RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 34, NUMBER 5 1 SEPTEMBER 1986

Minimal spanning tree: A new approach for studying order and disorder

C. Dussert, G. Rasigni, M. Rasigni, and J. Palmari
Departement de Physique des Interactions Photons-Matière, Faculté Sciences de Saint Jérôme, 13397 Marseille Cedex 13, France

A. Llebaria
Laboratoire d' Astronomie Spatiale, 13012 Marseille, France

(Received 7 July 1986)

We develop a new approach for studying order and disorder in sets of particles. This approach is based on a graph constructed from the set of points locating the positions of the particles. This graph, which is called the minimal spanning tree, allows us to deduce two parameters, namely, the average edge length m and the standard deviation σ, which are characteristic of the repartition to be studied. The method is applied to particles of an aggregated lithium thin film deposited on a dielectric substrate. These particles are found to be partially ordered. The use of a diagram involving both m and σ turns out to be a powerful tool for the determination of the degree of order in very various systems.

Data in the form of a set of points, scattered within a region of space, arise in many fields such as astronomy, crystallography, solid-state physics, biology, etc. It is often possible to consider the objects to be studied (stars, elementary particles, aggregates, proteins, etc.) as points and so to treat any such data set as a spatial point pattern. The classification of patterns as regular, random, or aggregated may be an oversimplification at an early stage of the analysis. Let us consider for instance very thin metallic deposits.[1] In general these deposits are not continuous but made up of a population of small particles. Are these particles distributed at random, are they clumped or are they set out according to a certain order that the eye cannot detect? One way to tackle these questions is to compute statistical moments that characterize the surface, particularly the second-order moment, known as the autocovariance function (ACF).[2-4] However, the ACF is sometimes difficult to interpret because it takes both the size of the particles and their spatial distribution into account. If we focus our attention only on the distribution of the particles, it is preferable to use tests based on pattern recognition techniques, applied to the particles considered as point objects. Generally these tests are statistical comparisons between the observed distribution of either interpoint distances or local density of points and corresponding distributions obtained from some well-characterized model such as a theoretical model or a Monte Carlo simulation.[5] This comparison helps to distinguish between two main distribution tendencies, clustered and regular. We have developed a new approach which is more informative than the above methods. It is based on a graph which is called the minimal spanning tree (MST).

Basic definitions of the graph theory may be found in Refs. 6–8. Let us recall that an edge-weighted linear graph $G=(X,E)$ is composed of a set of points $X=\{x_1,x_2,\ldots\}$ called nodes and a set of node pairs $E=\{(x_i,x_j)\}$ called edges, with a number called weight (in this paper the Euclidean distance) assigned to each edge. A tree is a connected graph without closed loops. A MST is a tree which contains all of the nodes and where the sum of the edge weights is minimal.

Figure 1(a) shows the electron photomicrograph of a shadowed surface replica of an aggregated lithium deposit. This deposit has been prepared using experimental techniques described in Ref. 9. It is possible to reconstruct the surface profile by means of a technique developed by Rasigni, Rasigni, Palmari, and Llebaria,[10] namely, a microdensitometer analysis of the micrograph of the surface-shadowed carbon replica. Then the quantized profile allows us to deduce the positions of particles [Fig. 1(b)] from the maxima seen in the profile.[11] Figure 1(b) may also be obtained from the micrograph density values. Such a simplified process is used if the knowledge of the quantized profile is not required for further computations. For convenience only $N=460$ particles have been represented in Fig. 1(b) whereas full computations were based on $N=2000$. Figure 1(c) shows the MST of Fig. 1(b) computed using Prim's algorithm.[12] In this algorithm the MST is grown from a single node by adding the closest node to the current tree at each stage along with the edge corresponding to that closest distance (smallest weight). Bentley and Friedman[13] have performed an accelerated version of Prim's algorithm. Depending on the starting point there may be more than one MST for a given set of points, but all of the MST's have the same edge-length histogram. It follows that statistical information deduced from the histogram, such as the average edge length m and the standard deviation σ may be used as characteristics for the corresponding distribution. The histogram of the MST of Fig. 1(c) is represented in Fig. 1(d). From the full computation the values $m=0.800$ and $\sigma=0.190$ are derived. These values have been normalized according to the process described by Hoffmann and Jain,[14] which may be summarized as follows. The area A of the sampling window is defined as the area H of the convex hull of the data normalized by means of the relation

