IL NUOVO CIMENTO Vol. 29 C, N. 4 Luglio-Agosto 2006
DOI 10.1393/ncc/i2006-10009-y

Parallel computation of biochemical processes for soil remediation through cellular automata simulation

M. Battaggia([1]) and G. Maino([2])([*])

([1]) *Politecnico di Torino - Torino, Italy*

([2]) *Università di Bologna and ENEA, Divisione di Fisica Applicata via G. Fiammelli 2, 40129 Bologna, Italy*

(ricevuto il 22 Febbraio 2006; approvato il 12 Maggio 2006; pubblicato online il 24 Luglio 2006)

Summary. — We present a generalization of a biochemical model of bacterial degradation of organic compounds in soils, based on genetic networks of cellular automata, and discuss the relevant parallel computation approach. The parallel implementation of the genetic model has been performed on a homogeneous cluster of personal computers in an MPI (Message Passing Interface) environment with high specific performances. By this way, it is possible to perform realistic three-dimensional simulations and derive useful information for *in situ* technological applications.

PACS 47.56.+r – Flows through porous media.
PACS 82.20.Wt – Computational modeling, simulation.
PACS 82.33.Ln – Reactions in sol gels, aerogels, porous media.
PACS 89.60.-k – Environmental studies.

1. – Introduction

In the eighties, after concern grew up in the society about the general state of health of the environment, scientists and technologists became interested in the use of microorganisms for a possible solution to the world-wide problem of contaminated sites. Widespread releases of contaminants into soil systems, ranging from surface soils to deep aquifers, represent a major challenge for biotechnology applications. Bioremediation has been proved to be a potentially invaluable tool [1-6] in order to afford and solve this problem. In particular, the *in situ* treatments are carried out locally, without removing the polluted soil, thus saving energy, time and money with respect to the traditional techniques. Moreover, the *in situ* bioremediation effectively removes the problem from

([*]) E-mail: giuseppe.maino@bologna.enea.it

its roots, by full degradation of the chemical pollutants, and implies less risk exposure to the concerned workers.

However, the detailed description of the processes occurring during a bioremediation of surface and subsurface contamination introducing into the soils different microorganism species, requires the knowledge of many physical, chemical and biological mechanisms and mutual interactions, in addition to the inherent difficulties of obtaining reliable experimental information (can the bugs eat the contaminant?) in order to plan the whole intervention. Therefore, we need suitable biophysical and biochemical models, describing these processes and providing the necessary insight for optimization of the whole procedure in each particular case. Bacteria have different behaviors and interactions *in vitro* and *in situ,* and extrapolating the laboratory observations to the realistic situation is a very difficult task. Finally, even if a realistic model has been developed, the computational requirements represent a major challenge: The relevant computing code has to be implemented on parallel architectures.

Modelling processes underlying bacterial degradation of organic compounds and soil bioremediation thus belongs to the typical class of complex systems with a nonlinear many-body dynamics of the mesoscopic and microscopic components. In the past, the relevant system has been described by means of phenomenological kinetic, transport and diffusion equations; recently [7], an alternative, promising modelling approach has been proposed, based on the random Boolean network of genetic networks [8], where the biomass is represented by a network of interacting genes.

With respect to the traditional simulations of diffusive systems based on molecular dynamics or Monte Carlo methods, this random-walk cellular automaton technique has an inherent computational efficiency and a form which is very suitable for high-speed massively parallel computers.

In this work, we address two main improvements with respect to the previous simulations:

- the random-walk cellular automaton code in C language is adapted to a simple yet low-cost and efficient parallel architecture, consisting of a network of biprocessor personal computers with Linux operating system and MPI procedure;
- the biophysical model is improved taking into account a more refined bacterial growth model [9] than the traditional one [10] (often adopted in simplified approximated forms) and including memory effects by means of algebraic and functional-analysis techniques, recently introduced [11,12] in order to provide a convenient yet computationally simple framework for the description of the dynamical behavior of macromolecules.

2. – The biochemical and physical model

A short description of our model for bioremediation is here outlined, since physical and chemical details as well as the mathematical formalism will be presented in a forthcoming article [13]. We consider two different layers in the present biophysical and chemical model, whose dynamical simulation is performed by means of suitable cellular automata [14]:

- A physical-chemical layer (PCL), describing the flowing of many fluids (water and/or other liquid phases) in the porous medium (the considered soil) and the mutual interactions of the chemical substances dissolved in them.

– A biological layer (BL), modelling the dynamical behavior of bacteria, their interactions and with the pollutant chemicals, the production of metabolites.

