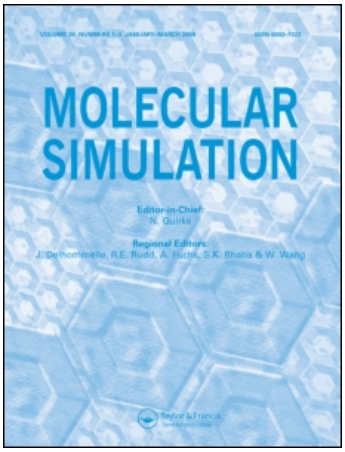

Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Parallel Computation of the Matrix of the Chemical Distances for Defective Graphs

Ottorino Ori[a]; Giulio Destri[b]; Paolo Marenzoni[c]
[a] Thinking Machines Corporation Cambridge, MA [b] Dipartimento di Ingegneria dell'Informazione Università di Parma, Parma, Italy [c] Dipartimento di Fisica Università di Parma, Parma, Italy

To cite this Article Ori, Ottorino , Destri, Giulio and Marenzoni, Paolo(1994) 'Parallel Computation of the Matrix of the Chemical Distances for Defective Graphs', Molecular Simulation, 12: 1, 49 — 56
To link to this Article: DOI: 10.1080/08927029408022535
URL: http://dx.doi.org/10.1080/08927029408022535

Molecular Simulation, 1994, Vol. 12(1), pp. 49–56
Reprints available directly from the publisher
Photocopying permitted by license only

PARALLEL COMPUTATION OF THE MATRIX OF THE CHEMICAL DISTANCES FOR DEFECTIVE GRAPHS

OTTORINO ORI

Thinking Machines Corporation
Cambridge, MA 02142-1214

GIULIO DESTRI

Dipartimento di Ingegneria dell'Informazione Università di Parma, Viale delle Scienze 43100 Parma (Italy)

PAOLO MARENZONI

Dipartimento di Fisica
Università di Parma, Viale delle Scienze 43100 Parma (Italy)

(Received April 1993, accepted April 1993)

We report for an efficient parallel SIMD algorithm, implemented on the Connection Machine, calculating the distance matrix on a large class of defective graphs (graphs with vacancies) representing a given chemical structure. The crucial algorithmic aspects are described in details. The first application of our method simulates the diffusion of vacancies in a periodic square lattice under the effects of a novel, pure *topological* potential: the Wiener number of the graph.

KEY WORDS: Distance matrix, defective graphs, Wiener index, Löwenstein rule, Connection Machine

1 INTRODUCTION

In a recent paper [1] we presented a parallel implementation on the Connection Machine (CM) [2] of a fast algorithm for calculating the entries of the matrix of the chemical distances D for a given chemical graph G. As usual, G is defined by giving a set of nodes (vertices) interconnected by a set of bonds (edges). Thinking for example of a crystalline lattice, the related graph can be easily built by identifying the lattice atoms with the graph vertices. The chemical bonds become the edges of the graph which accounts only for the topological features underlying the initial crystalline structure. All the geometrical data, first of all bond distances and angles, are neglected, and a set of topological methods can be conveniently used in order to derive a large class of physico-chemical data on the proposed system (see the review [3] and [4]). In particular, reference [1] shows an interesting application, based on the knowledge of the D's elements, allowing the partial determination of the ^{13}C NMR spectrum of flullerene molecules.

The current paper still deals with the parallel determination of D, this time being focused on *defective graphs* (DG's). With this term, we will refer to graphs generated from a pre-existing regular (without defects) graph G_r by inserting a given number of holes. From this point of view, *defective graph's* present a wide class of possible irregularities (see next paragraph) which force to find out computational tools able to handle a big variety of cases. We will show that, after minor changes, our algorithm [1] performs very well also in calculating D for *defective graph's*. The next paragraphs list both pathologies commonly encountered with *defective graph's* and adopted computational solutions.

We will close the article with a simple example of topological simulation of holes migrating in a lattice. This kind of theoretical study takes place on a set of consecutive *defective graph's* and generally implies big changes in their topology. In the reported example we will follow, within a periodic square lattice, the diffusion of a set of vacancies under the sole D's influence. This practically unexploited [5] way of describing the diffusion mechanism in a given chemical structure, relies only on the Wiener number:

$$W(G) = \frac{1}{2} \sum_{i,j=1}^{N} d_{ij} \tag{1}$$

of the involved *defective graph's*. In (1) the integers d_{ij}'s (expressing the number of bonds joining the two vertices i and j along the shortest path in G) are the elements of D. Under the action of this new kind of topological interaction, the system presents some physically interesting behaviors, which will be illustrated in detail.

