4. BALL-TYPE MODELS

4.1 Introduction

We shall start the introduction of the different types of DEM techniques with the BALL-type models which are definitely the most widespread and popular discrete element techniques. The name "BALL" was used for the software introduced by Cundall & Strack (1979). This was the first software which was widely applied for the simulation of granular mechanics problems. Though the first code of Peter A. Cundall was UDEC (see Section 5), he became world-famous for BALL.

BALL, and those several similar codes inspired by BALL, applied strictly circular elements in those first times. Towards the end of 1980ies the three-dimensional version TRUBAL (using spherical elements) was prepared by Cundall, who provided the two codes free to anyone who wanted to use, and several scientists all over the world took advantage of this possibility. The quick development of granular mechanics in the 1980ies and 1990ies was partly due to this.

BALL and TRUBAL are so simple from mechanical point of view that anyone with a little experience in computer programming is able to prepare a BALL-type code (of course, a user-friendly input and output system already needs some expertise). No wonder that several similar codes were born, differing from each other only in the element shapes, the constitutive models for the contacts the user is offered, and in the applicable boundaries. However, the principles of the mechanical and numerical modelling are the same. These common principles will be introduced in this Section.

In the 1990ies the commercial versions of BALL and TRUBAL, named PFC-2D and PFC-3D ("Particle Flow Code") were born. Their input and output systems were much more convenient, and – which is even more important – an option to "glue together" circles/sphere into irregularly shaped particles was introduced. As already mentioned above, other researchers developed different codes (mostly research softwares) with particle shapes like ellipses, ovals/ovoids, elements composed of cylinders and parts of spheres etc.

However, there are a few characteristics shared by all BALL-type models, so they are the criteria to consider a method a BALL-type model:

- they are time-stepping methods, and within this, they apply the method of central differences for the time integration,
- the elements are perfectly rigid,
- the elements form point-like, deformable contacts with each other. The wide variety of contact models (e.g. cohesional, Coulomb-type frictional, concrete-like, linear, Hertzian, arbitrary user-defined etc) allow the users to simulate a large range of different granular materials.

4.2 The elements

In the BALL-type models the perfectly rigid elements can have any shape which satisfy the condition that the contacts occur between separate, discrete points, and not along extended straight or planar surfaces. In the free software OVAL (Kuhn, 2003) for instance, the

elements are composed of parts of toroids and spheres. Ellipses are applied in the 2D model of Ting (Ting et al, 2003) and in (Ng, 2001). Hopkins (1996) uses elements consisting of cylinders and parts of spheres etc. Today's granular mechanics literature is rich in models applying very realistic particle shapes consisting of dozens of spheres and/or other elementary shapes. Figure 1. shows 2D models of sand (Matsushima and Saomoto, 2002):

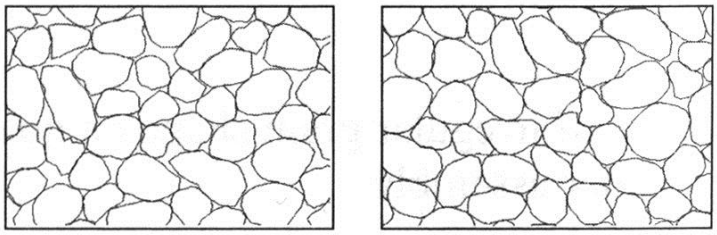

Figure 1.
Simulated "sand particles" of Matsushima and Saomoto

Irrespectively of how complicated their shape is, rigid elements in 3D have the same six degrees of freedom: their reference point translates (3 components) and the element rotates about the reference point (3 components). All existing models use the center of mass for the reference point. The displacement vector of element p is:

$$\mathbf{u}^p(t) = \begin{bmatrix} u_x^p(t) \\ u_y^p(t) \\ u_z^p(t) \\ \varphi_x^p(t) \\ \varphi_y^p(t) \\ \varphi_z^p(t) \end{bmatrix},$$

which, if being collected for all elements of the system, can be summarized into a hypervector consisting of as many 6-scalar blocks as N, the number of elements in the system:

$$\mathbf{u}(t) = \begin{bmatrix} \mathbf{u}^1(t) \\ \mathbf{u}^2(t) \\ \vdots \\ \mathbf{u}^N(t) \end{bmatrix}$$

These are the basic unknowns which should be determined being accumulated along a series of small Δt time steps.

