BIM中基于蒙特卡罗模拟的成本估算的研究

赵倩纯，胡伟松，陈宇斌，吴百锋

（复旦大学计算机科学技术学院，上海 201203）

**摘要：**近年来，随着经济水平和计算机水平的迅速发展，建筑信息模型（Building Information System）受到业内的广泛关注，BIM技术也成为“十二五”计划重点推广技术。成本控制是建筑行业的永恒主题，在同等条件下，对成本控制不力的企业必将会被市场所淘汰。建筑成本估算数据量大，而且受到一些风险因素的影响，现有的估算方法无法充分考虑各风险事件的影响。蒙特卡罗模拟方法是研究这类应用的强有力的工具，充分考虑各成本风险因素的不确定性和复杂性。该模拟过程涉及到大规模的数据密集型计算，对硬件平台的计算能力要求甚高，使得传统的多处理器系统难以胜任。而且每次模拟都是相互独立的，彼此之间没有数据依赖关系。新型众核GPU架构在处理大规模的密集型数据并行计算方面有着巨大的优势，实验证明，随着模拟规模的增大和处理器核数的增加，GPU可获得持续的加速比。

**关键词：**BIM，成本估算，蒙特卡罗，数据并行，GPU，数据密集

**~~Research on Cost Estimation Based on Monte Carlo in BIM~~**

~~ZHAO Qian-chun, HU Wei-song, CHEN Yu-bin, WU Bai-feng~~

~~（School of Compute Science,~~~~Fudan University, Shanghai 201203,China）~~

**项 目：**鲁班软件大学合作计划项目

**作者简介：**赵倩纯（1989-），女，硕士研究生，主要研究方向：计算机体系结构；胡伟松，硕士研究生；陈宇斌，博士研究生；吴百锋，教授。

**E-mail：**12210240003@fudan.edu.cn

**~~【Abstract】 With the rapid development of economy and computers, Building Information System(BIM) has been focused by industry, and BIM technology had been the key promotion technology of the 12~~~~th~~ ~~Five Year Plan. Cost control is the eternal theme of architecture industry. Enterprises which are weak for this under the same circumstance will be eliminated by the market. Data amount of architecture cost is very huge, and it is affected by some rick factors. The methods for estimation are lack of credibility because they just analyze a single value and are unable to take uncertainty of all cost elements into consideration. In this paper, we propose a new method that does the Monte Carlo simulation based on the Probability distribution of cost element, which makes the cost estimation run in a new data parallel way. This can better match the many-core processor architectures of GPU cluster for its beautiful parallelism. By the way we got the parallel performance and execution speed in our experiments.~~**

**~~【Key words】 BIM, cost estimation, Monte Carlo, data parallel, GPU cluster.~~**

**0** 引言

建筑信息模型BIM，是在建筑项目的规划设计、建造施工、运转维护的整个或某个阶段中，引用3D、4D或者5D信息技术，进行系统设计、协同施工、虚拟建造、工程量计算、造价管理、设施运行等的技术和管理手段，是以建筑工程项目的各项相关信息数据，如设计、施工、运营过程中的相关数据作为基础建立的模型[1] [2]。利用BIM技术，可以实现虚拟设计和智能设计，实现碰撞检测、能耗分析、成本预测等[2]。关于建筑领域成本预测的课题具有很大的意义，成本控制关乎低碳、环保、绿色建筑、自然生态、社会责任、福利等宏大叙事，是建筑行业永恒的主题。

和一般的产品生产过程不同，建筑项目的实现过程是在不确定性外部环境和条件下进行的，往往受到诸如自然、劳动生产率、施工管理水平、市场等众多不确定因素的影响。这些风险事件之间相互联系互相影响，使得成本预算过程非常复杂。另外，每一个施工阶段都牵涉大量种类繁多的材料、机械设备、工种、消耗等，要把每一种人工、材料、机械等消耗都统计清楚，而且牵扯到众多的部门和岗位，数据量十分巨大。

蒙特卡罗方法是研究不确定的复杂过程的强有力的工具。虽然不能确切的知道工程总的成本预算，但是我们可以统计总结出相关的风险事件，将影响项目工程成本构成的风险因素考虑进来，再通过应用随机数逐步模拟出一个可能的成本区间，进而估算出一个计划成本值。

从本质上讲，蒙特卡罗模拟是进行多次独立的重复试验的，是一项数据密集型计算；而且每次试验都是相互独立的，彼此之间没有数据依赖关系，具有天生的并行性，因此用蒙特卡罗法进行成本估算是一项大规模数据的并行过程。对于这种大规模密集数据的应用场景，为了满足计算结果的精度和准确度，将涉及到庞大的计算量，这对硬件平台的能力提出了很高的要求，使得传统的多处理器系统难以胜任。

拥有大规模并行计算能力的新型众核处理机的问世，为这类大规模数据并行的蒙特卡罗模拟提供了一种新的解决途径。在本文中，我们提出了面向众核计算平台的蒙特卡罗模拟估算解决方案，该方法基于大规模数据并行的思想，将蒙特卡罗模拟转化成了数据级并行的过程，随着模拟规模或处理器核数的增大，获得了可观的加速比，充分利用了众核处理器提供的并行计算能力，大大提高了蒙特卡罗模拟的计算速度。

