

Modelo de Paralelização para o Modelo de Repton*

Carlos Eduardo Souza Moreira[†], Ney Lemke, Gerson
Geraldo H. Cavalleiro

Programa de Ps-Graduação em Computação Aplicada – PIPCA
Centro de Ciências Exatas e Tecnológicas
Universidade do Vale do Rio dos Sinos
São Leopoldo – RS – Brasil
Tel. 590-3333, Fax 590-8162
{cadu, lemke, gersonc}@exatas.unisinos.br
Palavras Chave: Simulação, Método de Monte Carlo, Eletroforese

Introdução

Eletroforese é uma técnica experimental que permite separar misturas de cadeias de polímeros de acordo com sua massa. Ela é de grande importância na biologia molecular e na genética. A ideia básica da eletroforese é a separação de partículas carregadas, empurrando-as através de um gel estacionário através da aplicação de um campo elétrico. Polímeros maiores tendem a se mover de forma mais lenta assim ao iniciar-se o experimento com uma mistura de diferentes tipos de partículas, cada qual movendo-se com diferentes velocidades após um dado intervalo de tempo os polímeros mais leves terão se deslocado uma distância maior do que os mais pesados, permitindo a separação das partículas de acordo com sua massa. A técnica tem sido aplicada de forma generalizada em projetos genômicos na determinação da sequência de nucleotídeos que compõe o DNA de uma determinada espécie.

Neste artigo investiga-se o Modelo de Repton proposto por Rubinstein [1] como um modelo simples para a eletroforese de polímeros. O modelo de Rubinstein apesar de sua simplicidade é capaz de descrever de forma adequada os principais resultados experimentais. Ele pode ser simulado utilizando técnicas de Monte Carlo que em muitos casos de interesse possuem um custo computacional elevado. Visando amenizar este custo computacional, neste trabalho apresenta-se um modelo de paralelização para o modelo adequado para agregados de computadores onde cada nó é um SMP. Desta forma podemos explorar dois níveis de paralelismo inter e intra-nós.

O artigo está organizado na forma: inicialmente descreve-se o Modelo Repton e a técnica utilizado para simulá-lo, a seguir apresenta-se o modelo de paralelização proposto e os resultados preliminares e finalmente apresentam-se as perspectivas deste trabalho.

*Projeto Anahy – CNPq (55.2196/02-9)

[†]ITI-CNPq

Modelo Repton

No Modelo de Repton o DNA é representado como um polímero, que consiste em uma cadeia de N segmentos ou reptons. O movimento do DNA ocorre em um gel que em uma boa aproximação pode ser representado como um sistema bidimensional. Ao aplicar-se um campo elétrico sobre o sistema, o polímero começa a se mover como representado na figura 1, devido a natureza do polímero o movimento do sistema não é inteiramente livre. Visando simplificar o problema Rubinstein [1] propôs um modelo que considera apenas uma cadeia de polímeros se movendo ao longo de uma direção. A movimentação do polímero obedece as regras abaixo, que foram introduzidas para mimetizar o movimento do polímero em duas dimensões:

1. Um repton do interior da cadeia pode mover-se para uma posição vizinha, desde que um de seus vizinhos imediatos (à esquerda ou à direita) já esteja na posição para o qual ele irá movimentar-se, e o outro vizinho esteja na posição que o repton está deixando.
2. Os dois reptons dos extremos da cadeia podem mover-se para qualquer posição adjacente a sua atual, desde que tal movimento não o coloque mais de uma posição de distância de seu vizinho na cadeia.

A figura 1 representa esquematicamente estas regras onde cada ponto corresponde a um dos reptons.

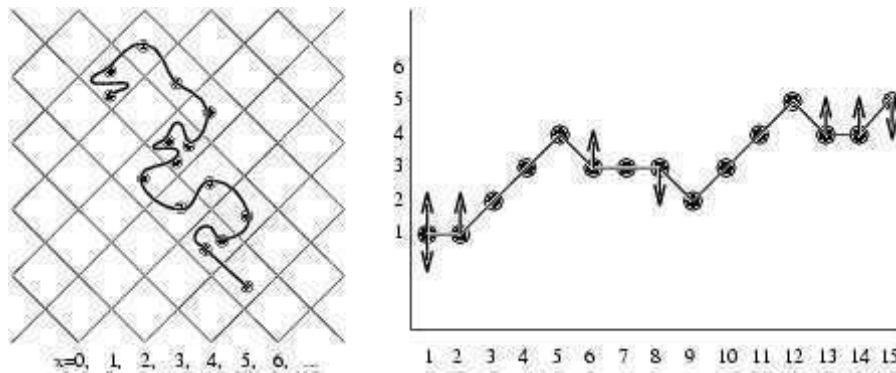


Figura 1: À esquerda representa-se o movimento de de um polímero em duas dimensões. À direita representa-se uma configuração típica do modelo de Repton. Se os reptons estão numerados ao longo da cadeia, iniciando com o fim no canto superior esquerdo, nós podemos caracterizar essa configuração plotando as posições de x como uma função do número de reptons.

