

# 改进的储层直接取样随机模拟方法及 GPU 实现

谢青<sup>1,2</sup>, 彭威<sup>1,2</sup>, 柳瑶阁<sup>1,2</sup>, 黄涛<sup>1,2</sup>, 卢德唐<sup>1,2</sup>

(1. 中国科学技术大学近代力学系, 安徽合肥 230027; 2. 中国科学技术大学石油与天然气研究中心, 安徽合肥 230027)

**摘要:** 讨论了储层随机模拟方法中的直接取样方法, 并对其中地质模式分量的选取方式进行了改进, 提出了结合空间相关关系模型结构化特性的方法. 针对模式子空间中的求解问题, 提出了基于统一计算设备构架 (compute unified device architecture, CUDA) 的并行策略. 实验结果表明, 模式分量选取方式的改进有效改善了两相河流沉积系统中河道的连续性, 并且模式子空间中的求解的并行方法具有较小的时间复杂度. 根据选取参数组的不同, 并行方法的计算速度比串行方法最低提速 10 倍, 最高提高了近 100 倍.

**关键词:** 随机模拟; 直接采样; 并行计算; 统一计算设备构架 (CUDA)

**中图分类号:** P628+.2; TE122.2+4 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2013.08.004

**引用格式:** Xie Qing, Peng Wei, Liu Yaoge, et al. Improvements for direct sampling stochastic simulation and GPU implementation[J]. Journal of University of Science and Technology of China, 2013, 43(8): 626-630.

谢青, 彭威, 柳瑶阁, 等. 改进的储层直接取样随机模拟方法及 GPU 实现[J]. 中国科学技术大学学报, 2013, 43(8): 626-630.

## Improvements for direct sampling stochastic simulation and GPU implementation

XIE Qing<sup>1,2</sup>, PENG Wei<sup>1,2</sup>, LIU Yaoge<sup>1,2</sup>, HUANG Tao<sup>1,2</sup>, LU Detang<sup>1,2</sup>

(1. Department of Modern Mechanics, University of Science and Technology of China, Hefei 230027, China;  
2. Research Center of Oil and Natural Gas, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** The direct sampling stochastic simulation method for reservoir modeling was discussed, the selection of geological pattern component was improved, and a method was presented which combined the structural characteristic information of the spatial relationship with the pattern component. CUDA (compute unified device architecture)-based parallel strategies were also proposed for obtaining an optimal solution within the pattern subspaces. Experimental results show that the proposition of the pattern component selection greatly improves the large-scale continuity of the sand channels in the two-facies sedimentary system. Further, the parallel computing method for solving the pattern subspace has small time complexity. The parallel computational efficiency on GPU shows a 10X to 100X improvement compared with the serial implementation with different computing parameters.

**Key words:** stochastic simulation; direct sampling; parallel computing; compute unified device architecture (CUDA)

收稿日期: 2013-03-29; 修回日期: 2013-04-17

基金项目: 国家自然科学基金重点基金 (10932010), 国家重大科技专项 (2011zx05009-006), 国家重点基础研究发展 (973) 计划 (2011CB707305) 资助.

作者简介: 谢青, 男, 1986 年生, 博士. 研究方向: 储层随机建模方法. E-mail: thirty@mail.ustc.edu.cn

通讯作者: 卢德唐, 教授. E-mail: dtlu@ustc.edu.cn

## 0 引言

研究地层中非均质储层的一个主要手段是对地下储层特征进行三维分布数字化建模. 传统上主要依赖于区域化变量和随机函数理论, 对储层中空间变量之间的相关关系进行计算, 得出描述两个随机变量间关系的变差模型, 在此基础上进行蒙特卡洛模拟来分析地下沉积相分布结构和储层物性的分布情况. 随着勘探开发技术的发展和各研究学科之间的进一步综合, 多分布多变量多维模拟逐渐成为研究地下储层油气分布的主要手段.

