

庞超, 周祖昊, 刘佳嘉, 等. 基于 GPU 加速的分布式水文模型并行计算性能[J]. 南水北调与水利科技(中英文), 2024, 22(1): 33-38.  
PANG C, ZHOU Z H, LIU J J, et al. Parallel computing performance of distributed hydrological model accelerated by GPU[J]. South-to-North Water Transfers and Water Science & Technology, 2024, 22(1): 33-38. (in Chinese)

# 基于 GPU 加速的分布式水文模型并行计算性能

庞超<sup>1,2</sup>, 周祖昊<sup>2</sup>, 刘佳嘉<sup>2</sup>, 石天宇<sup>2</sup>, 杜崇<sup>1</sup>, 王坤<sup>2</sup>, 于新哲<sup>2</sup>

(1. 黑龙江大学水利电力学院, 哈尔滨 150080; 2. 中国水利水电科学研究院流域水循环模拟与调控国家重点实验室, 北京 100038)

**摘要:** 针对具有物理机制的分布式水文模型对大流域、长序列模拟计算时间长、模拟速度慢的问题, 引入基于 GPU 的并行计算技术, 实现分布式水文模型 WEP-L(water and energy transfer processes in large river basins)产流过程的并行化。选择鄱阳湖流域为实验区, 采用计算能力为 8.6 的 NVIDIA RTX A4000 对算法性能进行测试。研究表明: 提出的基于 GPU 的分布式水文模型并行算法具有良好的加速效果, 当线程总数越接近划分的子流域个数(计算任务量)时, 并行性能越好, 在实验流域 WEP-L 模型子流域单元为 8 712 个时, 加速比最大达到 2.5 左右; 随着计算任务量的增加, 加速比逐渐增大, 当实验流域 WEP-L 模型子流域单元增加到 24 897 个时, 加速比能达到 3.5, 表明 GPU 并行算法在大尺度流域分布式水文模型计算中具有良好的发展潜力。

**关键词:** 基于 GPU 的并行算法; 物理机制; 分布式水文模型; WEP-L 模型; 计算性能

**中图分类号:** P333.9; TV121.7 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdqk.2024.0005

近年来, 分布式水文模型向大流域以及更加精细化的方向发展, 其发展要求大量数据信息的处理和要素多层次的物理计算过程, 因而也面临着模拟时间长、计算效率低的问题, 而并行计算正是解决耗时问题的有效技术之一<sup>[1]</sup>。

目前应用较为广泛的并行技术有基于 OPENMP 架构的 CPU 并行技术、基于消息传递的 MPI 并行编程模型以及基于 CUDA(compute unified device architecture)架构的 GPU 并行技术。利用 OPENMP 架构编程的相关研究, 如: 卢浩等<sup>[2]</sup>对水文分析若干过程进行了并行化算法的研究和实践, 双线程最大加速比达到 1.62; 秦泽宁等<sup>[3]</sup>在区域分解方法的基础上, 实现基于贪心算法的优化调度, 加速比最大达到 3.04; 刘军志等<sup>[4]</sup>基于栅格分层的逐栅格汇流算法并行得到较好的加速比。利用 MPI 编程的相关研究, 如: 向东等<sup>[1]</sup>基于 MPI 实现分布式水文模型(water and energy transfer processes in large river basins)WEP-L 并行, 加速比最大达到 4.8; Cui 等<sup>[5]</sup>以 Illinois 河流域为研究区, 实现了基于子流域的分

布式水文模型 r.water.fea 的并行化, 11 个进程计算的加速比为 4.5; 李兴等<sup>[6]</sup>基于 MPI 实现对超大时空尺度的地表水流数值模拟的并行实现, 具有较好的加速性能, 提升工程应用能力。目前由于 GPU 具有结构上的优势, 并拥有数量庞大的计算单元及高带宽<sup>[7]</sup>, 对于结构复杂的密集计算部分有着很高的计算效率, 因此 GPU 并行技术也被应用于分布式水文模型的并行计算。曾志强<sup>[8]</sup>实现了 Grid-Xin'anjiang 概念性水文模型产流计算的 GPU 并行化, 经 GPU 并行的产流和汇流计算效率得到显著提升, 拥有线程数量优势的 GPU 并行加速比明显高于 OpenMP 并行; 赵向辉等<sup>[9]</sup>基于 GPU 的并行算法实现了汇流并行分析, 实验结果表明, 提出的算法能提高系统的计算效率; 刘永和等<sup>[10]</sup>基于分布式新安江模型实现了产流计算的并行化, 同时提出一种与并行产流模块相结合的非并行的递归汇流执行方法; 宏瑾靓<sup>[11]</sup>基于 GPU 的技术实现了分布式新安江模型的并行计算, 研究结果表明模型计算效率明显高于基于 CPU 的并行技术。

