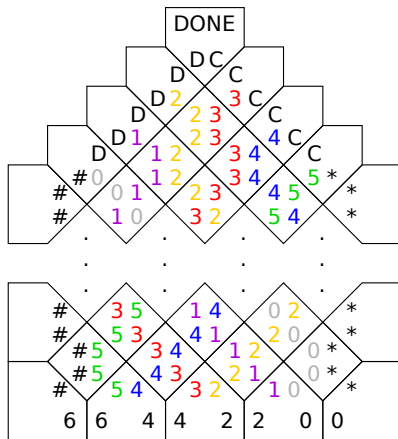
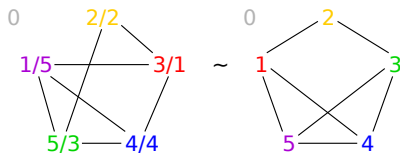
3-obarvení grafu

- NP-complete problém

Grafový izomorfismus

- je domníváno, že není NP-complete

$$Bnd \sim \frac{5}{4}n^2, \quad Ti \sim 8e^2 - 4en^2 + n^4, \quad Gl \sim n^2.$$



Reference



Leonard M Adleman.

Molecular computation of solutions to combinatorial problems.
Science - New York then Washington, pages 1021–1024, 1994.



Matthew Cook, Yunhui Fu, and Robert Schweller.

Temperature 1 self-assembly: Deterministic assembly in 3d and probabilistic assembly in 2d.

In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 570–589. SIAM, 2011.



Richard P Feynman.

There's plenty of room at the bottom.

Engineering and Science, 23(5):22–36, 1960.



Erik Winfree.

Algorithmic self-assembly of DNA.

PhD thesis, California Institute of Technology, 1998.