目前对于地下油气储层随机模拟, 主要有模拟退火法<sup>[1-2]</sup>、Gibbs 取样法<sup>[3-4]</sup>、单范式方程法<sup>[5]</sup>、空间分类过滤法<sup>[6]</sup>、直接取样法<sup>[7]</sup>等. 其中直接取样法对解空间的模式特征进行筛选, 采用分类带权函数计算模式差异进行模拟结果求解. 由于直接取样法不需要进行迭代, 对于约束条件没有数量和种类的限制, 因此在地下油气储层和含水层的建模应用中, 使用非常广泛<sup>[8-10]</sup>.

但是, 随着应用的深入, 直接取样方法也呈现出若干不足之处. 一方面, 储层中相关地质特征及地质变量具有复杂性, 尤其在空间上表现出各向异性的特征. 如典型河流沉积相组、断层、渗透率等在平面和垂向呈现分布和数量上的较大差异. 直接取样法求解时对于地质模式分量的选取没有考虑这些因素, 因而在重构大规模地质特征时存在缺陷. 另一方面, 储层中相关地质变量具有多样性, 很大程度上增加了分类带权函数的计算量. 目前, 有必要对现有直接取样方法进行相关改进, 以扩展该方法的实用性和应用领域.

近年来, 随机模拟领域的并行方法已经成为了研究的热点<sup>[11-12]</sup>. 本文采用了空间相关模型的相关参数, 对直接取样法的解空间进行了约束, 并给出了改进的分类带权函数. 对于模拟方法中最耗时的解空间搜索步骤, 我们基于主从策略, 利用 GPU 众核高度数据并行的特点, 加速了模式空间中最优解的求解过程. 采用河流两沉积相的地质模式空间, 进行了串行和改进并行方法的模拟. 实验结果表明, 改进的解空间约束条件有效改善了模拟结果中河流相的连续性特征, 且并行方法取得了 10~100 倍的加速比.

## 1 理论和方法

### 1.1 序贯模拟方法

直接取样法采用序贯模拟方法<sup>[13]</sup>, 该模拟方法

沿随机路径求出各模拟点的条件概率累积函数, 对于  $N$  个随机变量  $Z(x)$  的条件联合概率模型可以表示为

$$F_N[z_1, z_2, \dots, z_N | (n)] = \text{Prob}\{Z_i \leq z_i, i = 1, 2, \dots, N | (n)\} \quad (1)$$

可以通过  $N$  个步骤来得到式(1)中的  $N$  元联合概率模型, 假设各个随机变量独立, 可知:

$$F_N[z_1, z_2, \dots, z_N | (n)] = \text{Prob}\{Z_1 \leq z_1 | (n)\} \cdot \text{Prob}\{Z_2 \leq z_2 | (n+1)\} \cdot \dots \cdot \text{Prob}\{Z_N \leq z_N | (n+N-1)\} \quad (2)$$

通过下列  $N$  个随机变量的条件累积分布函数可唯一确定联合分布

$$\left. \begin{aligned} Z_1 &\sim \text{Prob}\{Z_1 \leq z_1 | (n)\} \\ Z_2 &\sim \text{Prob}\{Z_2 \leq z_2 | (n+1)\} \\ \dots \\ Z_N &\sim \text{Prob}\{Z_N \leq z_N | (n+N-1)\} \end{aligned} \right\} \quad (3)$$

### 1.2 直接取样方法

假设地质模式空间  $S_{\text{pat}}$ , 空间大小  $|S_{\text{pat}}| = N$ . 随机选取子空间  $S_{\text{pat}}^{\text{sub}}$ ,  $|S_{\text{pat}}^{\text{sub}}| = M = f * N, 0 < f < 1, f$  为子空间的选取比例因子.  $S_{\text{pat}}^{\text{sub}}$  中包含一系列地质模式  $\text{pat}_i, 1 \leq i \leq M$ , 直接取样方法从  $S_{\text{pat}}^{\text{sub}}$  中选取一个模式  $\text{pat}_j$ , 和模型中的待模拟地质变量  $z$  的地质模式  $\text{pat}_z$  进行比较,  $z$  的解必须满足下列条件之一:

