translation from the crystal, operation L represents the axis naming/direction convention used by the detector manufacturer (eight possibilities), K is an operation scaling pixels to millimetres, D is a detector distortion function and $\mathbf{B}$ represents the beam position on the detector surface.

Equations (11.4.2.1)–(11.4.2.8) fully describe the existence and position of the diffraction peaks, which is all that is needed for the autoindexing procedure.

11.4.3. Autoindexing

Among the number of autoindexing algorithms proposed (Vriend & Rossmann, 1987; Kabsch, 1988; Kim, 1989; Higashi, 1990; Leslie, 1993), the method based on periodicity of the reciprocal lattice tends to be the most reliable (Otwinowski & Minor, 1997; Steller *et al.*, 1997).

Autoindexing starts with a peak search, which results in the set of $\{p, q, i\}$ triplets, where i is the number of the image in which the peak with position $\{p, q\}$ was found. The program takes advantage of the fact that for any rotation matrix

$$([R]\mathbf{S}) \cdot ([R]\mathbf{a}) = \mathbf{S} \cdot \mathbf{a}. \qquad (11.4.3.1)$$

When

$$[R] = [R_3(-\varphi_3)][R_2(-\varphi_2)][R_1(-\varphi_1)], \qquad (11.4.3.2)$$

equation (11.4.3.1) applied to equation (11.4.2.2) becomes:

$$([R]\mathbf{S}) \cdot \mathbf{a}_o = h, \qquad (11.4.3.3)$$

where $\mathbf{a}_o$ is a three-dimensional vector with as yet unknown components. Note that the matrix $[R]$ represents crystal rotation when the crystal is in the diffraction condition defined by the existence of the solution to equations (11.4.2.1)–(11.4.2.4), described by vector $\mathbf{S}$. For data collected in the wide oscillation mode* the angle at which diffraction occurs is not known *a priori*; however, it can be approximated by the middle of the oscillation range of the image. Combining the peak position $\{p, q\}$ with equations (11.4.2.5) and (11.4.2.8) provides an estimate of the vector $\mathbf{S}$. So, we expect that equation (11.4.3.3) and similar equations for k and l are approximately (owing to approximation and experimental errors) satisfied. The purpose of autoindexing is to determine the unknown vectors $\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0$ and the $\{h, k, l\}$ triplet for each peak. To accomplish this, three equations (11.4.3.3) for each peak must be solved. *DENZO* introduced a method based on the observation that the maxima of the function

$$\sum_i \cos[2\pi([R]\mathbf{S}_i) \cdot \mathbf{a}_0] \qquad (11.4.3.4)$$

are the approximate solutions to this set of equations (11.4.3.3). To speed up the search for all significant maxima, a two-step process is used. The first step is the search for maxima of function (11.4.3.3) on a three-dimensional uniform grid, made very fast owing to the use of a fast Fourier transform (FFT) to evaluate function (11.4.3.4). Function (11.4.3.4) is identical to structure-factor calculations in the space group $P1$, which allows the use of the crystallographic FFT. Because the maxima at the grid points (*HKL* uses a $96 \times 96 \times 96$ grid) only approximate the maxima of function (11.4.3.4), the vectors resulting from a grid search are optimized by the Newton method. Function (11.4.3.4) has maxima not only for basic periodic vectors $\mathbf{a}_0, \mathbf{b}_0$ and $\mathbf{c}_0$, but also for any integer linear combination of them. Any set of three such vectors with a minimal nonzero determinant can be used to describe the crystal lattice. Steller *et al.* (1997) describe the algorithm that finds the most reliable set of three vectors. This set needs to be converted to the one conventionally used by crystallographers, as defined in *IT* A (1995).

To generate the conventional solution, two steps are used. Step 1 finds the reduced primitive triclinic cell. *IT* A provides the algorithm for this step. Subsequently, step 2 finds conventional cells in Bravais lattices of higher symmetry.