$$A=H/[1-(f/N)] \ , \tag{1}$$

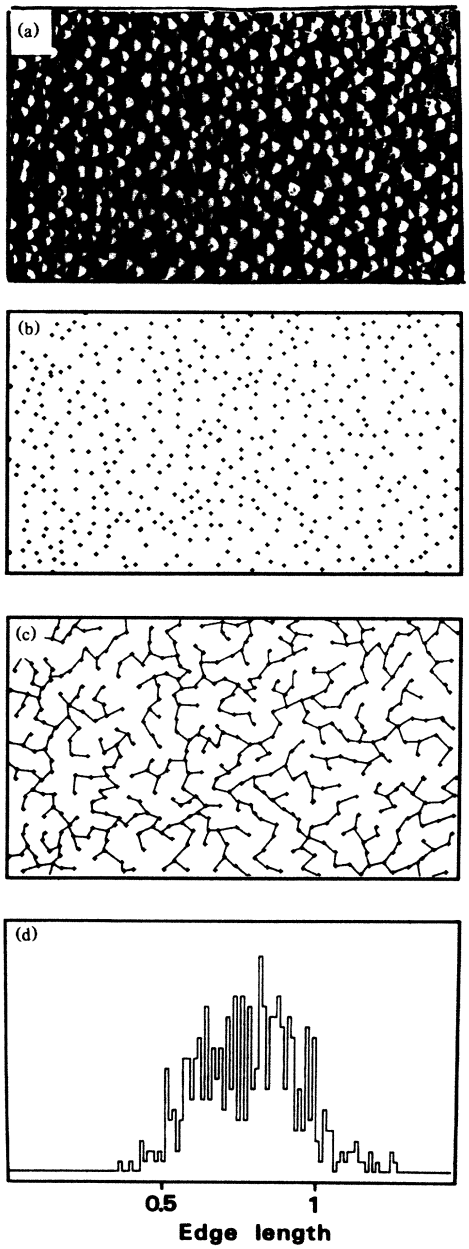

FIG. 1. (a) Electron photomicrograph of shadowed surface for an aggregated lithium thin film deposited on a dielectric substrate. W-Pt shadow casting at an angle of 50°. The line represents 0.5 μm. (b) Position of lithium particles considered as points. (c) Minimal spanning tree (MST) concerning the distribution of (b). (d) Histogram of edge lengths for the MST of (c).

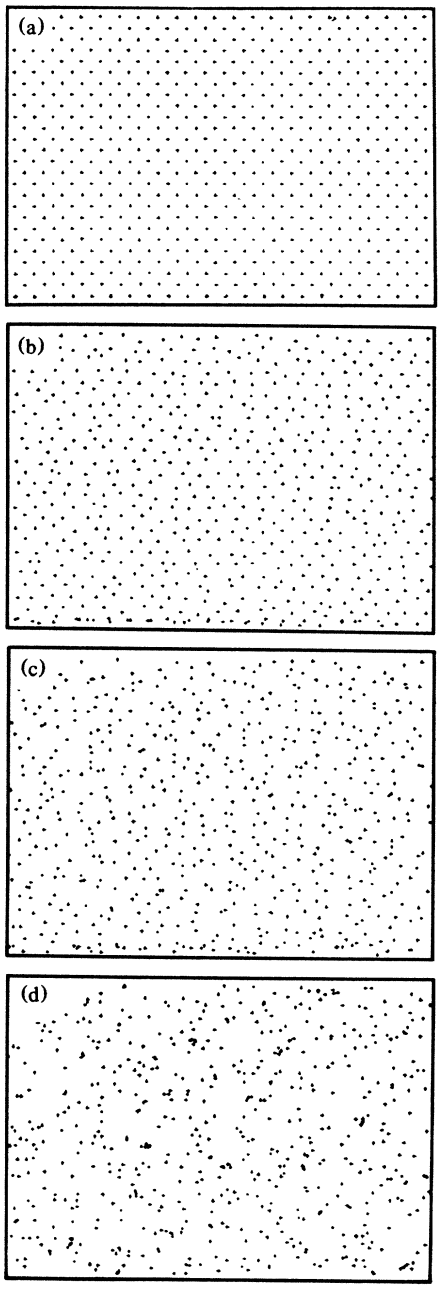

FIG. 2. (a) Set of 500 points ordered at the nodes of a perfectly triangular lattice. (b), (c), and (d) Randomization of arrangement of (a) giving each point a new position deduced from its previous one by means of a Gaussian distribution having the standard deviation ω. The ω values are 0.1 and 0.2 for (b) and (c), respectively. (d) corresponds to the random distribution.

