ARRAY COMPUTING BASED IMPLEMENTATION OF WATER REINJECTION IN GEOTHERMAL STRUCTURE

Simeon Kostyanev¹, S. Kocsárdi², Z. Nagy³, P. Szolgay³, S. Akin⁴, Velislav Stoyanov¹

¹ University of Mining and Geology “St. Ivan Rilski”, 1700 Sofia, E-mail: simeon44@yahoo.co.uk; velis1avs@yahoo.com
² Analogic Computers Inc., Budapest, Hungary
³ Budapest University, Hungary
⁴ Ankara Technical University, Turkey

ABSTRACT: A method is given here to model the processes of filtration and heat transfer in a three-dimensional geothermal medium. The main goal of the analysis was to find and proof the correlation between pressure and temperature in a geothermal structure. The Cellular Neural /Nonlinear Networks paradigm is a natural framework to describe the behavior of locally interconnected dynamical systems which have an array structure. Emulated digital implementation of the CNN-Universal machine can use space variant templates (weighted interconnections) and multi-layered structures on different array processing architectures.

We want to find the optimal computational architecture which is satisfying the functional requirements, using the minimal precision and nevertheless achieving maximum computing power. To meet these requirements we want to process the spatial-regions with the highest possible parallelism.

Introduction

The Cellular Neural Network is a non-linear dynamic processor array. Its extended version, the CNN Universal Machine (CNN-UM), was invented in 1993 [T. Roska and L. O. Chua, 1993]. The CNN paradigm is a natural framework to describe the behavior of locally interconnected dynamical systems which have an array structure. So, it is quite straightforward to use CNN to compute the solution of different PDEs [R. Carmona, F. Jiménez-Garrido, R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, Z. Nagy and P. Szolga, 2003; P. Szolga, G. Vóro and Gy. Erőss, 1997] – but in practical realizations the result cannot be used because of the limitations of the analog CNN-UM chips such as low precision or the restricted usability by applications operating with space-variant templates in a multi-layered structure – and to use the highest order of time derivatives. By implementing the CNN-UM architecture on an array processor it is possible to modify the cell model and simulate its behavior in very short time.

A method is given here to model the processes of filtration and heat transfer in a three-dimensional geothermal medium. The main goal of the analysis was to find and proof the correlation between pressure and temperature in a geothermal structure. Knowing the partial differential equations of a valid geothermal problem it is necessary to compute complex spatio-temporal dynamic.

We want to find the optimal computational architecture which is satisfying the functional requirements, using the minimal required precision and nevertheless achieving maximum computing power. To meet these requirements we want to process the spatial-regions with the highest possible parallelism.

The geothermal model

A mathematical model of filtration and thermal processes of the surveyed region, the “Kazichene – Ravno pole” has been developed with a view to producing geothermal energy. The
mechanism of thermodynamic processes is strictly defined by Darcy’s law of filtration and Fourier’s law of heat transfer [Pentland, Gitirana and Fredlun, 2001], so it is expressed by differential equations, supplemented with initial and boundary conditions, conformable to the specific problem.

**Boundary value problem of filtration process**

For underground water movement the main differential equation referring to the filtration in the surveyed stratum, is

\[
\frac{\partial}{\partial x} \left( T \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left( T \frac{\partial H}{\partial y} \right) = 0
\]

where \( H \) is water pressure measured from unspecified reference plane, \( T \) is stratum conductivity \( T = k \cdot m \) which is function of co-ordinates \( x \) and \( y \), \( k \) is filtration coefficient and \( m \) is stratum thickness which is a function of the co-ordinates \( x \) and \( y \). The filtration rate in each point of the filtration field is:

\[
V_x = -k_j \frac{\partial H}{\partial x}, \quad V_y = -k_j \frac{\partial H}{\partial y}
\]

**Boundary value problem of heat transfer**

As the heat transfer takes place in three layers of different hydro-geological and thermo-physical characteristics, the model is described by three differential equations. The structure of the layers is shown in Figure 1.

