[LS90] B. D. Lubachevsky and F. H. Stillinger, Geometric properties of random disk packings, *J. Statistical Physics* **60** (1990), 561–583.

[MFP95] C. D. Maranas, C. A. Floudas, P. M. Pardalos, New results in the packing of equal circles in a square, *Discrete Mathematics* **142** (1995), 287–293.

[MS95] H. Melissen and P. C. Schuur, Packing 16, 17 or 18 circles in an equilateral triangle, *Discrete Math.* **145** (1995), 333–342.

[M97] H. Melissen, *Packing and Covering with Circles*, Ph.D. Thesis, 1997, Netherlands, ISBN 90-393-1500-0.

[NO96] K. J. Nurmela and P. R. J. Östergård, Packing up to 50 equal circles in a square, *Discrete & Computational Geometry*, **18** (1997), 111-120.

[O61] N. Oler, A finite packing problem, *Canad. Math. Bull.* **4** (1961), 153–155.

[R75] G. E. Reis, Dense packings of equal circles within a circle, *Math. Mag.* **48** (1975), 33–37.

[S71] J. Schaer, On the packing of ten equal circles in a square, *Math. Mag.* **44** (1971), 139–140.

[S79] K. Schlüter, Kreispackung in Quadraten, *Elem. Math.* **34** (1979), 12–14.

[V89] G. Valette, A better packing of ten circles in a square, *Discrete Math.* **76** (1989), 57–59.

experiments which show that the hard disk algorithm finds the best packing much easier than the soft disk algorithm when there are many local close-by optima: on the way to those optima configuration space $\{Z\}_d$ disconnects at a very late stage of the expansion. Relative chance for the trajectory to stay in the configuration subvolume that leads to the global optimum is very high. It should be also noted that as the number of disks n increases, the presence of multiple local optima near the global optimum become a rule rather than exception; thus the hard disk packing method becomes progressively more advantageous than the soft disk packing method.

5. Discussion

Given the advantage of the hard disk method why do researchers use the soft disk procedure, instead? Probably because it is a simpler algorithm. The structure of the soft disk program is straightforward: given configuration $Z(t)$, compute configuration $Z(t + \Delta t)$ by adding uniformly to all coordinates (disk positions) $z_i(t)$ a similarly computed increment $\Delta z_i(t)$. This $\Delta z_i(t)$ may be the ith coordinate of the gradient of $P(Z)$ prorated with a certain step. By contrast, the hard disk algorithm must discern complex combinatorics of asynchronous disk collisions and as such is not so easy to program. It involves examining the states only of nearby disks for processing a collision of a disk. Various data structures are employed to make the computations efficient [L91].

References

[CS93] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd ed., Springer-Verlag, New York, 1993.

[GS90] R. L. Graham and N. J. A. Sloane, Penny-packing and two-dimensional codes, *Discrete Comput. Geom.* **5** (1990), 1-11.

[GMPW91] C. de Groot, M. Monagan, R. Peikert, and D. Wurtz, Packing circles in a square: a review and new results, in *System Modeling and Optimization* (Proc. 15th IFIP Conf. Zurich 1991), 45–54.

[FG69] J. H. Folkman and R. L. Graham, A packing inequality for compact convex subsets of the plane, *Canad. Math. Bull.* **12** (1969), 745–752.

[GL95] R. L. Graham and B. D. Lubachevsky, Dense packings of equal disks in an equilateral triangle: from 22 to 34 and beyond, *The Electronic Journ. of Combinatorics* bf 2 (1995), #A1.

[GL96] R. L. Graham and B. D. Lubachevsky, Repeated patterns of dense packings of equal disks in a square, *The Electronic Journ. of Combinatorics* **3** (1996), #R16.

[GLNO98] R. L. Graham, B. D. Lubachevsky, K. J. Nurmela, and P. R. J. Östergård, Dense packings of congruent circles in a circle, *Discrete Mathematics*, **181** (1998), 139-154.

[G70] M. Goldberg, The packing of equal circles in a square, *Math. Mag.* **43** (1970), 24–30.

[L91] B. D. Lubachevsky, How to simulate billiards and similar systems, *J. Computational Physics* **94** (1991), 255–283.

$\{Z\}_0$ has degenerated in the figure to one dimension. $Z^{(1)}$ and $Z^{(2)}$ are the two configurations that yield the optima; $Z^{(2)}$ yields the global optimum. To make comparison more in favor of

