

基于自回避行走数值模拟的并行计算*

高和蓓¹, 李洪^{2,3}, 龚斌²

(1. 温州职业技术学院 信息系, 浙江 温州 325035; 2. 山东大学 计算机科学与技术学院, 济南 250100; 3. 温州大学 计算机系, 浙江 温州 325035)

摘要: 提出一种蒙特卡洛方法并行计算高分子的单链性质, 高分子链采用自回避行走方法生成初始状态, 链节点的随机运动通过键长涨落方法进行模拟。自回避链的配分函数没有解析解, 研究高分子链的性质往往要计算大量的样本。利用样本之间的独立性进行并行计算达到线性加速比, 使蒙特卡洛模拟高分子链的并行计算时间缩短到科学计算可以接受的时间范围。

关键词: 并行计算; 线性加速比; 自回避行走; 计算机模拟

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1001-3695(2014)04-1039-04

doi:10.3969/j.issn.1001-3695.2014.04.020

Parallel computing based on numerical simulation of self-avoiding walk

GAO He-bei¹, LI Hong^{2,3}, GONG Bin²

(1. Dept. of Information Science, Wenzhou Vocational & Technical College, Wenzhou Zhejiang 325035, China; 2. School of Computer Science & Technology, Shandong University, Jinan 250100, China; 3. Dept. of Computer Science, Wenzhou University, Wenzhou Zhejiang 325035, China)

Abstract: This paper presented a parallel Monte Carlo method for the calculation of a single polymer chain. It used self-avoiding walk method to generate the initial state of polymer chain, and simulated the random motion of the chain by the bond-fluctuation method. For no analytical solution may calculate chain partition function, properties of polymer chains tend to be calculated by large number of samples. It developed parallel program for computing the independent samples to reach the linear speedup. Parallel execution time was acceptable for parallel Monte Carlo simulation of polymer chains.

Key words: parallel computing; linear speedup; self-avoiding walk(SAW); computer simulation

0 引言

计算机产生的动力是因为人们想发明一种能够用于科学计算的机器。从1946年第一台计算机诞生到现在,短短的六十多年,计算机经历了飞速发展,遵循摩尔定律,晶体管数量每18个月翻番的同时,计算性能也会增加一倍。近二十年来,并行计算在实际应用中越来越多,已经被广泛应用于生物信息、材料科学、流体力学、药物设计、计算化学、气象科学等学科领域的科学计算,而且随着研究的开展与深入对计算能力需求有增无减。开发和应用数学建模、数据分析、计算机仿真技术等进行众多应用领域的高性能计算。在这些应用领域中,数据量和复杂性不断增长,单单依靠理论分析和实验已难以应付。因此,必须依靠大规模计算机模拟技术,从海量数据中提取最有用的信息,形成许多应用领域与计算机科学的交叉学科——计算科学,应用高性能计算机提高各自的应用学科(如物理、化学或工程等学科)中最先进的理论和技术水平^[1,2]。

高分子的计算理论与建模也是随着计算机的发展而成为计算化学中的一个重要分支,高分子链在界面吸附成为了高分子、生物科学等方面的一个研究热点^[3-6]。往往借助于计算机

模拟进行研究,线型链是高分子常用的一种简化模型,由大量相同结构的单元串联而成,一般高分子链因热运动使分子在空间中的形状不停地发生变化形成各种构象。高分子的分子量往往非常大,因此高分子链的构象数也十分大。于是对单个高分子链的构象统计也是一个复杂的计算问题,高分子链构象统计通常需要较长的计算时间,采用串行计算只能局限于计算一些短链或局部的性质,近些年常采用并行计算进行研究^[7,8]。

本文对高分子单链的数值模拟进行并行计算研究,首先采用自回避行走方法产生,然后利用键长涨落方法模拟无规热运动,考虑高分子链与表面接触有吸附作用,而且相互作用强度与温度有关。本文采用并行 Monte Carlo 方法模拟高分子链在吸附表面的性质,在高性能集群上利用多核计算独立的样本,计算结果再进行归约,可以达到理想的线性加速比。

