CLOSED ORBIT CORRECTION IN SYNCHROTRONS

D.Dinev

Bulgarian Academy of Sciences

Institute for Nuclear Research and Nuclear Energy, Bulgaria

Algorithms for closed orbit correction in synchrotrons as well as related topics as error sources, statistical characteristics of the orbit, etc., are discussed. The review covers both traditional methods for orbit correction: beam-bump, harmonical correction, etc., and the newest developments: eigenvector correction, SVD, algorithms for optimum positioning of dipoles, etc. The use of expert system and artificial neural systems is described as well. The last chapter is devoted to the first turn steering.

Обсуждаются алгоритмы для коррекции замкнутой орбиты в синхротронах и связанные с ними вопросы — источники ошибок, статистические свойства орбиты и т.д. Обзор охватывает как традиционные методы коррекции: метод локального искажения орбиты, гармонический метод и т.д., так и новейшие алгоритмы: коррекцию с применением собственных векторов, сингулярное разложение матрицы связи, алгоритмы для оптимального расположения дипольных магнитов и т.д. Обсуждается применение экспертных систем и искусственных нейронных сетей для коррекции орбиты. Последний раздел посвящен коррекции первого оборота.

1. INTRODUCTION

The transverse motion of the particles in a cyclic accelerator is a superposition of a motion along a closed trajectory (equilibrium or closed orbit) and betatron and radial-phase oscillations around this curve. In a real magnetic structure both the closed orbit and the oscillations around it are distorted due to errors in the magnetic fields, displacements and tilts of the elements from their designed positions, stray fields, and ground movements. As far as the closed orbit is concerned, perturbations cause its deformation. The maximum deviation of the distorted orbit from the reference orbit reaches to some ten millimeters. Such a large deviation does not allow for the accelerator aperture to be used effectively and hampers the work of the injection and extraction systems.

In storage rings the beam lifetime and the maximum current of the accumulated particles depend on the accuracy of the orbit.

An important point is the orbit stability. The time changes of the orbit increase the dynamic aperture. In synchrotron light sources an unstable orbit will increase the effective emittance thus reducing the effective brightness of the photon beams.

This is the reason why in accelerators special systems of either small correcting magnets or additional correcting coils in the main dipoles or controlled displacements of the quadrupoles are used.

The purpose of the orbit correction consists in choosing the proper strengths and positions of the correction elements so that the smallest possible deviations from the reference orbit may be achieved.

A special problem is the first turn steering which we discuss in the last chapter.

In this paper a survey of both the major orbit correction algorithms used in synchrotrons and storage rings and original results on the orbit correction obtained by the author during his work on the superconducting synchrotron NUCLOTRON in JINR (Dubna) and on the cooler-synchrotron COSY (Julich) is done.

2. TRANSVERSE PARTICLE MOTION UNDER LINEAR PERTURBATIONS

We shall begin our discussion of the closed orbit distortion and correction with a brief survey of the transverse particle motion in cyclic accelerators in the presence of the so-called linear perturbations.

In a curvilinear coordinate system (s, x, y) used in cyclic accelerators where s is directed along the reference orbit; x, along the main normal vector; z, along the binormal vector, the Lagrangian of the transverse particle motion can be presented in the form [1]

$$\mathcal{L}(x, x', z, z', s) - p \sqrt{\left(1 + \frac{x}{\rho}\right)^2 + x'^2 + z'^2} + e\left[\left(1 + \frac{x}{\rho}\right)A_s + x'A_x + z'A_z\right]. \quad (2.1)$$

In (2.1) the distance along the reference orbit s was taken to be the independent variable; $()'$ denotes differentiation with respect to s; p and e are the particle momentum and charge; ρ is the radius of the curvature of the reference orbit; and A_s, A_x, A_z are the vector potential components.