\[
\frac{\partial}{\partial t} \left( \lambda_1 \frac{\partial t_1}{\partial x} \right) + \frac{\partial}{\partial y} \left( \lambda_1 \frac{\partial t_1}{\partial y} \right) + \frac{\partial}{\partial z} \left( \lambda_1 \frac{\partial t_1}{\partial z} \right) = \rho_1 c_1 \frac{\partial t_1}{\partial \tau}
\]

\[
\frac{\partial}{\partial t} \left( \lambda_2 \frac{\partial t_2}{\partial x} \right) + \frac{\partial}{\partial y} \left( \lambda_2 \frac{\partial t_2}{\partial y} \right) + \frac{\partial}{\partial z} \left( \lambda_2 \frac{\partial t_2}{\partial z} \right) - m \rho c V_x \frac{\partial t_2}{\partial x} = m \rho c V_y \frac{\partial t_2}{\partial y}
\]

\[
\frac{\partial}{\partial t} \left( \lambda_3 \frac{\partial t_3}{\partial x} \right) + \frac{\partial}{\partial y} \left( \lambda_3 \frac{\partial t_3}{\partial y} \right) + \frac{\partial}{\partial z} \left( \lambda_3 \frac{\partial t_3}{\partial z} \right) = \rho_3 c_3 \frac{\partial t_3}{\partial \tau}
\]

where the symbols are as follows:

\( t_1, t_2 \) and \( t_3 \) denote temperatures in the 1st, 2nd and 3rd area and are functions of \( x, y \) and the time \( \tau \).

\( \lambda_1, \lambda_2 \) and \( \lambda_3 \) are the coefficients of conductivity in the same areas and are functions of \( x, y \) and \( z \).

\( c_1, c_2 \) and \( c_3 \) mean heat capacities of the rocks.

\( \rho_1, \rho_2 \) and \( \rho_3 \) are the rock densities in the respective areas.

The initial and boundary conditions are as follows:

\[
t_1(x, y, 0) = \varphi(x, y, z) \]

\[
t_2(x, y, 0) = \psi(x, y, z) \]

\[
t_3(x, y, 0) = \gamma(x, y, z)
\]

At the disruption \( \lambda \), conjugation conditions are satisfied:

\[
\lambda_1 \frac{\partial t_1}{\partial z}(x, y, \tau) = 0
\]

\[
\lambda_2 \frac{\partial t_2}{\partial z}(x, y, \tau) = 0
\]

Here the square brackets indicate the difference of the boundary value at the different sides of the surfaces, with disturbed continuity of parameter \( \lambda \).

On the boundary \( z = 0 \) the temperature is constant:

\[
t_1(x, y, 0) = t_1 = \text{const}
\]

The boundary condition of area 3 is:

\[
\lambda_3 \frac{\partial t_3}{\partial z}(x, y, \tau) \big|_{z=0} = q_0 = \text{const}
\]

**Discretisation of PDEs in time and space**

The process described by the governing equation of filtration is a truly boundary value problem, which does not depend from the time and although by our computations the filtration terms are space dependent, but constant values in time.

In implementation of the governing equations of heat transfer on different hardware units, it is necessary to discretise the system of equations both in accordance with space and time.

The first order forward Euler formulation has been used to perform this operation and result wise a set of explicit, coupled finite-difference equations have been derived corresponding to equation (3)-(5) and can be described by the following formulas:

\[
\frac{d}{dx} \left( \frac{\lambda}{k} \frac{d t_1}{d x} \right) + \frac{d}{dy} \left( \frac{\lambda}{k} \frac{d t_1}{d y} \right) = \rho c \frac{d t_1}{d \tau}
\]

\[
\frac{d}{dx} \left( \frac{\lambda}{k} \frac{d t_2}{d x} \right) + \frac{d}{dy} \left( \frac{\lambda}{k} \frac{d t_2}{d y} \right) - m \rho c V_x \frac{d t_2}{d x} = m \rho c V_y \frac{d t_2}{d y}
\]

\[
\frac{d}{dx} \left( \frac{\lambda}{k} \frac{d t_3}{d x} \right) + \frac{d}{dy} \left( \frac{\lambda}{k} \frac{d t_3}{d y} \right) = \rho c \frac{d t_3}{d \tau}
\]
\begin{align}
\Delta t_{x,y,i}^{n+1} &= \Delta t_{x,y,i}^{n} + \frac{\Delta t}{\rho c_{i}} \Delta x^{2} \left( \alpha_{x,y,i}^{n} - \rho c_{i} \Delta x \right) \\
&\quad + \frac{\Delta t}{\rho c_{i}} \Delta x^{2} \left( \alpha_{x,y,i}^{n} - \rho c_{i} \Delta x \right) + \rho c_{i} \Delta x^{2} \left( \alpha_{x,y,i}^{n} - \rho c_{i} \Delta x \right) \\
&\quad + \frac{\Delta t}{\rho c_{i}} \Delta x^{2} \left( \alpha_{x,y,i}^{n} - \rho c_{i} \Delta x \right) + \rho c_{i} \Delta x^{2} \left( \alpha_{x,y,i}^{n} - \rho c_{i} \Delta x \right)
\end{align}

(10)

The SPEs are SIMD only processors which are designed to handle streaming data. Therefore they do not perform well in general purpose applications and cannot run operating systems. FlexIO interfaces where one of them (the Broadband Interface (BiF)) is coherent and makes it possible to directly connect two Cell processors.