- ①  $\Phi(\text{pat}_i, \text{pat}_z) \leq t, \text{pat}_i \in S_{\text{pat}}^{\text{sub}}$  且  $j = \min\{i | \Phi(\text{pat}_i, \text{pat}_z), \text{pat}_i \in S_{\text{pat}}^{\text{sub}}\}$
- ②  $\Phi(\text{pat}_i, \text{pat}_z) > t, \text{pat}_i \in S_{\text{pat}}^{\text{sub}}$  且  $\Phi(\text{pat}_j, \text{pat}_z) = \min\{\Phi(\text{pat}_i, \text{pat}_z), \text{pat}_i \in S_{\text{pat}}^{\text{sub}}\}$

式中,  $t$  为模式筛选门限值.  $\Phi(\text{pat}_j, \text{pat}_z)$  为模式筛选的分类带权函数, 通常采用如下形式:

$$\Phi(\text{pat}_j, \text{pat}_z) = \sum_{n=1}^N \omega_n | \text{comp}_n^{\text{pat}_j} - \text{comp}_n^{\text{pat}_z} | \quad (4)$$

式中,  $n$  为地质模式中地质分量的个数;  $\text{comp}_n^{\text{pat}_j}, \text{comp}_n^{\text{pat}_z}$  分别为地质模式  $\text{pat}_j, \text{pat}_z$  中的分量;  $\omega_k$  为各地质变量的权值.

串行算法需要在模式子空间  $S_{\text{pat}}^{\text{sub}}$  中顺序对目标模式进行求解. 模拟的计算量主要取决于模式子空间的选取比例因子  $f$ , 地质分量的个数  $n$  和模式筛选门限值  $t$ . 当进行大规模多变量模拟时, 模式子空间大小  $S_{\text{pat}}^{\text{sub}}$  和分量个数  $n$  较大且分量  $\text{comp}_n^{\text{pat}_j}, \text{comp}_n^{\text{pat}_z}$  形式复杂, 计算效率很低.

## 2 方法的改进及并行化

### 2.1 模式子空间分量的约束

在选取取样子空间  $S_{pat}^{sub}$  中模式  $pat_j$  的  $n$  个分量时,原始直接取样算法采用随机选取的方式.这种方法没有考虑到地质变量所反映的地质特征在空间上的变化规律.考虑到变差模型是表征区域化变量空间变化特征的重要工具,我们通过求取每个方向上的变差函数进而提取各向异性变差模型,从而反映出地质模式空间  $S_{pat}$  各向异性特征.然后利用变差模型提供的结构信息,来改进结合地质特性的子空间选取方式.

首先求取整体模式空间  $S_{pat}$  各向异性的变差模型,各向异性的变差模型的结构特征可用下列转换矩阵表示:

$$M_{trans} = \begin{bmatrix} 1/r_x & 0 & 0 \\ 0 & 1/r_y & 0 \\ 0 & 0 & 1/r_z \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \cdot \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

式中,  $r_x, r_y, r_z$  为 3 坐标主轴的缩放系数,  $\alpha, \beta, \theta$  为 3 坐标主轴的旋转角.对  $pat_j$  中所有原始分量通过下式进行转换:

$$h' = [h'_x, h'_y, h'_z]^T = M_{trans} \cdot [h_x, h_y, h_z]^T \quad (6)$$

最后通过转换后各分量距模式中心的距离  $|h'|$  来确定模式  $pat_j$  的  $n$  个分量.图 1 中显示了模式空间  $S_{pat}$  中地质模式的各向同性分量和经过转换后的各向异性分量

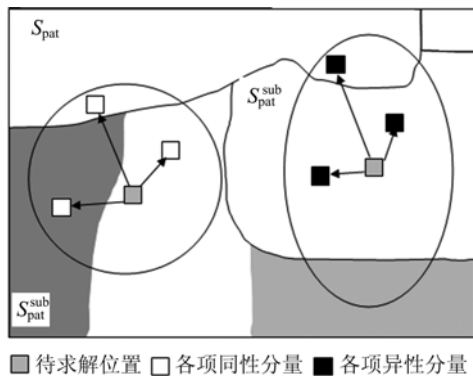


图 1 模式空间中地质模式的各向同性分量和经过转换后的各向异性分量

Fig. 1 The isotropic components and the transformed anisotropic components of the geological pattern in the pattern space

各向异性分量.