收稿日期: 2023-08-16 修回日期: 2023-12-06 网络出版时间: 2024-01-03

网络出版地址: <https://link.cnki.net/urlid/13.1430.tv.20231229.1605.002>

基金项目: 国家重点研发计划项目(2021YFC3000205); 江西省水利科技重大项目(202023ZDKT03); 水利部重大科技项目(SKR-2022056)

作者简介: 庞超(1998—), 男, 河南南阳人, 主要从事分布式水文模型并行算法研究。E-mail: 1799336883@qq.com

通信作者: 周祖昊(1975—), 男, 湖北武汉人, 正高级工程师, 博士, 主要从事流域水循环及伴生过程模拟与调控研究。E-mail: zhzh@iwhr.com



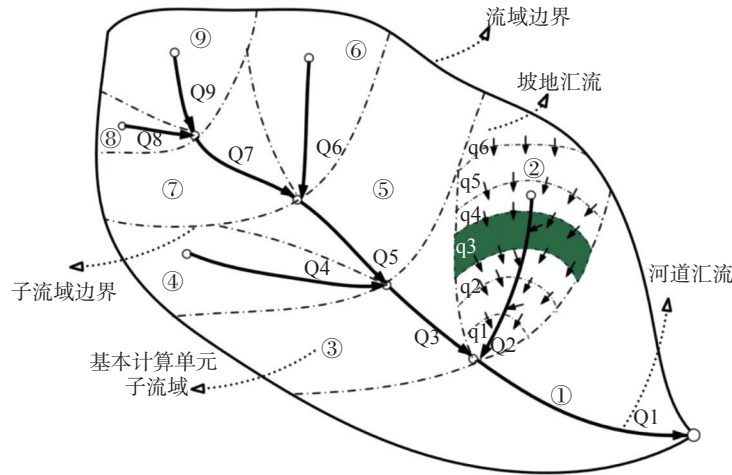


图 1 WEP-L 模型平面结构  
Fig. 1 Plane structure of WEP-L model

本文所使用的分布式水文模型 WEP-L 已经完成模块化的优化,模型主要模块包括输入输出参数文件读取模块、产流模拟模块、汇流模拟模块和参数初始化模块等,各模块通过全局变量相互关联,满足并行化的基础需求。

WEP-L 模型产流模块计算包括两大循环:时间循环,按年、月、日的顺序进行,并逐日进行计算统计;空间循环,在子流域单元上的循环。对于同一个模拟单元来说,连续时间序列的模拟具有前后依赖关系,必须进行串行计算。从空间角度看,WEP-L 模型将流域划分为多个子流域单元,每个子流域的产流计算都是相互独立的,这为空间上的并行计算提供了可能,在考虑模拟单元之间计算依赖关系的基础上,可以将不同模拟单元的计算任务分配到多个计算单元上进行的并行计算,而汇流模块计算中坡面汇流各子流域彼此之间相互独立。但河网汇流过程中依赖上游子流域的输入,不满足 GPU 并行的相互独立性要求。因此对该模型的产流过程基于子流域进行并行化优化,并通过调用核函数,将子流域计算任务分配在 GPU 上进行计算。

### 2.3.3 WEP-L 模型产流过程并行性实现

分布式水文模型 WEP-L 产流过程程序并行化实现:利用 GPU 生成多条并行线程,每个线程控制一个子流域的计算,通过线程索引分配后逐子流域遍历循环模拟计算。当计算全部完成后,统计当前的子流域参数信息,并赋值状态数组,通过 PCIE 线从 GPU 赋值到 CPU 以备后续汇流计算。在 GPU 并行化实现中有频繁在 CPU 和 GPU 之间传递的中间变量,可通过 `cudaMemcpy` 函数实现双向拷贝;对