From (2.1) the following expression about the transverse motion Hamiltonian can be deduced

$$\mathcal{H}(x, p_x, x, p_z, s) -$$

$$-\left(1 + \frac{x}{\rho}\right)\sqrt{p^2 - (p_x - eA_x)^2 - p_z - eA_z)^2} - e\left(1 + \frac{x}{\rho}\right)A_s. \quad (2.2)$$

We shall consider here the cases of a sector magnetic field and of a quadrupole field, which are the most important for particle accelerators. For these cases substituting in (2.2) the relevant expressions for the vector potential A and expanding the kinematical part in a power series, retaining only the major first few members, one can obtain that

$$
\mathcal{H} \approx \mathcal{H}_0 = \begin{cases} \dfrac{1}{2}\dfrac{p_x^2}{p_0} + \dfrac{1}{2}\dfrac{p_z^2}{p_0} + p_0\dfrac{(1-n)}{2}\left(\dfrac{x}{\rho}\right)^2 + p_0\dfrac{n}{2}\left(\dfrac{z}{\rho}\right)^2, \\ \qquad\qquad\qquad \text{for a sector magnetic field} \\[2mm] \dfrac{1}{2}\dfrac{p_x^2}{p_0} + \dfrac{1}{2}\dfrac{p_z^2}{p_0} - \dfrac{eg}{2}(x^2 - z^2), \\[2mm] \qquad\qquad\qquad \text{for a quadrupole field.} \end{cases}
\tag{2.3}
$$

In (2.3) we denoted with $\mathcal{H}_0$ the Hamiltonian of the linearized transverse motion; $p_0 = -eB_0\rho$ is the particle momentum corresponding to the reference trajectory; $B_0\rho$ is the beam rigidity; $n = -(\rho/B_0)(\partial B/\partial x)$ is the field index; and g is the quadrupole gradient.

We shall consider in this paper correcting magnets as short magnets, with small $(B_c \ll B_0)$ and uniform $(n = 0)$ field. Thus the linearized Hamiltonian for the corrections is

$$
\mathcal{H}_0 = \frac{1}{2}\frac{p_x^2}{p_0} + \frac{1}{2}\frac{p_z^2}{p_0} - eB_c x.
\tag{2.4}
$$

Let's now introduce in the Hamiltonian the so-called linear perturbations:
a) field errors $\Delta B = B - B_0$,

b) magnetic element misalignments, Δx and Δz,

c) dipole tilts around the axis s, θ,

d) stray magnetic field, ΔB.

These errors cause linear about x and z members to appear in the Hamiltonians

$$
\mathcal{H} - \mathcal{H}_0 + \delta\mathcal{H}^{(1)},
\tag{2.5}
$$

where

$$
\delta\mathcal{H}^{(1)} =
$$

$$
= \begin{cases} -e\Delta B_z x + \dfrac{p_0}{\rho}\theta z + \dfrac{p_0(1+n)}{\rho^2}x\Delta x - \dfrac{p_0 n}{\rho^2}z\Delta z + \left(1+\dfrac{x}{\rho}\right)\Delta p, \text{ for dipoles} \\ egx\Delta x - egz\Delta z, \text{ for quadrupoles.} \end{cases}
\tag{2.6}
$$

The quadratic part $\mathcal{H}_0$ of the Hamiltonian (2.5) describes particle oscillations, as long as the linear part $\delta\mathcal{H}^{(1)}$ causes an internal force depending on the variable s. This is the well-known classical mechanics case of forced small oscillations. Because of the internal force, the orbit is distorted. The new orbit is a periodical solution of the Hamilton equations

$$\frac{dx_{c0}}{ds} = \frac{\partial\mathcal{H}}{\partial p_{x,c0}}, \qquad \frac{dp_{x,c0}}{ds} = -\frac{\partial\mathcal{H}}{\partial x_{c0}},$$

$$x_{c0}(s + 2\pi R) = x_{c0}(s),$$

$$p_{x,c0}(s + 2\pi R) = p_{x,c0}(s). \tag{2.7}$$

In (2.7) R is the mean radius of the accelerator.

Equations equivalent to (2.7) are valid for the vertical plane as well.