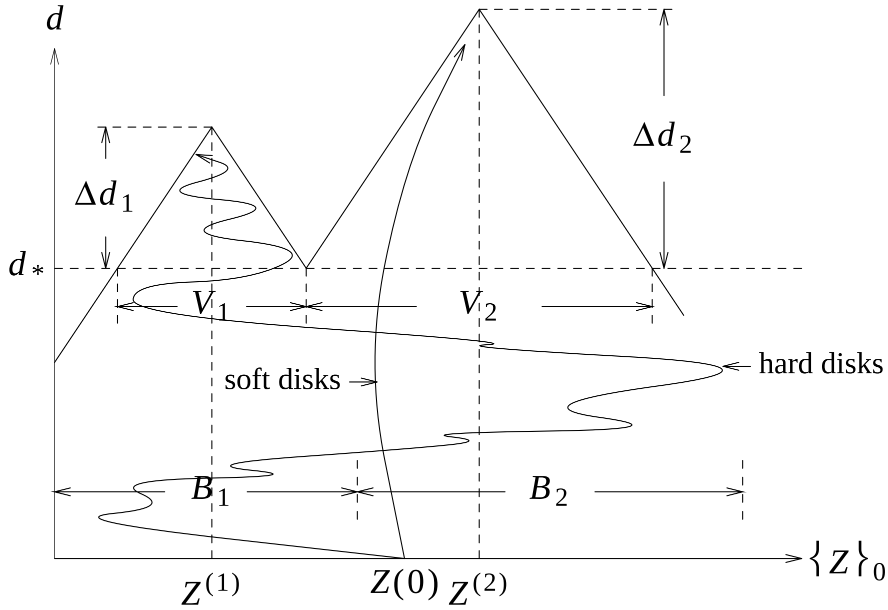

Figure 4.1: Comparing the chances of hard and soft disk trajectories to reach the optimum

the soft disk procedure, we ignore the fact that the latter finds optima only approximately. The initial configuration space $\{Z\}_0$ is partitioned into two basins $\{Z\}_0 = B_1 \bigcup B_2$ corresponding to attractors $Z^{(1)}$ and $Z^{(2)}$, respectively, in the gradient field of $P(Z)$. Depending on whether initial configuration $Z(0)$ belongs to B_1 or B_2 the final configuration will be $Z^{(1)}$ or $Z^{(2)}$, respectively. Under the uniform choice of $Z(0)$ the chance to get to the global optimum $Z^{(2)}$ is the fraction which the volume of B_2 constitutes in the total volume of $\{Z\}_0 = B_1 \bigcup B_2$.

Under the hard disk procedure, rather than taking into account the initial disk configuration at $d = 0$ we consider the configuration just before the time when the set of reachable configurations disconnects into two subsets, V_1 and V_2. This can be done because the memory of the initial configuration will be quickly erased along the ergodic trajectory. At the disconnect time we have $d = d_*$. The chance to get to the global optimum $Z^{(2)}$ is the fraction which the volume of V_2 constitutes in the union $\{Z\}_{d_*} = V_1 \bigcup V_2$.

In Figure 4.1 it is not clear why the hard disk trajectory is finding the global optimum easier than the soft disk trajectory: the fraction of volume of V_2 in $\{Z\}_{d_*}$ seems to be about the same as the fraction of volume of B_2 in $\{Z\}_0$. This illusion of equivalence is caused by $\{Z\}_0$ being one-dimensional. Suppose $\{Z\}_0$ is a k-dimensional space and like in Figure 4.1 let the remaining increase of diameter from the disconnect point d_* up to the optima be Δd_1 and Δd_2. We will also assume, that the topology of the approaches toward the optima are similar, e.g., as in Figure 4.1 the angles at the optima cones are equal. Then volumes of sets V_1 and V_2 would relate as $(\Delta d_1)^k$ to $(\Delta d_2)^k$. Considering high dimensionality k of the disk configuration Z, the volume of set V_1 on approach to the local optimum $Z^{(1)}$ will be overwhelmingly smaller than that of V_2 on approach to the global optimum $Z^{(2)}$ and so will be the chance for the trajectory to converge to $Z^{(1)}$ rather than to $Z^{(2)}$. The advantage of the hard disk trajectory will be higher in the cases when disconnect diameter d_* is larger. This conclusion is in perfect agreement with the where more than two disks come into contact. Across such manifolds the velocities of disks change discontinuously (as functions of initial conditions; as functions of time, the velocities are discontinuous even at binary collisions).

Note that the fact that the hard disk algorithm can only treat binary collisions and recognizes no triple collisions is not essential in the consideration above. (Triple collisions practically never occur anyway.) The essential were that when disturbed the triple collision splits into different sequences of binary collisions and that the after-collision disk velocity changes in a discontinuous fashion when the collision sequence thus changes.

At the final stages of the hard packing procedure, when the configuration is close to being tight (like in Figure 2.5), the "almost" triple and multiple collisions occur very frequently because the disks are very close to each other. The velocity discontinuity in the hard-packing procedure coupled with the roundoff noise creates intensive probing micro motions within the shrinking available configuration volume as the disk positions converge. The velocities of the disks do not converge, in particular, do not converge to zero. The disks continue to randomly vibrate, so to say. By contrast, at the final iterations of the soft packing procedure, the gradient in the steepest descent (analogous to the velocity in the hard packing procedure) converges to zero as the disk center positions converge.

