Nuclear Instruments and Methods in Physics Research A 620 (2010) 56–62

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

Hierarchical agglomerative sub-clustering technique for particles management in PIC simulations

Giacomo Grasso [a], Michele Frignani [b], Federico Rocchi [a], Marco Sumini [a,*]

[a] Nuclear Engineering Laboratory (LIN) of Montecuccolino, University of Bologna, v. dei Colli 16, 40136 Bologna, Italy
[b] Ansaldo Nucleare corso Perrone 25, 16161 Genova, Italy

ARTICLE INFO

Available online 29 January 2010

Keywords:
Clustering
Merging
Variance reduction technique
Particle in Cell

ABSTRACT

The effectiveness of Particle-In-Cell (PIC) codes lies mainly in the robustness of the methods implemented, under the fundamental assumption that a sufficient number of pseudo-particles is concerned for a correct representation of the system. The consequent drawback is the huge increase of computational time required to run a simulation, to what concerns the particles charge assignment to the grid and the motion of the former through the latter. Moreover the coupling of such methods with Monte-Carlo-Collisional (MCC) modules causes another expensive computational cost to simulate particle multiple collisions with background gas and domain boundaries.

Particles management techniques are therefore often introduced in PIC-MCC codes in order to improve the distribution of pseudo-particles in the simulation domain: as a matter of facts, the aim at managing the number of samples according to the importance of the considered region is a main question for codes simulating a local phenomenon in a larger domain or a strongly collisional system (e.g.: a ionizing plasma, where the number of particles increases exponentially).

A clustering procedure based on the distribution function sampling applied to the 5D phase space (2D in space, 3D in velocity) is here proposed, representing the leading criterion for particles merging and splitting procedures guaranteeing the second order charge moments conservation.

Applied to the study of the electrical breakdown in the early discharge phase of a Plasma Focus device, this technique is shown to increase performances of both PIC kernel and MCC module preserving the solution of the electric field and increasing samples representativeness in stochastic calculations (with respect to more traditional merging and splitting procedures).

1. Introduction

Particle In Cell (PIC) codes are very flexible and robust tools for the microscopic simulation of many-body systems. The effectiveness of such codes mainly lies in the simplicity of the method: a finite number of simulation particles is moved along a simulation grid, which carries the field interaction. It is clear that the main drawback is the representativeness of the finite number of simulation particles with respect to the physical system to be investigated: even exploiting the most powerful parallel computers, the number of simulation particles is destined to remain several orders of magnitude below the actual particles population. On the other side, the computational time of a PIC module is proportional to the sum of the number of particle N_p and $N_g^D \log N_g^D$, N_g being the number of grid points in D dimensions.

As widely explained in Refs. [1,2], for a simulation to be representative, the number of pseudo-particles per numerical grid cell should be sufficiently high to avoid strong fluctuations in the particles-field solution/interaction: the higher the number of particles in a cell, the more stable the method to what concerns the behavior of particles and all the processes they induce. Furthermore, plasma dynamics simulations could lead to a not uniform distribution of simulation particles in the domain of interest: some cells might undergo a net depletion of particles, resulting in a very poor statistics, while other cells may result over-populated, even far above the aimed number of simulation particles.

A good strategy to get to a compromise between a sufficient representativeness in the whole domain and a computationally acceptable number of pseudo-particles consists in trying to obtain the most uniform simulation particles density in the whole domain as possible, artificially reducing the number of particles in overpopulate regions (merging) and increasing it in depleted sub-domains (splitting).

The key point of such an optimizing strategy is represented by the localization of highly/poorly populated regions (i.e., the

* Corresponding author.
E-mail addresses: giacomo.grasso@mail.ing.unibo.it (G. Grasso), marco.sumini@unibo.it (M. Sumini).

0168-9002/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2010.01.060

definition of a proper rules set for the partitioning of the phase space into clusters of similar particles), where pseudo-particles will be merged/splitted. It is then clear how the choice of the clustering criterion directly affects the actual effectiveness of the method. On the other hand, it should be also noticed that the criterion itself could easily introduce biases in the simulation whether particles to be "rearranged" belong to an excessively wide phase space region.

2. The HASC technique

The effectiveness of the merging and splitting techniques is based on the idea of an equivalence principle: the starting and final sets of particles have to equally contribute to the grid moments and to sample the same distribution function in the phase space. This condition is also recognized sufficient to minimize the bias introduced by the two methods, in favor of the simulation consistency. In order to do this, the overall approach chosen is to identify local closed groups of pseudo-particles in the phase space: within each cluster, and according to the reciprocal groups distribution, both particles merging and splitting procedures will be applied.

