Petroleum Geoscience Big Data and GPU Parallel Computing

Fei Han; Sam Z. Sun
Lab for Integration of Geology and Geophysics (LIGG),
China University of Petroleum
Beijing,China
hanfei0066@126.com

***Abstract*: Petroleum geoscience big data is defined in this paper. CPU/GPU hybrid system is used to try to accelerate computing speed of petroleum geoscience big data using chaotic quantum particle swarm optimization (CQPSO) inversion method as an example, and the computing time of CQPSO is reduced significantly.**

Keywords: Petroleum geoscience big data; GPU parallel computing

Ⅰ. Introduction

The recent advances in petroleum and computer techniques give birth to the explosive growth of petroleum geoscience data. The huge volume of petroleum geoscience data has reached PB even ZB. However, it is very difficult to process the huge volume of petroleum geoscience big data on CPU in a short time[1,2,3]. This paper computes petroleum geoscience big data on CPU/GPU hybrid system to increase computational efficiency significantly.

Ⅱ. What is "Petroleum Geoscience Big Data"

The "Petroleum Geoscience Big Data" refers to the volume and velocity of data that outstrip the storage and computing capacity, as well as the variety and complexity.

A. A variety of petroleum geoscience big data

There are many different kinds of petroleum geoscience data, and the seismic data accounts the largest proportion, 85% of the petroleum data. The schematic of proportion is shown in Fig.1.

Fig.2 shows the multi-fold schematic of seismic data. S is a shot point and R is a receiver point which can receive seismic wave. Assuming that a single trace of seismic data contains 3501 sampling points, storage S_1 of a single trace can be obtained as follows:

$$S_1 = 3501 \times 4 \text{ Byte} = 14 \text{ KB}. \quad (1)$$

For a small work area of 225Km^2, with 15×15 surface element and 500 coverage, storage S_2 of original data is:

$$S_2 = 1000 \times 1000 \times 500 \times S_1 = 6 \text{ TB}. \quad (2)$$

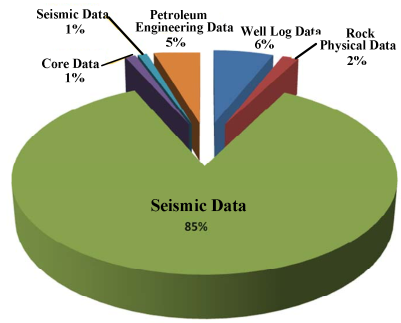
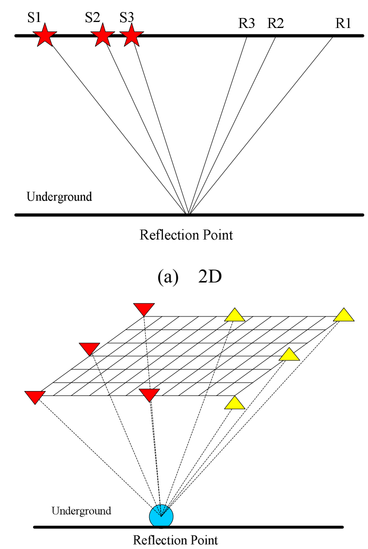


Fig.1. Proportions schematic of petroleum geoscience big data

B. The huge volume of petroleum geoscience big data



(a) 2D

(b) 3D

Fig.2. Multi-fold schematic of seismic data

At present, the seismic attributes proposed by an expert have reached more than 200. Assuming that there are 50 experts, storage S_3 of total data is:

$$S_3 = 200 \times 50 \times S_2 = 6 \text{ PB}. \quad (3)$$

If the area is bigger and folds increase, the collected data volume will increase many times.

Ⅲ. Petroleum geoscience big data computing on GPU

Sun and Liu proposed a new non-linear AVO

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.59

inversion method [4], with an employment of chaotic quantum particle swarm optimization (CQPSO) to solve non-linear problems. For the specific inversion process of CQPSO, refer to paper 4.

This paper accelerates this method on hybrid CPU/GPU supercomputing system.

A. Parallelism analysis on CQPSO algorithm

CQPSO is an algorithm based on population models, as same as other evolutionary algorithms. There are three parallelism characteristics in CQPSO algorithm.