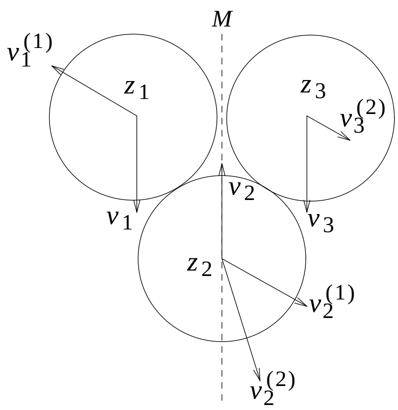

Figure 3.1: The outcome of the mirror symmetric triple collision is not mirror symmetric if treated as a sequence of two binary collisions

4. Advantages of the hard disk procedure

Suppose the expansion rate E in the hard disk procedure is very small and that the trajectory $Z(t)$ generated under such E is ergodic for sufficiently large t. The ergodicity means that given a measurable subset S in the available for a given disk diameter $d = d(t)$ connected configuration set $\{Z\}_d$ (which is also assumed to be measurable), at a random time t the configuration $Z(t)$ is found in S with probability equal to the fraction of volume of S in the total volume of $\{Z\}_d$. The ergodicity means, in particular, totally "forgetting" the initial configuration $Z(0)$.

Proving ergodicity in this situation as in many other cases is a very difficult task. We assume the ergodicity holds and see which advantage it gives to the procedure. Consider a very schematic rendering in Figure 4.1 where possible trajectories of the hard and soft disk procedures are drawn when there are two local optima. The high dimensional configuration space disks, like that in Figure 2.4, the soft packing procedure would not even produce approximation good enough to draw the "approximate" picture.

3. Implicit randomness of the hard disk method

As mentioned in the introduction, both hard and soft disk packing methods generate a configuration trajectory, which can be formally considered as deterministic. If the computations were performed with the infinite precision, we could in both cases start with the configuration Z resulted after some amount of processing, and by reversing the computations, obtain the initial configuration Z_0 (for example, the disk collision diagram in Figure 2.2 is time reversible). Real computations are subject to the roundoff noise which makes the trajectories non-reversible and non-deterministic. We argue, that the amounts of the non-determinism are substantially different in these two cases. While the soft method generates essentially a deterministic trajectory, which "remembers" the initial state, we argue that the trajectories generated by hard method are ergodic, with many probing motions, virtually random; the memory of the initial state is lost very rapidly.

An indication of the difference is in the continuity of the trajectory with respect to the initial condition. First consider the soft packing. Let $Z(0)$ be initial configuration of disks and $Z(t)$ be the configuration at time (or after step) t. Potential $P(Z)$ have the only singularities when $z_i = z_j$ for $i \neq j$ and it is usually assured that the trajectory $\{Z(\tau), 0 \leq \tau \leq t\}$ (or sequence of configurations $Z(0), Z(1), ...Z(t)$) stays well separated from the singularities. Because of smoothness of the potential, $Z(t+1)$ which results from $Z(t)$ in a steepest descent or another gradient method continuously depends on $Z(t)$. Hence given t and $\epsilon > 0$ we can always find sufficiently small $\delta > 0$ such that if we disturb $Z(0)$ within the δ-neighborhood of its non-disturbed value, then $Z(t)$ will stay within the ϵ-neighborhood of its non-disturbed value. In other words, a finite segment of the trajectory of the soft packing algorithm is continuous with respect to the initial conditions.

Under the hard packing algorithm such continuity generally does not hold. To see that consider Figure 3.1, which depicts a triple collision. Three disks, 1, 2, and 3 simultaneously arrive at the corresponding positions z_1, z_2, and z_3 with the pre-collision vector velocities v_1, v_2, and v_3, respectively. At the collision time t_*, disks 1 and 2 are in contact, as well as disks 2 and 3, but not disks 1 and 3. The pre-collision position $Z \doteq \{z_1, z_2, z_3\}$ and velocity $V \doteq \{v_1, v_2, v_3\}$ configurations are mirror symmetric with respect to the middle vertical line M. There are two possible orders of processing the triple collision. In one order, disks 1 and 2 collide first and obtain new velocities $v_1^{(1)}$ and $v_2^{(1)}$, and then disks 2 and 3 collide and obtain new velocities $v_2^{(2)}$ and $v_3^{(2)}$. The initial velocity of disk 2 for the second pairwise collision is $v_2^{(1)}$ as if the second collision occurred later than the first one. (The after-collision velocities can be obtained using the method depicted in Figure 2.2.) As a result of this order of processing, the after-collision velocity configuration V^{new} is *not* mirror-symmetric. If 2 and 3 collide first and then 1 and 2 collide, the conclusion is the same, but the V^{new} resulted in the latter order of processing is the mirror-image of the V^{new} under the former order.

Because of the non-symmetry, it is clear that if we make a δ-small disturbance to the initial position configuration Z so that, in particular, pairs 1-2 and 2-3 do not come to their collision sites simultaneously, then, no matter how small δ, the configurations at finite time $t > t_*$ will be substantially different. Thus, continuity of the dependence of the trajectory with respect to initial condition is broken in the configuration space $\{Z\}$ across the manifolds mented with a "tightening" procedure. The input to the latter includes a list of pairs $\{(i,j)\}$ of disks, such that $|z_i - z_j| = d$, and a list $\{i\}$ of disk indices, such that $|z_i| = 1$, in other words, the lists of contacts. Those contacts should be derived from examining numerically the steady-state configuration Z. The contacts are written as equations and an iterative procedure is set to find a solution to this usually overdefined system. The approximate configuration Z is used as a starting point for the iterations. Disk diameter does not enter the equations as a known value but if a solution is found, the diameter can be derived thereof. The tightening procedure is implemented with an arbitrary precision; for example, it can be run with 100 decimal digits. Considering such a precision, if the procedure numerically converges, we conclude with a justifiable confidence that the configuration exists, in particular, all contacts are guessed correctly.