**1 BIM中的工程造价模型**

通常成本模型可以分为两类：确定成本模型和风险成本模型。确定成本模型的前提是价值可以归因于所有的变量，假设这些变量都是可以已知或可以预测的，这种模型已经用于建设行业传统的成本估算中。确定性成本模型预测成本的公式是：

其中*C*表示元素成本，*Q*表示元素数量，*R*表示元素单价，所以*n*个元素的总成本可表示为：

在实际项目实施过程中，所发生的活动时间及结果具有一定的不确定性，这类不确定性构成成本估算的风险事件。在一定的时间段内，风险事件综合构成成本风险值，因此成本风险值如下：

公式中表示成本风险值，表示*X*的期望，表示在研究时间段内风险事件所带来的成本值。当一段时间内有多个风险事件发生时，上述公式表达为：

公式中可线性叠加是由各风险事件的无关性来保证的。

为了判断某一风险事件能否作用于以及作用于的程度，引入判别函数，其中*t*表示时刻和*d*表示作用程度，定义如下所示：

则可进一步表达为：

在公式中，相关的各风险事件是相互独立的，但是实际情况下，风险事件之间存在着或多或少的联系。所以，大部分情况下，必须对风险事件进行解耦合分析处理，下面讲述风险事件的解耦方法。

风险事件（Risk Event， RE）能否发生及其可能导致的后果是由风险因素（Risk Factor，RF）的性质和测度相互作用共同决定的。风险因素由多个性质互不相同的风险因子（Risk Gene, RG）组成，RF的性质是由相关因子集在特定时间和环境条件下的共同性质决定。风险因子可定义为：。其中因子性质（Risk Gene Character，RGC）表示因子在构成因素的过程中起到主导作用的那一方面的性质；因子测度（Risk Gene Measure, RGM）表征在因子构成因素的过程中某一性质起到的作用的大小。

在一段时间内，风险因子会随一个或者多个因子测度值的变化发生变化，我们引入自生长函数来描述这个性质：，并有下面的公成立：

因子的性质、对应测度、与其他测度之间的关系可下面两个公式进行表述：

综合上述两个公式可得：

在变化的外界环境的影响下，风险因子自身的生长也有一定变化。定义风险因子*RG*对应的环境参数为，对环境作用函数为：，并且有成立。对进行分解：

可得出

综合公式可得：

其中为对的矩阵可乘式转换。

构成因素的因子之间具有三种相互作用方式：削弱、互斥和增强，这种相互作用的规则是由因子测度构成因素评价值（Risk Factor Value, RFV）时选用的权值决定的。定义风险因素，其中RFC（Risk Factor Character）为风险因素性质，，*P*为*RFV*对*RG*的评价值，布尔量表示*FI*为*RF*对*RG*的筛选值。则有：

公式中， *RF*包含*n*个RG，RG的因子性质最多有*m*个，*RFV*由*RGM*加权构成。

风险事件包含事件性质（Risk Evevt Character, REC）、发生条件（Risk Event Qualification, REQ），风险事件的重要属性是事件发生时间t、事件性质REC以及事件对成本值的影响程度\。

对于性质为的风险事件来说，当时将预期发生，且其风险因素评价值为，假设其当前对应的风险因子的因子测度初值为，可推算不等式成立的时刻，该时刻表示风险事件发生的预期时间。根据不等式可求出*、t*，这里我们引入风险事件评估值及评估函数*，*则有以下表达公式成立：

综上所述，风险成本值可以通过上述公式得到完整的描述。

**2 蒙特卡罗模拟**

蒙特卡罗（Monte Carlo）方法是一种数值计算方法，在1947年由NMetropolis命名，并于1949年正式使用[148]。也称为随机模拟（Random Simulation）法，有时也被称为随机抽样（Random Sampling）法。其基本思想为：当所求问题的解是某个事件的概率或者某个随机变量的数学期望时，通过某种实验的方法，得出该事件发生的频率或者该随机变量，通过求若干个具体观察值的算术平均值而得到问题的解。因此，可以将蒙特卡罗方法一般地理解为用随机实验的方法计算积分[3]。也就是说，将索要计算的积分看作服从某种分布密度函数的随机变量的数学期望。通过多次重复的独立实验，从分布密度函数中抽取*n*个子样本， 将相应的 *n*个随机变量的值的算术平均值作为积分的估计值（近似值）：

当N趋向无穷大时。可得到如下公式：

利用大数定理，得到如下公式：

根据公式可知，期望值 *x* 的估计值能用 *g* 来表示。

蒙特卡罗模拟成本估算的一般计算过程可分为以下几个步骤，这些步骤可用图1进行表示：



图1蒙特卡罗模拟一般过程

3 面向众核体系结构的实现

3.1 众核GPU体系结构

众核GPU和通用多核CPU之间的性能具有很大差距，比如，2009年众核GPU与多核CPU在峰值浮点运算的吞吐量方面的比率竟然高达10:1，这是因为这两种处理器采用了完全不同的设计理念。NVIDIA Tesla K20m GPU就有2496个处理器核而Intel core i7只有4个处理器内核。如图2所示：