于长期贮存在 GPU 中的变量、数组等则在计算开始时统一定义到 GPU 中,程序运行结束后,统一清理。

## 3 鄱阳湖流域 WEP-L 模型单元划分

基于汇流累积方法提取模拟河网,并将模拟河网的各河段进行唯一编码,其中将所有流入同一河段的流域栅格作为同一子流域,以此将鄱阳湖流域划分为若干子流域<sup>[21-22]</sup>,基于 Pfafstetter 规则<sup>[23]</sup>确定各子流域的上下游关系。图 2 为划分的子流域单元。

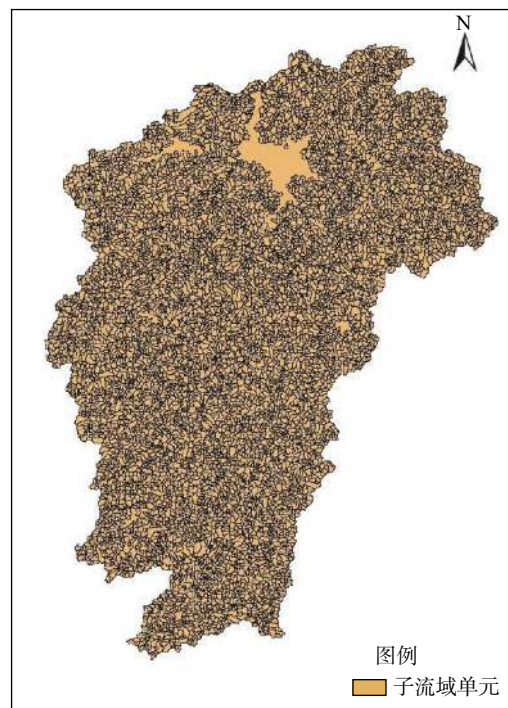


图 2 鄱阳湖流域子流域单元划分  
Fig. 2 Sub-basin units of Poyang Lake basin

## 4 结果与讨论

### 4.1 并行计算性能分析方法

针对不同的核函数配置、不同的计算任务量(子流域划分数量),测试基于 GPU 加速的分布式水文模型 WEP-L 产流过程并行计算算法的加速性能。并行计算的加速比计算公式为

$$S_{\text{speedup}} = \frac{t_1}{t_2} \quad (1)$$

式中:  $t_1$  表示串行计算时间,  $s$ ;  $t_2$  表示并行计算时间,  $s$ ;  $S_{\text{speedup}}$  为并行计算的加速比。

### 4.2 不同核函数参数配置下加速性能分析

针对划分的子流域个数为 8 712 个,通过核函数设置调用 GPU 的线程块为 8,线程数分别为 4、8、16、32、64、128、256、512、1 024,对模型加速性能进行分析,图 3 为串行、并行计算时间对比图。通常,提高并行性(增加线程数),计算时间减少。分析发现,随着线程数增加,模型计算时间逐渐减少,加速比逐渐增大。当线程总数越接近划分的子流域个数时,产流并行模拟计算的时间越短,最短时间为 1 290 s,最大加速比为 2.54,符合基于 GPU 并行计算中随着并行度的增大,并行时间逐渐减少,加速比增大的规律。

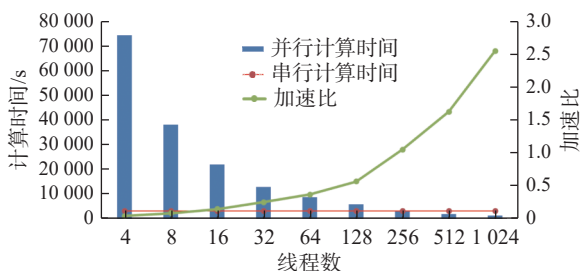


图 3 GPU 不同线程数对应并行计算时间对比

Fig. 3 Comparison chart of parallel computing time for different numbers of GPU threads