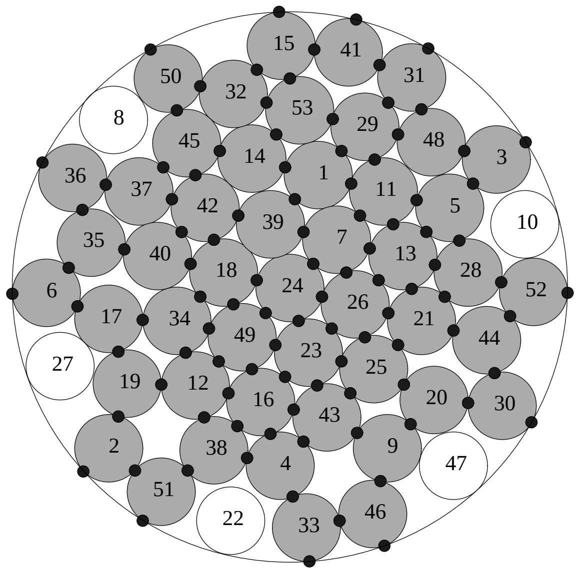

Figure 2.5: The best found packing of 53 equal disks in a circle. All contacts are marked with black dots. An absence of a dot indicates a positive gap.

For a large number of disks it may be difficult to guess the contacts, though. Figure 2.5 presents a case in point. The scale and resolution of the picture are not sufficient for correct identification of the contacts. Numerical examining of the configuration obtained under the hard disk algorithm allows one to split the suspected contacts into sets A and B. In A, a gap between the disks in a pair or the boundary and a disk does not exceed 10^{-12} fraction of a disk diameter, d. The smallest gap in B is between disks 34 and 40 and it is equal $d \cdot 6.5.. \cdot 10^{-8}$.

This gives a clear identification of the contacts for the tightening procedure. After the procedure converges, the resulted configuration has all the contacts (set A) confirmed as such to within $d \cdot 10^{-98}$. All gaps that look like contacts (set B) are confirmed as gaps and the sizes of the gaps are also confirmed. By contrast, the soft disk packing, although it generates a similarly looking picture (except for contact identification), does not separate clearly the contacts from the gaps.

One can apply the tightening procedure for each possible subset A chosen as guessed contacts until the procedure converges thereby confirming the guess. In the given case there are 2^{115} ways to choose set A, because $|A \bigcup B| = 115$. Too many to try! That is how the soft packing method fails here in favor of the hard packing method. For much larger number of

