Longitudinal phase space tomography with space charge

S. Hancock and M. Lindroos
CERN, Geneva, Switzerland

S. Koscielniak
TRIUMF, Vancouver, British Columbia, Canada

Tomography is now a very broad topic with a wealth of algorithms for the reconstruction of both qualitative and quantitative images. In an extension in the domain of particle accelerators, one of the simplest algorithms has been modified to take into account the nonlinearity of large-amplitude synchrotron motion. This permits the accurate reconstruction of longitudinal phase space density from one-dimensional bunch profile data. The method is a hybrid one which incorporates particle tracking. Hitherto, a very simple tracking algorithm has been employed because only a brief span of measured profile data is required to build a snapshot of phase space. This is one of the strengths of the method, as tracking for relatively few turns relaxes the precision to which input machine parameters need to be known. The recent addition of longitudinal space charge considerations as an optional refinement of the code is described. Simplicity suggested an approach based on the derivative of bunch shape with the properties of the vacuum chamber parametrized by a single value of distributed reactive impedance and by a geometrical coupling coefficient. This is sufficient to model the dominant collective effects in machines of low to moderate energy. In contrast to simulation codes, binning is not an issue since the profiles to be differentiated are measured ones. The program is written in FORTRAN 90 with high-performance FORTRAN extensions for parallel processing. A major effort has been made to identify and remove execution bottlenecks, for example, by reducing floating-point calculations and recoding slow intrinsic functions. A pointer-like mechanism which avoids the problems associated with pointers and parallel processing has been implemented. This is required to handle the large, sparse matrices that the algorithm employs. Results obtained with and without the inclusion of space charge are presented and compared for proton beams in the CERN protron synchrotron booster. Comparisons of execution times on different platforms are presented and the chosen solution for our application program, which uses a dual processor PC for the number crunching, is described.

I. INTRODUCTION

The underlying principle of tomography is to combine the information in a sufficiently large number of projections to be able to reconstruct unambiguously the fuller picture with the extra dimension reinstated. Thus, for example, many one-dimensional profiles of x-ray transparency taken from different angles can give doctors an image of a two-dimensional slice through a patient.

The application of tomography to longitudinal phase space in an accelerator becomes obvious once it is realized that a bunch of particles performing synchrotron motion is analogous to a patient rotating in a stationary body scanner. On each turn around the machine, a longitudinal pickup provides a "snapshot" of the bunch projected at a slightly different angle. It only remains to combine such profiles tomographically to obtain a two-dimensional picture of phase space density [1,2].

The nonlinearities of synchrotron motion are taken into account by tracking test particles in order to build maps which describe the evolution of phase space. The maps are used to reconstruct iteratively a distribution whose projections converge towards the measured bunch profiles. The tracking can be made arbitrarily complex. In particular,

collective effects due to the interaction of the beam with a wideband reactive impedance are readily included [3] since the wakefield may be modeled in terms of the derivative of bunch shape and this is known from the measured data.

One objective is to provide on-line measurements of rms emittance and momentum spread, for example, without making any assumptions about longitudinal matching.

II. RECONSTRUCTION

Back projection is a key process by which the contents of the bins of a one-dimensional histogram are redistributed over the two-dimensional array of cells which comprise the reconstructed image. Given no *a priori* knowledge of the original two-dimensional distribution, the contents of one bin are shared over all the cells that could have contributed to that bin. The back projection of all bins of all profiles yields a first approximation to the original distribution.

Algebraic reconstruction techniques (ART) [4] exploit the fact that the coefficients for sharing bins in back projection can also be used to project the contents of cells into those bins. Hence a set of projections can be obtained from the first approximation. Back projection of the bin-by-bin difference between the original set of profiles and this new

one yields an improved approximation. Further iterations converge more rapidly if any cell whose contents have become negative is reset to zero.

The problem with conventional ART is that its strategies for estimating the redistribution coefficients are based on straight-line back projection geometry. This implies either rigid, circular motion of the two-dimensional distribution or that its projections be measured simultaneously. An alternative approach is to consider how the contents of one cell get projected into the bins of a particular profile. By launching a small number of test particles which, initially, are uniformly distributed within the cell, the calculation of coefficients becomes a simple matter of counting how many particles end up in each bin at the particular instant when the profile was measured. Thus, a hybrid algorithm which combines particle tracking with ART allows large-amplitude synchrotron motion to be taken into account since the trajectories of the test particles need not be assumed circular. Although iteration proceeds as before, there is a price to be paid: a large map of (projection) coefficients must first be built and its inverse (for back projection) derived for every profile in the set of measured data. On the other hand, since most of the computational effort is invested in building the maps, it becomes trivial to repeat a calculation with fresh data taken under the same conditions.

