

# Modely sebeskládajících DNA nanostruktur

Vypracoval: Jakub Klemsa

Školitel: Ing. Štěpán Starosta, Ph.D.

Fakulta jaderná a fyzikálně inženýrská  
Matematická informatika

19. června 2014

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

# Úvod

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

- hledání Hamiltonovské cesty (HC) orientovaným grafem
- rozhodovací problém existence HC je NP-úplný

# Úvod

## Výhody

- paralelizmus – ve zkumavce až  $10^{18}$  „větších“ molekul
- energetická efektivita (Adleman [1])

# Úvod

## Výhody

- paralelizmus – ve zkumavce až  $10^{18}$  „větších“ molekul
- energetická efektivita (Adleman [1])

## Nevýhody

- na  $10^{18}$  operací stačí hrubá síla (řádově dny na clusteru s tisíci jader)
- pravděpodobnostní povaha
- chybovost

# Úvod

## Výhody

- paralelizmus – ve zkumavce až  $10^{18}$  „větších“ molekul
- energetická efektivita (Adleman [1])

## Nevýhody

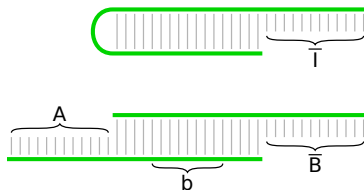
- na  $10^{18}$  operací stačí hrubá síla (řádově dny na clusteru s tisíci jader)
- pravděpodobnostní povaha
- chybovost

## Pole studia

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

# Souvislost s Chomského hierarchií

Lineární vlákna  $\leftrightarrow$  regulární jazyky (Winfree [5])

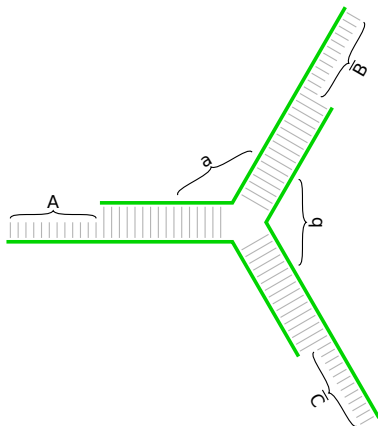


Obrázek: Vlákno pro iniciální symbol  $I$  a vlákno pro pravidlo  $A \rightarrow bB$ .

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

# Souvislost s Chomského hierarchií

Stromové struktury  $\leftrightarrow$  bezkontextové jazyky (Winfree [5])

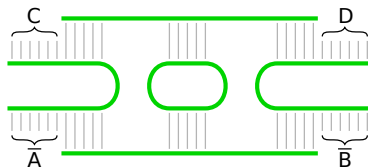


Obrázek: Struktura odpovídající pravidlu  $A \rightarrow aBbC$ .



# Souvislost s Chomského hierarchií

Dvojkřížené molekuly  $\leftrightarrow$  Turingův stroj (TS) (Winfree [5])

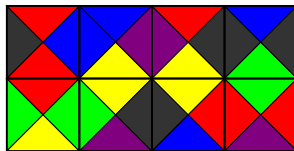


Obrázek: Dvojkřížená molekula, velmi stabilní (Seeman, Fu [4]).

# Wangovo dláždění

Čtvercové dláždění roviny (její části), kde

- dlaždice mají na hranách barvu z konečné množiny barev (lepidel)
- jsou orientované (zakázáno rotovat nebo překlápět)
- sousedit 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 sousedit se všemi
- dlážďení se utváří
  - z iniciální dlaždice
  - po jedné dlaždici
  - součet právě připojených lepidel musí být větší nebo roven zadané hodnotě (tzv. *teplota*, ozn.  $\tau$ )

# Studované složitosti I

## Biostep complexity $Bs(n)$

- počet laboratorních procedur (popsané v Adleman [2], Winfree [5])
- 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

# 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 aTAMu

aTAM je Turingovsky univerzální (TU)

- ve 2D při teplotě  $\tau = 2$ , Winfree [5]
- ve 3D při teplotě  $\tau = 1$ , Cook [3]

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
- v práci str. 15-16



# Meze studovaných složitostí

## Lemma

*Studované složitosti v tomto systému jsou omezené:*

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

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

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

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

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

BPP je třída jazyků rozhodnutelných pravděpodobnostním Turingovým strojem (PTS) v polynomiálním čase, považuje se za proveditelnou

Z předchozího lemmatu plyne:

## Důsledek

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

## Poznámka

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

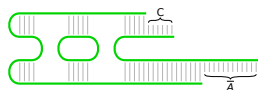
# Přizpůsobení aTAMu řešení NP problémů

Odvozen z Winfreeho ukázky řešení problému Hamiltonovské cesty – avšak srozumitelnější

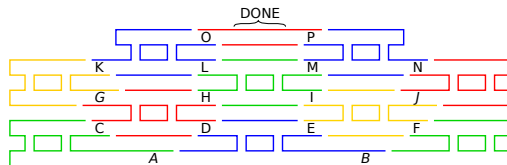
- $\tau = 2$
- 5 dalších typů dlaždic včetně daných sil lepidel
- pevně nastavená počáteční  $t_0$

## Poznámka

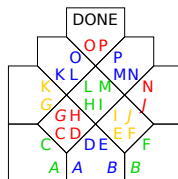
Tento model lze snadno simulovat klasickým aTAMem.



(a) Rohová molekula



(b) Schéma sebeskladu

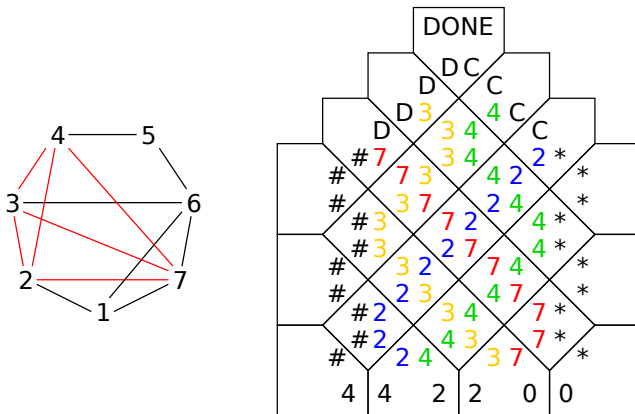


(c) Model

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



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

# Simulace v xgrow

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

- skriptem (na CD) k zadanému grafu generuji potřebné dlaždice
- se zapnutím kinetiky není jednoduché dosáhnout bezchybného dlážďení

# Další vyřešené problémy

## 3-obarvení grafu

- NP-úplný problém

## Grafový izomorfismus

- domnívá se, že není NP-úplný

# Reference I



Leonard M Adleman.

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



Leonard M Adleman.

On constructing a molecular computer.  
1995.



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.



Tsu Ju Fu and Nadrian C Seeman.

Dna double-crossover molecules.  
*Biochemistry*, 32(13):3211–3220, 1993.



## Reference II



Erik Winfree.

*Algorithmic self-assembly of DNA.*

PhD thesis, California Institute of Technology, 1998.