### 2.2 模式筛选的并行处理

由前述直接取样方法的分析可知,随机模拟的工作量主要是对各地质变量分量的转换及搜索,以及在模式子空间  $S_{pat}^{sub}$  中对模拟值的求解即目标模式的选取.由于算法的随机性及各个参数的取值的影响,导致了子空间  $S_{pat}^{sub}$  内最优解分布的随机.

目前,基于共享内存和消息传递模型的并行算法在科学计算领域得到了迅速发展<sup>[14]</sup>.一个通用的策略就是将求解子空间  $S_{pat}^{sub}$  分解为一系列独立的单元进行并行求解,并合并求解结果.但是,在传统的并行模式下,求解过程中数据传递、执行同步等额外的计算开销,会降低并行算法的效率,甚至使并行算法的效率低于串行算法.

针对上述问题,我们采用了主从式(master-slave)并行策略和统一计算设备构架(compute unified device architecture, CUDA)计算技术<sup>[15]</sup>,利用众核(many-core)GPU 高度数据并行化的特点,将求解子空间  $S_{pat}^{sub}$  内执行的一系列筛选的分类带权函数的计算分配到独立的 GPU 线程.利用改进的并行规约算法完成各个求解单元的求解及求解结果筛选的工作,使得串行算法的时间复杂度  $O(N)$  降低为  $O(\log(N))$ .我们编制了相应的计算软件,该软件包含前处理、后处理、数据显示,计算等模块,可以完成串行和并行方法的随机模拟和结果分析.并行算法仅仅在计算的前处理和后处理阶段进行 CPU 和 GPU 之间的数据交换,数据传输的时延得以最小化,使得通信的开销可以忽略不计,提高了计算的并行度.

## 3 实验结果和分析

本文所使用的计算服务器采用双核 Intel i3540 3.07 GHz, 4 GB 内存,计算显卡采用 2 GB 显存的 GeForce GTX 680 卡, CUDA 驱动版本号为 301.42.

使用典型两相河流沉积系统来进行随机离散模拟测试.对于串行和改进的并行方法,本文使用 3 组不同的数据进行对比.方法参数见表 1.

表 1 直接取样方法试验的参数组合

Tab. 1 The parameter combinations of stochastic simulation method

| 参数        | 一组   | 二组    | 三组    |
|-----------|------|-------|-------|
| 子空间选取比例因子 | 0.01 | 0.005 | 0.005 |
| 模式筛选门限值   | 0.02 | 0.1   | 0.02  |
| 地质分量个数    | 30   | 30    | 50    |

为验证空间分量选取方式的改进效果,我们把以上 3 组数据的串行和改进并行方法的模拟结果进行对比,如图 2 所示. 图中黑色为背景泥岩相,白色为河道砂岩相.

