



УДК 514.86

СПРОЩЕНИЙ АЛГОРИТМ ПРОСТОРОВОГО ХЕШУВАННЯ ДЛЯ ЗАДАЧ МОЛЕКУЛЯРНОЇ ДИНАМІКИ

Ніщин О. Ю., д.т.н.,

Дашкевич А. О., к.т.н.,

Охотська О. В.,

Національний технічний університет «Харківський політехнічний інститут»

Тел. (057)707-64-31

Мацулевич О. Є., к.т.н.

Таврійський державний агротехнологічний університет

Анотація - розглянуто алгоритми знаходження найближчих сусідів у багатовимірних просторах. Запропоновано новий алгоритм, на основі просторового хешування, який дозволяє суттєво скоротити використання машинного часу та кількість операцій, що виконуються на відміну від методу повного перебору. Наведено основні етапи використання алгоритму.

Ключові слова: алгоритм найближчих сусідів, просторове хешування, метод повного перебору.

Постановка проблеми. Для багатьох задач молекулярної динаміки суцільного середовища найбільш витратним є знаходження найближчих сусідів у багатовимірних просторах. На сьогоднішній час немає відомих точних алгоритмів для рішення цієї проблеми ніж лінійний пошук. Наближені алгоритми забезпечують більш швидкий пошук, з незначною втратою точності, але наведені в них параметри не належать корекції та уточненню. Тому стає необхідним наближений алгоритм пошуку найближчих сусідів, який буде приймати будь-які необхідні параметри та бажану ступінь точності без втрати швидкості розрахунку.

Аналіз останніх досліджень. У наступний час використовують алгоритми основані на наступних методах:

- метод зв'язаних чарунок (LC – linked cells);
- метод сортування взаємодії (IS – interaction sorting);
- метод Верле таблиць (VT – Verlet table);
- метод періодичного переупорядкування атомів (LCR –



linked cell based reordering);

– метод рандомізованих kd-дерев (K-D Tree).

Основним недоліком наведених методів є те, що з ростом кількості атомів (точок) зростають вимоги до оперативної пам'яті тому швидкість розрахунків зменшується. Крім того атоми рухаються швидше, ніж найбільша відстань, що допустима між атомами, що робить список сусідів недійсним. Параметри, що входять до складу цих методів є остаточними, та не можуть бути змінені. Усе це робить моделювання неефективним [1-5].

Формулювання цілей статті. Розробка спрощеного алгоритму пошуку найближчих сусідів в масиві часток у Евклідовому просторі для задач молекулярної динаміки.

Основна частина. Найбільш точним методом віднаходження найближчих сусідів є метод повного перебору усіх пар часток, але це потребує великих витрат часу на обчислення, адже необхідно розраховувати квадратний корінь, що є витратною операцією з точки зору машинних обчислень. Інший напрямок включає використання спеціальних структур даних (kd-дерев, r-дерев та ін.) або алгоритмів зменшення розмірності (просторове хешування). Такі алгоритми дозволяють значно скоротити час, необхідний для обчислень за рахунок точності. В роботі [6] розроблено метод просторового хешування, в якому запропоновано замінити координати x , y та z деякої точки на одну координату — хеш, який запропоновано розраховувати наступним чином:

$$h = \frac{x}{c} 2^k + \frac{y}{c} 2^m + \frac{z}{c} 2^n$$

де c – розмір сітки;

k , m та n – деякі константи, $k > m > n$ (або $k < m < n$).

В роботі [7] запропоновано наступний метод обчислення хешу для точки:

$$h = ((x \cdot p_1) XOR (y \cdot p_2) XOR (z \cdot p_3)) \bmod n,$$

де $p_1 = 73856093$, $p_2 = 19349663$, $p_3 = 83492791$ – великі прості числа;

XOR – побітова операція “Виключне АБО”;

$\bmod n$ – операція віднаходження залишку від ділення на n ;

n – загальна кількість точок.

Наведені методи дозволяють швидко розраховувати найближчих сусідів, але потребують або експериментального підбору констант, або розрахунки з великими числами, що може бути корисним



при використанні графічних процесорів (відеокарт). Таким чином вказані алгоритми не є універсальними з точки зору часу.