11.4.3.1. *Lattice symmetry*

The relationship between a higher-symmetry cell and the reduced primitive triclinic cell can be described by

$$[A] = [M][P], \qquad (11.4.3.5)$$

where $[A]$ and $[P]$ are matrices of the type $\{a_0, b_0, c_0\}$, with $[P]$ representing the reduced triclinic primitive cell, and $[M]$ is one of the 44 matrices listed in *IT* A.† If $[A]$ is generated using equation (11.4.3.5) from an experimentally determined $[P]$, owing to experimental errors it will not exactly satisfy the symmetry restraints. *DENZO* introduced a novel index that helps evaluate the significance of this violation of symmetry. This index is based on the observation that from $[A]$ one can deduce the value of the unit cell, apply symmetry restraints to the unit cell and calculate any matrix $[A']$ for the unit cell that satisfies these symmetry restraints. If $[A]$ satisfies symmetry restraints, the matrix $[U]$, where

$$[U] = [A][A']^{-1}, \qquad (11.4.3.6)$$

will be unitary and

$$[U]^T - [U]^{-1} = 0. \qquad (11.4.3.7)$$

The index of distortion printed by *DENZO* is

$$\left\{\sum_i \sum_j \left([U]_{ij}^T - [U]_{ij}^{-1}\right)^2\right\}^{1/2} \Big/ 6, \qquad (11.4.3.8)$$

where i and j are indices of the 3×3 matrix $[U]$.

The value of this index increases as additional symmetry restraints are imposed, starting from zero for a triclinic cell. Autoindexing in *DENZO* always finishes with a table of distortion indices for 14 possible Bravais lattices, but does not automatically make any lattice choice.

11.4.3.2. *Lattice pseudosymmetry*

The cell-reduction procedure cannot determine lattice symmetry, since it cannot distinguish true lattice symmetry from a lattice accidentally having higher symmetry within experimental error (*e.g.* a monoclinic lattice with $\beta \simeq 90°$ is approximately orthorhombic). If one is not certain about the lattice symmetry, the safe choice is to assume space group $P1$, with a primitive triclinic lattice for the crystal, and to check the table again after the refinement of diffraction-geometry parameters. A reliable symmetry analysis can be done only by comparing intensities of symmetry-related reflections, which is done later in *SCALEPACK* or another scaling program.

11.4.3.3. *Data-collection requirements*

The total oscillation range has to cover a sufficient number of spots to establish periodicity of the diffraction pattern in three dimensions. It is important that the oscillation range of each image is small enough so that the lunes (rings of spots, all from one reciprocal plane) are resolved. One should note that the requirement

* This is when the crystal oscillation angle during the measurement of a single diffraction pattern is larger than the angular reflection width.

† It should be noted that No. 17 contains an error (Kabsch, 1993).

for lune separation is distinct from the requirement for spot separation. If lunes overlap, spots may have more than one index consistent with a particular position on the detector.

The autoindexing procedure described above is not dependent on prior knowledge of the crystal unit cell; however, for efficiency reasons, the search is restricted to a reasonable range of unit-cell dimensions, obtained, for example, from the requirement of spot separation. In *DENZO*, this default can be overridden by the keyword 'longest vector', but the need to use this keyword is a sign of a problem that should be fixed. Either the defined spot size should be decreased or data should be recollected with the detector further away from the crystal.

11.4.3.4. *Misindexing*

Autoindexing is sensitive to inaccuracy in the description of the detector geometry. The specified position of the beam on the detector should correspond to the origin of the Bragg-peaks lattice (Miller index 000). Autoindexing will shift the origin of the lattice to the nearest Bragg lattice point. An incorrect beam position will result in the nearest Bragg lattice point not having the index 000. In such a situation, all reflections will have incorrectly determined indices. Such misindexing can be totally self-consistent until the intensities of symmetry-related reflections are compared. This dependence of the indexing correctness on the assumed beam position is the main source of difficulties in indexing (Gewirth, 1996; Otwinowski & Minor, 1997). The beam position has to be precise, as the largest acceptable error is one half of the shortest distance between spots.

Indexing limited to determining h, k, l triplets is not very sensitive to other detector parameters. Errors by a degree or two in rotation or by 10% in distance are unlikely to produce wrong values of h, k and l. Sometimes even a very large error, such as the distance being too large by a factor of 5, will still produce the correct h, k, l triplets. The detector position error will be compensated by an error in the lattice determined by autoindexing. For this reason, the accuracy of the lattice is not a function of the autoindexing procedure, but depends mainly on the accuracy of the detector description. By the same token, the distortion of the lattice also depends on the accuracy of the detector parameters.