As far as the latter BL modelling is concerned and discussed in ref. [7], the total biomass can be considered as a set of interacting genes, due to the large genetic exchange between different bacterial cells, in addition to the genetic similarity between different species. The interaction between genes is simply parameterized, while a distinction is introduced between different species and the effects of intracellular reactions within the cell membranes, whose edge effect is explicitly taken into account. Moreover, the dynamics of chemical processes induced by existing contaminants in the soil or supplied nutrients and intermediate metabolites is introduced in the biochemical model.

The parallel implementation of the model—resulting in fast simulations—allows us also to deal with time intervals so long as to include genetic evolution and possible modifications (namely, crossover, stochastic mutations, etc.) in the gene ecology.

It is worth recalling that, in the standard Kauffman model [8], a single cell is considered since the main interest is appointed to the genetic intracellular dynamics. The generalization of this model, developed in ref. [7], introduces a number of different cells, thus allowing the treatment of intercellular dynamical interactions, while a further quantity must be considered, namely the activity of a given gene, closely related to the number of genes of that type present at each given time in the biophysical system.

The considered interactions among different genes are then approximated by Boolean functions, according to Kauffman's original idea [8]. The resulting random Boolean network can describe the genetic evolution through a suitable input matrix defining how each gene influences all the others and randomness is extensively used in the choice, for instance, of the input genes in the whole population that interact with the considered one. The relevant networks thus evolve according to a synchronous dynamics at each time step, where the state function of each gene is computed by applying the Boolean function to the state of its input nodes (interacting genes).

As previously outlined, the significant quantity in the simulation process is the gene activity, that is proportional to the number of genes of given type existing at each time step, obtained by suitable averages on the statistical ensemble.

Memory effects have been introduced in this work, where the value of the activation at a given time depends on both the activation values of the input nodes and higher-order terms taking into account the previous history of the whole system. Our approach is essentially based on the analytical expressions developed in ref. [11]. In addition, we can easily deal with a suitable interaction among different genes and the chemicals introduced in the system by means of a suitable discretization procedure of the diffusion and chemical reaction equations [15].

The former PCL model describes how the liquid and gaseous phases, namely water and air, respectively, diffuse in a porous medium such as the soil, the chemical compounds in solution are conveyed and interact with the walls of pores through adsorption/desorption processes. In particular, this part of the model deals with percolation effects and the induced contamination in the inner water by the chemical pollutants [16, 17]. The PCL is then approximated by *macroscopic* cellular automata which are a generalization of the classical cellular automata, when the state variables can assume a very large number of different values and even—in some cases—to be continuous. We adopt a very simple cubic patterns for the cellular automata simulations, where each cell borders with four cells and is characterized by means of four variables, namely i) the volume of water present in the cell at each time step, ii) the contents of polluted water, iii) the quantity

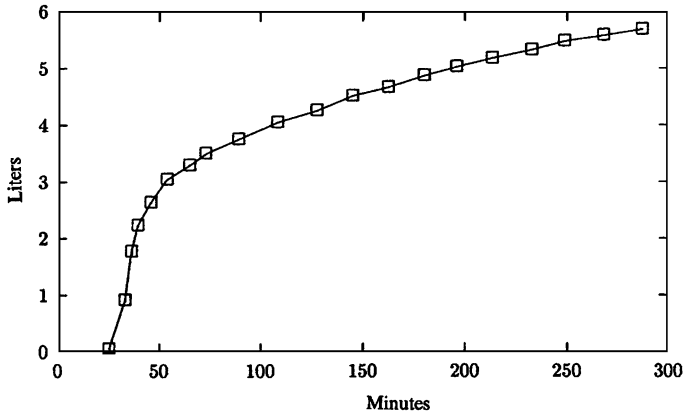

Fig. 1. – Total amount of water filtered through 1 m^3 soil *vs.* time variable.

of pollutant melted in solution, and iv) the quantity of pollutant adsorbed by the pores' walls or, equivalently, the mass of pollutant present at the interface between the solid and liquid phases.

The mathematical equations and algorithms of the present model will be presented and discussed in a forthcoming paper. In this article, we show, as an example, a few results of the performed numerical simulations: Let us consider a 1 m^3 soil, described by a $100 \times 100 \times 100$ grid, which is flooded by 10 liters of polluted water (the concentration of the melted chemical pollutant is 1 g/l), then fig. 1 represents the total amount of water filtered through the considered soil structure. In fig. 2, the average concentrations of con-

