

Accelerating of the reservoir simulation by the GPU

Yusuke Ishizawa,¹ Keita Matsumoto,¹ Kazuki Horie,¹ Takahiro Sakaki,¹ Natsuki Koyama,¹
Mitsuyoshi Shoji,¹ Satoshi Kuge,¹ Kozo Sato,² Komei Okatsu,³ Yuji Miyake³

¹Progress Technologies,Inc., ²University of Tokyo, ³JOGMEC



Abstract

For statistical evaluation ... we aim to accelerate reservoir simulator TOUGH2/ECO2N.
As the result, we achieved over 5 times speed up by implementing GPU.



Prior to CCS project

It is very important to evaluate the integrity of CO2 storage site.

→ Modeling of the site and Simulating CO2 behavior.

The computational model include much **uncertainty** (porosity, permeability etc.).

To consider the uncertainty, generally-used a statistical evaluation Monte Carlo approach.

This approach requires a number of simulation ⇒ The time cost is unrealistic (more than one month).

In this study, we accelerate the reservoir simulation by GPU.

TOUGH2/ECO2N

TOUGH2 is a numerical simulator for flows of multicomponent, multiphase fluids in porous and fractured media.
ECO2N module is designed for simulating the geologic sequestration of CO2 in saline aquifers.

The program structure is shown in Fig.1. (Equation of State is ECO2N)

Higher cost processes of TOUGH2/ECO2N are

- Solution of linear equations
(We use Preconditioned BiCGSTAB as solver)
- Assembly of balance equations
- Equation of State

These processes account for 80~90% of execution time.