通过图 3 还发现,当模型并行度较低时,并行程序的计算时间比串行计算时间(3 280 s)还慢。这是由于 GPU 的核心频率较低,单个线程的计算效率远低于 CPU,所以 GPU 在低并行度下的计算效率较低。GPU 在高并行度下能够充分利用其计算核心多的优势,克服核心频率较低的限制,并行计算性能显著提升。

### 4.3 不同计算任务量下加速性能分析

GPU 并行计算的性能不仅仅受并行度的影响,还受计算任务量的影响。将鄱阳湖流域分别划分为 8 712、16 083、19 459、24 897 个子流域,并统一

设置 GPU 并行计算的核函数为<<<8,1024>>>来分析不同的任务量对并行加速效果的影响。表 2 为设置不同计算单元数下的加速比计算结果,图 4 为不同计算单元数与计算效率的关系。从图 4 可以直观地看出,随着子流域单元个数的增加,串行以及并行计算的时间都逐渐增加,但并行计算时间增加的幅度小于串行计算时间增加的幅度,所以加速比增加。但是随着并行度的增加,模型的加速比增加的幅度变小,最后达到 3.5 左右。

表 2 不同任务量下的加速比  
Tab. 2 Speedup ratio for different workloads

核函数配置	子流域单元数/个	串行时间/s	并行时间/s	加速比
8×1 024	8 712	3 280	1 290	2.54
	16 083	8 057	2 652	3.04
	19 459	9 478	2 722	3.48
	24 897	12 472	3 559	3.50

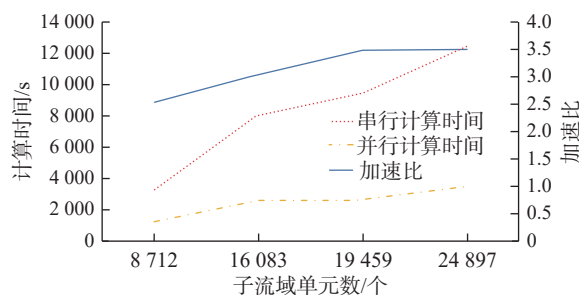


图 4 不同任务量与计算效率的关系

Fig. 4 Relationship between different workloads and computational efficiency

## 5 结论

研究针对具有物理机制的分布式水文模型 WEP-L 产流过程计算时间长的问题,将基于 GPU 的并行技术应用到 WEP-L 分布式水文模型产流过程中。以鄱阳湖流域为例,对模型并行性能进行分析,得出以下结论:

基于 GPU 的并行技术对具有物理机制的分布式水文模型具有较好的加速比,能够显著减少计算时间,当线程总数越接近划分的子流域个数时,产流并行模拟计算的时间会越短,当鄱阳湖流域 WEP-L 模型划分的子流域个数为 8 712 个时,最短计算时间为 1 290 s,最大加速比为 2.54。

随着计算任务量的增加,串行和并行的计算时间都增加,由于并行计算时间增加的幅度远小于串行计算方法,所以加速比逐渐增大,但是增速减小。

当实验流域 WEP-L 模型子流域单元增加到 24 897 个时,加速比能达到 3.5,表明 GPU 并行算法在具有物理机制的大尺度流域分布式水文模型计算中具有良好的发展潜力。

在下一阶段工作中,可以从以下 4 个方面深入开展研究:优化任务分配,通过任务负载均衡减少等待时间;优化模型参数传递以及模型框架,减少同步等待时间和传输数据次数;在产流并行算法的基础上,耦合开发汇流并行算法模块;将基于 CUDA 架构的 GPU 并行技术和基于 Openmp 架构的 GPU 并行技术结合起来,研发 CPU-GPU 异构并行算法。