in which N is the number of points in the window, and f is the number of faces on the convex hull. The normalized values of m and σ of the MST constructed from a given set of data are obtained by dividing the original ones by the expression

$$(NA)^{1/2}/(N-1) \ . \tag{2}$$

Indeed the expected length of a randomly chosen edge of a MST related to N uniformally distributed points in a sampling window of area A, is asymptotically proportional to expression (2).[15]

All distributions can be plotted in the (m,σ) plane and easily compared with well-characterized distributions (for instance, perfectly ordered or random ones) due to the normalization process. For example, consider [Fig. 2(a)] the set of points ordered at the nodes of a lattice (triangular for instance). The edge-length histogram of the corre-

sponding MST provides the normalized values $\sigma=0$ and $m=1.075$. The null value of σ is obviously characteristic of a periodic lattice whereas the m value may vary according to the selected periodic lattice. For instance, m is equal to 1 for the square lattice. In addition to the two previous lattices, there may be other arrangements in the plane for which $\sigma=0$, in particular, the arrangements called mosaics, which are composed of an infinite set of regular polygons.[16-18] Figure 3 shows the eleven possibilities permitting the arrangement of a set of points on a plane along regular or semiregular mosaics. The m values corresponding to each model are also reported in this figure. The arrangement in Fig. 2(a) may be randomized by giving each point a new position deduced from its previous position using a Gaussian distribution with a standard deviation of ω. Figures 2(b) and 2(c) show two steps corresponding to two increasing values of ω. For a value of ω of the order of the initial m value, the uniform random distribution is reached as it can be seen in Fig. 4 which shows the variations of m and σ vs ω. The corresponding values of this uniform random distribution, namely, $m=0.662$ and $\sigma=0.311$, have been corroborated by Monte Carlo simulations of random distributions generated using the linear congruential method.[19]

The values of (m,σ) for different values of ω [from $\omega=0$ (perfect arrangement) to ω much greater than the initial m value (random arrangement)], in the case of the triangular lattice, are plotted (solid line) in Fig. 5. If we had chosen a different initial arrangement we would have obtained another path leading to the same random arrangement. This is demonstrated in Fig. 5 by representing

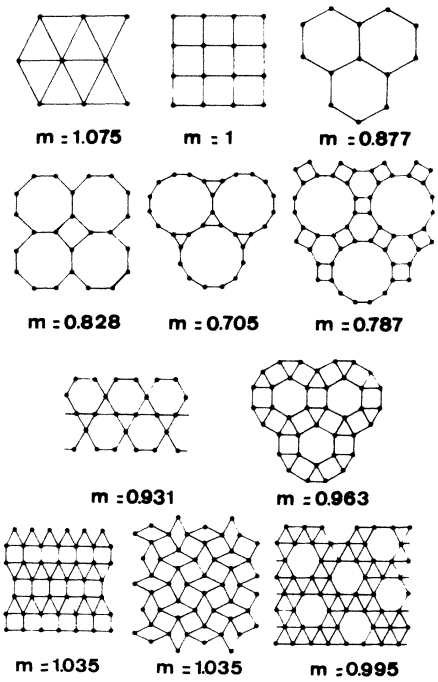

FIG. 3. Eleven possibilities permitting the arrangement of a set of points on a plane along regular or semiregular mosaics. The m values corresponding to each model are also reported on this figure.