11.4.3.5. *Twins*

Special care has to be taken if more than one crystal contributes to the diffraction image. When there is a large disproportion between volumes (*e.g.* the presence of a satellite crystal), autoindexing may work without any modifications. In the case of similar volumes, the manual editing of weaker reflections and resolution cuts can make the proportion of reflections from one crystal in the peak-search list large enough for the autoindexing method to succeed. If the crystals have a similar orientation, using only very low resolution data may be the right method. In the case of twinned crystals, autoindexing sometimes finds a superlattice that results in integer indices simultaneously for both crystals. In such a case, *DENZO* solves the problem of finding the best three-dimensional lattice that incorporates all of the observed peaks. Unfortunately, for a twinned crystal, this is a mathematically correct solution to an incorrectly posed problem.

11.4.4. Coordinate systems

There are four natural coordinate systems used to describe a diffraction experiment, defined by the order in which the data are stored, the beam and gravity, or the beam and the goniostat axes (spindle or 2θ). These coordinate systems will be called, respectively, *data*, *beam–gravity*, *beam–spindle* and *beam–2θ*.

11.4.4.1. *Beam–gravity*

To visualize a diffraction pattern, beam–gravity is the coordinate system clearly preferred by human physiology. The universal preference to relate to the gravity direction is revealed by the observation that people generally perceive an image in a mirror as inverted left–right rather than top–down. Hence *XdisplayF* uses the beam–gravity coordinate system, except when diffraction data cannot be related to gravity.*

11.4.4.2. *Data*

The first (1983) *DENZO* implementation used the data coordinate system to describe the beam position on the detector and to define the integration box. This is still the case in order to keep backward compatibility.

11.4.4.3. *Beam–spindle*

Until 1998, *DENZO* supported only a single-axis goniostat and used a beam–spindle coordinate system to define crystal and detector orientation and polarization. Initially, the goniostat spindle axis was assumed to be horizontal, so the direction perpendicular to the beam and spindle was described by the keyword *vertical*, which in reality may not relate to the gravity direction for some goniostats. The keyword *rotx* relates to rotation around the spindle axis, *roty* around the *vertical axis* and *rotz* around the beam axis. The definition of the orientation matrix in the communication file between *DENZO* and *SCALEPACK* uses an unintuitive convention: the letter *y* in *roty* relates to the first element of the vector, *x* in *rotx* relates to the second and *z* in *rotz* to the third. However, the matrix always has a positive determinant, so this convention has no impact on the handedness of the coordinate system. This unfortunate choice of convention, preserved for backward compatibility reasons, appears only in the communication file and has no significance for anybody who does not inspect the matrix.

11.4.4.4. *Beam–2θ*

The recent addition of a general goniostat introduced a conceptual change in the *DENZO* coordinate system. The data-collection axis can be oriented in any direction, so in principle *rotx*, *roty* and *rotz* no longer need to be defined relative to the data-collection axis. However, to keep the useful correlation between refinable parameters (*crystal rotz* and *detector rotz* being close to 100% correlated), one real and two virtual goniostats are used simultaneously in *DENZO*. Refinable crystal parameters (*crystal rotx*, *roty*, *rotz*) are still defined, as in the past, by the data-collection axis and the beam. This means that the directions of rotations defined by *fit crystal rotx*, *roty* and *rotz* do not rotate around the data-collection axis as the program advances from one image to another. This coordinate system changes with the change in direction of the data-collection axis. Crystal orientation is defined by three constant, perpendicular axes, which, in the current version, no longer have to be aligned with the physical crystal goniostat. However, the so-called *2 theta* rotation has a fixed axis, and, if it exists, it defines the *DENZO* coordinate system together with the beam axis. Thus the current coordinate system in *DENZO* should be called beam–2θ. Fortunately for the user, the conversions between different coordinate systems are handled transparently. For example, the refined change in the crystal orientation is converted from the refined goniostat to the crystal-orientation goniostat. The

* There may occasionally be an exception to this when the experimental system is not known.