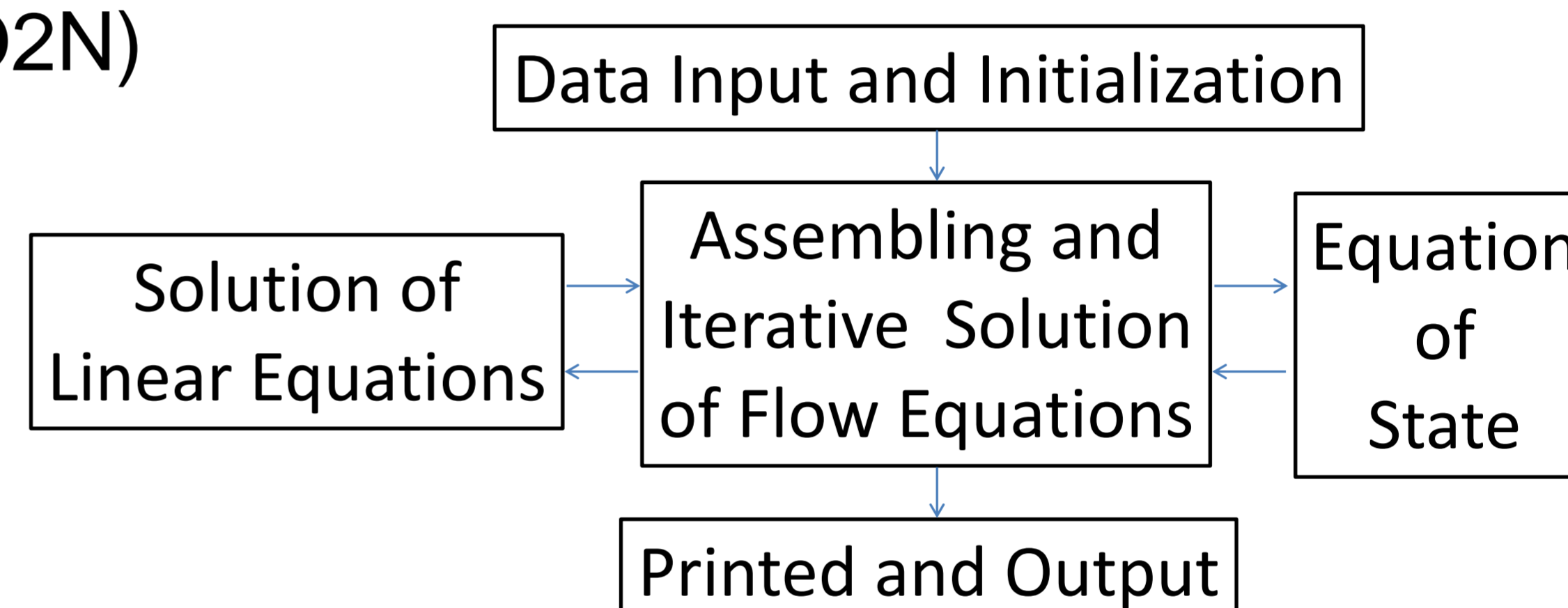


Fig.1. TOUGH2/ECO2N program structure

General Purpose GPU (GPGPU)

GPU is many-core architecture.

ex) GeForce GTX580 has 512 cores

available highly parallel computing

To implementing GPU, we used CUDA platform.

Acceleration method by GPU

Basic theory

Multi-thread parallelization (loop unrolling)

Memory access optimization (coalescing)

+

For preconditioned BiCGSTAB

BiCGSTAB has data-dependent and list access.

Cyclic MultiColor + Cuthill Mckee ordering

An element data is independent among the same color.



MultiColor (Color#=4)

Cuthill-Mckee

Cyclic MC-CM
load balance is good

Sparse matrix storage format is ELL-R

+ optimized for CPU

Evaluation

Hardware specification

- Intel Xeon E5540 (Nehalem)
- NVIDIA GeForce GTX 580 (Fermi)



Experimental model

- Size : 129x122x7
- Total of CO2 injection : 11000 t
- Simulation term : 100 years

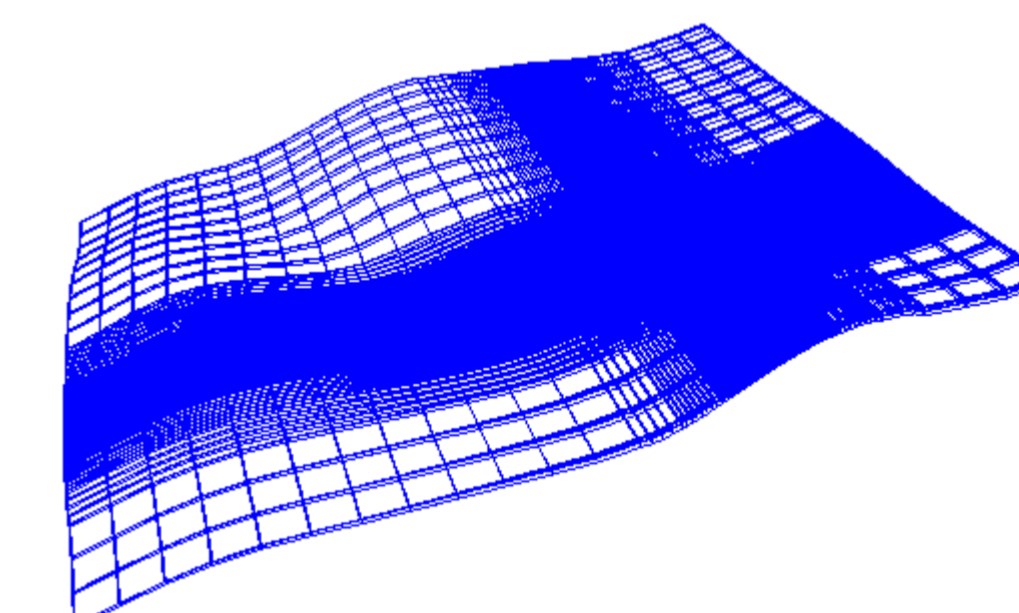


Table1. Acceleration performance

	CPU		CPU+GPU		Acceleration
	Time [sec]	Cost [%]	Time [sec]	Cost [%]	
Solution of linear equations	9371.40	78.2	1295.68	60.7	X7.2
Assembly of balance equations	1679.95	14.0	680.76	31.9	X2.4
Equation of State	924.88	7.7	146.14	6.8	X6.3
Total	11986.62	100	2133.99	100	X5.6

Acceleration up to 5.6 times (11986 → 2133 sec)

Conclusion

We implemented GPU to TOUGH2/ECO2N and achieved 5.6 times speedup (110,000 grid blocks).