- The calculation of particle individual position is parallel with the evaluation and update of particle best position p_{best}.
- The evaluation and update of group global best position g_{best} are in parallel. After the evaluation and update of p_{best} of a particle, it will immediately be compared with the current g_{best} and update g_{best}.
- The generation processes of next groups are in parallel. The next position of a particle depends on the current p_{best}, g_{best} and individual's own inertia.

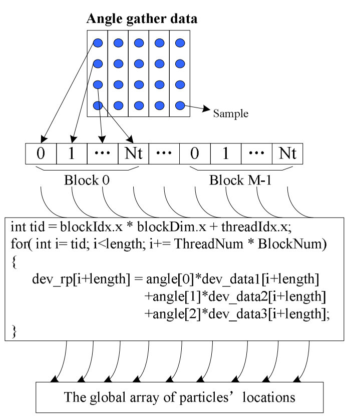


Fig.3. The parallel computing model of CQPSO on GPU

B. CQPSO algorithm optimization

In this paper, MPI+CUDA parallel technology is used to optimize the computational efficiency. MPI can easily achieve interprocess communication among multiple hosts. CUDA can achieve parallel computing by creating many CUDA threads on GPU.

C .Field data test

This paper applies CPU algorithm and GPU algorithm to actual seismic data in Tahe area in Tarim Basin in China to verify their correctness and computational efficiency. The parameters of two platforms are shown in Table 1.

TABLE 1. Parameters of CPU cluster and GPU cluster

Test platform	**CPU platform (Intel® Xeon® E5420)**	**GPU platform 1 (Tesla C2075)**	**GPUplatform 2 (Kapler 40)**
Memory	4GB	6 GB	12 GB
Number of cores	8	448	2880
Quantity	100	4	4
CUDA Capbility	--	2.0	3.5

GPU algorithm is executed on GPU cluster and a total of 60 processes are created in 10 nodes on a CPU cluster to test CPU algorithm. The running time of CPU and GPU are shown in Fig.4.

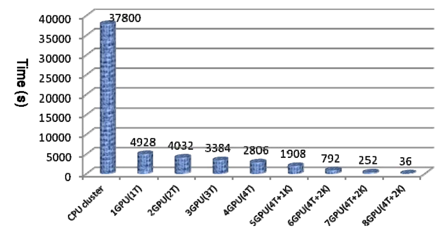


Fig.4. Comparison diagram of computing time (CPU vs. GPU)

From Fig.4, we know that the computing time of GPU algorithm is significantly shorter than that of CPU algorithm. The computing speed of eight GPUs is 137 times faster than that of a GPU cluster with 10 nodes and 1050 times faster than that of a single-core CPU. Then we use eight GPUs to compute whole work area with 48GB at 3.1h.

Therefore, the computing time can be reduced greatly using GPU and it is very beneficial for petroleum geoscience big data processing.

Acknowledgments

This research is supported by Lab for Integration of Geology and Geophysics, China University of Petroleum, Beijing. The authors wish to express their appreciation to Lifeng Liu, Pei Yang, Hongyu Yu for many discussions.

References

[1] Yan Ma, Haiping Wu, Lizhe Wang, Bormin Huang, Rajiv Ranjan, Albert Zomaya, Wei Jie. Remote sensing big data computing: Challenges and opportunities. Future Generation Computer Systems, 2014.

[2] S. Habib, V. Morozov, N. Frontiere, H. Finkel, A. Pope, K. Heitmann, Hacc: Extreme scaling and performance across diverse architectures, in: High Performance Computing, Networking, Storage and Analysis (SC), 2013 International Conference for, Nov. 2013, pp. 1–10.

[3] Min Xie, Yutong Lu, Kefei Wang, Lu Liu, Hongjia Cao, Xuejun Yang, Tianhe-1a interconnect and message-passing services, IEEE Micro, (2012), 32 (1) :8–20.

[4] Sam Zandong Sun, Lifeng Liu. A numerical study on non-linear AVO inversion using chaotic quantum particle swarm optimization, Journal of Seismic Exploration, 2014,23(4).