Let u and p_u be the conjugate variables describing the oscillations around the orbit

$$x = x_{c0} + u, \qquad p = p_{x,c0} + p_u. \tag{2.8}$$

If we now perform a canonical transformation from the variables x, p_x to u, p_u using as a generating function

$$f_2(x, p_u, s) = (p_{x,c0}(s) + p_u)x - x_{c0}(s)p_u, \tag{2.9}$$

we shall obtain for the new Hamiltonian

$$K(u, p_u, v, p_v, s) = \mathcal{H}(x_{c0}(s) + u, p_{x,c0}(s) + p_u, z_{c0}(s) + v, p_{z,c0}(s) + p_v, s) =$$

$$= \mathcal{H} - \left(\frac{\partial\mathcal{H}}{\partial x_{c0}}u + \frac{\partial\mathcal{H}}{\partial p_{x,c0}}p_u + \frac{\partial\mathcal{H}}{\partial z_{c0}}v + \frac{\partial\mathcal{H}}{\partial p_{z,c0}}p_v\right). \tag{2.10}$$

The Hamiltonian K does not contain linear in u, p_u, v, p_v members, i.e., it describes only the particle oscillations around the closed orbit.

From the Hamiltonians (2.3), (2.6) one can deduce the equations of the transverse particle motion. In Newton's form they are

$$\frac{d^2x}{ds^2} + k_x(s)x = F_x(s),$$

$$\frac{d^2z}{ds^2} + k_z(s)z = F_z(s), \tag{2.11}$$

where

$$k_x(s) = \begin{cases} \dfrac{(1-n)}{\rho^2}, & \text{for dipoles} \\ 0, & \text{for correctors} \\ \dfrac{-g}{B_0\rho}, & \text{for quadrupoles,} \end{cases} \qquad (2.12)$$

$$k_z(s) = \begin{cases} \dfrac{n}{\rho^2}, & \text{for dipoles} \\ 0, & \text{for correctors} \\ \dfrac{g}{B_0\rho}, & \text{for quadrupoles,} \end{cases} \qquad (2.13)$$

$$F_x(s) = \begin{cases} -\dfrac{\Delta B_z}{B_0\rho} - \dfrac{(1+n)}{\rho^2}\Delta x, & \text{for dipoles} \\ -\dfrac{B_c}{B_0\rho}, & \text{for correctors} \\ -\dfrac{g\Delta x}{B_0\rho}, & \text{for quadrupoles,} \end{cases} \qquad (2.14)$$

$$F_x(s) = \begin{cases} -\dfrac{n}{\rho^2}\Delta z - \dfrac{\theta}{\rho}, & \text{for dipoles} \\ -\dfrac{B_c}{B_0\rho}, & \text{for correctors} \\ -\dfrac{g\Delta x}{B_0\rho}, & \text{for quadrupoles.} \end{cases} \qquad (2.15)$$

3. CLOSED ORBIT

The equations (2.11) are Hill's equations with a non-zero right-hand side. Its general solution can be represented as a sum of the general solution of the homogeneous equation and a particular solution of the non-homogeneous one. The latter can be taken as periodic, keeping in mind the accelerator symmetry. This periodic particular solution will describe the closed orbit as long as the general solution of the homogeneous equation describes the particle oscillation.

The particle oscillations in an accelerator can be described by the Twiss's amplitude function $\beta(s)$ [2]

$$x(s) = a \sqrt{\beta(s)} \cos \left(Q \int_0^s \frac{ds}{Q\beta(s)} \right),$$ (3.1)

where Q is the number of the betatron oscillations per turn.

Let us introduce the new variables:

a) generalized azimuth

$$\varphi = \int_0^s \frac{ds}{Q\beta(s)},$$ (3.2)

b) normalized deviation

$$\eta = \frac{x}{\sqrt{\beta(s)}}.$$ (3.3)

In these new variables the oscillations (3.1) can be written as

$$\eta(\varphi) = a \cos(Q\varphi).$$ (3.4)

By implication we shall describe the closed orbit using the variables (3.2), (3.3).