4.3 The contacts

4.3.1 Description of the contact state

Contacts in the BALL-type models are born when two elements intersect with each other. Apart from unrealistic cases, the common (intersected) domain is small in comparison to the size of the elements, so the contacts can be considered point-like. A concentrated force and perhaps a concentrated moment can be transmitted in this point between the two elements.

To every contact, e.g. to contact c formed by the elements p and q, assign a local coordinate frame (n, t, w). The axis n is the normal direction of the common tangent plane of

the two surfaces at the contact point. Considering *p* as the first and *q* as the second element forming the contact, *n* points outwards of element *p*. The axes *t* and *w* are perpendicular to each other and they are inside the common tangent plane. It is important to emphasize that as the elements move, the local frame changes.

In this coordinate frame the force and moment acting on the first element expressed by the second element can be summarized into a single vector:

$$\mathbf{S}^c = \begin{bmatrix} N^c \\ T_t^c \\ T_w^c \\ M_n^c \\ M_t^c \\ M_w^c \end{bmatrix} .$$

N^c denotes the normal force in the contact (positive if tension, negative if compression), T_t^c and T_w^c are the two components of the tangentional force, M_n^c is the twisting moment between the two elements (rotating in the tangent plane), finally M_t^c and M_w^c are the two components of the bending moment (they are perpendicular to each other and both of them lie in the tangent plane).

The increment of the relative displacement of the two small material points forming the contact can be written in the local frame of the contact as

$$d\boldsymbol{\delta}^c = \begin{bmatrix} d\delta_n^c \\ d\delta_t^c \\ d\delta_w^c \\ d\theta_n^c \\ d\theta_t^c \\ d\theta_w^c \end{bmatrix} = \begin{bmatrix} du_n^{qc} \\ du_t^{qc} \\ du_w^{qc} \\ d\varphi_n^{qc} \\ d\varphi_t^{qc} \\ d\varphi_w^{qc} \end{bmatrix} - \begin{bmatrix} du_n^{pc} \\ du_t^{pc} \\ du_w^{pc} \\ d\varphi_n^{pc} \\ d\varphi_t^{pc} \\ d\varphi_w^{pc} \end{bmatrix} .$$

Here the first three components are the increments of the relative translations in the normal and in the two tangential directions, while the last three components are the increments of the relative rotations around *n*, *t* and *w* (in the twisting and in the two rolling directions).

4.3.2 Constitutive relations

The constitutive relations of a contact describe how the contact forces and moments depend on the relative displacements, they give limitations to the contact forces ad moments which can be resisted by the contact, and they may also specify how the contact behaves after these limits are reached (e.g. frictional sliding, plastic deformations etc). According to the logics of DEM, after the displacement increments during the actual timestep are determined, from them the contact forces and moments are calculated afterwards. So the basic role of the constitutive relations is to provide a calculation recipe to determine the actual contact forces from the displacement history of the contact.

The simplest and most widely applied contact model is the ***linearly elastic*** contact with ***Coulomb friction***. In this case the normal force can only be compression, and its magnitude is proportional to the *n*-directional "overlap" between the two elements:

$$N^c = k_N^c \delta_n^c ; \quad N^c \le 0$$

It is also valid in an incremental sense (with no limitation on the sign of the force increment):

$$dN^c = k_N^c d\delta_n^c \ .$$

The normal force cannot become positive: when the overlap disappears, the contact does not exist anymore.