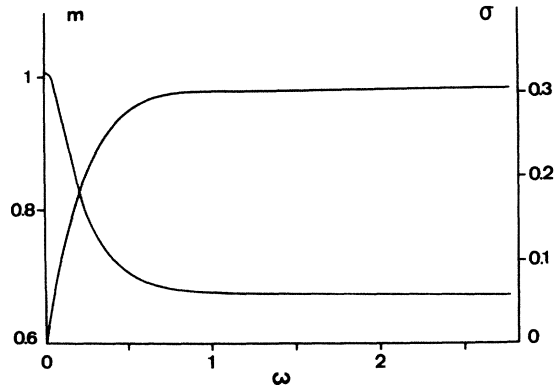

FIG. 4. Average edge length m and standard deviation σ vs ω, for the MST's of the different arrangements obtained by randomizing the set of points ordered at the nodes of a triangular lattice.

the values of $(m,\sigma=0)$ corresponding to the models represented in Fig. 3 and two paths whose initial m values are 0.705 and 1. It has to be noted that the path leading to the disorder also varies slightly with the uniform randomization process.

Actually the sizes of particles may not be small compared to their separations. So we have considered an improved model called the inhibition model.[5] It consists of randomly distributed points with the constraint of a minimal interpoint distance d, related to the diameter of a hard disk that represents the particles. The dashed curve in Fig. 5 represents the values of m and σ deduced from this model for increasing values of d.

If we plot the values of m and σ found for the lithium particles repartition in Fig. 5, we conclude that these parti-

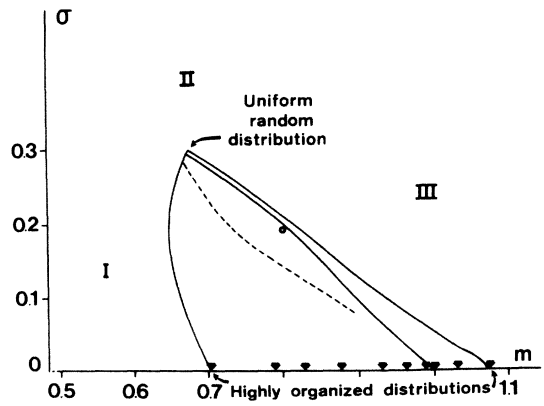

FIG. 5. Solid lines: values of (m,σ) for several values of ω ranging from $\omega=0$ (perfect arrangement) to ω much greater than the initial m value (uniform random distribution). The arrows on the m axis are related to the m values reported on Fig. 3. Dashed line: values of (m,σ) deduced from the inhibition model for increasing normalized values of the minimal interpoint distance d ranging from $d=0$ (random distribution) to $d=0.780$ (which corresponds to a packing intensity $\tau=0.48$). The point ○ characterizes the lithium particles distribution.

cles are not randomly distributed. Moreover, we note that the corresponding point appears to be far from the inhibition curve. That means that the partial order observed on this distribution is not due to the constraints dictated by the sizes of particles. This result may be significant in the context of a more general study of thin-film nucleation.

The method presented above offers several advantages when compared to conventional statistical approaches. In particular, the existence of fast MST algorithms and the ease with which the two parameters m and σ can be derived lead to a low computation time. This allowed us to investigate a large number of points ($N > 2000$) easily and thus to reduce statistical errors. Moreover, the use of normalized variables permits the comparison of different distribution types regardless of sample density and size. Lastly, the use of two distinct parameters, namely, m and σ, leads to a more informative two-dimensional classification which then permits the comparison of distributions by taking a simple reading in the m,σ plane. The previous study of this plane is far from exhaustive. For instance, the areas marked I, II, III on the diagram of Fig. 5 correspond, respectively, to cluster structure[8] (small $m,\sigma \neq 0$), gradients of concentration (large σ), and two-dimensional quasiperiodic tilings[20,21] (large $m,\sigma \neq 0$). Accurate simulations to quantisize these areas are currently under way.

Though this paper demonstrates the utility of the MST method only for lithium particles, we believe that it could be applied more generally and be proved to be a powerful tool for the study of various physical, chemical, and biological structures.

We are indebted to Dr. M. Cadhilac for fruitful discussions.