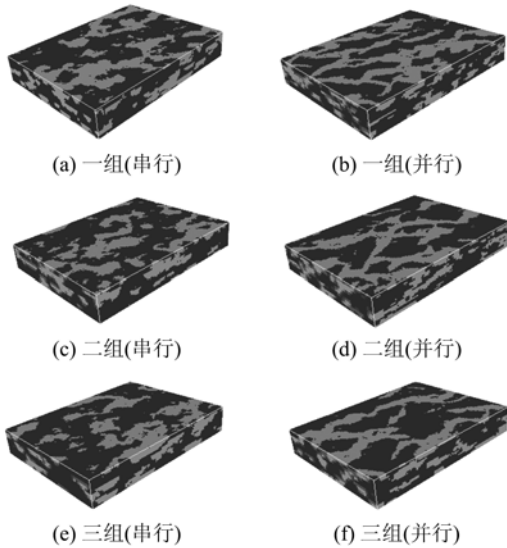


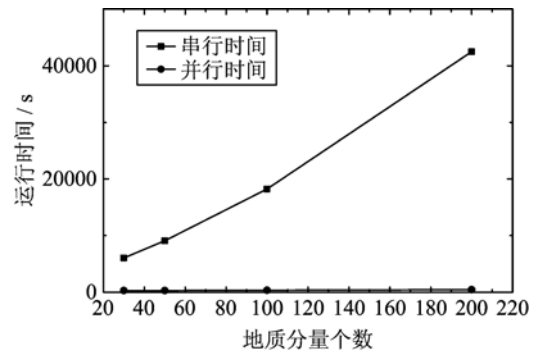
图 2 两相河流沉积系统串行计算和并行计算的模拟结果

Fig. 2 The simulation results of the two-facies fluvial system using the serial and parallel methods

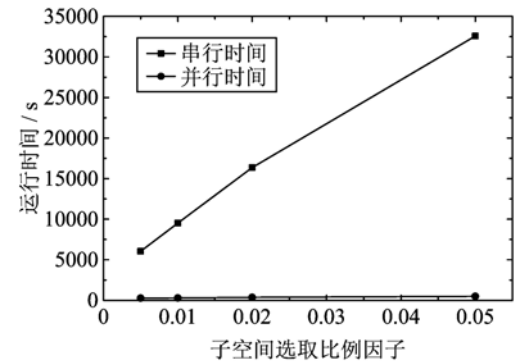
从图 2 的模拟结果中可以看出,对于每一组模拟参数,并行方法中因为在选取模式地质分量中考虑了河道分布的空间关系,使得模拟结果中河道大范围的连续性相比串行结果有了很大改善. 并且,改进并行方法的模拟结果基本反映了初始地质模式库中的结构性,再现了河道砂体的几何形态.

为了进一步测试并行方法执行速度的提高效果,我们采用更多的参数组合,来对比串行方法和并行方法的计算时间. 图 3 中给出了计算时间随 3 个主要参数变化的对比.

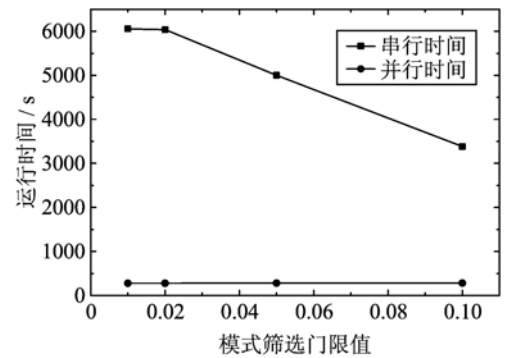
从图 3 的各主要参数对模拟时间的敏感性分析对比中可以看出,串行方法受输入参数影响较大. 较大的子空间选取比例因子、地质分量个数和较小的模式筛选门限值会影响求解空间的大小及模式筛选的分类带权函数的复杂度,从而使得串行方法的计算时间急剧增大. 图 3(c)中虽然串行方法的运行时间可以随着模式筛选门限的增大而明显减少,但在模式筛选门限值大于 0.1 时,河流砂体已经出现了明显的中断现象. 由于并行方法高效地分解了求解任务,使得各个 GPU 线程能够同步地完成相对轻量级的计算任务,从而使得各个参数的变化对于并



(a) 地质分量个数对模拟时间的影响



(b) 子空间选取比例因子对模拟时间的影响



(c) 模式筛选门限值对模拟时间的影响

图 3 随 3 个主要参数变化的计算时间

Fig. 3 The relation between the parameters of the stochastic method and the simulation time

行方法的模拟时间影响不大.