2 THE ALGORITHM FOR DEFECTIVE GRAPHS

Figure 1 summarizes the salient features of our parallel computation of D for a connected graph with N vertices. The only input data the algorithm needs are the so called *adjacency lists* (AL's) which store in a very simple manner the connectivity information. AL's consist in a set of N monodimensional arrays v_i, containing in the first place the number n_i of vertices connected to the ith node. In the remaining n_i positions the labels $j_1, j_2 \ldots j_{n_i}$ of these vertices are stored (see Figure 2a). These v_i's are conveniently positioned in the memory of the serial computer working as CM's front end, avoiding in this manner time costly inter-processors communications [1]. The d_{ij}'s evaluation is reached following a *pulling* approach: starting from the knowledge of the first coordination shell (the set of nearest neighbours) of a given node i and assuming completed the calculation of its kth coordination shell, we can fill the $(k+1)$th one, simply checking, among the remaining not connected nodes j of the graph, if one of these j's has at least a nearest neighbour already belonging to the kth shell of i. If this is the case, $d_{ij} = k + 1$. Our parallel code performs the calculation of D a row at a time, the task being done when all d_{ij}'s are known.

As we have stressed above, this algorithm has a natural flexibility and it can be used for getting D on a general graph. The first good feature of this method, is that the value n_i does not need to be equal for all the lattice nodes i. In particular n_i can be set to zero when the ith node is hosting a hole. Hence we can kill vertices of a starting graph G, keeping our parallel computation of D perfectly working,

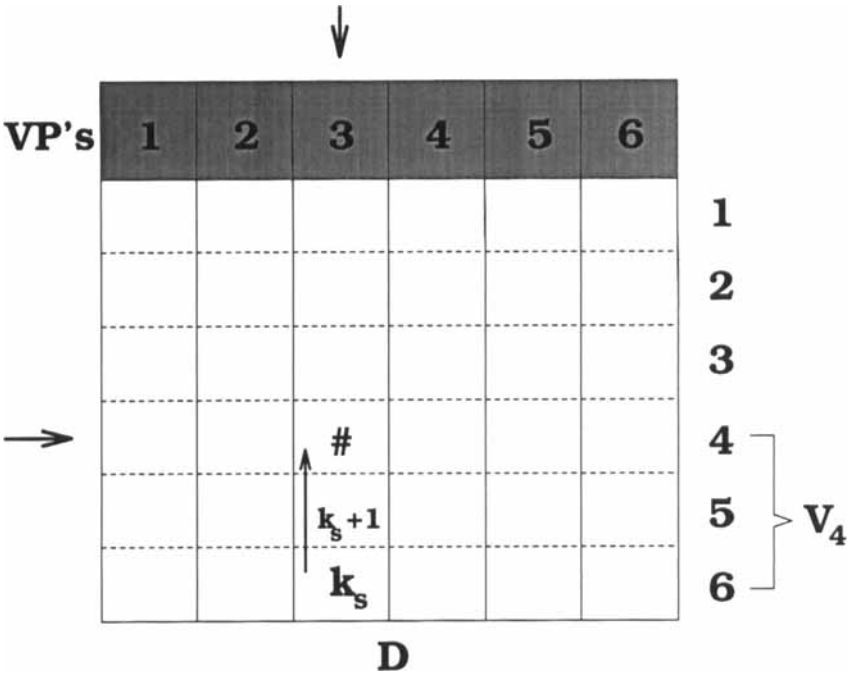

Figure 1 The D's columns are stored in the local memory of the CM's virtual processors (VP's). The code gets a D's row at a time. Computing for example on a graph with six nodes the fourth D's row, the d_{43} entry is determined following a double test: i) d_{43} has to be unknown (usually a negative integer # marks these elements); ii) one of its first neighbours (the 6th node in our example) has to be already connected to the 4th one, with $d_{46} = k_s$ (k_s being the value of the previously completed coordination shell). v_4 contains the information on the connectivity of the 4th node. When all the #'s disappear the calculation is done.

the only effort being the implementation of an efficient method for updating AL's. Putting N_h holes on G, we generate two "symmetric" effects: i) we lose the n_i bonds centered on i (being i one among the N_h vacancies); ii) we reduce by one the connectivity of the n_i atoms belonging to the i's first coordination shell. Another situation to take care of is the possibility that, with a sufficiently high number of holes, one gets a non-connected graph (a graph in-which one or more sites axe disconnected from the remaining ones). If this is the case, some of the d_{ij}'s will remain indetermined at the end of the D's calculation.