1 模型

高分子链的模型分为无格点模型和有格点模型。虽然无格点模型较真实地反映了高分子链的结构和构象,但是计算机模拟较复杂,主要表现在链重叠检测和向量化计算。因此,高分子链计算中也经常会采用有格点模型。本文模拟空间建立

收稿日期: 2013-06-15; **修回日期:** 2013-07-27 **基金项目:** 国家自然科学基金资助项目(61070017,11304231);温州市科技计划项目(G20110045)

作者简介: 高和蓓(1979-),女,浙江温州人,讲师,硕士,主要研究方向为计算机模拟、图像处理与科学计算(bogolyx@163.com);李洪(1978-),男,浙江缙云人,副研究员,在职博士,主要研究方向为并行计算、高分子模拟计算;龚斌(1964-),男,山东济南人,教授,博导,博士,主要研究方向为网格与高性能计算、图形与图像处理等。

在一个简立方体网格 $L_x \times L_y \times L_z$, 链节点的坐标值 (x, y, z) 限制在网格格点上取值, 格子的边长 a 通常取为单位长度 1。在 $z=0$ 放置一个不可被高分子链所穿越的无穷大的平面, 链节点全部位于平面的上方 ($z>0$), 在 x, y 方向采用周期性边界条件^[9-11]。

1.1 链的生长算法

在自回避行走(SAW)模型中, 空间中的每个格点只能被一个链节点占有, 链长为 N 的高分子链, 相当于一步一步地行走 N 步所生成。二维 SAW 链的示意图如图 1 所示, 它是自回避免链生长后的初始状态, 其中链长 $N=20$ 。高分子链在某一时刻为何种构象完全是随机的, SAW 链的配分函数没有精确的函数表达式可以直接解析, 通常采用 Monte Carlo 方法模拟高分子 SAW 链的静态和动态性质^[5-7]。计算方法用类 C 语言形式化描述如下:

a) 初始化参数

```
//用#define 宏定义链长 N、模拟空间的规模 Lx, Ly, Lz 等参数
a[Lx][Ly][Lz] = {0}; /* 空间格点是否有链节点, 0 表示空格点, 1 表示 A 类节点, 2 表示 B 类节点 */
chain[N][3] = {0}; //记录线性链的坐标
b[6][3] = {{0,0,1}, {0,0,-1}, {0,1,0}, {0,-1,0}, {1,0,0}, {-1,0,0}};
//沿坐标轴的六个可行走方向
```

b) 固定链的首节点坐标在平面中心位置上 ($L_x/2, L_y/2, 1$), 即

```
a[Lx/2][Ly/2][1] = 1; //在三维数组中标志第一个链节点
chain[0][0] = Lx/2;
chain[0][1] = Ly/2;
chain[0][2] = 1; //给第一个链节点的坐标赋值
```

c) 从节点 $1 \sim N-1$ 通过随机行走产生新节点

```
for(i=0; i<N-1; i++)
do{
    k = (int) 6 * rand(); //随机产生一个运动方向
    x = chain[i][0] + b[k][0];
    y = chain[i][1] + b[k][1];
    z = chain[i][2] + b[k][2];
} while (a[x][y][z] == 0 && z > 0);
//直到有最近邻格点为空, 且位于表面上方
for (j=0; j<3; j++)
chain[i+1][j] = chain[i][j] + b[k][j];
//通过节点 i 随机行走一步产生第 i+1 个节点坐标
a[chain[i+1][0]][chain[i+1][1]][chain[i+1][2]] = 1; //置该空间格点被链节点占有的标志 1-A, 2-B
}
```