Cell Processor Architecture

The Cell Broadband Engine Architecture (CBEA) is designed to achieve high computing performance with better area/performance and power/performance ratios than the conventional multi-core architectures. The CBEA defines a heterogeneous multi-processor architecture where general purpose processors called Power Processor Elements (PPE) and SIMD processors called Synergistic Processor Elements (SPE) are connected via a high speed on-chip coherent bus called Element Interconnect Bus (EIB). The CBEA architecture is flexible and the ratio of the different elements can be defined according to the requirements of the different applications. The first implementation of the CBEA is the Cell Broadband Engine (Cell BE or Cell), being designed for the Sony Playstation 3 game console, contains 1 PPE and 8 SPEs. The block diagram of the Cell is shown in Figure 2.

The PPE is a conventional dual-threaded 64-bit PowerPC processor which can run existing operating systems without modification and can control the operation of the SPEs. To simplify processor design and achieve higher clock speed instruction reordering is not supported by the PPE. The EIB is not a bus as suggested by its name but a ring network which contains 4 unidirectional rings where two rings run counter to the direction of the other two. The dual-channel Rambus XDR memory interface provides very high 25.6GB/s memory bandwidth. I/O devices can be accessed via two Rambus

![Cell Processor Architecture](image)

Fig. 2. Model of the examined region

Cell Systems

The third generation blade system is the IBM Blade Center QS22 equipped with new generation PowerXCell 8i processors manufactured by using 65nm technology. Double precision performance of the SPEs are significantly improved providing extraordinary computing density – up to 6.4 TFLOPS single precision and up to 3.0 TFLOPS double precision in a single Blade Center house. These blades are the main building blocks of the world’s fastest supercomputer at Los Alamos National Laboratory which first breaks through the “petaflop barrier” of 1,000 trillion operations per second.

Solution on a CNN Architecture

To model the process of reinjection on emulated digital CNN architecture [8] a space-variant CNN model has been developed based on equation (10)-(12), which is operating with 3.5 dimensional templates. The second equation which describes the behavior of the water saturated transitional layer contains two additional parts which were derived from the time-independent filtration equation and make the connection between the process of filtration and heat transfer.

By the process of filtration only the temperature values have to calculated and updated during the iterations, so it can be used zero-input CNN templates using the given initial values as initial state of the template running. To design space-variant, non-linear template for the three-dimensional medium we have designed 3 coupled 2D templates using an r=1 neighborhood for every three physical layers, so every feedback template-threefold is containing 27 elements.
The structure of the coupled templates for one physical layer can be seen in Figure 3., where \( n \) denotes the described physical layer.

![Fig. 3. Structure of coupled template](image)

The relatively small local memory of the SPEs does not allow to store all the data required for the computation, therefore an efficient buffering method is required to save memory bandwidth. In our solution a belt of 6 slices is stored in the local memory from the array: 3 slices are required to form the local neighborhood of the currently processed row, one slice is required for data synchronization, and two slices are required to allow overlap of the computation and communication as shown in Figure 4. During implementation the environment of the CNN simulation kernel was used [Z Nagy, Zs. Vörösházi, P. Szolgay, 2006]. Template operations are optimized according to the discretized equations (10)-(12) to improve performance. The optimized kernel requires about 32KB memory from the local store of the SPE leaving approximately 224KB for the slice buffers. Therefore the size of the buffer is maximum 3584 grid points (59x59 array) while the number of slices is only limited by the size of the main memory.

![Fig. 4. Local store buffers](image)

The space-variant templates for the first and third physical layers can be determined similarly, there only need to be used the appropriate \( \rho \) and \( c \) multiplier coefficients.

By using the previously described discretization method a C based solver is developed which is optimized for the SPEs of the Cell architecture.

The large (128-entry) register file of the SPE makes it possible to store the neighborhood of the currently processed cell during the solution of the governing equations. The number of load instructions can be decreased significantly. Since the SPEs cannot address the global memory directly, the user’s application running on the SPE is responsible to carry out data transfer between the local memory of the SPE and the global memory via DMA transactions.