III. TRACKING

Consider the time at which a general particle crosses the rf gap in a synchrotron with a single cavity

$$t_{i,m+1} = t_{i,m} + \frac{2\pi}{\omega_{i,m}}. \tag{1}$$

Here, $t_{i,m}$ is the arrival time of the ith particle at the start of the mth turn, i.e., at the downstream end of the gap, and $\omega_{i,m}$ is its angular revolution frequency on that turn.

Using subscript zero to denote the synchronous particle,

$$t_{i,m+1} = t_{i,m} + \frac{2\pi}{\omega_{0,m}}\left(\frac{1 + \Delta R_{i,m}/R_{0,m}}{1 + \Delta\beta_{i,m}/\beta_{0,m}}\right), \tag{2}$$

where $\Delta R_{i,m}$, $\Delta\beta_{i,m}$ are radial position and relativistic velocity differences with respect to the synchronous particle and $h\omega_{0,m}$ is the rf frequency on harmonic h. Rewriting in terms of energy differences, the relative rf phase at the arrival time of the ith particle is

$$\Delta\phi_{i,m+1} = \frac{\omega_{0,m}}{\omega_{0,m-1}}\Delta\phi_{i,m}$$
$$+ 2\pi h\left(\frac{1 + \alpha_p\beta_{0,m}^{-2}\Delta E_{i,m}/E_{0,m}}{1 + \gamma_{0,m}^{-2}\beta_{0,m}^{-2}\Delta E_{i,m}/E_{0,m}} - 1\right), \tag{3}$$

since

$$h\omega_{0,m}(t_{0,m+1} - t_{0,m}) = 2\pi h. \tag{4}$$

Hence, to a good approximation,

$$\Delta\phi_{i,m+1} = \Delta\phi_{i,m} - 2\pi h\left(\frac{\eta_{0,m}\Delta E_{i,m}}{\beta_{0,m}^2 E_{0,m}}\right), \tag{5}$$

since momentum compaction

$$\alpha_p = \gamma_{0,m}^{-2} - \eta_{0,m}, \tag{6}$$

where $\gamma_{0,m}$ is the relativistic energy of the synchronous particle.

Assuming negligible modification of the synchronous phase due to self-fields, the corresponding energy increment at the end of the mth turn yields

$$\Delta E_{i,m+1} = \Delta E_{i,m} + q[V_{\mathrm{rf},m+1}(\phi_{0,m+1} + \Delta\phi_{i,m+1})$$
$$- V_{\mathrm{rf},m+1}(\phi_{0,m+1})$$
$$+ V_{\mathrm{self},m+1}(\phi_{0,m+1} + \Delta\phi_{i,m+1})], \tag{7}$$

where q is the charge carried by the particle, ϕ_0 is the synchronous phase, and V_{rf} and V_{self} are the applied rf and self-field voltage functions, respectively. The latter may be written [5] in terms of the line charge density $q\lambda$ along the bunch

$$V_{\mathrm{self},m}(\phi) = qh^2\omega_{0,m}\left[\frac{gZ_{\mathrm{vacuum}}}{2\beta_{0,m}\gamma_{0,m}^2} - \left|\frac{Z_{\mathrm{wall}}}{n}\right|\right]\frac{d\lambda_m(\phi)}{d\phi}. \tag{8}$$

The factor in the square brackets is the effective impedance seen by the beam and comprises a direct space charge term (which is expressed in terms of a geometrical coupling coefficient g and the impedance of free space Z_{vacuum}) and the distributed impedance of the vacuum chamber $|Z_{\mathrm{wall}}|$ (divided by the mode number n).

Equations (5) and (7) together constitute the turn-by-turn tracking used in the code. However, since the line charge density is not necessarily known at every turn, the self-field voltage is evaluated from the mean of the nearest two bunch profile measurements. Smoothing and differentiation are achieved using a Savitzky-Golay filter [6] of order 4. This has the advantage over conventional low-pass filtering of not increasing the apparent bunch length.