表 2 实验得出的各参数变化情况下的加速比

Tab. 2 The maximum and minimum speedups in the experiments

| 参数        | 最小加速比 | 最大加速比 |
|-----------|-------|-------|
| 地质分量个数    | 22    | 97    |
| 子空间选取比例因子 | 22    | 65    |
| 模式筛选门限值   | 12    | 22    |

从表 2 可以看出,在我们的模拟测试中,根据参数选取的不同,并行方法相对于串行方法的加速比

达到一个数量级(12 倍)以上,最高可达将近两个数量级(97 倍)。

## 4 结论

本文讨论了储层随机模拟方法中的直接取样随机模拟方法,针对其中地质模式分量的选取方式没有考虑地质区域化变量相关特征的不足,提出了结合空间相关关系模型结构化特性的方法.对于模式子空间中的最优解的求解问题,提出了基于 CUDA 的并行策略,并开发了相应的并行模拟软件.实验结果表明,考虑空间区域化变量相关关系的模式分量选取方式改善了两相河流沉积系统中河道的连续性.并且模式子空间中的求解的并行方法相比于串行方法,时间复杂度从  $O(N)$  降低为  $O(\log(N))$ ,运用模拟方法进行两相河流沉积系统三维建模时,运行速度提升了 10~100 倍,为重建和模拟地下结构和地层参数提供了更为快捷的计算工具.

### 参考文献(References)

- [1] Deutsch C V. Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data [D]. Stanford: Stanford University, 1992.
- [2] Deutsch C V, Wen X H. Integrating Large-Scale Soft Data by Simulated Annealing and Probability Constraints[J]. *Mathematical Geology*, 2000, 32(1): 49-67.
- [3] Srivastava R M. Iterative methods for spatial simulation [R]. Stanford: Stanford Center for Reservoir Forecasting, 1992: 24.
- [4] Srivastava R M. An overview of stochastic methods for reservoir characterization[C]// Yarus J M, Chambers R L. *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies, Volume ( I )*. Tulsa: American Association of Petroleum Geologists, 1994: 107-125.
- [5] Strebelle S, Journel A G. Sequential simulation drawing structures from training images[R]. Stanford: Stanford University, 2000.
- [6] Zhang T, Switzer P, Journel A. Filter-based classification of training image patterns for spatial simulation[J]. *Mathematical Geology*, 2006, 38(1): 63-80.
- [7] Mariethoz G, Renard P, Straubhaar J. The Direct Sampling method to perform multiple-point geostatistical simulations [J]. *Water Resour Res*, 2010, 46(11): W11536.
- [8] Haugen M, Mariethoz G, Mukerji T. Exploring Direct Sampling and Iterative Spatial Resampling in History Matching [C]// Proceedings of 23rd SCRF Annual Affiliates Meeting. Stanford: Stanford University, 2010.
- [9] Mariethoz G, Kelly B F J. Modeling complex geological structures with elementary training images and transform-invariant distances[J]. *Water Resources Research*, 2011, 47(7): W07527.
- [10] Mariethoz G, Renard P, Straubhaar J. Extrapolating the Fractal Characteristics of an Image Using Scale-Invariant Multiple-Point Statistics [J]. *Mathematical Geosciences*, 2011, 43(7): 783-797.
- [11] Huang T, Lu D T, Li X, et al. GPU-based SNESIM implementation for multiple-point statistical simulation [J]. *Computers & Geosciences*, 2013, 54: 75-87.
- [12] Tahmasebi P, Sahimi M, Mariethoz G, et al. Accelerating geostatistical simulations using graphics processing units ( GPU ) [ J ]. *Computers & Geosciences*, 2012, 46: 51-59.
- [13] Deutsch C V, Journel A G. *GSLIB: geostatistical software library and user's guide*[M]. Oxford: Oxford University Press, 1992.
- [14] 陈国良. 并行计算: 结构·算法·编程[M]. 北京: 高等教育出版社, 2003.
- [15] NVIDIA Corporation. *NVIDIA CUDA Programming Guide v4.0*[R]. NVIDIA Corporation, 2011.