1.2 链的运动

高分子链运动通过键长的涨落来实现, 相邻的两个链节点当且仅当位于在同一个单位长度的立方体上的两个不同顶点, 三维格点模型中键长允许取值为三种情况, 即单位边长 (本文取单位长度 1)、单位正方形上的对角线 ($\sqrt{2}$) 和单位立方体上的对角线 ($\sqrt{3}$)。由于键长可涨落性, 三维情况下每个格点的配位数 z 为 26, 二维配位数 z 为 8。每个与表面距离为 1 的节点有作用能 E 。自回避免链经过随机布朗运动后, 构象会发生变化, 在不考虑链节点间以及链与表面之间相互作用的情况下 (当温度 T 取无穷大时), SAW 链经过一定时间运动后的示意图如图 2 所示。

高分子链不停地做布朗运动, 随机选择一个链节点, 随机选择延坐标轴 x, y, z 的六个方向之一运动, 该尝试是否被接受还要同时满足下面条件:

a) 随机选择第 i 个链节点运动, $i = (\text{int}) N \times \text{rand}()$ 。若 $i=0$ 表示第一节点, 重新选择。

b) 检测第 i 个链节点沿坐标轴六个方向上的最近邻格点是否有空格点, 并记录相应的序列, 从序列中随机选择一个方向 k 。

e) 新格点位置位于表面上方 ($\text{chain}[i][2] + b[k][2] > 0$) 是否满足。

d) 链节点的新位置与前后最近邻链节点是否满足键长的允许条件 ($1, \text{sqrt}(2), \text{sqrt}(3)$)。

e) ΔE 是运动前后能量迁移量, 如果 $\Delta E > 0$, 那么波尔茨曼因子 $\exp(-\Delta E/k_B T)$ 是否大于一个新产生的随机数 $\text{rand}()$ 。

若 a) ~ e) 都满足, 则链节点成功运动一次, 若平均每个链节点尝试运动一次, 则为一个蒙特卡洛时间单位 MCS (Monte Carlo step)。

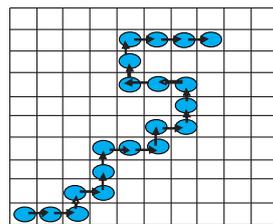


图1 二维格点SAW链生长的初始状态示意图

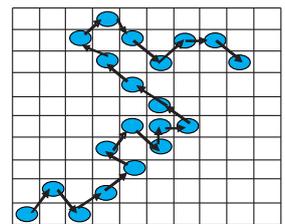


图2 二维格点SAW链运动后状态的示意图

1.3 并行 Monte Carlo 方法

高分子链的静态和动态性质都与链节点与表面的相互作用强度有关, 本文中链节点运动到与表面的距离为 1 时 (即 $z=1$) 可获得作用能 $E = -1$, 表示链与表面有相互吸引作用, 强度还与温度有关。为了研究链性质与温度之间的关系, 尤其是确定临界吸附点, 需要通过计算大量的样本进行抽样统计。每计算一个独立样本, 重新生长 SAW 链, 退火法模拟计算不同温度下的高分子构象、热力学和动力学性质^[5,6,9]。初始状态先从一个高温 $T=8$ 开始, 然后以步长 ΔT 减小温度 T (本文在临界点附近取 $\Delta T=0.05$, 其他温度 ΔT 取值稍大一些)。在同一个温度 T , 链构象随运动时间不断发生变化, 每个样本先运行足够的松弛时间 $\tau = 2.5N^{2.13}$ MCS, 然后再运行 100τ MCS 计算各参量平均值, 每隔 0.1τ 记录一个数据, 每个温度 T 运行 10^3 个独立样本求平均。每个计算结果有 10^6 个数据进行平均。每个独立样本可以通过高性能计算中心机群并行计算, 本文通过消息传递接口 (message passing interface, MPI) 来实现, 并行算法用流程图描述, 如图 3 所示。

2 结果分析与讨论

2.1 加速性能

自回避免链的数值模拟在超大规模、超高性能科学计算应用的高性能服务器集群上进行了测试与计算。测试的环境为: 基于刀片集群的天梭 TS10000 高性能集群系统, 80 个 NX560T 刀片, 采用 32 nm 制程工艺英特尔至强 5600 系列处理器, 共有 960 个计算核数, 计算峰值达 10 万亿次。机群的互连网络是采用 40 Gbps 的 Infiniband。采用先进的水冷系统提供了一个稳定、高效的运行环境。为了测试相同链长在不同计算核数的加速比, 采用的参数为链长 $N=100$, 样本数为 480, 计算的蒙特卡洛时间为 10^6 MCS。在相同参数下, 不同处理器核数下并行