В роботі запропоновано модифікований метод просторового хешування для швидкого віднаходження найближчих сусідів, в якому хеш розраховується наступним чином (для двомірного випадку):

- 1) простір розбивається на сітку з розміром $C \times C$;
- 2) для координат усіх точок (x, y) розраховують індекси j – номер колонки, i – номер рядку у сітці клітинки, у якій знаходиться точка;
- 3) за заданими індексами розраховують значення хешу:

$$h = j \cdot 10^n + i \cdot 10^m,$$

де n та m є невеликими цілими числами, значення яких залежать від кількості розрядів у j та i , відповідно.

У загальному випадку:

$$n = m + 1, m = 0.$$

Для практичного використання даного хешу додатково вводяться дві структури даних:

- 1) список номерів вихідних точок, та їх хешів;
- 2) список хешів, та відповідних їм точок.

Для пошуку найближчих точок необхідно знайти усі точки з таким же хешем, та точки, у яких хеш відрізняється на одиницю в меншу або більшу сторону.

За результатами роботи було проведено тестування для заданої множини точок, які було згенеровано випадково (рис. 1), серед яких обиралась деяка довільна точка (рис. 2) і проводився розрахунок найближчих сусідів за методами повного перебору та просторового хешування. Результати роботи методу повного перебору продемонстровано на рис. 3, результати роботи методу модифікованого просторового хешування — на рис. 4.

Можна побачити, що метод повного перебору найбільш точно знаходить усі точки, що знаходяться біля заданої, але для такої точності необхідно перебирати усі пари точок, тобто кількість операцій пропорційна квадрату від кількості вихідних точок. На відміну від повного перебору, запропонований метод знаходить усі точки, які належать до клітинки з вихідною точкою, та у суміжних з нею клітинках. Це призводить до визначення надлишкових точок, від яких необхідно позбавлятися при подальшому розрахунку.

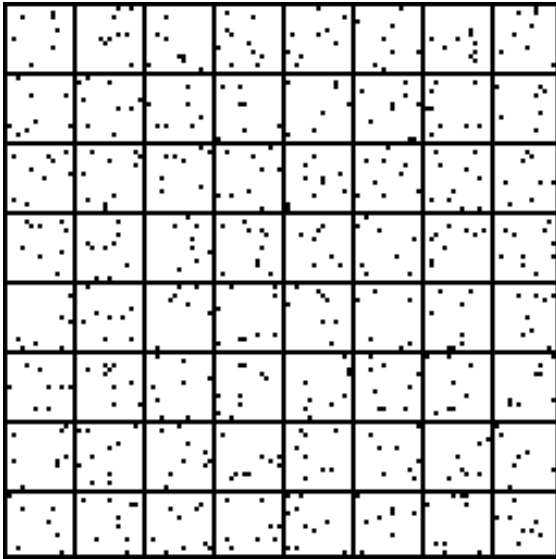
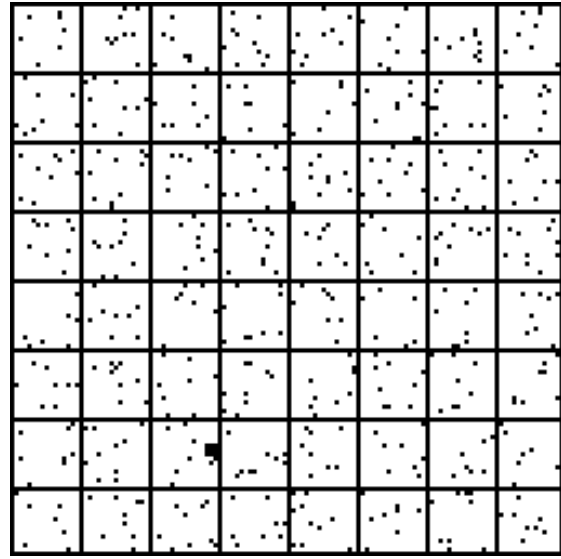
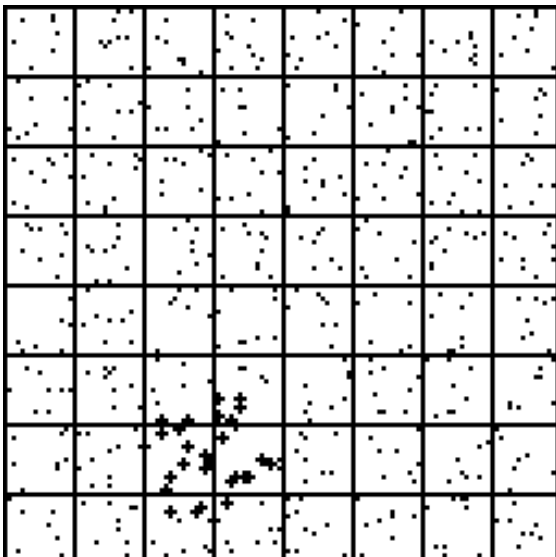
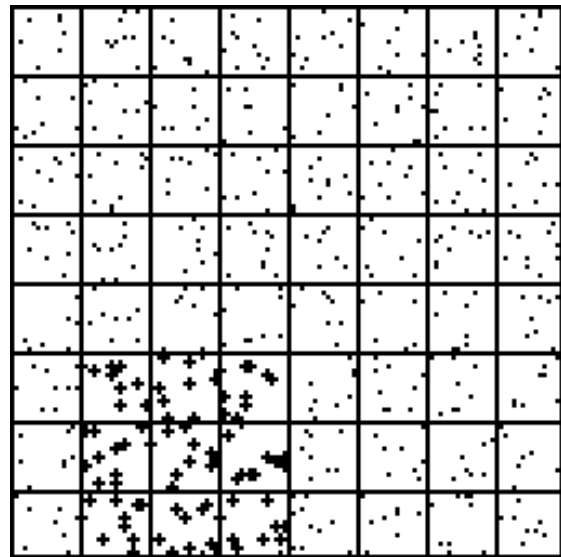