Using the properties of the β function it is possible to transform the equation (2.11) to an equation of forced oscillations

$$\frac{d^2\eta}{d\varphi^2} + Q^2\eta = Q^2 f(\varphi),$$ (3.5)

where

$$f(\varphi) = \beta^{3/2}(\varphi)F(\varphi).$$ (3.6)

Using the method of varying the integration constants the following periodical particular solution of (3.5) can be obtained [2]

$$\eta = \frac{Q}{2 \sin \pi Q} \int_\varphi^{\varphi + 2\pi} f(t) \cos Q(\varphi + \pi - t)dt.$$ (3.7)

The integral representation (3.7) is the base formula for the closed orbit description.

There exist two approaches to the closed orbit treatment:

a) A matrix approach

As the perturbations are equal to zero out of the magnetic elements ($f(\varphi) = 0$) and the elements are short enough compared to the accelerator circumference ($\Delta\varphi \ll 2\pi$), the integral (3.7) can be transformed to the sum

$$\eta(\varphi_i) = \frac{Q}{2 \sin \pi Q} \sum_{\substack{j=1 \\ \varphi_i \le \varphi_j \le \varphi_i + 2\pi}}^{M+L} \bar{f}_i \, \Delta \varphi_j \cos Q(\varphi_i + \pi - \varphi_j). \tag{3.8}$$

Here M is the total number of the dipoles, L is the number of quadrupoles and the bar above a symbol means averaging over the magnetic element.

It is convenient to put (3.8) in the matrix form

$$\eta_i = \sum_{\substack{j=i \\ \varphi_i \le \varphi_j \le \varphi_i + 2\pi}}^{M+L} A_{ij} \delta_j, \tag{3.9}$$

where

$$A_{ij} = \cos Q(\varphi_i + \pi - \varphi_j) \tag{3.10}$$

and

$$\delta_j = \begin{cases} -\dfrac{Q \, \overline{\beta_j^{3/2} \Delta \varphi_j}}{2 \sin \pi Q} \Delta B_j - \dfrac{Q \, \overline{\beta_j^{3/2} \eta_j \Delta \varphi_j}}{2 \sin \pi Q \rho_j^2} \Delta x_j, & \text{for dipoles} \\[4ex] -\dfrac{Q \, \overline{\beta_j^{3/2} \Delta \varphi_j}}{2 \sin \pi Q} B_{cj}, & \text{for correctors} \\[4ex] -\dfrac{Q \, \overline{\beta_j^{3/2} g_j \Delta \varphi_j}}{2 \sin \pi Q B \rho} \Delta x_j, & \text{for quadrupoles} \end{cases} \tag{3.11}$$

$$\delta = \frac{\sqrt{\beta}}{2 \sin \pi Q} \varepsilon, \tag{3.12}$$

ε being the kick in the element.

The reason for introducing the generalized perturbations δ_i is that they have the same dimensions as the normalized orbit η, i.e., $m^{1/2}$.

b) Harmonic analysis approach

The orbit $\eta(\varphi)$ is periodic with a period of 2π. Let us expand it in a Fourier series

$$\eta(\varphi) = \frac{u_0}{2} + \sum_{k=1}^{\infty} (u_k \cos k\varphi + v_k \sin k\varphi). \tag{3.13}$$

Let us also expand the perturbations $f(\varphi)$ in a Fourier series

$$f(\varphi) = \frac{f_0}{2} + \sum_{k=1}^{\infty} (f_k \cos k\varphi + g_k \sin k\varphi). \tag{3.14}$$

From the equation (3.5) the following relation between the Fourier coefficients of the orbit and the perturbations can be deduced

$$u_0 = f_0$$

$$u_k = \left(\frac{Q^2}{Q^2 - k^2} \right) f_k, \qquad v_k = \left(\frac{Q^2}{Q^2 - k^2} \right) g_k. \qquad (3.15)$$

Of course, the matrix and the harmonic orbit treatments give the same results which can be demonstrated using the formula

$$\cos Q(\varphi_i + \pi - \varphi_j) = (\sin \pi Q) \frac{2Q}{\pi} \left(\frac{1}{2Q^2} - \sum_{k=1}^{\infty} \cos k \, \frac{(\varphi_i - \varphi_j)}{k^2 - Q^2} \right). \qquad (3.16)$$

4. ERROR SOURCES

The linear perturbations causing the closed orbit distortion can be summarized in the following way.

a) Constant errors.

   i) Errors in the coercive force.