计算得到的加速比如图 4 所示,并行计算的加速比达到线性加速比。

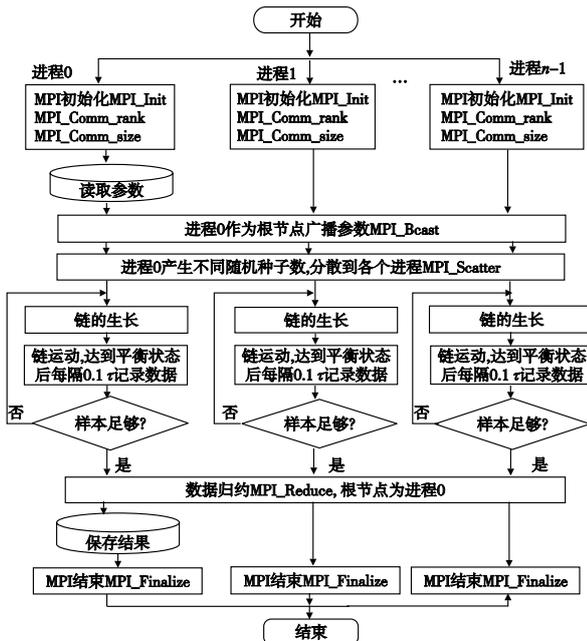


图3 并行程序的执行流程

对不同链长 N 所需要的计算时间 t 进行了实测与估计。链长 N 为 10、20、50、100、200 和 400, 计算样本数为 240, 每个样本从高温 $T=5$ 到低温 $T=0.1$ 共计算 25 个温度值, 每个温度值计算的蒙特卡罗时间为 $100 \tau \text{MCS}$, 其中 $\tau = 2.5N^{2.13}$ 。测试中采用 240 核计算不同链长得到的计算时间如图 5 所示。图 5 中纵坐标左边的时间单位为 s, 右边的时间单位为 d。测试数据在计算时间 t 和链长 N 的双对数图中呈线性递增关系。根据不同链长在 240 核上的计算结果, 单个链节点运动一次的平均执行时间约为 $4 \times 10^{-7} \text{ s}$ 。以此估计不同链长需要的计算时间, 每个样本运行 25 个退化温度, 每个温度运行时间 $100 \tau \text{MCS}$, 得到的估计时间与测试所得到的时间相一致。随着链长 N 的增加, 实际测试时间与推算时间之间的误差越来越不明显, 如图 5 所示, 测试时间与推算时间随链长 N 增加几乎重合。

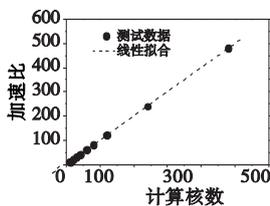


图4 数值模拟计算的加速比

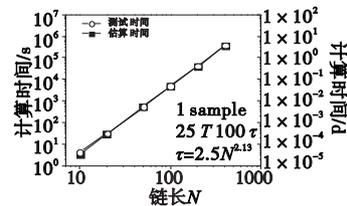


图5 模拟不同链长计算时间

预测同一链长在不同核数上计算所需的并行计算时间, 可以选择合适的核数来进行科学计算。依据是不同链长的一个链节点运动一次所需的平均时间几乎相同, 以此来估计长链在不同处理器核数的执行时间。通常模拟中需要运行约 1 000 样本求平均。根据高性能计算集群的硬件环境, 确定计算样本数为 960, 每个样本计算 25 个温度值, 每个温度值计算的蒙特卡罗时间为 $100 \tau \text{MCS}$, 其中 $\tau = 2.5N^{2.13}$, 链长 $N=400$ 在不同核数上所需的计算时间 t 的估计值如表 1 所示, 在单处理器上计算需要 10.61 y, 这个时间数量级对于该领域的科学计算是难以接受的。如果在 960 核上计算时间大约只要 4 d, 对于科学计算来说, 这个时间是大家所容易接受的。