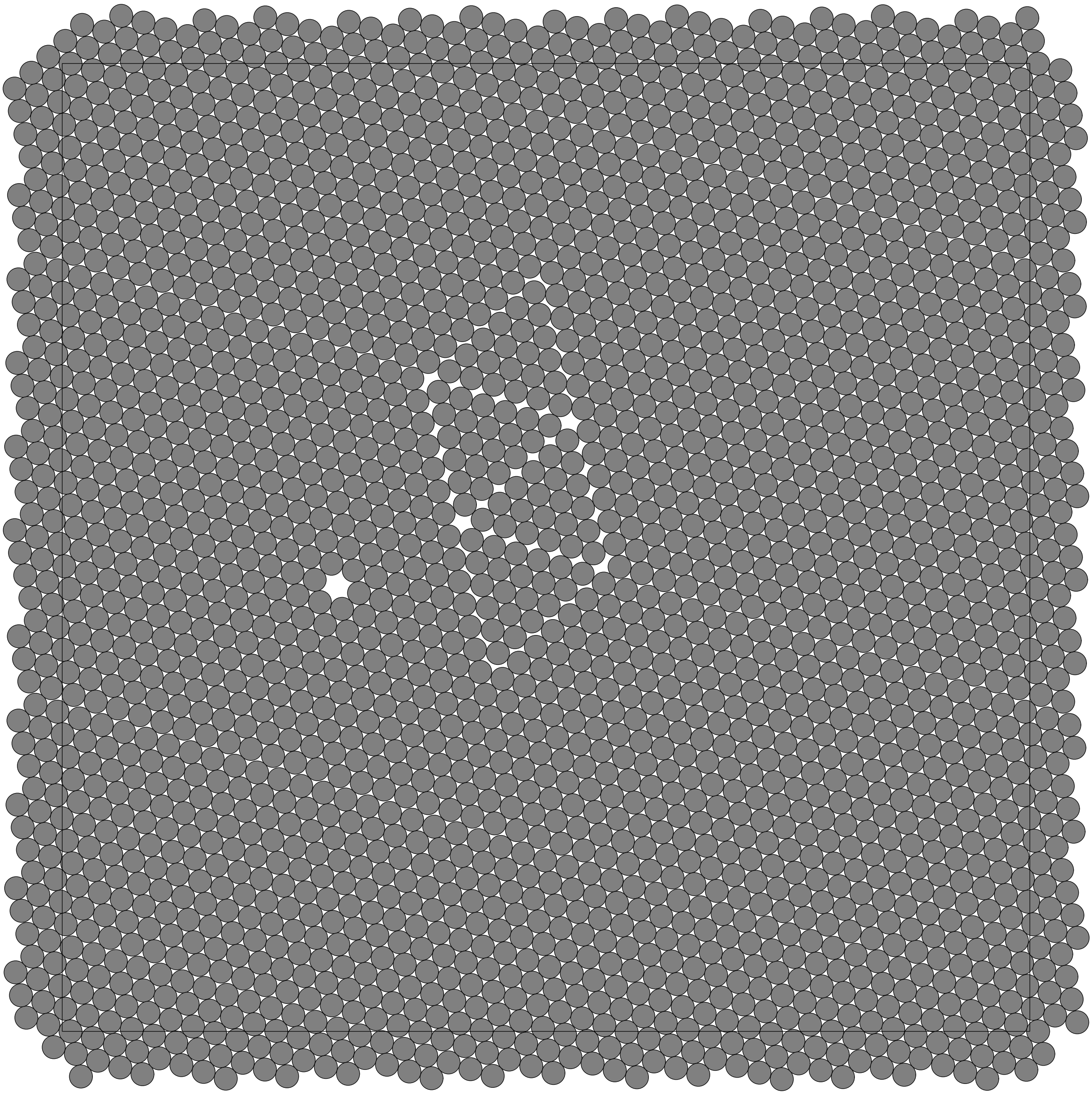

Figure 2.4: Another packing of 2000 disks in a square with periodic boundary obtained under a slow disk expansion, $E = 0.001$. If the monovacancy near the center is filled with the 2001-st disk, the obtained packing seemingly becomes perfectly symmetric. Might that be the best packing of 2001 equal disks in a square with periodic boundary? Its experimentally computed density (when the 2001-th disk is inserted) is 0.901635...

For a fixed m, a steady-state of this motion (a Z such that the sum of repulsion forces is zero for each disk center; such Z is also a local minimum of $P(Z)$) only approximates a local maximum of $d(Z)$ of the original packing problem. One needs to increase m in the course of the system evolution to "harden" the disks. In the limit $m = \infty$ local minima of $P(Z)$ are the same as local maxima of $d(Z)$. The "hardened" disk diameter, when $m \to \infty$, is λ, so the λ must increase together with m or be guessed sufficiently close to the optimum value of d.

Because of its approximate nature, the soft disk packing procedure in [GLNO98] is compli-