The action of a phase loop is not included in the tracking. Typically, closed-loop conditions do not affect the bunch during a measurement span unless its dipole motion or the filamentation of a badly matched distribution would otherwise have shifted the barycenter of the observed profiles.

Equation (5) takes the ratio of synchronous revolution periods on consecutive turns to be unity, which is a good approximation except at very low energies. Furthermore, the orbit expansion is made only to first order in fractional energy offset, so that reconstructing near transition should be avoided. This is true since the lack of phase space motion near transition precludes tomography.

Since it is not dissipative, a pure reactive impedance cannot alter the average energy of the bunch nor, in the absence of coherent motion, is there any modification of the synchronous phase. Equation (7) takes the self-field voltage to be zero at ϕ_0. That is, the center of motion about

which particles are tracked is the aggregate one determined by considerations of energy balance in the absence of collective effects. This simplification guarantees the convergence of the root-finding algorithm that is used to evaluate ϕ_0 and it assumes that the average energy of the bunch is in equilibrium at E_0. Typically, this implies only a small error with respect to the true center of individual particle motion and the method is known to be very tolerant of such errors. No resistive (in-phase) component of the self-field voltage is considered.

For a circular beam of radius a in a circular pipe of radius b, the coupling coefficient of the particle ensemble may be estimated [7] as $g = 0.5 + 2\ln(b/a)$. In the absence of cylindrical symmetry, the situation is more complicated, but the direct space charge component can still be expressed in terms of this single input parameter.

IV. DISCREPANCY

Discrepancy [4] expresses in a single figure of merit the residual bin-by-bin differences between the projections of a reconstructed distribution and the original profiles,

$$d = \sqrt{\frac{1}{M} \sum_{\forall i \ni N_i > 0} (e_i - r_i)^2 / N_i} . \tag{9}$$

Here, e_i and r_i are, respectively, the measured and reconstructed contents of the ith bin and M is the number of terms in the summation. The weighting factor N_i is the number of image pixels that project into the ith bin. However, since each e_i constitutes an independent measurement whose variance is dominated by noise and is therefore the same for all i, the expression can be modified slightly so that d^2 becomes more like the mean χ^2 per bin. Thus,

$$d' = \sqrt{\frac{1}{M'} \sum_{i=1}^{M'} (e_i - r_i)^2} , \tag{10}$$

where M' is the total number of bins in all profiles. It is this form of discrepancy that is implemented in the code for monitoring convergence.
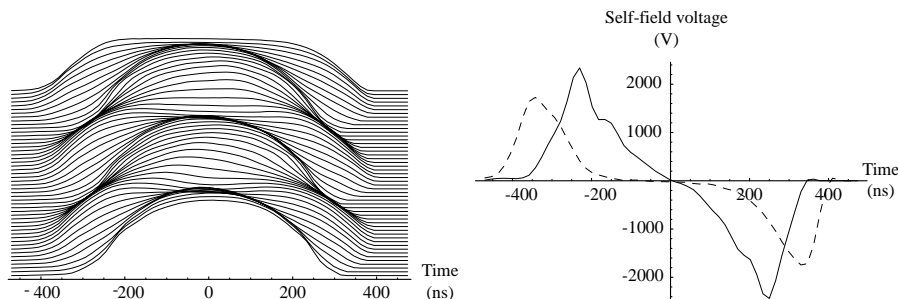
V. SOME RESULTS

The mountain range data of Fig. 1 are tomographically reconstructed in Fig. 2 with and without the inclusion of space charge. The images correspond to the time of the first measured profile, i.e., to a minimum of bunch length, but the reconstructed distribution is only fully upright when space charge is taken into account. The dashed bucket separatrix illustrates the loss of acceptance. The coupling coefficient was estimated as $g = 1.8$ from beamscope [8] measurements of transverse beam size. In comparison, $g = 2.0$ produced the best reconstructed image (see Fig. 3). Since the beamscope is known to overestimate the horizontal size of the beam, the larger value of g was adopted. This corresponds to a space charge impedance of more than 700 Ω. Since the inductive wall impedance of the protron synchrotron booster (PSB) is considerably less than this, it was simply taken to be zero.