表 1 估计不同核数下并行计算时间

核数	960	480	240	120	60	24	12	1
t/d	4.035	8.071	16.14	32.28	64.6	161	323	3874
t/y	0.011	0.022	0.044	0.088	0.18	0.44	0.88	10.61

2.2 计算结果分析

在自回避链的模型中, 本文模拟了两嵌段共聚高分子 $A_{N/2}B_{N/2}$ 的性质, 其中 N 为链长, 链节点由两种节点类型构成, 分别标记为 A 和 B , 链节点 A 与表面存在吸附作用 $E_{AS} = -1$, 链节点 B 与表面无吸附作用 $E_{BS} = 0$ 。本文考虑 A 的长度和 B 的长度相等的情况。在表面上方的 $(Lx/2, Ly/2, 1)$ 位置固定一个不可移动的节点 X , 链的一端固定在 X 节点上。链由两个链段构成, 分别为 $N/2$ 个节点的 A 段和 $N/2$ 个节点的 B 段。若 A 链段固定在 X 上, 该两嵌段共聚高分子链表示为 $XA_{N/2}B_{N/2}$, 若 B 链段固定在 X 上则表示为 $XB_{N/2}A_{N/2}$ 。分别模拟了它们的构象性质与温度的关系, 如图 6 所示, 其中链长 $N=400$, 温度 T 分别为 2 和 0.2。图中深灰色实心圆表示 A 节点, 浅灰色实心圆表示 B 节点, X 节点用 \star 号表示。在高温 $T=2$ 时, 除了 X 固定在表面上外, 几乎没有其他节点被吸引到表面上。在低温 $T=0.2$ 时, 在链的两种不同序列 $XA_{N/2}B_{N/2}$ 和 $XB_{N/2}A_{N/2}$ 中, A 节点都几乎吸附在表面上, $XA_{N/2}B_{N/2}$ 中的 B 链段伸展在三维空间中, $XB_{N/2}A_{N/2}$ 中的 B 链段的一端固定在 X 上, 一端被吸附在表面上的 A 链段牵制着, B 链段的中间部分节点仍在三维空间上。模拟结果表明, 两嵌段共聚高分子链, 不同链段的一端拴在表面上在低温时表现出不同的构象性质。

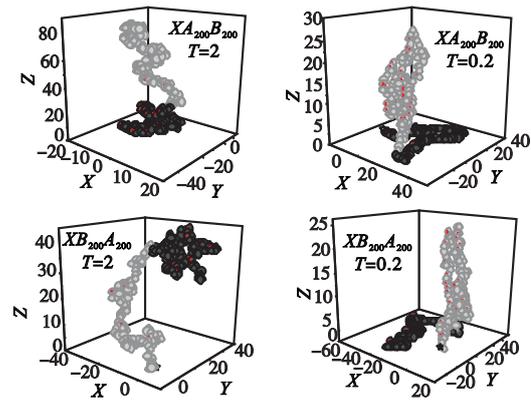


图6 自回避链在不同温度的构象

在模拟中计算了高分子链与表面接触的平均吸附节点数 $\langle M \rangle$ 与温度 T 之间的关系, 如图 7 所示, 分别计算了链长 N 为 10、20、40、50、80、100、150、200、300 和 400 的情况, 其中图 7 (a) 是 $XA_{N/2}B_{N/2}$, 图 7 (b) 是 $XB_{N/2}A_{N/2}$ 。由图中可以看出, 对于不同的链长 N 都存在一个临界吸附点 T_c 。高温 $T > T_c$ 时被吸附的节点数平均值 $\langle M \rangle \approx 0$, 这是由于节点与表面的相互吸附作用不强, 另外平面存在空间上排斥作用, 无论是节点 A 还是节点 B 都伸展在空间中而很少与表面相互吸引。在低温 $T < T_c$ 时被吸附的节点数平均值 $\langle M \rangle \neq 0$, $\langle M \rangle$ 随着温度 T 降低而增大, 较低温度时 A 节点几乎被吸附在表面上, 即 $T \rightarrow 0$ 时, $\langle M \rangle = N/2$ 。临界吸附温度通过不同链长在不同温度下的 M 值来标度, 通过公式

$$M = N^\nu (a_0 + a_1(T - T_c)N^{1/\delta} + O((T - T_c)N^{1/\delta})^2)) \quad (1)$$