The tangential force and the tangential relative translation are related to each other in an incremental sense, and a limit depending on the compression is given for the magnitude of the tangential force:

$$\begin{bmatrix} dT_t^c \\ dT_w^c \end{bmatrix} = k_T^c \begin{bmatrix} d\delta_t^c \\ d\delta_w^c \end{bmatrix}; \quad \sqrt{\left(T_t^c\right)^2 + \left(T_w^c\right)^2} \le -\boldsymbol{\nu} \cdot N^c$$

When the tangential force reaches this value, the relative tangential translations increase beside a constant tangential force (i.e. with zero stiffness), as long as the direction of relative translations do not change.

Another widely applied contact model is the Mindlin-Hertz-approximation. The well-known Hertz theory (Hertz, 1881) describes the compressional behaviour of two spheres. For equal radii, for instance, the stiffness of the contact depends on the normal force in the following way:

$$k_N = \frac{3}{2} \frac{\left(R \cdot N\right)^{1/3}}{\left(\frac{3(1-\boldsymbol{\mu})^2}{4E}\right)^{2/3}}$$

(here μ is the Poisson-coefficient and *E* is the Young-modulus of the material of the spheres). The meaning of this formula is illustrated in Figure 2. Similar, though of course more complicated, relations apply to a few other cases like unequal spheres; otherwise the above formula is a basis for approximative relations.

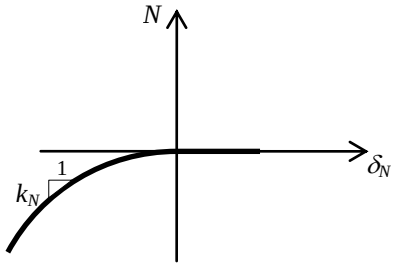

Figure 2.
The Hertzian relation between
contact normal force and normal deformation

Based on this, during a small Δt time step the increment of the compression force and the increment of relative translation in normal direction are related to each other as

$$\Delta N^c = k_N^c \Delta \delta_n^c \; ; \quad N^c \le 0 \; .$$

The tangentional behaviour can be approximated based on the theories proposed by Cattaneo (1938), Mindlin (1949) and then completed by Mindlin és Deresiewicz (1953). The theory describes the case of two equal, linearly elastic – perfectly plastic spheres compressed into each other, and gradually displaced along each other in the tangentional direction. The validity of the very complicated formulas is rather limited, so in those practical cases like unequal spheres, non-spherical elements in contact etc. approximations are derived from them. In general, the application of these simplified approximations is computationally rather expensive.

Cohesion means that the normal force can be positive (i.e. tension): the contact survives if the two elements diverge from each other, until the contact gets to the breakage limit. Cohesion is usually accompanied by linearly elastic force-deformation-law both I normal and in tangentional directions in the different DEM codes. Since a cohesional contact suffers relatively small deformations until being broken, it is not necessary to make a difference between incremental and accumulated force-displacement relations. The constitutive relations are:

$$N^c = k_N^c \delta_n^c \; ; \quad N^c \le N_{\max}$$
$$\begin{bmatrix} T_t^c \\ T_w^c \end{bmatrix} = k_T^c \begin{bmatrix} \delta_t^c \\ \delta_w^c \end{bmatrix} ; \quad \sqrt{\left(T_t^c\right)^2 + \left(T_w^c\right)^2} \le T_{\max}$$

The contact breaks either of the tension exceeds the prescribed limit, $N_{\max}$, or the shear force reaches $T_{\max}$. (In the latter case in most DEM models the contact becomes frictional, but cohesionless.)

Concrete, granular rocks etc. are granular materials where the contacts between the grains resist relative rotations in addition to relative translations, which means that moments are transmitted between the grains. To simulate such materials with DEM, the constitutive model of the contacts have to express how the twisting and bending moments depend on the relative rotations. Usually this is done with the help of constant rotational stiffnesses. The constitutive relations are completed with the limit moments which can be resisted by the contact.