The deliberately mismatched bunch generates a varying self-field voltage (see Fig. 1) which can therefore be distinguished from a mere calibration error of the rf voltages. When space charge was included, discrepancy minima as functions of rf voltage were obtained in good agreement with the measured cavity voltages on both harmonics.

VI. COMPUTER CODE

A. Program philosophy

The algorithm was originally developed in MATHEMATICA™, a choice mainly motivated by the rich set of built-in functions for graphics and for the manipulation of arrays. However, the objective was merely to establish a proof of principle. Consequently, with the aim of reducing the execution time by a factor of at least 100, the code was translated into FORTRAN 90/HIGH PERFORMANCE FORTRAN (HPF). This was done with the view to use parallel architectures. Mathematical tool kits with the associated definition modules from Ref. [9] were integrated in the program from the outset to avoid unnecessary duplication of standard numerical routines. The entire issue of providing a user interface for the demanding control room environment of the CERN accelerator complex was separated out as an independent development.

FIG. 1.   Left: bunch shape oscillations of 6.5×10^{12} protons measured after an abrupt reduction in the second-harmonic component of a stationary dual-harmonic bucket at 100 MeV in the CERN PSB. Right: corresponding self-field voltage functions obtained from the mean derivative of the first two (solid line) and last two (dashed line) profiles.
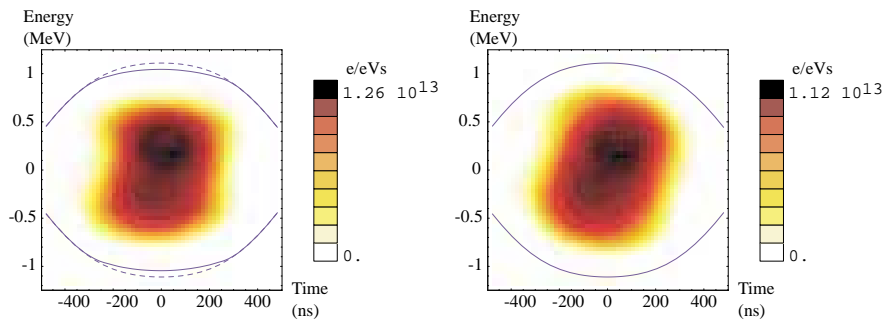
FIG. 2. (Color) Phase space reconstructions, with (left) and without (right) space charge. Note the different density scales.
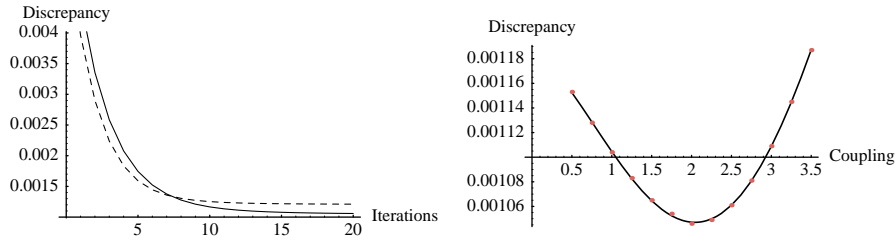

FIG. 3.   Left: convergence for the two cases of Fig. 2; the solid line is with space charge included. Right (Color): discrepancy (after 50 iterations) versus geometrical coupling.
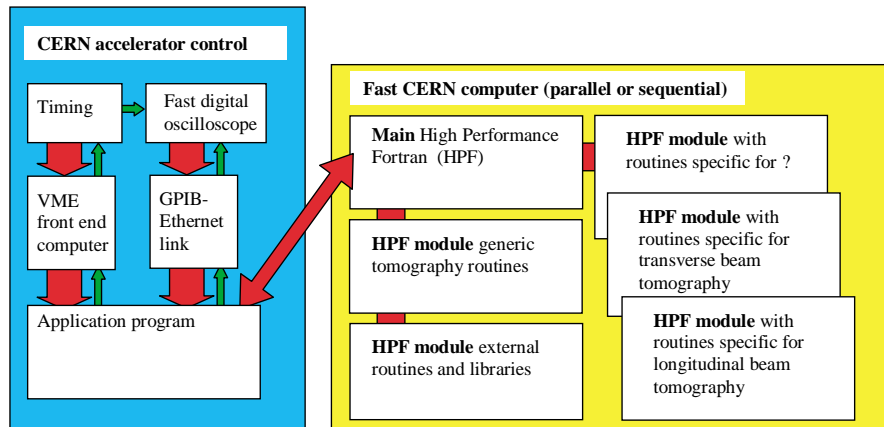

FIG. 4. (Color) A schematic view of a future on-line tomography system. To the left, the hardware and graphical user interface (GUI) configuration is drawn, and, to the right, the structure of the present numerical part of the code. The data flow is indicated with broad arrows and the command flow with narrow arrows.