计算, 从模拟结果发现两嵌段高分子链 $XA_{N/2}B_{N/2}$ 和 $XB_{N/2}A_{N/2}$ 的相变点相同 $T_c \approx 1.6$, 而且与均聚物的临界温度相一

致^[5,6,9]。本文通过并行计算不同样本的高分子构象,大大减少了计算时间。

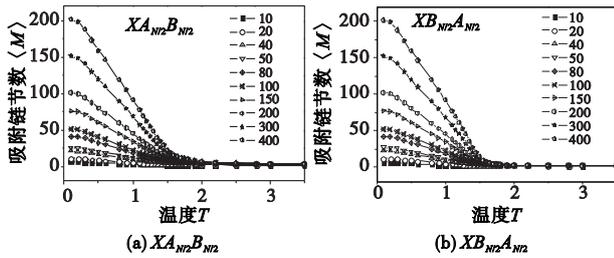


图7 表面接触的节点数与温度 T 之间的关系

3 结束语

本文采用蒙特卡洛并行计算高分子链的生长和运动,自回避行走方法模拟生成高分子链的初始状态,通过链长涨落模型计算链节点运动。高分子链的静态和动态性质往往需要计算大量的样本进行抽样统计,提出了蒙特卡洛并行计算方法把不同的独立样本采用并行计算达到线性加速比,尤其是链长 $N = 400$ 在单处理器计算需要 10.61 y,通过并行高性能计算大约只要 4 d 的时间。

本文从独立样本上进行并行计算,使高分子链的蒙特卡洛模拟的计算时间缩短到科学计算可以接受的时间范围。如果为了进一步减少并行计算时间或计算更长链的性质,可以进一步挖掘问题的并行性,如一个独立样本中各链节点运动的并行计算,通过 MPI + OpenMP 或者 MPI + GPU CUDA 的模式进一步优化计算模型,使之能够在上万个处理器或核的更大规模并行计算系统上运行。

参考文献:

[1] NEDELCO S, WERNER M, LANG M, *et al.* GPU implementations of the bond fluctuation model[J]. *Journal of Computational Physics*, 2012, 231(7):2811-2824.

[2] MARINS J L, YUN Xiao-ping, BACHMANN E R. An extended Kalman filter for quaternion-based orientation estimation using MARG sensors[C]//Proc of International Conference on Intelligent Robots and Systems. 2001:2003-2011.

[3] IVANJKO E, VASAK M, PETROVIC I. Kalman filter theory based mobile robot pose tracking using occupancy grid maps[C]//Proc of International Conference on Control and Automation. 2005:869-874.

[4] 马瑞平,魏东,张明廉.一种改进的自适应卡尔曼滤波及在组合导航中的应用[J]. *中国惯性技术学报*, 2006, 14(6):37-40.

[5] HAN Ping, GAN Hao-liang, HE Wei-kun, *et al.* Aircraft attitude estimation based on central difference Kalman filter[C]//Proc of the 11th IEEE International Conference on Signal Processing. 2012:294-298.

[6] OU Yang, XIA Yuan-qing, FU Meng-yin. A modified method of nonlinear attitude estimation based on EKF[C]//Proc of the 12th International Conference on Control Automation Robotics & Vision. 2012:901-906.

[7] MARINS J L. An extended Kalman filter for quaternion-based attitude estimation[D]. Monterey:Naval Postgraduate School, 2000.

[8] 秦永元. 惯性导航[M]. 北京:科学出版社, 2005.