Several other types of contact behaviour can be included in a DEM code (e.g. time dependent, plastic, loosing etc. models). In some codes (e.g. in PFC or in YADE) the user can prepare his or her own constitutive model. Such an option is very useful both for researchers and for practicing engineers.

4.4 The equations of motion and their time integration

BALL-type models use the equations of motion shown in Section 2. for rigid elements:

$$\mathbf{M}^p(t)\mathbf{a}^p(t) = \mathbf{f}^p(t, \mathbf{u}(t), \mathbf{v}(t)) \quad .$$

This equation consists 6 scalar equations for element p (3 for translational acceleration components and 3 for rotational acceleration). According to the method of central differences (see Section 3), the calculation of a (t_i, t_{i+1}) interval whose length is Δt is based on the following data, which have to be known when starting the calculation of the time step:

- → position of every element at t_i , and the displacement increments during the last time step;
- → average velocities belonging to the last time step, $\mathbf{v}_{i\text{-}1/2}$-et,
- → external forces acting at t_i.

From the positions, displacements and contact constitutive relations, the actual contact forces (i.e. those belonging to t_i) are determined. Then the actual contact forces all the other forces are reduced to the reference points. For element p the equations of motion can be discretized as

$$\mathbf{M}_i^p \frac{\mathbf{v}_{i+1/2}^p - \mathbf{v}_{i-1/2}^p}{\Delta t} = \mathbf{f}_i^p$$

or as

$$\mathbf{v}_{i+1/2}^p = \mathbf{v}_{i-1/2}^p + \Delta t \cdot \left(\mathbf{M}_i^p\right)^{-1} \mathbf{f}_i^p \quad .$$

All quantities on the right are known, so the average velocities belonging to the analysed interval can be calculated, and from them, the new position of the element is received:

$$\mathbf{u}_{i+1}^p = \mathbf{u}_i^p + \Delta t \cdot \mathbf{v}_{i+1/2}^p \quad .$$

Then a new time step can follow.

It is important to emphasize that the calculations do not take into account that within a single time step the contact forces change. Velocities in the analysed interval are calculated from the accelerations valid at t_i. However, when an element displaces, the deformations in its contacts will change, and it modifies the contact forces. This effect, and any other changes of forces acting on the elements, is expressed by the stiffness matrix of the system (i.e. the Jacobian matrix of the reduced force vector) in the different displacement methods of structural analysis. Since the stiffness matrix is missing from the equations of motion of BALL-type models, the calculations may significantly over-estimate the displacement increments during the time step. (The larger is Δt, the more significant is the over-estimation.) The too large displacement increments cause too large internal forces against the displacements, and these too large forces too strongly push back the elements in the next time step. The calculated behaviour oscillates around the exact solution. When the exact solution is, for instance, an equilibrated state, the calculated results vibrates around the equilibrium state. A too large timestep length may even cause a numerical breakdown.

To decrease this problem, BALL-type models set a limitation to the length of the applicable time step, and the different commercial codes usually do not allow the user to arbitrarily choose Δt. The maximally allowed length can be estimated with the help of the largest eigenfrequency of the individual elements:

$$\Delta t \le \min_{(p)} \left\{ \min \left(\sqrt{\frac{m^p}{k_{trans}^p}}, \sqrt{\frac{I^p}{k_{rot}^p}} \right) \right\} \quad .$$

This expression is based on the numerical stability condition of the central difference scheme: in the case of a perfectly elastic system the time step should not exceed the value T/π where T is the minimum eigenperiod of the whole system (Bathe and Wilson, 1976). A BALL-type system of discrete elements is, however, not necessarily elastic – the contacts may frictionally slide, and different damping effects may also significantly influence the behaviour; in addition, even in the very unusual perfectly elastic case the large number of elements would make the exact determination of the eigenperiod a rather time-consuming task. Hence, Cundall proposed to find a critical timestep length separately for every degrees of freedom of every particle, as if the each particle would be supported against all possible individual motions by elastic springs attached to its neighbours being perfectly fixed. Then take the smallest value of these critical timestep lengths. So, in the above formula the stiffnesses k^p_{trans} and k^p_{rot} denote the translational and rotational stiffnesses of element p (these can be compiled from the contact stiffnesses), and I^p is the rotational inertia belonging to the direction of largest rotational resistance.