The fact that the method could be applicable to processes other than synchrotron motion implied a division of the code into generic modules and modules specific to longitudinal beam dynamics. Furthermore, all parameters characterizing the considered process are passed to the program as input at execution time, making the code, to a large extent, data driven. A schematic view of the structure of the program is presented in Fig. 4.

B. Hardware

The raw data were initially acquired with a dedicated digital oscilloscope and transferred to the tomography application program using the standard CERN-PS control system. This meant passing sizable amounts of data (typically 50 kbytes) over a front-end computer designed for transferring short control sequences. The same front-end computer is used to control all the hardware, including timing. In the latest version, the control and data transfers are performed over a separate general purpose instrumentation bus (GPIB)-Ethernet link. This has improved the performance by at least a factor of 10, and a typical data acquisition can now be performed in seconds.

C. Coding considerations

In the original MATHEMATICA™ code there is extensive use of floating-point operations and many very large and

very sparse matrices. In the FORTRAN 90 version, the number of floating-point operations is reduced by the use of integer representation until the final step in each number manipulation, while a pointerlike reallocatable vector representation was created to deal with the sparse matrices. The sparse matrices are stored in an array with sufficient depth to contain most of the data. Where additional depth is needed the excess data are stored in a supplementary array and the index to the array element used is stored in the last element of the original matrix. The supplementary array is reallocatable and this procedure can obviously be repeated any number of times. Consequently, the actual depth of the matrix is limited only by the available memory during execution. The built-in pointer facility of FORTRAN 90 was avoided because it would make the efficient use of parallel architectures with shared memory impossible.

A fundamental part of the algorithm is the construction of the forward and backward projection maps by tracking a small number of test particles launched from each cell in the two-dimensional image. The individual test particles are tracked without considering any particle-to-particle interaction. Consequently, the launching and tracking of all particles for all cells can be done in parallel, a fact which was taken into consideration from the beginning and is fully exploited in a parallel version.

A complication in the space charge code is that the self-field voltage is evaluated from the measured profiles as a function of bins, while tracking is performed in terms of rf phase relative to the synchronous particle. This makes it necessary to perform a coordinate transformation at each tracking step adding a considerable overhead to the tracking subroutine. Furthermore, since some particles may get tracked outside the limits of the profiles, cyclical boundary conditions are necessary in the coordinate transformation.

A first execution time analysis showed that more than 90% of the time was spent in the tracking procedure long-track, 75% in evaluating sine functions (despite the use of fast libraries). The sine function evaluation was therefore replaced by a look-up procedure over the range -2π to 2π. Tests with a table of 8192 values per quadrant produced satisfactory results

D. Parallelization

Although the optimizations described above have proved efficient, the use of multiple processors with HPF

TABLE I. CPU time in seconds for the reconstruction without space charge of one well-resolved image using different processors and different versions of the sine function. A fast FORTRAN library for the sine function was used on the power PC (PPC).

Version	PPC Type 604	PC single Pentium III	PC dual Pentium III
F90/HPF code	99	38	22
Optimized code	33	26	15

TABLE II. CPU time in seconds for the reconstruction with space charge of one well-resolved image using different processors and different versions of the sine function. A fast FORTRAN library for the sine function was used on the PPC.

Version	PPC Type 604	PC single Pentium III	PC dual Pentium III
F90/HPF code	113	44	24
Optimized code	n.a.	41	23

directives and the parallel FORALL statement was pursued. A strong motivation for this comes from the availability of very cheap dual-processor PCs running under LINUX. Consequently, the code was ported to a dual-processor LINUX environment and an HPF version of the CPU-demanding tracking routine for both the original and space charge versions of the code was developed. Furthermore, the Portland Group's [10] HPF compiler also makes a best possible attempt to parallelize the standard FORTRAN 90 code (e.g., WHERE loops). Comparing the single-processor execution times in Tables I and II to the dual-processor execution times shows that the code parallelizes well. The fixed overhead (this should be added to the dual-processor numbers) which is needed at execution time to set up the parallelism is approximately 5 s. The present price and performance of multiple-processor PCs promises a favorable performance for the numerical part of the code. In the near future, it should be possible to acquire, calculate, and display a well-resolved tomogram in much less than a minute.