Apesar de sua simplicidade não existe solução analítica para este modelo, exceto para cadeias pequenas. Para que se possa investigar o modelo urge a utilização de técnicas de Monte Carlo. Inicialmente seleciona-se randômicamente um dos N reptons e uma das duas possíveis direções (para cima ou para baixo). Se o movimento é permitido, de

acordo com as regras descritas acima, então o movimento é validado, caso contrário o movimento é rejeitado. Para simular a eletroforese deve-se incluir o efeito do campo elétrico, introduzido através do parâmetro E . Movimentos na direção do campo são sempre aceitos enquanto movimentos opostos são aceitos com probabilidade $\exp(-E/2)$ (em uma situação análoga ao algoritmo de Metropolis).

Modelo de Paralelização

Dadas as características do modelo de Repton: pouco sincronismo e comunicação, podemos propor um modelo eficaz de paralelização para a arquitetura de interesse. Além disso, devido a natureza estocástica do modelo, a obtenção de resultados de interesse implica na simulação de amostras com diferentes sementes de números aleatórios, o que nos permite explorar a paralelização entre os nós de um agregado.

Como pudemos perceber na análise das regras de movimentação de um repton, citadas na seção anterior, a nova posição do repton depende exclusivamente, da direção sorteada aleatoriamente e da posição atual de seus vizinhos. A partir disso podemos pensar na implementação concorrente para o problema, que pode consistir em responsabilizar um fluxo de execução por um grupo de reptons composto por, no mínimo, um repton e seus vizinhos diretos. As posições dos reptons das fronteiras do grupo, todavia, são necessárias para a determinação da posição de outros reptons não incluídos no grupo, gerando a necessidade de sincronização entre fluxos de execução para que o polímero possa avançar.

Para solucionar esta dependência de dados, pode-se utilizar uma seção crítica. Este mecanismo garante que apenas um fluxo de execução está acessando determinado dado, naquele instante de tempo. Se um repton eleito pertencer a fronteira do grupo o fluxo de execução bloqueará até que o repton selecionado termine seu cálculo e saia da seção crítica.

Além disso utilizando a interface MPI pode-se realizar simultaneamente diversas simulações de uma determinada configuração. As simulações consomem da ordem de vários minutos, e são inteiramente independentes. Uma vez realizadas as simulações procede-se a um processo de análise que consome uma fração desprezível do custo computacional. Atualmente a infra-estrutura para a realização destas simulações está em construção.

A figura 2 apresenta resultados preliminares de nossa simulação, no detalhe apresentamos a dependência da posição com o tempo para um polímero com 10 unidades e $E = 0.6$ realizamos médias sobre 100 amostras. O gráfico maior representa a velocidade v do polímero em função do campo E , os eixos estão reescalados para facilitar a comparação com resultados experimentais. Os dados foram obtidos usando 2 fluxos de execução e médias sobre 10 amostras.

Perspectivas

Como continuação deste trabalho espera-se avaliar o desempenho da arquitetura proposta investigando a influência no desempenho do tamanho do polímero, N e do número p de diferentes fluxos de execução. Espera-se com isto determinar um valor

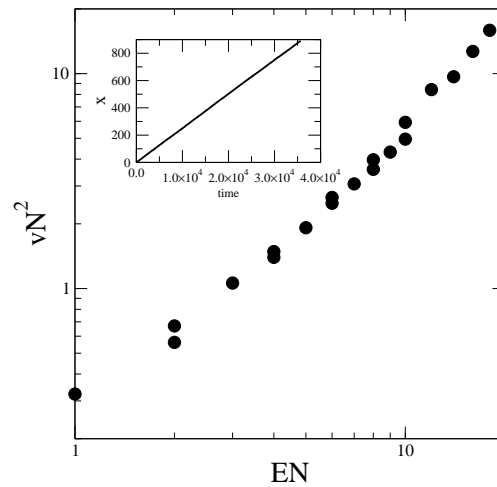


Figura 2: O detalhe apresenta a posição de um polímero em função do tempo para $N = 10$ e $E = 0.6$. A figura maior apresenta a dependência funcional da velocidade com o campo aplicado. Os gráficos foram reescalados para facilitar a comparação com resultados experimentais e outras simulações.

ótimo de p que garanta o máximo desempenho. Os testes serão realizados em máquinas utilizando o sistema operacional GNU/LINUX e a versão 2.5 do núcleo de execução, que permite desempenho superior neste tipo de situação que a versão 2.4.

A arquitetura proposta neste trabalho pode ser facilmente estendida para qualquer simulação de Monte Carlo baseada em rede. Espera-se desenvolver uma API que facilite a implementação deste tipo de simulação permitindo ao usuário abstrair os detalhes da arquitetura que possui.

Referências

- [1] Rubinstein M, "Discretized model of entangled-polymer dynamics", *Phys. Rev. Lett.*, vol. 59, pp 1946-1948, 1987.