[9] 董冀. MTi 微惯性航姿系统/GPS 组合技术研究[D]. 哈尔滨:哈

[2] UHLHERR A. Parallel Monte Carlo simulations by asynchronous domain decomposition [J]. *Computer Physics Communications*, 2003, 155(1):31-41.

[3] JEPPESEN C, WONG J Y, KUHL T L, *et al.* Impact of polymer tether length on multiple ligand-receptor bond formation [J]. *Science*, 2001, 293(5529):465-468.

[4] LUETTNER-STRETHMANN J, RAMPF F, PAUL W, *et al.* Transitions of tethered polymer chains; a simulation study with the bond fluctuation lattice model[J]. *Journal of Chemical Physics*, 2008, 128(6):064903/1-15.

[5] LI Hong, QIAN Chang-Ji, LUO Meng-bo. Conformational properties of a polymer tethered to an interacting flat surface[J]. *Polymer Journal*, 2010, 42:383-385.

[6] LI Hong, QIAN Chang-ji, SUN Li-zhen, *et al.* Simulation of a flexible polymer tethered to a flat adsorbing surface[J]. *Journal of Applied Polymer Science*, 2012, 124(1):282-287.

[7] LEE J H, KIM S Y, LEE J. Parallel algorithm for calculation of the exact partition function of a lattice polymer [J]. *Computer Physics Communications*, 2011, 182(4):1027-1033.

[8] GROSS J, JANKE W, BACHMANN M. Massively parallelized replica-exchange simulations of polymers on GPUs [J]. *Computer Physics Communications*, 2011, 182(8):1638-1644.

[9] LI Hong, QIAN Chang-ji, WANG Chao, *et al.* Critical adsorption of a flexible polymer confined between two parallel interacting surfaces[J]. *Physical Review E*, 2013, 87(1):012602/1-7.

[10] GAO He-bei, LI Hong, QIAN Chang-ji. Parallel Monte Carlo simulation of single polymer chain[J]. *Applied Mechanics and Materials*, 2013, 263-266:3317-3320.

[11] GAO He-bei. Computer simulation of a flexible polymer chain in a slit [M]. *High Performance Networking, Computing, and Communication Systems, Communications in Computer and Information Science*. [S. l.]:Springer, 2011:417-424.

(上接第 1038 页)

[10] TITERTON D H. 捷联惯性导航技术[M]. 王秀萍,等译. 2 版. 北京:国防工业出版社, 2007.

[11] 秦勇,臧希. 基于 MEMS 惯性传感器的机器人姿态检测系统的研究[J]. *传感技术学报*, 2007, 20(2):298-301.

[12] 秦永元,张洪钺,汪叔华. 卡尔曼滤波与组合导航原理[M]. 西安:西北工业大学出版社, 1998.

[13] 黄卫权,董冀. 基于 MTi 微惯性航姿系统的卡尔曼滤波器设计[J]. *自动化技术与应用*, 2009(1):54-57.

[14] KARIMI M, BOZORG M, KHAYATIAN A R. A comparison of DVL/INS fusion by UKF and EKF to localize an autonomous underwater vehicle [C]//Proc of the 1st RSI/ISM International Conference on Robotics and Mechatronics. 2013:62-67.

[15] GROSS J N, YU Gu, RHUDY M B, *et al.* Flight-test evaluation of sensor fusion algorithms for attitude estimation [J]. *IEEE Trans on Aerospace and Electronic Systems*, 2012, 48(3):2128-2139.

[16] 傅军,张晓峰,朱涛. 基于扩展卡尔曼滤波的加速度计特性估计方法[J]. *武汉理工大学学报:交通科学与工程版*, 2008, 32(3):255-258.

[17] 赵龙,吴康. 新型自适应 Kalman 滤波算法及其在应用[J]. *压电与声光*, 2009, 31(6):908-911.

[18] 乔相伟,周卫东,吉宇人. 用四元数状态切换无迹卡尔曼滤波器估计的飞行器姿态[J]. *控制理论与应用*, 2012, 29(1):97-103.

尔滨工程大学, 2009.