E. User interface

The graphical user interface (GUI) is being written in C with the execution of the numerically demanding part of the code in a parallel environment. The call to the parallel environment is done through remote job submission, where the job is passed as well as the input and output data.

F. Usage notes

Successful reconstruction requires that the measured data span an interval of the order of at least half a synchrotron period. In addition, normalization requires that each profile encompass the same number of particles. Together, these constraints suggest that no particles exist outside the largest closed phase space trajectory that can be drawn inside the image width. Consequently, in order to reduce computation time, map coefficients are derived only for those cells of the image that lie within this limiting trajectory at the reconstruction time. The cells of interest can be further restricted to a range of columns between upper and lower limits supplied as input parameters. This effectively declares the profile bins that lie outside this range to be empty at the reconstruction time, although they may well be populated at other times. In the event that the closed trajectory is too restrictive to permit

the reconstruction of the entire distribution, a flag may be set to extend the region of interest up to a fixed off-energy limit for all columns in the specified range. If this flag is not set and there is no second-harmonic rf component, the cell height is chosen such that small-amplitude trajectories in the reconstructed image are circular.

The parabolic fit option to find automatically the synchronous point has been superseded by a foot-tangent method, as this is more robust and works equally well for dual-harmonic bunch shapes. The (initial) depth of the map elements is no longer an input parameter and is now calculated according to the length of the profiles. Instead, the number of test particles that are tracked per cell has become a free parameter, so that the accuracy of the maps can be traded off against execution time. In the case of a strong space charge reduction of acceptance, some test particles may be launched outside the bucket. This is not a problem because the normalization of map elements takes into account any particles that get tracked outside the image width. However, when self-fields are included, particles are tracked modulo an integer number of rf periods so any that do leave the image width may also return.

VII. CONCLUSIONS

An algorithm has been developed for longitudinal phase space tomography in which the contents of the reconstructed array is effectively rotated instead of inclining profile bins in order to make a projection. This allows a different mapping to be applied to each cell in the array so that rigid, circular motion of the phase space distribution need not be assumed.

The code has been refined to include collective effects due to direct space charge and reactive wall impedance.

A poorly known parameter in the physical model of the algorithm may be estimated by maximizing the resultant image quality as a function of that parameter. The space charge impedance of the CERN PSB has effectively been measured in this way under conditions contrived to induce a strong space charge effect.

The algorithm is a hybrid one and, consequently, arbitrarily complex rf systems can be catered for by modifying a small part of the code. Likewise, the method may be extended to cover other nonrigid bodies whose deformation is governed by a known model.

The advances in the optimization of the numerical part of the code are very promising. It seems likely that a GUI in C together with number crunching in parallel structures will yield response times of much less than 1 min for a single image, making on-line machine optimization with tomography possible.

[1] http://tomograp.home.cern.ch/tomograp/

[2] S. Hancock, M. Lindroos, E. McIntosh, and M. Metcalf, Comput. Phys. Commun. **118**, 61 (1999).

[3] S. Hancock, S. Koscielniak, and M. Lindroos, in *Proceedings of the European Particle Accelerator Conference, 2000,* http://accelconf.web.cern.ch/AccelConf/e00/index.html

[4] R. Gordon, IEEE Trans. Nucl. Sci. **21**, 78 (1974).

[5] S. Hansen, H. G. Hereward, A. Hofmann, K. Hubner, and S. Myers, IEEE Trans. Nucl. Sci. **22**, 138 (1975).

[6] A. Savitzky and M. J. E. Golay, Anal. Chem. **36**, 1627 (1964).

[7] R. Baartman, TRIUMF Technical Report No. TRI-DN-92-K206, 1992.

[8] H. Schonauer, CERN Technical Report No. CERN/PS 92-10(HI), 1992.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing* (Cambridge University, Cambridge, England, 1996).

[10] http://www.pgroup.com/