#### 参考文献:

- [1] 向东,周祖昊,袁胜,等.基于MPI的分布式水循环模型并行计算性能研究[J].水文,2020,40(5):36-40,27. DOI: 10.19797/j.cnki.1000-0852.20190201.
- [2] 卢浩,王少华,李绍俊,等.基于OpenMP的并行化水文分析算法研究与实现[J].测绘与空间地理信息,2013,36(S1):7-10. DOI: 10.3969/j.issn.1672-5867.2013.z1.002.
- [3] 秦泽宁,黎曙,周祖昊,等.分布式水文模型区域分解并行计算方法及其应用[J].水电能源科学,2020,38(10):1-4,12.
- [4] 刘军志,朱阿兴,秦承志,等.分布式水文模型的并行计算研究进展[J].地理科学进展,2013(4):538-547. DOI: 10.11820/dlkxjz.2013.04.006.
- [5] CUI Z T, VIEUX B E, NEEMAN H, et al. Parallelisation of a distributed hydrologic model[J]. Computer Applications in Technology, 2005, 22(1): 42-52. DOI: 10.1504/IJCAT.2005.006802.
- [6] 李兴,魏清福,张景新.自由表面水流数值模拟程序并行化的预处理[J].水动力学研究与进展 A 辑,2023,38(2):222-231. DOI: 10.16076/j.cnki.cjhd.2023.02.008.
- [7] 曾志强.流域水文模拟的GPU并行化方法研究[D].武汉:华中科技大学,2021. DOI: 10.27157/d.cnki.ghzku.2021.006317.
- [8] 覃金帛,曾志强,梁藉,等.GPU并行优化技术在水利计算中的应用综述[J].计算机工程与应用,2018,54(3):23-29,63.
- [9] 赵向辉,苗青,付忠良,等.基于CUDA的汇流分析并行算法的研究与实现[J].计算机应用研究,2010,27(7):2445-2447,2451. DOI: 10.3969/j.issn.1001-3695.2010.07.011.
- [10] 刘永和,冯锦明,徐文鹏.分布式水文模型的GPU并行化及快速模拟技术[J].水文,2015,35(4):20-26. DOI: 10.3969/j.issn.1000-0852.2015.04.004.
- [11] 宏瑾靓.基于分布式水文模型的GPU并行化及快速模拟技术研究[J].黑龙江水利科技,2018,46(4):15-18,68. DOI: 10.14122/j.cnki.hskj.2018.04.005.
- [12] JIA Y, WANG H, ZHOU Z, et al. Development of the WEP-L distributed hydrological model and dynamic assessment of water resources in the Yellow River basin[J]. Journal of Hydrology, 2006, 331(3-4):606-629. DOI: 10.1016/j.jhydrol.2006.06.006.
- [13] ZHOU Z, JIA Y, QIU Y, et al. Simulation of dualistic hydrological processes affected by intensive human activities based on distributed hydrological model[J]. Journal of Water Resources Planning and Management, 2018, 144(12): 04018077. DOI: 10.1061/(ASCE)WR.1943-5452.0000990.
- [14] CAI J, ZHOU Z, LIU J, et al. A three-process-based distributed soil erosion model at catchment scale on the Loess Plateau of China[J]. Journal of Hydrology, 2019, 578: 124005. DOI: 10.1016/j.jhydrol.2019.124005.
- [15] 周祖昊,刘佳嘉,严子奇,等.黄河流域天然河川径流量演变归因分析[J].水科学进展,2022,33(1):27-37. DOI: 10.14042/j.cnki.32.1309.2022.01.003.
- [16] 刘昕,贾建伟,王栋.鄱阳湖地区河湖与含水层的水交换应用研究[C]//中国水利学会.中国水利学会2021学术年会论文集第五分册.黄河水利出版社,2021:215-221. DOI: 10.26914/c.cnkihy.2021.057226.
- [17] 赵贵章,董锐,王赫生,等.近30年鄱阳湖与洞庭湖水文变化与归因[J].南水北调与水利科技(中英文),2020,18(5):74-83. DOI: 10.13476/j.cnki.nsb-dqk.2020.0094.
- [18] 贾仰文,王浩,倪广恒.分布式流域水文模型原理与实践[M].北京:中国水利水电出版社,2005.
- [19] 张峰,寥卫红,雷晓辉,等.分布式水文模型子流域划分方法[J].南水北调与水利科技,2011,9(3):101-105. DOI: 10.3724/SP.J.1201.2011.03101.
- [20] 张哲.基于GPU异步流的遥感影像加速处理方法研究[D].开封:河南大学,2021. DOI: 10.27114/d.cnki.ghnau.2020.001890.
- [21] 刘佳嘉,周祖昊,贾仰文,等.基于DEM河网干支拓扑关系的子流域编码规则[J].河海大学学报(自然科学版),2013,41(4):288-293. DOI: 10.3876/j.issn.1000-1980.2013.04.002.