Рис. 1 Вихідна множина точок

Рис. 2 Точка для пошуку
найближчих сусідівРис. 3 Найближчі сусіди за
методом повного переборуРис. 4 Найближчі сусіди за
методом просторового хешування

Висновки. За результатами моделювання встановлено, що запропонований метод дозволяє встановити усі точки, що знаходяться в близьких клітинках сітки, що є надлишковим з точки зору кількості операцій. Наприклад, для вихідної множини з 640 точок, методом повного перебору для довільної точки було знайдено 29 найближчих точок, а запропонованим методом — 89 точок. Таким чином алгоритм потребує додаткового сортування точок за параметром необхідної відстані, наприклад, з допомогою алгоритму повного перебору для відібраних точок.

Література

1. Фомин Э. С. Сортировка массивов коротких последовательностей



- на GPU для метода сортировки взаимодействий в молекулярной динамике / Фомин Э. С. // Вестник УГТУ - Уфа : УГАТУ, 2013 Т. 17, № 2 (55). с. 75–84
2. *Nister D.* Scalable recognition with a vocabulary tree./ Nister D. Stewenius H. // CVPR, 2006, pp.2161–2168
 3. *Liu T.* An investigation of practical approximate nearest neighbor algorithms./ Liu T., Moore A., Gray A., and Yang K. // Neural Information Processing Systems, 2004
 4. *Schindler G.* Cityscale location recognition./ Schindler G., Brown, M., and Szeliski, R // CVPR, 2007, pp.1–7
 5. *Lowe D. G.* Distinctive image features from scaleinvariant keypoints./ Lowe D. G. // Journal of Computer Vision// 2004, pp.91–110
 6. *Nguyen H.* (ed.) GPU Gems 3 / Nguyen H. // Addison Wesley, 2007, 1008 p.
 7. *T. Ize* Asynchronous BVH Construction for Ray Tracing Dynamic Scenes on Parallel Multi-Core Architectures / T. Ize, I. Wald, and S. G. Parker // Proceedings of the 2007 Eurographics Symposium on Parallel Graphics and Visualization, 2007.

УПРОЩЕННЫЙ АЛГОРИТМ ПРОСТРАНСТВЕННОГО ХЕШИРОВАНИЯ ДЛЯ ЗАДАЧ МОЛЕКУЛЯРНОЙ ДИНАМИКИ

А.А. Дашкевич, А.Ю. Ницын, Е.В. Охотская, А.Е. Мацулевич

Аннотация — рассмотрены алгоритмы определения ближайших соседей в многомерных пространствах. Предложен новый алгоритм на основе алгоритма пространственного хеширования, который позволяет существенно уменьшить затраты машинного времени и количество операций на поиск по сравнению с методом полного перебора. Приведены основные этапы работы алгоритма.

SIMPLIFIED SPATIAL HASHING ALGORITHM FOR MOLECULAR DYNAMICS PROBLEMS

A. Dashkevich, A. Nitsyn, E. Ohotskaya, A Matsulevich.

Summary

Nearest neighbors algorithm in multidimensional spaces is described. The modified spatial hashing algorithm, that allows to reduce time and operation number is proposed. Main stages of algorithm is shown.