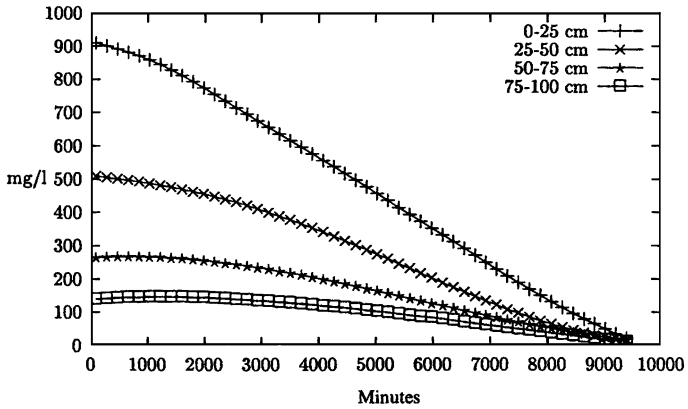

Fig. 2. – Average concentrations of chemical pollutants at different depths as a function of time.

TABLE I. – *Execution times, in seconds, for the simulation described in the text.*

No. of processors	$16 \times 16 \times 16$	$32 \times 16 \times 16$	$64 \times 16 \times 16$	$32 \times 16 \times 32$
1	705.4	1421.9	2844.9	2869.7
2	458.5	818.0	1538.6	1560.4
4	284.2	463.8	823.9	836.1
8	194.8	284.2	464.1	471.3

taminants at different depths are depicted and the final effect of bioremediation is evident. An interesting result is that the pollutant disappears in each layer at the same time.

3. – Parallel issues

The parallel implementation of the theoretical model, previously described, has been essentially based on a simple domain decomposition of the ensemble of genetic network simulations. The statistical ensemble which allows us a determination of *macroscopic* quantities such the gene activity and the time evolution of the relevant population, is then partitioned into a number of subdomains, each assigned to different processors. Therefore, the exchange of data among different blocks is reduced to a minimum, with a significant improvement in the computing performances. At each time iteration, only a communication step is needed in order to update the ensemble statistics and perform the relevant averages.

The parallel algorithm has been developed on a homogeneous cluster of eight personal computers, each node of the hardware structure consisting of a dual Pentium II processor (300 MHz) with 128 MBytes DIMM RAM and 4.5 GBytes HD SCSI UW. A fast ethernet connection at 100 Mbit/sec links the nodes together and a typical peak performance of 194 MFlops per cycle and single processor has been observed on sample calculations [18].

The adopted operating system is Linux RedHat 5.0 and the cellular automaton code has been written in standard C language. Finally, an MPI environment has been adopted for the parallel simulation of the statistical ensemble [19]. The whole procedure can be immediately extended to a larger number of processors or applied to a heterogeneous cluster of workstations with the same software characteristics.

As for the evaluation of the parallel performances of the proposed computational strategy, with respect to sequential calculations, we considered the classical speed-up

TABLE II. – *Speed-up factors for the same simulation as in table I.*

No. of processors	$16 \times 16 \times 16$	$32 \times 16 \times 16$	$64 \times 16 \times 16$	$32 \times 16 \times 32$
1	1.00	1.00	1.00	1.00
2	1.54	1.74	1.85	1.84
4	2.48	3.07	3.45	3.43
8	3.62	5.00	6.13	6.09

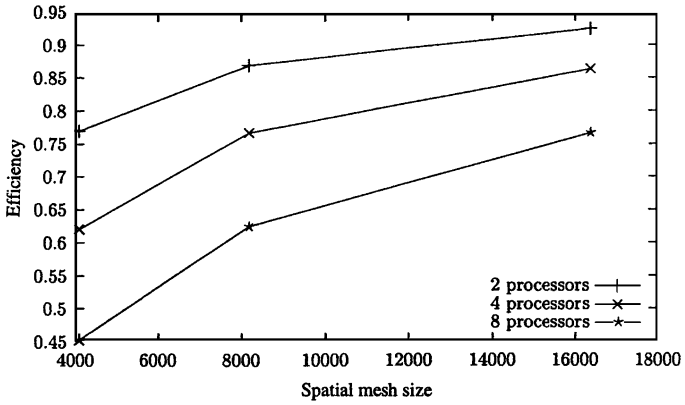

Fig. 3. – Computational efficiency as a function of the number of parallel processors.

and efficiency parameters. Tables I and II list the total computing running times and the speed-up factors, respectively, for different numbers of parallel processors and three-dimensional mesh sizes. One-hundred time steps have been considered, roughly corresponding to 50 minutes of the real process.

The results of table I confirm that a parallel algorithm is clearly needed in order to perform realistic calculations for bioremediation processes ranging from a few weeks to one month or more, with reasonable and affordable computing times. The speed-up factors of table II approach the ideal behavior (speed-up equal to the number of processors) when the size of the spatial mesh increases, thus exhibiting good scalability properties of the model and the relevant code. Finally, the efficiency performances (defined as the ratio between the speed-up and the number of parallel processors) are shown in figs. 3 and 4 as a function of the number of processors, the mesh size being constant, and the spatial mesh sizes (in this case, the number of processors is kept fixed), respectively.