This kind of errors can be estimated approximately by

$$\Delta B \approx \mu_0 \frac{l_{st}}{l_a} \Delta H_c, \qquad (4.1)$$

where l_{st} is the magnet core length and l_a is the aperture.

   ii) Errors due to eddy currents.

   iii) Stray magnetic fields.

   iv) Earth magnetic field.

b) Errors proportional to the main magnetic field.

   i) Permeability errors. Approximately

$$\frac{\Delta B}{B} \approx \left(\frac{l_{st}}{\mu_r^2 l_a} \right) \Delta \mu. \qquad (4.2)$$

   ii) Errors in the magnet core length

$$\frac{\Delta B}{B} \approx - \frac{1}{\mu_r l_a + l_{st}} \Delta l_{st}. \qquad (4.3)$$

iii) Aperture errors

$$\frac{\Delta B}{B} \approx -\frac{1}{\dfrac{l_{st}}{m_r} + l_a} \Delta l_a.$$ (4.4)

iv) Adjustment errors — element misalignments, median plane displacements, dipole tilts, quadrupole magnetic centre displacements, errors in the coil positions, ets.

v) Ground movement.

c) Errors appearing only in high fields — these are errors due to saturation.

d) Fast noise: these are ground movements and power supplies ripples in the 1—100 Hz bandwidth which cause fluctuations of the orbit.

5. STATISTICAL CHARACTERISTICS OF THE PERTURBATIONS AND THE ORBIT

The perturbations ΔB and Δx are random functions of the generalized azimuth φ. Keeping in mind that they are non-zero only within the magnetic elements and that these elements are short enough compared to the accelerator circumference ($\Delta \varphi < 2\pi$), one can represent perturbations as sums of elementary random functions — $\Delta B_i \, \Pi_i(\varphi)$ and $\Delta x_i \, \Pi_i(\varphi)$, where ΔB_i and Δx_i are random values and $\Pi_i(\varphi)$ are single rectangular pulses of length $\Delta \varphi_i$. The random values ΔB_i and Δx_i are uncorrelated, normally distributed, with a zero mathematical expectation and with standard deviations $\sigma_{\Delta B}$, $\sigma_{\Delta x}$ equal for all elements.

Let's expand the perturbations in a Fourier's series

$$\Delta B(\varphi) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\,\varphi + b_k \sin k\,\varphi).$$ (5.1)

Fourier's amplitudes a_k and b_k are random values. Their mathematical means vanish and their variances are

$$D(a_0) = \sigma_{\Delta B}^2 \sum_{j=1}^{M} \left(\frac{\Delta \varphi_j}{2\pi}\right)^2,$$

$$D(a_k) = \sigma_{\Delta B}^2 \sum_{j=1}^{M} \left(\frac{\cos k\,\varphi_j \, \Delta \varphi_j}{\pi}\right)^2,$$

$$D(b_k) = \sigma_{\Delta B}^2 \sum_{j=1}^{M} \left(\frac{\sin k\, \varphi_j\, \Delta\, \varphi_j}{\pi} \right)^2. \tag{5.2}$$

In uniform magnetic structures consisting of p equal periods, the formulae (5.2) are simplified leading to a «white» spectrum

$$D(a_k) = D(b_k) = \frac{1}{2} D(a_0) = \frac{\sigma_{\Delta B}^2 P}{2\pi^2} \sum_{j=1}^{M_p} \Delta\, \varphi_j^2. \tag{5.3}$$

In (5.3) M_p is the number of dipoles per period. Using the accelerator symmetry it is possible to demonstrate as well that $E(a_k b_k) = 0$, i.e., that a_k and b_k are non-correlated. Analogical results can be obtained for Δx.

Let us now discuss the statistical characteristics of the orbit itself.