Summarizing, our experience with *defective graph's*, did suggest four basic modifications of our standard procedure, itemized as follow:

- When a vacancy lies on a node i, the n_i value (the number of i's bonds) is set to zero. As usual, n_i is stored in the first v_i's place (see Figure 2).
- For each hole i, we change the connectivity of its neaxest neighbours. If j is one of these vertices, we have to decrement n_i by one and we have to modify v_j also, taking into account that the bond with the site i has been removed. Shifting the n_j surviving bonds in the places $v_j(2)$, $v_j(3)$ ... $v_j(1 + n_j)$, is the simplest solution of that problem (see Figure 2b).

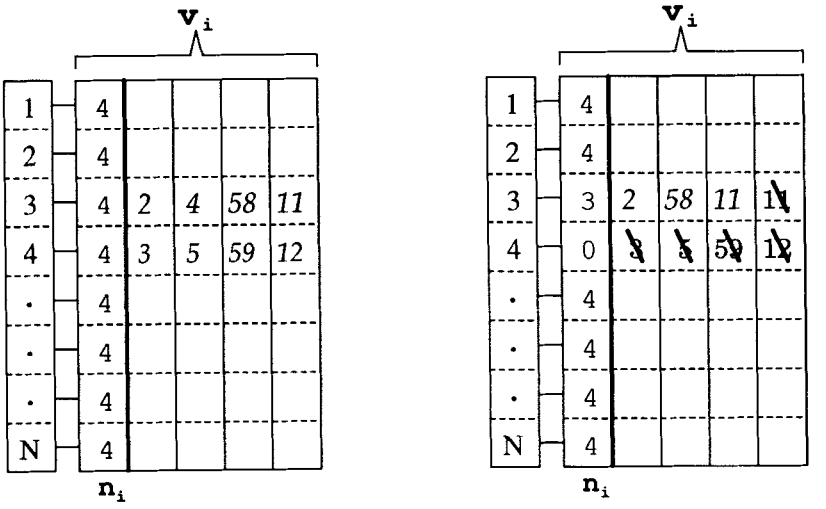

Figure 2a AL's for a regular 4-connected graph with N nodes are shown (for clarity only two of them are completely reported). Figure 2b. Now the 4th node is occupied by a hole. One has not to generate a new pile with $N - 1$ nodes, but simply: i) the connectivity of that node is cancelled; ii) its nearest neighbours (for example the 3th node) reduce the number of their bonds and adjust coherently their AL's (the crossed locations are no longer used during the subsequent D's calculation).

Thanks to the above two, items, starting from the N AL's related to a regular graph G_r, the algorithm is able to manage successively generated *defective graph's*, containing N_h holes in different positions, without wasting time in compressing the surviving AL's in a new pile of $N - N_h$ monodimensional arrays (Figure 2b). This is a crucial point, which reveals all its importance when for example the N_h holes iteratively migrate along G_r, generating a large set of consecutive *defective graph's* and thus a large number of D's computations.

On a connected graph, we remember the D matrix itself works as a sort of logical mask during the calculation, reducing both the overall computational time and the memory occupation. In fact, during the percolation along the coordination shells of the ith node, the sole d_{ij}'s algorithmically active in trying to pull connectivity are the unreached ones, which we initially label with a negative integer (see in Figure 3a the initial setup of D for a regular graph, where the diagonal zero elements mark the inactive D's positions). For *defective graph's* also, the starting values of D have special relevance for speeding up the calculation:

- For *defective graph's* the computationally inactive D's positions (initially set to zero) are not only the diagonal ones, but also the d_{ij}'s placed along the rows and the columns corresponding to the N_h holes (Figure 3b). In the presence of a connected *defective graph's* the D's computation is over when all the computationally active (non-zero) elements are filled.

When disconnected graphs occur (for example, Figure 4d shows how the presence of four holes can bisect the graph corresponding to a periodic square lattice), further

	1	2	3	4	5	6
1	0	#	#	#	#	#
2	#	0	#	#	#	#
3	#	#	0	#	#	#
4	#	#	#	0	#	#
5	#	#	#	#	0	#
6	#	#	#	#	#	0

a

	1	2	3	4	5	6
1	0	#	#	0	#	#
2	#	0	#	0	#	#
3	#	#	0	0	#	#
4	0	0	0	0	0	0
5	#	#	#	0	0	#
6	#	#	#	0	#	0

b

Figure 3 The initial D's values for a regular graphs (a) and for a *defective graphs* with a hole placed on the 4th node (b), are given.