The large (128-entry) register file of the SPE makes it possible to store the neighborhood of the currently processed cell during the solution of the governing equations. The number of load instructions can be decreased significantly. Since the SPEs cannot address the global memory directly, the user’s application running on the SPE is responsible to carry out data transfer between the local memory of the SPE and the global memory via DMA transactions.

![Fig 5. Generation of the left neighborhood](image)

This limitation can be removed by slicing the cell array into 4 vertical stripes and rearranging the cell values. In the above case, the 4-element vector contains data from the 4 different slices as shown in Figure 6. This makes it possible to eliminate the shift and shuffle operations to create the neighborhood of the cells in the vector. The rearrangement should be carried out only once, at the beginning of the computation and can be
carried out by the PPE. Though, this solution improves the performance of the simulation data, data dependency between the floating-point instructions may still cause pipeline stalls. In order to eliminate this dependency the inner loop of the computation must be rolled out. Instead of waiting for the result of the first floating-point instruction, the computation of the next group of cells is started. The level of unrolling is limited by the size of the register file.

![Rearrangement of the state values](Image)

Figure 6. Rearrangement of the state values

To utilize the power of the Cell architecture computation work should be distributed between the SPEs. In spite of the large memory bandwidth of the architecture the memory bus can be easily saturated. Therefore an appropriate arrangement of data between SPEs can greatly improve computing performance. One possible solution is to distribute grid data between the SPEs. In this case each SPE is working on a narrow horizontal slice of the grid as shown in Figure 7. Communication between the SPE is required only during the computation of the first and last row of the slice (gray areas), which can be efficiently carried out by a single DMA transaction.

![Data distribution between SPEs: slicing](Image)

Fig. 7. Data distribution between SPEs: slicing

However the above data arrangement is well suited for the architecture of the array processors and simplifies the inter-processor communication, the SPEs are accessing main memory in parallel which might require very high memory bandwidth. If few instructions are executed on large data sets then memory system is saturated resulting in low performance. Static timing analysis of the optimized CFD solver kernel showed that a grid point can be updated in approximately 63 clock cycles. Each update requires movement of 24byte data (1x4byte state value, 1x4byte updated state value, 4x4byte mask value) between the main memory and the local store of the SPE. The Cell processor is running on 3.2GHz clock frequency therefore in an ideal case the computation kernel using one SPE is 203 million update/s. The estimated memory bandwidth requirement is 4.8Gbyte/s which is less than 20% of the available memory bandwidth. Therefore grid data can be distributed between 5 SPEs and each of them can work on its own slice in parallel without running into a memory bottleneck.

Results, performance

To show the efficiency of our solution a complex test case was accomplished in a model having 77x77x17 grid points.

![Input map for transitional layer](Image)

Fig. 8. Input map for transitional layer

![Temperature map after simulation](Image)

Fig. 9. Temperature map after simulation

The input and the result of 32 bit floating-point computation of water saturated layer after 1 year of simulation time with 2048s timestep is shown on Figure 8. and Figure 9. The computation time was 221 seconds on Intel Core 2 Duo 2GHz microprocessor processor. This is equivalent to approximately 6.9 million cell updates.

The performance of the IBM Cell based emulated digital solution of the model is very encouraging. The previous computation takes just 7.5s using 1 SPE and 4.14s using all the 8 SPEs of the Cell processor, so the computation has been accelerated by 29.4 and 194.2 times compared to the performance of Core 2 Duo.

Conclusion

The state equation of water reinjection in geothermal reservoirs with realistic input parameters was solved using a CNN architecture. The cell model of the emulated digital CNN-UM processor was modified according to the governing equations of model of heat transfer and filtration processes. The CNN-UM solution can be efficiently accelerated using reconfigurable devices also [Celoxica Ltd. Homepage, Xilinx Products Homepage, S Kocsár, Z. Nagy, S Kostianev, P. Szolgy, 2006]. The proposed architecture was implemented on IBM Cell processor based QS22 architecture.

Our solution was optimized according to the special requirements of the Cell architecture. Performance comparison showed that about 29-time speedup can be achieved with respect to a high performance microprocessor in the single SPE solution, while the speedup is 194-time higher when all the 8 SPEs are utilized. In the future we are planning to extend our solution to use multiple Cell processors.
References


Recommended for publication of Editorial board