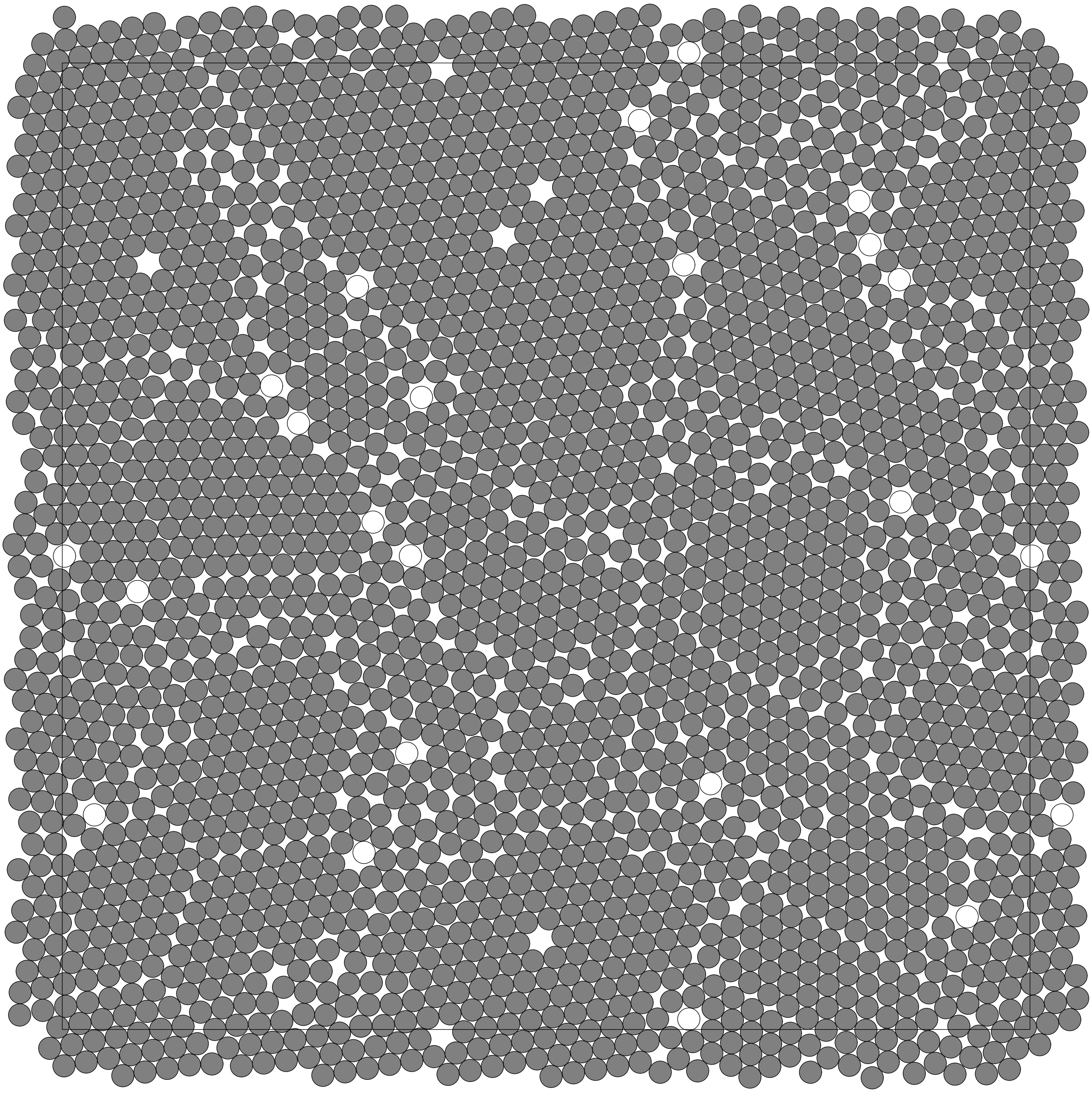

Figure 2.3: A packing of 2000 disks in a square with periodic boundary. The packing is obtained under a fast disk expansion, $E = 100$. The packing consists of crystalline grains with many rattlers (non-constrained disks; they are unshaded) concentrated along the grain boundaries. Monovacancies occur within the hexagonally packed grains.

consider the boundary, which can be introduced in several ways, for example by adding in (1) components that repel disk centers from approaching and penetrating it. Potential $P(Z)$ gives rise to a certain gradient procedure of descent which results in a sequence $Z(0), Z(1), Z(2)...$ of configurations that can be roughly thought of as representing a deterministic motion of the configuration vector Z under the repulsion forces generated by the potential.

proportionally change both the disk expansion speed and linear disk velocities. We normalize the simulation input by assuming the unity mean initial disk velocity. Thus E becomes the rate of disk expansion in relation to the mean disk velocity.

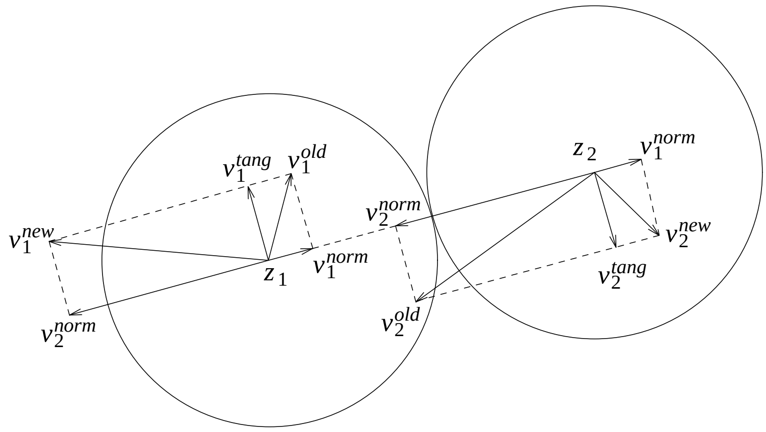

Figure 2.2: Change of velocities of two disks at their collision

Figures 2.3 and 2.4 present two out of many packings thus obtained. The resulting structure depends on many factors, including initial disk center positions and initial velocities. By far, the most important parameter is the disk expansion rate E. For smaller E, final configurations are more regular and have larger final disk diameter d and density, defined as fraction of the region area which is covered by disks. Experiments show that as $E \to 0$, while other parameters stay the same, the chance to obtain the best packing increases. But smaller E also means longer computing. For example, the packing in Figure 2.4 of 2001 disks obtained under a very slow disk expansion $E = 0.001$ (with one extra disk placed into the vacancy near the packing center) seems so symmetric, that one may conjecture its optimality.

While the original problem is that of packing hard disks, most researchers prefer to treat disks as soft ones, wherein the potential imposed is such that a disk center "feels" the presence of other centers at distances larger than those at which disks touch each other. To be specific, consider disk packing in a circle. It is convenient, as is done in [GLNO98], to substitute the original problem of densest packing of n equal disks in a circle of a given radius with the problem of sparsest spread of n points inside a circle of unity radius centered at the origin. In the latter problem we are required to find configuration Z of n points on the plane, $Z = \{z_1, z_2, ...z_n\}$, so that $|z_i| \leq 1$ and the quantity $d = d(Z) \doteq min_{i \neq j}|z_i - z_j|$ is as large as possible. The two formulations are equivalent: a sparsest spread Z can be also considered as a densest packing. The configuration of disk centers in the packing would coincide with Z, the disk diameter would be $d(Z)$, and the enclosing circular region would be centered at the origin and would have diameter $2 + d(Z)$.