From (3.9) it follows that the orbit deviation η_i as a sum of normally distributed values is normally distributed itself, with an expected value of zero and a variance

$$D(\eta_i) = \left(\frac{Q}{2B_0\, \rho\, \sin \pi Q} \right)^2 \left[\sigma_{\Delta B}^2 \sum_{j=1}^{M} \beta_j^3\, \Delta\, \varphi_j^2\, \cos^2 Q(\varphi_i + \pi - \varphi_j) + \right.$$

$$\left. + \sigma_{\Delta x}^2 \sum_{j=1}^{L} g_j^2\, \beta_j^3\, \Delta\, \varphi_j^2\, \cos^2 Q(\varphi_i + \pi - \varphi_j) \right]. \tag{5.4}$$

In uniform magnetic structures, (5.4) is simplified

$$D(\eta_i) = \left(\frac{Q}{2B_0\, \rho\, \sin \pi Q} \right)^2 p\, \sigma_{\Delta B}^2 \sum_{j=1}^{M_p} \frac{\beta_j^3\, \Delta\, \varphi_j^2}{2} + $$

$$+ \left(\frac{Q}{2B_0\, \rho\, \sin \pi Q} \right)^2 p\, \sigma_{\Delta x}^3 \sum_{j=1}^{L_p} \frac{g_j^2\, \beta_j^2\, \Delta\, \varphi_j^2}{2}. \tag{5.5}$$

In the same way the statistical characteristics of the orbit divergence $\eta'(\varphi)$ can be calculated and one obtains that for uniform structures

$$D(\eta') = Q^2 D(\eta). \tag{5.6}$$

Successful accelerator operation requires that the maximum orbit deviation η_{max} and its statistical characteristics be known.

As for every accelerator from a large group of accelerators with different random error distributions, the maximum orbit deviation η_{max} appears at different points φ_{max}; η_{max} is not normally distributed.

It is very difficult to obtain the statistical characteristics of η_{max} in the general case. As a rough approximation let us restrict ourselves to the major $k \approx Q$ harmonic in the orbit expansion (3.12)

$$\eta(\varphi) \approx u_k \cos k\,\varphi + v_k \sin k\,\varphi = A \cos(Q\,\varphi + \alpha_k). \tag{5.7}$$

It is easy to see that in this approximation $E(\eta, \eta') = 0$, i.e., η and η' are statistically independent.

We have for the two-dimensional probability [3]

$$p\left(\eta, \frac{\eta'}{Q}\right) = \frac{1}{2\pi\sigma_\eta^2} \exp\left(-\frac{A^2}{2\sigma_\eta^2}\right). \tag{5.8}$$

Obviously the orbit amplitude probability $p(A)$ can be obtained by integrating (5.8) along a circle with radius A in the plane $\left(\eta, \dfrac{\eta'}{Q}\right)$

$$p(A) = \int_0^{2\pi} p\left(\eta, \frac{\eta'}{Q}\right) A d\varphi = \frac{A}{\sigma_\eta^2} \exp\left(-\frac{A^2}{2\sigma_\eta^2}\right). \tag{5.9}$$

Formula (5.9) represents Rayleigh's probability distribution.

$$p(A) = \frac{2A}{\sigma_A^2} \exp\left(-\frac{A^2}{\sigma_A^2}\right) \tag{5.10}$$

with

$$\sigma_A = \sqrt{2}\,\sigma_\eta = \sqrt{2}\,\sigma_{u_k} = \sqrt{2}\,\sigma_{v_k}. \tag{5.11}$$

The integral distribution function for the Rayleigh distribution is:

$$\phi(A) = 1 - \exp\left(-\frac{A^2}{\sigma_A^2}\right). \tag{5.12}$$

A better estimate will be achieved if two contributing harmonics $k \langle Q \langle (k+1)$ are taken into account.