- [22] 刘佳嘉, 周祖昊, 贾仰文, 等. 分布式水文模型子流域编码方法对比分析[J]. *河海大学学报(自然科学版)*, 2017, 45(1): 22-29. DOI: [10.3876/j.issn.1000-1980.2017.01.004](https://doi.org/10.3876/j.issn.1000-1980.2017.01.004).
- [23] 秦泽宁, 周祖昊, 刘明堂, 等. 分布式水文模型时空离散化并行计算方法研究[J]. *人民黄河*, 2020, 42(8): 15-20. DOI: [10.3969/j.issn.1000-1379.2020.08.004](https://doi.org/10.3969/j.issn.1000-1379.2020.08.004).

## Parallel computing performance of distributed hydrological model accelerated by GPU

PANG Chao<sup>1,2</sup>, ZHOU Zuhao<sup>2</sup>, LIU Jiajia<sup>2</sup>, SHI Tianyu<sup>2</sup>, DU Chong<sup>1</sup>, WANG Kun<sup>2</sup>, YU Xinzhe<sup>2</sup>

( 1. School of Hydraulic & Electric-power, Heilongjiang University, Harbin 150080, China; 2. State Key Laboratory of Simulation and Regulation of River Basin Water Cycle, China Institute of Water Resources and Hydropower Research (IWHR), Beijing 100038, China )

**Abstract:** With the development of distributed hydrological models towards larger watersheds and finer granularity, computational efficiency gradually became a bottleneck, and parallel computing technology emerged as an effective solution to this challenge. In the realm of parallel computing for distributed hydrological models, most of the existing studies have primarily focused on CPU-based parallel techniques, with relatively limited research on GPU-based parallel methods. Furthermore, investigations on distributed hydrological models incorporating physical mechanisms remain scarce.

This study centered around the physically-based distributed hydrological model WEP-L ( water and energy transfer processes in large river basins ) and explored the utilization of GPU-based parallel computing techniques. From a spatial perspective, the WEP-L model divides the watershed into numerous sub-basin units, where each unit's runoff calculations are independent, offering spatial parallelism. The interdependencies between simulation units were taken into account while allocating jobs to several computer units for parallel execution. Consequently, the runoff process of the model was parallelized based on sub-basins, dividing the Poyang Lake basin into 8,712 sub-units, and employing GPU threads to execute parallel computations through kernel functions.

It is founded that the distributed hydrological model's suggested GPU-based parallel approach significantly accelerated the process. With an increase in GPU thread count, the parallel computing time steadily reduced. The parallel performance was most efficient when the total thread count closely approached the number of divided sub-basins. In the experimental Poyang Lake basin with 8,712 sub-basin units in the WEP-L model, the maximum speedup reached around 2.5. Secondly, the performance of GPU parallel computing was influenced not only by the degree of parallelism but also by the computational workload. With an increase in computational workload, both serial and parallel computation times increased. However, due to the smaller rate of increase in parallel computation time compared to the serial method, the speedup gradually increased, albeit at a diminishing rate. When the number of sub-basin units in the experimental WEP-L model increased to 24,897, the speedup ratio reached 3.5, indicating the considerable potential for GPU parallel algorithms in the computation of physically-based large-scale watershed distributed hydrological models.

In conclusion, GPU-based parallel algorithms showed great promise for computing large-scale, physically-based, watershed-distributed hydrological models. The results indicated that the enhancement of parallel efficiency was contingent not only on the number of parallel threads activated but also on the size of the computational workload. The parallel calculation time decreased gradually as the number of GPU threads rose. As the computing demand rose, the speedup ratio increased correspondingly. GPU-based parallel computing represents the current trend in parallel computing. This study could provide valuable experience for other researchers exploring GPU parallel algorithms, contributing to the facilitation of interdisciplinary collaboration between computer science and water resources engineering.

**Key words:** GPU-based parallel algorithm; physical mechanism; distributed hydrological model; WEP-L model; computational performance.