[1]M. Rasigni and G. Rasigni, J. Opt. Soc. Am. **67**, 510 (1977).
[2]Y. W. Lee, *Statistical Theory of Communication* (Wiley, New York, 1960).
[3]G. Rasigni, F. Varnier, M. Rasigni, J-P. Palmari, and A. Llebaria, Phys. Rev. B **25**, 2315 (1982).
[4]G. Rasigni, F. Varnier, M. Rasigni, J-P. Palmari, and A. Llebaria, J. Opt. Soc. Am. **73**, 222 (1983).
[5]P. J. Diggle, *Statistical Analysis of Spatial Point Patterns* (Academic, New York, 1983).
[6]C. Berge, *Théorie des Graphes et ses Applications* (Dunod, Paris, 1967).
[7]F. Harary, R. Z. Norman, and D. Cartwright, *Structural Models* (Wiley, New York, 1965).
[8]C. T. Zahn, IEEE Trans. Comput. C **20**, 66 (1971).
[9]M. Rasigni and G. Rasigni, J. Opt. Soc. Am. **62**, 1033 (1972).
[10]M. Rasigni, G. Rasigni, J-P. Palmari, and A. Llebaria, J. Opt. Soc. Am. **71**, 1124 (1981).
[11]G. Rasigni, J Palmari, M. Rasigni, F. Varnier, J-P. Palmari, A. Llebaria, and F. Marty, Biophys. J **47**, 431 (1985).
[12]R. C. Prim, Bell Sys. Tech. J. **36**, 1389 (1957).
[13]J. L. Bentley and J. H. Friedman, IEEE Trans. Comput. C **27**, 97 (1978).
[14]R. Hofmann and A. K. Jain, Patt. Recognit. Lett. **1**, 175 (1983).
[15]J. Beardwood, J. H. Halton, and J. M. Hammersley, Proc. Cambridge Philos. Soc. **55**, 299 (1959).
[16]W. W. R. Ball, *Mathematical Recreations and Essays* (MacMillan, London, 1959).
[17]A. F. Wells, *The Third Dimension in Chemistry* (Oxford Univ. Press, New York, 1956).
[18]P. S. Stevens, *Les formes dans la Nature* (Editions de Seuil, Paris, 1978).
[19]D. E. Knuth, *The Art of Computer Programming* (Addison-Wesley, New York, 1969), Vol. 2.
[20]R. Penrose, Math. Intelligencer **2**, 32 (1979).
[21]M. Duneau and A. Katz, Phys. Rev. Lett. **54**, 2688 (1985).

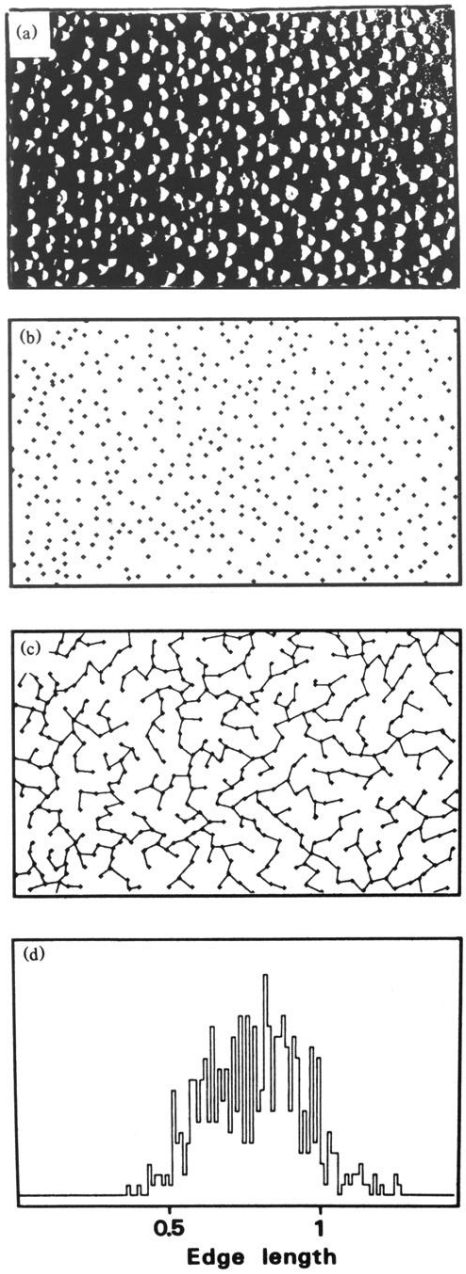

FIG. 1. (a) Electron photomicrograph of shadowed surface for an aggregated lithium thin film deposited on a dielectric substrate. W-Pt shadow casting at an angle of 50°. The line represents 0.5 μm. (b) Position of lithium particles considered as points. (c) Minimal spanning tree (MST) concerning the distribution of (b). (d) Histogram of edge lengths for the MST of (c).