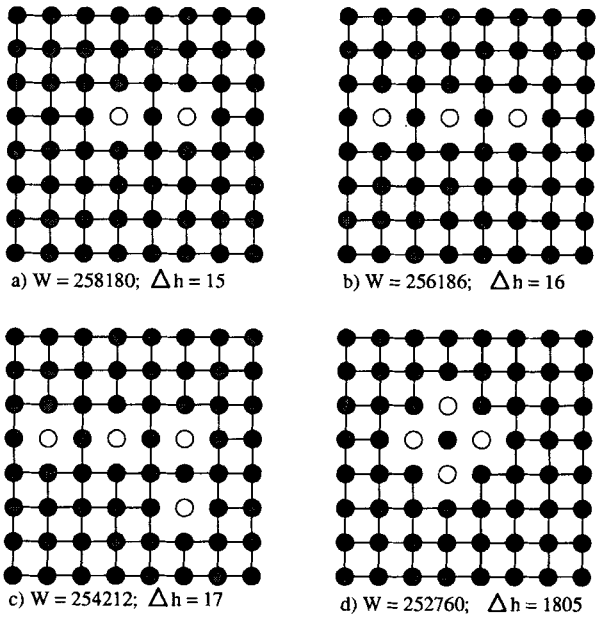

Figure 4 For a $L=16$ periodic square lattice, G_{min} and E_{min} are reported for $N_h = 2, 3, 4$. For simplicity, only a little section of the lattice is depicted. The ideal periodic square lattice has $W=262144$. Δ_h represents the energy barrier separating the proposed C_{min}'s from other configurations. a) and b) respectively give the most compact configuration when 2 or 3 holes are present. For the $N_h = 4$ case, two G_{min} are given: c) the lattice is still a connected one; d) the holes break the lattice in two disconnected parts. In this last case, we made the assumption that the indeterminate D's elements do not contribute to W.

complications arise. Labeling with i and j two nodes lying in two distinct subgraphs, the related distance d_{ij} can not be computed (there is no paths connecting those vertices). Thus we have a new criterion for considering the D's computation terminated:

- For disconnected *defective graph's* the calculation of the ith row of D finishes when the value of the current shell k exceeds a given cut-off M.

The possibility of defining such a cut-off is intuitively obvious (M corresponds to the length of the longest path in G), and gets theoretical support from our previous results relating M to the structure (number of nodes N, dimensionality d and topology) of the underlying ideal graph G_r. For example, for d-dimensional tori we found:

$$M_d(N) = \frac{dN}{2} \tag{2}$$

whereas for a Sodalite unit cell embedded in a Sodalite lattice made by $L \times L \times L$ cells one has:

$$M_{\text{Sod}}(L) = 3L + 1 \tag{3}$$

The four algorithmic "tricks" listed above, do not substantially alterate the data parallel layout of Figure 1, being therefore easily implemented on the CM.

3 APPLICATION TO A PERIODIC SQUARE LATTICE

The importance in having for *defective graph's* a general purpose tool for deriving the distance matrix D, is outlined in the following investigation, where the diffusion of a set of holes in a periodic square lattice (with edge L) is treated by mean of a Monte Carlo (MC) simulation [6] based on the knowledge of the sole $W(G)$ (1).

In our system, $N = L^2$ nodes constitute the reference ideal 4-connected graph G_N in which a set of the N_h vacancies can migrate adjusting their position in order to minimize the Wiener number of the resulting defective graphs G_{N_h}. We can thus define the energy of the system simply as $E = JW(G)$, where the coupling constant J will be set to one in the current paper. When N_h holes are present, $W(G_{N_h})$ takes contribution from the paths interconnecting $N - N_h$ occupied nodes. Starting from a given configuration $G_{N_h}(0)$, where N_h holes are randomly introduced on G_n, the initial energy value is computed as $E_0 = W(G_{N_h}(0))$. Following the usual MC technique, one can choose randomly, for each hole, one among its occupied nearest neighbours (their number ranges between zero and four), moving there the hole. In this way a new graph $G_{N_h}(1)$ is generated. With respect to $G_{N_h}(0)$, $G_{N_h}(1)$ presents some of the N_h holes located on a different lattice node, implying both the reconstruction and the deletion of some of the $G_{N_h}(0)$'s bonds. For this proposed new configuration, we carry out the calculation of $E_1 = W(G_{N_h}(1))$ and, invoking the well known Metropolis rule, the new configuration is accepted or rejected according to its Boltzmann weight, that is:

$$e^{-\beta \Delta E} = e^{-\beta J(E_1 - E_0)} \tag{4}$$

where $\beta \propto T^{-1}$ takes into account the temperature effects on the present system.