4.5 Damping

A system of perfectly rigid, undeformable discrete elements with elastic-frictional contacts dissipates energy through frictional sliding only. In physical reality, however, the situation is different: colliding particles are damaged and yielded in the small neighbourhood of the newborn contacts, and even the already existing contacts yield (hence dissipate energy) with increasing compression. In addition, when the grains float in some kind of a liquid, velocity-dependent drag forces dissipate the kinetic energy of the grains.

In addition to the effort to simulate these effects, in the BALL-type models damping has another role: it should help the convergence of the explicit solver, and ensure that when the model simulates a system converging to equilibrium, the model should also arrive to equilibrium instead of oscillating around it. So the role of damping in a BALL-type model is twofold: it should overcome the non-physical nature of contact models, and at the same time, it should stabilize the time integration.

O'Sullivan (2011) gives an extended overview on most usual types of damping; we shall focus now on the two most important types in BALL-type models, called ***local damping*** and ***contact viscous damping***.

Local damping means adding a damping-force term to the right side of the equations of motion: a force vector whose components point opposite to the velocity components is added to the reduced force vector. Each component is α–times the corresponding component of the reduced force vector.

As an example, consider the x-component of the translation of element p. Without damping, the equation of motion in the central difference scheme is

$$v^p_{x,i+1/2} = v^p_{x,i-1/2} + \Delta t \cdot \frac{1}{m^p} f^p_{x,i} \ .$$

Local damping modifies this equation in the following way:

$$v^p_{x,i+1/2} = v^p_{x,i-1/2} + \Delta t \cdot \frac{1}{m^p} \left(f^p_{x,i} - \alpha \cdot \left| f^p_{x,i} \right| \cdot \frac{v^p_{x,i-1/2}}{\left| v^p_{x,i-1/2} \right|} \right) \ .$$

Another example is the last scalar equation of motion of element *p* (i.e. the one describing the rotational acceleration about axis *z*):

$$\omega^p_{z,i+1/2} = \omega^p_{z,i-1/2} + \Delta t \cdot \frac{1}{I^p_z} m^p_{z,i} \ ,$$

which becomes:

$$\omega^p_{z,i+1/2} = \omega^p_{z,i-1/2} + \Delta t \cdot \frac{1}{I^p_z} \left(m^p_{z,i} - \alpha \cdot \left| m^p_{z,i} \right| \cdot \frac{\omega^p_{z,i-1/2}}{\left| \omega^p_{z,i-1/2} \right|} \right) \ .$$

The value of the coefficient α is optional; a usual value is, for instance, 0.70. (It means that the every non-zero component of the reduced load vector is cut back by 70%, hence the accelerations will be less by 70% than without local damping.)

Local damping has several interesting and advantageous features (see, for instance, O'Sullivan, 2011 or Itasca, 2008). Only accelerating motions are influenced: steady-state solutions are not affected, so – as an example – an equilibrated flow with constant velocity can be correctly simulated, and this kind of damping does not distort the flow kinematics. Those parts of the system are damped most which have the largest equilibrium errors (hence the name "local"). In addition, the constant α is a non-dimensional proportion, which means that the user should not worry about the scaling of α.