In this approximation

$$\eta(\varphi) \approx u_k \cos k\,\varphi + v_k \sin k\,\varphi + u_{k+1} \cos(k+1)\,\varphi + v_{k+1} \sin(k+1)\,\varphi =$$

$$= (u_k + u_{k+1} \cos \varphi + v_{k+1} \sin \varphi) \cos k\,\varphi +$$

$$+ (v_k - u_{k+1} \sin \varphi + v_{k+1} \cos \varphi) \sin k\,\varphi, \tag{5.13}$$

i.e., we have a harmonic oscillation with a slowly changing amplitude $A(\varphi)$

$$A^2(\varphi) = (u_k + u_{k+1} \cos\varphi + v_{k+1} \sin\varphi)^2 +$$

$$+ (v_k - u_{k+1} \sin\varphi + v_{k+1} \cos\varphi)^2. \tag{5.14}$$

From (5.14) the maximum amplitude can be calculated

$$A = \max A(\varphi) = r_k + r_{k+1}, \tag{5.15}$$

where

$$r_k^2 = u_k^2 + v_k^2; \quad r_{k+1}^2 = u_{k+1}^2 + v_{k+1}^2. \tag{5.16}$$

As it has already been demonstrated above, τ_k and τ_{k+1} which are the random amplitudes of the kth and $(u+1)_{th}$ harmonics have Rayleigh's distributions. These amplitudes are statistically independent. From these assumptions the following expression for the variance of the maximum amplitude can be derived

$$\sigma_A^2 = R_k^2 + \frac{\pi}{2} R_k R_{k+1} + R_{k+1}^2, \tag{5.17}$$

where

$$R_k^2 = 2\sigma_{u_k}^2; \quad R_{k+1}^2 = 2\sigma_{u_{k+1}}. \tag{5.18}$$

In Ref.4 it has been also shown that the probability for the orbit amplitude to be greater than A is

$$F(A) = 2\sqrt{\pi} \, \frac{R_k R_{k+1}}{R_k^2 + R_{k+1}^2} \, \frac{A}{\sqrt{R_k^2 + R_{k+1}^2}} \exp\left(-\frac{A}{R_k^2 + R_{k+1}^2}\right). \tag{5.19}$$

In Ref.4 it has been also shown that the best agreement between the analytical estimates and the results of the computer simulation of the closed orbit can be achieved if the major three or four harmonics of the perturbations are taken into account. As the exact solution in this approximation is accompanied by great mathematical difficulties only an approximate estimation for the orbit amplitude variance is carried out

$$\frac{\sigma_A^2}{\Sigma R_j^2} = \left(1 + \frac{\pi}{2} \frac{\overset{i \neq j}{\underset{i,j}{\sum}} R_i R_j}{\Sigma R_j^2}\right), \tag{5.20}$$

$$F(A) = (4\pi)^{m-1/2} \prod_{j=1}^{m} \left(\frac{R_j}{\sqrt{\Sigma R_j^2}} \right) \exp\left(-\frac{A^2}{\Sigma R_j^2} \right) \times$$

$$\times \sum_{i=0}^{m-1/2} \frac{(-1)^i (m-1)!}{2^{2i} i! (m-1-2i)!} \left(\frac{A}{\sqrt{\Sigma R_j^2}} \right)^{m-1-2j}, \qquad (5.21)$$

where m is the number of used harmonics.

6. A GENERAL DESCRIPTION
OF THE ORBIT CORRECTION METHODS

In accelerators as a rule the number of beam position monitors (N) is less than the numbers of dipoles (M) and quadrupoles (L) — $N < M + L$. This means that from the readings of the BPMs we can calculate only a part of the perturbations correction with uncertainty.