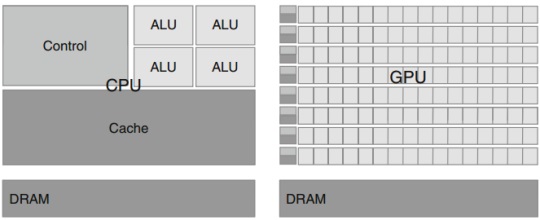


图2 CPU和GPU不同的设计理念

GPU普遍都采用多线程来提升运算速度与吞吐量，有着片上计算单元密集、存储带宽高效、性价比高等鲜明特点，因此更适合大规模的密集型数据并行计算。NVDIA推出CUDA（Compute Unified Device Architecture）这一可扩展的并行编程模型，可在应用程序中充分利用CPU和GPU各自的优点，通过将海量线程自动调度到相应的处理器核执行，实现高速运行。

3.2 CUDA程序执行过程

依据成本模型和蒙特卡罗方法的一般过程，我们的模拟过程可大致分为以下几个逻辑步骤：风险因素解耦合、生成随机数种子、随机数序列生成、并行模拟、归并等。本文选用CUDA编程模型来实现成本预算的蒙特卡罗模拟，执行过程如图3所示：



图3 MC模拟成本预算的CUDA执行过程

CUDA编程模型中的kernel在GPU上以SIMD的方式执行，即kernel中的每条指令被并行线程执行多次，但每次均处理不同的数据，达到了数据级并行。单个线程串行完成分配到该线程上的子任务：如风险因素解耦合极端的和的计算、随机数生成和模拟计算等。

4 实验及结果分析

4.1 实验环境

试验用到的服务器为DELL PowerEdge R720型号，并配有两个Intel Xeon E5-2603四核CPU，16G内存和两个NVIDIA Tesla K20m GPU，每个GPU含有2496个处理器核。服务器上运行64位CentOS 6.4操作系统， CUDA Toolkit版本为5.5，用于开发的GCC版本为4.4.7。

4.2 实验结果与分析

为了试验GPU的并行性能，我们进行了两次对比的模拟试验，实验一比较蒙特卡罗模拟过程在CPU和GPU（2496个处理器核）两个计算平台上的执行速度，实验二验证模拟加速比与处理器核数的关系。实验一中串行模拟和并行模拟的次数与所需时间的关系如表1所示。

|  |  |  |  |
| --- | --- | --- | --- |
| 模拟次数 | GPU（2496核）time (ms) | CPU time (ms) | 加速比 |
| 102400 | 3.22 | 3149 | 978 |
| 1024000 | 5.77 | 6456 | 1119 |
| 2560000 | 13.4 | 15860 | 1184 |
| 5120000 | 26.82 | 32743 | 1221 |
| 7680000 | 40.36 | 50934 | 1262 |
| 10240000 | 53.86 | 68776 | 1277 |

表1 模拟次数与加速比

从图4以看出，相对于CPU串行过程，GPU众核处理器对蒙特卡罗模拟具有更快的执行速度。而且，随着模拟规模的增大，加速比也有相应的增加。



图4 随着模拟次数增加的加速比

实验二以单块GPU（2496核）的执行时间为基准，计算出在其他处理器核数的执行情况下的加速比。由图5可以看出，随着处理器核数的增加，加速比以近似线性的方式增长。实验结果如图所示：



图5随着处理器核数增加的加速比

**5 结束语**

计算密集型应用一直是高性能计算(HPC)领域的一个重要话题。众核架构作为处理器的发展方向，将会逐渐成为计算密集型应用的主流应用平台。建筑领域中用蒙特卡罗模拟成本估算是一个典型的计算密集型问题，对其过程进行分解进而转化成完全的数据级并行处理，充分契合了众核处理器的体系结构，随着模拟规模和处理器核数的增大，都能获得一定的加速比。拥有大规模并行计算能力的新型众核处理机，为BIM中成本估算提供了一个良好的平台，为计算密集型应用注入了新的活力。

## 参考文献

1. ~~张建平. 建筑业应尽快推行建筑信息模型(BIM)技术.建筑技术[J].2011, 42(1): 9-13.~~
2. ~~孔嵩. 建筑信息模型BIM研究.建筑电气[J]. 2013, 4.~~
3. ~~张建平. BIM技术的研究与应用[J]. 施工技术· 资讯, 2011.1, 15-18.~~
4. ~~宫野. 计算多重积分的蒙特卡罗方法与数论网格法[J]. 大连理工大学学报, 2001, 41(1): 20-23.~~
5. ~~孙咏梅. 建筑工程造价. 北京大学出版社.~~
6. ~~Metropolis N, Ulam S. The Monte Carlo method [J]. Journal of the American statistical association, 1949, 44(247): 335-341.~~
7. ~~肖刚, 李天柁. 系统可靠性分析中的蒙特卡罗方法[M]. 科学出版社, 2003;~~