As the number of grid block is more, the performance would be highly.

Using our simulator, we expect that the Monte-Carlo simulation finishes in realistic time.

View of execution time

As the number of grids increases, the acceleration is higher.
In case of our study,

> 20000 grids : faster than CPU speed
= 20000 grids : equal to CPU speed
< 20000 grids : slower than CPU speed

In our model size (about 100000 grids),

The execution time depends on the memory bandwidth than the clock rate.

So, using higher memory bandwidth GPU, the simulation is more accelerating.

For any problem, the acceleration would be similar performance by tuning.

Table. The detail of experimental model

Size	4400m x 3880m x 923m
Injection rate	10000 ton/year
Injection term	1.5 year
Absolute permeability	average 3 mDarcy
porosity	average 22 %

View of GPU

GPU is a specialized circuit that is capable of rendering graphics much more quickly than a CPU

Modern GPU is just beginning to be applied to computing in CPU such as flow simulations.

The GPU has many cores than CPU and wider memory bandwidth ⇒ be able to execute in parallel.

In general, the cost performance of CPU+GPU machine is better than only CPU machine.

There is Tesla series and GeForce series.

	GeForce GTX 580	Tesla C2050
Cuda core processor	512	448
Processor clock rate	1.544 GHz	1.15 GHz
Amount of memory	1.536 GByte	3 GByte
Memory bandwidth	192.4 GByte/sec	144 GByte/sec

* Both GPU are Fermi architecture

Tesla is that the stability is higher,
memory bandwidth is narrower,
cost is more expensive
than GeForce.

In our study,

GPU that we used is GeForce GTX580.

The acceleration depends on memory bandwidth, the simulation using GeForce is faster than using Tesla.

View of acceleration method

■ Cyclic MultiColor Cuthill Mckee ordering method

A ordering method is used to solve the linear equations in parallel by making independent data.

The ordering method typically influences the convergence of iteration method.

⇒ There are **trade-offs** between the parallel rate and the convergence.

In case of Multicolor ordering method,

As the number of color increases, parallel rate is higher. But the convergence is lower.

In case of Cuthill Mckee ordering method,

It makes independent data without the trade-offs. But the load balance isn't so good and parallel rate is low.

↓ combining advantage of Multicolor and Cuthill-Mckee ordering

Cyclic Multicolor Cuthill Mckee ordering method has good convergence by using Cuthill-Mckee ordering
good load balance by using Cyclic Multicolor ordering.

■ ELL-R format

ELL-R format is used to storing the value of a matrix to data array.

The matrix values and the related values are stored to 3 arrays as shown.

$$A = \begin{pmatrix} 11 & 12 & 0 & 0 & 15 \\ 21 & 22 & 0 & 0 & 0 \\ 0 & 0 & 33 & 0 & 35 \\ 0 & 0 & 0 & 44 & 0 \\ 51 & 0 & 53 & 0 & 55 \end{pmatrix} \rightarrow \begin{array}{l} \text{Value} = 11, 21, 33, 44, 51, 12, 22, 35, 0, 53, 15, 0, 0, 0, 55 \\ \text{index1} = 1, 1, 3, 4, 1, 2, 2, 5, 0, 3, 5, 0, 0, 0, 5 \\ \text{index2} = 3, 2, 2, 1, 3 \end{array}$$

The array *Value* is the matrix values stored along a column except zero values.

The array *index1* and *index2* respectively indicate the column number of non-zero value and the total number of non-zero value of each column.

By using this format, GPU is able to coalescing access and the memory access speed makes more faster.

Our Company is

Progress Technologies, Inc.

provides the process improvement consulting and the support about development of products.
We began a study about CCS to contribute to prevention of global warming.

This study was carried out as part of the collaborative research with University of Tokyo.



PROGRESS URL : <http://www.progresstech.jp>
TECHNOLOGIES email : y-ishizawa@progresstech.jp