Thus, in a setting where the optimal spread is sought, [GLNO98] proposes potential

$$P(Z) \doteq \sum_{1 \leq i < j \leq n} \left(\frac{\lambda}{|z_i - z_j|} \right)^{2m} \tag{1}$$

where $\lambda > 0$ is a scaling factor and $m > 0$ controls the speed at which interdisk repulsion forces diminish with distance, smaller m softer the disks. For simplicity of the discussion, we do not

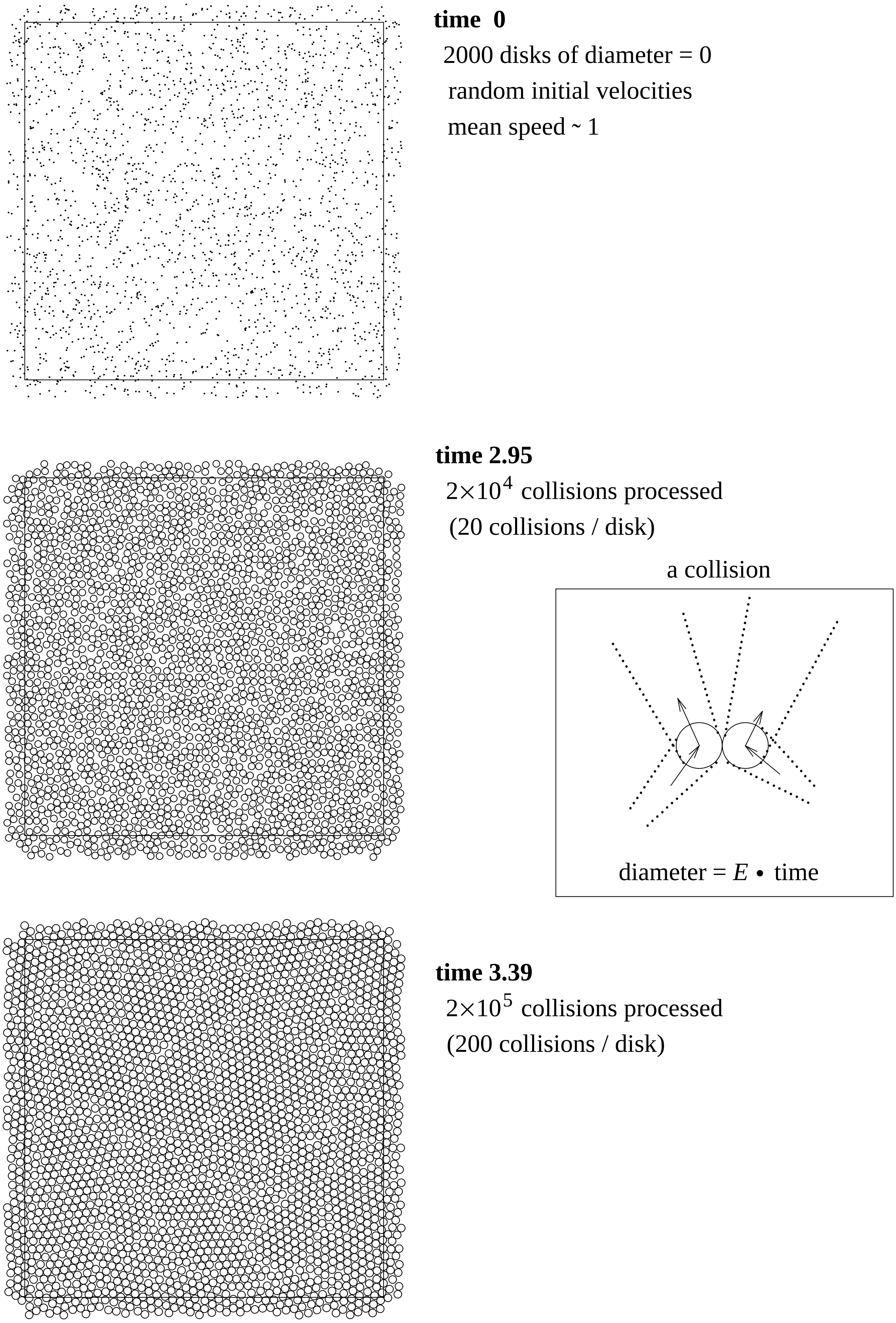

Figure 2.1: Successive stages in an instance of disk packing

space available while being subject to dynamic interactions with a certain force potential. The potential is chosen so that disk centers repel each other thereby expressing the optimization objective. Depending on the form of the potential the procedures can be split into two classes: packing soft disks and packing hard disks.

Hard disks pack easier than soft ones. The resulting configurations are consistently better: for values of n when both methods work, hard packings deliver larger d. In addition, hard disk method delivers a stable packing for substantially larger values of n than its soft counterpart, which gets stuck for such n. This note tries to explain this experimental observation.