The Metropolis rule states that when $E_1 > E_0$ the new configuration is accepted if the above Boltzmann weight exceeds a number r randomly generated between zero and one. The opposite case ($E_1 < E_0$) is systematically accepted. When β is big "enough" (for our model this means $\beta \approx 0.5$) the system tends toward the configuration C_{min}, described by the graph G_{min}, reaching in this way the energy minimum. On the other hand, simulations done with lower values of β ($\beta \approx 0.1$) force the system to change configuration describing large oscillations around C_{min}, and allow the determination of both G_{min}'s topology and the energy hypersurface profile [6].

Figure 4 reports the results (G_{min} and E_{min}) of a sets of calculations carried out for a periodic square lattice with $L = 16$ and $N_h = 2,3,4$. For a given N_h, the Δh represents the energy barrier separating the proposed C_{min}'s from other configurations. The existence of such a barrier, implies that, for each N_h, the holes reach the stable ($\beta = 0.5$) configuration C_{min} in which a well defined order appears. Although this finding will be investigated in future works for understanding the nature of this apparent "phase transition", some general consideration can be done. Under the action of the Wiener energy E, the $N - N_h$ occupied nodes dynamically vary their connectivity to maximize their global compactness (minimizing the sum over all the d_{ij}'s). In doing this, configurations where two or more holes are nearest neighbours are energetically unfavorable, as if they were implicitly interacting in a repulsive way. This may reflect the Löwenstein rule, a well known constraint regarding the available configurations that aluminum atoms have when they are replacing silicon on the nodes of a zeolitic lattice [7]. Löwenstein rule states that Al atoms can not occupy adjacent positions, and it is usually explained in terms of the tendency of the system to reduce the electrostatic repulsive interaction among Al^{3+} ions. Topologically, the relationship between a hole and an impurity is very close: both can be in fact simulated, considering their bonds having different length with respect to the remaining G's bonds. In particular, the configurations of Figure 4, can be reproduced when we consider the N_h holes as graph nodes having *very long* bonds (instead of the adopted picture presenting the holes as void vertices without any bond). This similarity between holes and impurities, together with the current results, points out that an alternative, topological foundation of the Löwenstein rule can be found, and future work on this topic will be carried out.

The last remark has a computational nature. For the systems with $N_h = 4$ depicted in Figure 4c and 4d the computational time on a 8K CM2 is about 500 seconds for a MC simulation with 256 iterations.

4 CONCLUSIONS

We generalized a previously proposed algorithm for calculating the matrix of the chemical distance for a large class of graphs with vacancies. Its parallel implementation performs really well in dealing with the various *defective graph's* encountered simulating the diffusion of holes in a crystalline lattice. We think this tool can be usefully applied for studying, for example, the stability of bulk zeolites when impurities (Al, Ti and so on) substitute silicon in the tetrahedral sites.

Acknowledgements

The authors are indebted with Prof. Gianni Conte of Parma University who made the local Connection Machine generously available throughout the course of this work.

References

[1] O. Ori and P. Marenzoni, "Parallel Computation of the Matrix of the Chemical Distancies on the Connection Machine" *Mol. Sim.* **10**, 00 (1993).
[2] *Connection Machine CM-200 Series Technical Summary*, Thinking Machines Corporation, Cambridge, Massachusetts, 1991.
[3] D.H. Rouvray, "Predicting Chemistry from Topology", *Scientific American*, 36-43 (October 1986).
[4] N. Trinajstic, *Chemical Graph Theory*, CRC Press, 1983.
[5] D. Bonchev, early 80's, proposed the study of the Wiener number for, defective square and cubic lattices. D. Bonchev, private communication.
[6] N.E. Cusack, *The Physics of Structurally Disordered Matter. An introduction* Adam Hilger, Bristol 1988.
[7] G. Melegari and O. Ori, "Computer-aided Topological Analysis of the Faujasite Lattice *I*: Exact Solution for Zeolite-X", *Mol. Sim.* **3**, 235-250 (1989).