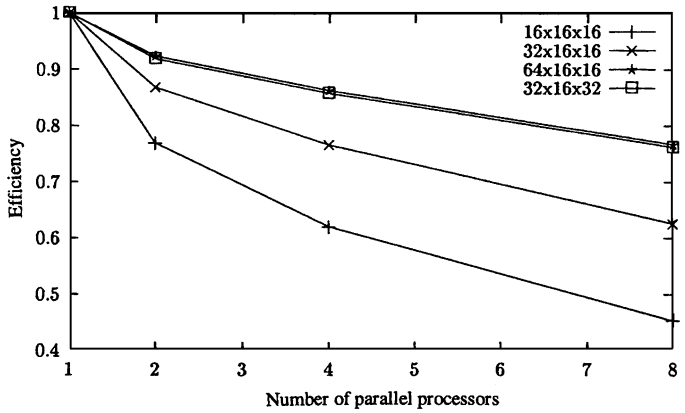

Fig. 4. – Computational efficiency as a function of the size of the spatial mesh.

4. – Concluding remarks

Theoretical improvements of the model—in addition to those previously described in this work—are currently in progress; however, the obtained results prove that a suitable parallel implementation of the cellular automata simulation allows us a realistic macroscopic description of physical diffusive processes and biochemical reactions in a large three-dimensional spatial mesh over long times of more than one week. The parallel code can be used for extensive calculations, also in order to check different forms of both intracellular and intercellular interactions in the gene population, thus providing a convenient tool for planning *in situ* bioremediation interventions.

Finally, such a realistic model of gene dynamics and ecology will made a comparison possible between the resulting bacterial population dynamics from the statistical average of gene activities and the usual models of bacteria evolution in laboratory and real environments, thus providing a useful insight into the microscopic foundations of phenomenological models of bacterial dynamics such as those proposed in refs. [9, 10].

REFERENCES

[1] Bajpai R. K. and Zappi M. E. (Editors), *Bioremediation of Surface and Subsurface Contamination*, *Annals of the New York Academy of Sciences*, Vol. **829** (New York) 1997.
[2] *Biotechnology for Soil Remediation* (C.I.P.A., Milan) 1998.
[3] Alleman B. C. and Leeson A., *In situ and on site Bioremediation* (Battelle Press, Columbus) 1997.
[4] Baker K. H. and Herson D. S., *Bioremediation* (McGraw-Hill, New York) 1994.
[5] Norris R. D. *et al.*, *Handbook of Bioremediation* (Lewis Publishers, Boca Raton) 1994.
[6] Rittman B. E. (Editor), *In situ Bioremediation* (National Reserach Council, National Academy Press, Washington D.C.) 1993.
[7] Serra R. and Villani M., *J. Theor. Biol.*, **189** (1997) 107.
[8] Kauffman S. A., *The Origins of Order* (Oxford University Press, Oxford) 1993.
[9] Ageno M., *La macchina batterica* (Lombardo Editore, Rome) 1992.
[10] Monod J., *Recherches sur la croissance des cultures bactériennes* (Hermann, Paris) 1958.
[11] Bonifazzi C., Maino G. and Paganetto G., *Nuovo Cimento D*, **20** (1999) 2041.
[12] Fini F., Maino G., Vaccari M., Casadio R. and Fariselli P., *Neural networks for the prediction of protein secondary structure and functional-analytical methods for their dynamics*, in *Proceedings of the Symposium on Statistical Mechanics and Computational Thermodynamics: from the Foundations to the Applications in the Fields of Engineering, Environmental Sciences and Diagnostics for the Cultural Heritage, IV Congresso Nazionale della Società Italiana di Matematica Applicata e Industriale (SIMAI), Giardini Naxos, June 1-5, 1998, SIMAI, Rome* (1998), p. 276.
[13] Battaggia M. and Maino G., submitted for publication.
[14] Wolfram S., *Rev. Mod. Phys.*, **55** (1983) 601.
[15] Gibson D. T., *Microbial degradation of Organic Compounds* (Marcel Dekker, New York) 1984.
[16] Freze R. A. and Cherry J. A., *Groundwater* (Prentice Hall, New York) 1979.
[17] Bear J., *Hydraulics of Groundwater* (McGraw-Hill, New York) 1979.
[18] Rossi P., private communication.
[19] Gropp W., Lusk E. and Skjellum A., *Using MPI: Portable parallel programming with the Message-Passing Interface* (MIT Press, Cambridge, Mass.) 1994.