It is argued below that while both methods generate formally deterministic disk trajectories, an operating factor of the hard packing method is its implicit randomization. The trajectory generated by the hard packing is ergodic. This becomes the method's key advantage over its soft packing counterpart when the global optimum is hidden among multiple close-by local optima. It should be noted that the issues discussed involve hard mathematical problems. A frontal attack on those problems does not seem feasible, so the discussion below is mostly heuristic, based on plausible assumptions and on examples, drawn from experience.

2. Hard and soft packings in comparison

Figure 2.1 illustrates the work of a hard disk algorithm while packing 2000 equal disks in a square with periodic boundary (torus). The initial stage at time $t = 0$ is represented on the top square in the figure where 2000 points are randomly scattered. To each point an initial velocity vector is randomly assigned (not shown). Some points lie outside the square; they are periodic images of the corresponding points inside. When $t > 0$, the points grow into disks, and all disks at each time t have common diameter $d = Et$, where $E > 0$ is a fixed expansion parameter. The growth continues until the configuration "jams," at which time one expects to have a packing.

At $t = 0$ disks do not overlap because their sizes are zero. By definition, the hard disk potential causes no force between two disks with centers z_1 and z_2 if $|z_1 - z_2| > d$. Thus, at $t > 0$, for as long as the distance from the disk to any other disk remains larger than d, the disk moves along a straight line with a constant velocity. At some time $t_* > 0$ the first contact occurs: $|z_1 - z_2| = d$ for some disks 1 and 2. An *infinite* repelling force is generated between the two disks once $|z_1 - z_2| \leq d$. As a result, the time duration is zero when the disks are staying in contact. During the contact an ideal elastic collision of two particles of equal masses is simulated. Both disks instantaneously change their velocity vectors so as to conserve total momentum and energy according to the rule of elastic collision.

The insert in Figure 2.1 shows such a collision. The details of velocity change for a pairwise disk collisions are shown in Figure 2.2 in the case of expansion rate $E = 0$. For $i = 1, 2$ we have decompositions $v_i^{old} = v_i^{tang} + v_i^{norm}$ and $v_i^{new} = v_i^{tang} + v_{3-i}^{norm}$ of the old and new vector velocities, respectively. The velocities are represented as the sums of the normal and tangential components with respect to the tangential line of impact. The exchange of normal velocity components occur: the new normal velocity component of disk i is the old normal velocity component of disk $3 - i$. It can be easily shown (and is seen in Figure 2.2) that at $t > t_*$ the two disks are moving away from each other. When $E > 0$, normal components of the after-collision velocities are augmented, so as to assure the non-overlap. This motion continues until the next collision of at least one of the two participants, at which time a new velocity exchange occurs and so on. Note that the evolution of the disk configuration does not change if we

Why Hard Disks Pack Easier

Boris D. Lubachevsky
bdl@bell-labs.com
Bell Labs Innovations for Lucent Technologies
Murray Hill, New Jersey 07974, USA

Abstract

Two methods for generating dense disk packings inside 2D containers are compared in their efficiency. Both methods simulate a dynamic system of moving disks with disks being subject to a repulsion force potential. Depending on the form of the potential, the disks are treated as hard particles in one method, or soft ones in the other. Whereas the simulated trajectory of disk configuration is formally deterministic in both methods, it is argued that under the hard disk potential, the trajectory is ergodic as if the disk motion was random, which creates an advantage over the truly deterministic method under the soft disk potential. The hard disk method is experimentally shown to be able to discern a delicate distinction of the global optimum from multiple close-by local optima whereas the soft disk method usually gets stuck when the global optimum is hidden among multiple local optima.

Key words: dense disk packing, optimization, hard spheres, steepest descent, ergodic

1. Introduction

Consider the task of fitting without overlaps a given number n of equal diameter disks entirely inside a given 2D container so that the disks have the largest diameter possible. A "recreational" simplicity of the formulation of the optimal disk packing problem hides a mathematical challenge. Over the years the problem has attracted a substantial attention of mathematicians (see, e.g., [FG69] [G70] [GMPW91] [O61] [S71] [S79] [V89]). In physics and material science, packings of disks (and spheres in 3D) have been advanced as representing the short-range atomic order that exists in solids or the arrangements of particles adsorbed on smooth surfaces and colloidal suspensions [LS90]. There are also connections between optimal packings and optimal codes [CS93] [GS90].

Mathematicians tried to improve packing records using paper, pencil, and perhaps, round coins, as the primary tools. Because the difficulty of the task increases with n, but the power and resolution of manual and mechanical methods (one of such is described in [R75]) are bounded, the stream of new published disk packings ceased in about mid '70s. The values of n in most cases were below $n = 20$. Recently many new interesting disk packings have been discovered for substantially larger n. The recent surge of disk packing activity is based on computer procedures [MS95] [M97] [NO96] [MFP95] [GL96].

Simulating a dynamic system of moving disks whose steady states correspond to disk packings is a popular computing method for solving the problem. Such a procedure starts with a random placement of n disk centers inside the container. Then the disks negotiate the limited