Contact viscous damping (e.g. Itasca, 2008) can be imagined as small dashpots being placed at each component of the contact forces. These dashpots act in parallel with the existing contact model. Numerically it means that a damping force is added to the contact force, of which the magnitude of the normal component is, for instance, given by

$$\left| D^c_N \right| = c_N \cdot v^c_{N,i-1/2} \ .$$

Here c_N is the damping constant (*N* stands for "normal"), and $v_N{}^c{}_{,i\text{-}1/2}$ is the normal component of the relative velocity at contact *c*. This kind of damping is similar to physical reality, to the real way of energy dissipation in the contacts of touching grains.

When viscous damping is used in a dynamic problem, such as a simulation of bouncing particles, appropriate viscous damping constants should be specified for the simulation to reproduce a realistic response. In a quasi-static problem the a damping constants may be chosen such that the calculation converges to equilibrium more quickly. However, the user should check that the influence of viscous damping on the results is acceptable, e.g. by comparing the results gained with different damping coefficients (Itasca, 2008).

4.6 Applications

BALL-type models played a fundamental role in the science of granular mechanics; their importance cannot be overestimated. They have been also applied in practice-oriented problems like pharmaceutical industry (e.g. Foo et al, 2004), soil mechanics (e.g. Calvetti et al, 2004), simulation of snow or ice blocks floating on a river (e.g. Hopkins et al, 1996), asphalt and railway ballast behaviour (e.g. Lu és McDowell, 2007), silo problems, or even the behaviour of a crowd in panic (Helbing et al, 2000).

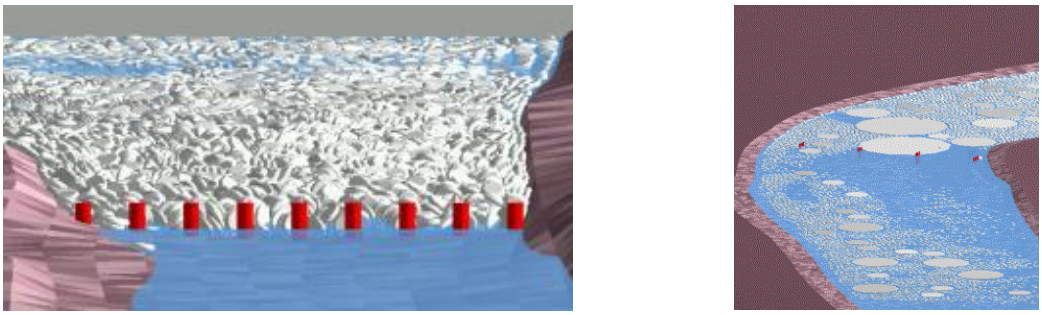

Figure 3.
BALL-type modelling of floating ice blocks on a river. Hopkins et al, 1996

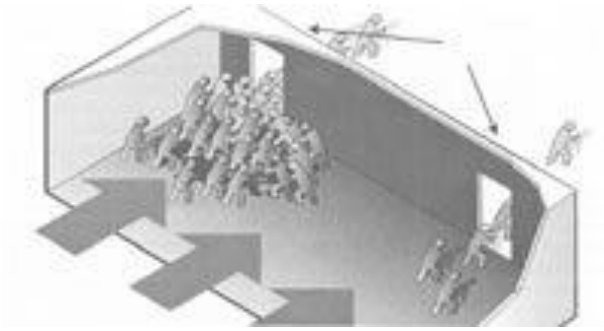

Figure 4.
Motion of a crowd in panic. Helbing et al, 2000

Questions

4.1. Under what conditions does a discrete element model belong to the BALL-type models?
4.2. What quantities are contained in the vector of contact forces, and in the vector of relative displacement increments belonging to a contact in a BALL-type model?
4.3. What types of contact models do you know? Shortly describe them!
4.4. Explain the calculation of a single time step in the BALL-type models! Why is it important to set a limit to the length of the time step, and how can this limit be estimated?
4.5. Why should damping be used in BALL-type models? Introduce local damping and contact viscous damping!