To identify particles clusters in the phase space, data clustering algorithms have been exploited. Such algorithms can be hierarchical or partitional: the former find successive clusters using previously established clusters, whereas the latter determine all clusters at once. Hierarchical algorithms can be in their turn agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms begin with each element as a separate cluster and merge them in successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters. For the specific task of interest the proposed Hierarchical Agglomerative Sub-Clustering (HASC) procedure has been envisaged: it is based on the partitioning of the starting set of simulation particles into clusters identified by subsequent agglomeration of closest particles within the phase space through a complete-linkage clustering strategy.

In general, if each element in a set of N objects is represented by a set of D measurements (or attributes), each object can be represented by a pattern, or D-place vector. The set itself can be therefore viewed as a $N \times D$ patterns matrix, each row of which defines a pattern and each column a feature (or measurement). One can figure the D features as a set of orthogonal axes defining a D-dimensional patterns space. A cluster can be thought as a collection of patterns which are close one another or which satisfy some spatial relationships. The task of a clustering algorithm is to identify such natural groupings in spaces of many dimensions. Clustering methods require the establishment of an index of proximity (*alikeness*, *affinity* or *association*) between pairs of patterns. This index can be computed from a patterns matrix or from raw data. The proximity index between the i-th and j-th patterns, $d_{i,j}$, must satisfy the following three properties:

1. $d_{i,i} = 0 \quad \forall i$
2. $d_{i,j} = d_{j,i} \quad \forall i,j$
3. $d_{i,j} \geq 0 \quad \forall i,j$

Generally speaking the geometrical domain coordinates may be considered as features of the D attributes. However, in the frame of a PIC code, the spatial grid for the solution of the field equations can be also used to separate particles into local sets by cell basis. For this task a hybrid counting sort algorithm can be exploited, since it is also known to increase performances avoiding cache trashing [3]; the method is based on an out-of-place sorting separated in three main loops, two over the particles (the first to count them and the second to sort them), which can be hybridized in the charge accumulation and particle push loops, and an intermediate one, over the cells to build an allocation table.

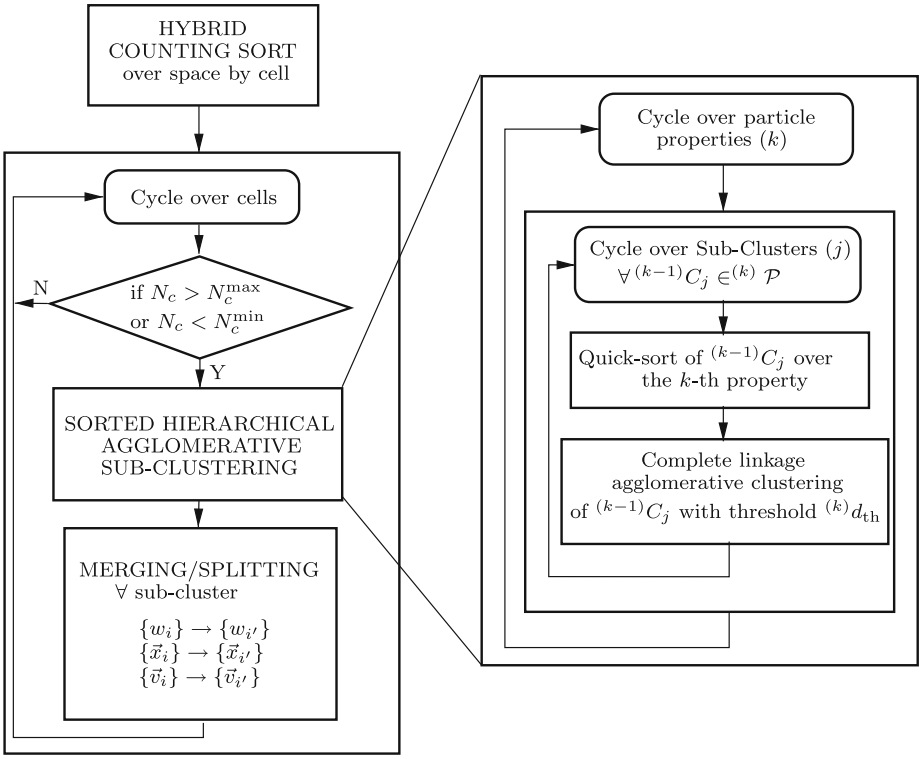

Fig. 1. Scheme of the sorted HASC merging/splitting technique.