The general block diagram of the orbit correction is shown in Fig.1. Figure 2 gives a classification of the orbit correction methods.
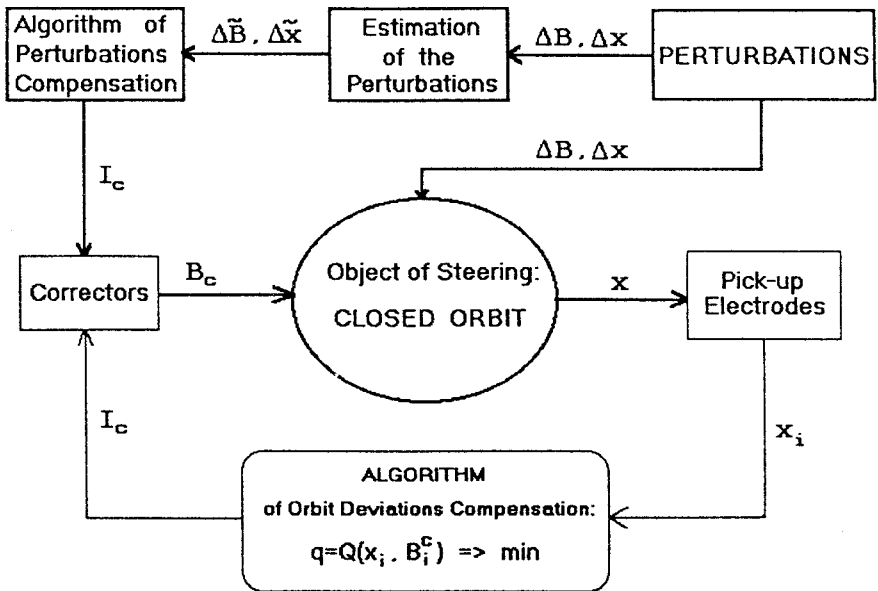

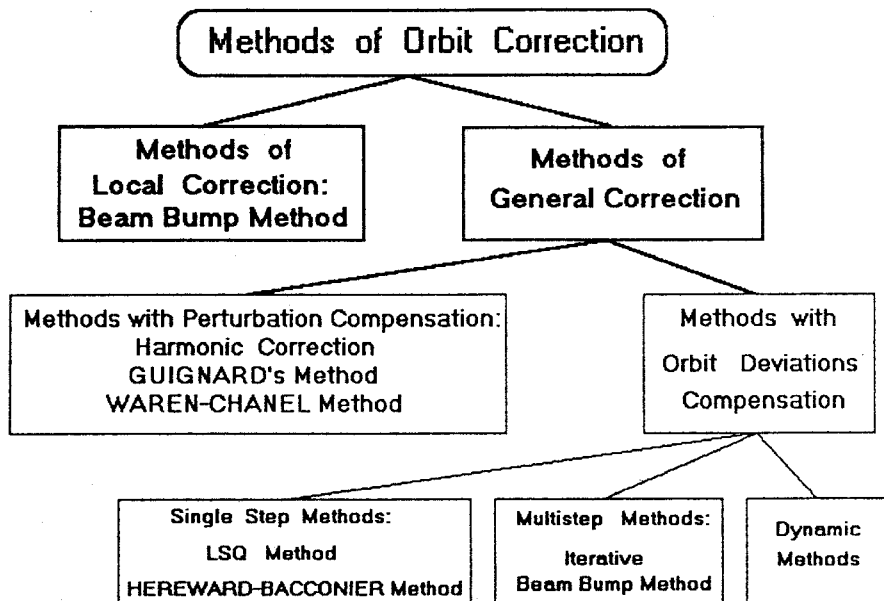Fig.1. Block diagram of orbit correction

```
┌─────────────────────────────────────┐
│      Methods of Orbit Correction     │
└─────────────────────────────────────┘
```

Methods of Local Correction: Beam Bump Method

Methods of General Correction

Methods with Perturbation Compensation:
Harmonic Correction
GUIGNARD's Method
WAREN–CHANEL Method

Methods with Orbit Deviations Compensation

Single Step Methods:
LSQ Method
HEREWARD–BACCONIER Method

Multistep Methods:
Iterative
Beam Bump Method

Dynamic Methods

Fig.2. Classification of orbit correction methods

The orbit correction methods can be divided into two main groups: methods for local correction and methods for correction of the orbit over the whole ring (general correction).

The local correction methods correct the orbit only in a part of the accelerator circumference. Outside of this region, the orbit is unchanged.

The methods for general correction cover methods that try to compensate perturbations directly and methods in which a kind of goal function, depending on the orbit deviations and corrector strengths, is minimalized.

The methods with perturbation compensation try to assess the perturbation strengths.

In the harmonic correction, this is achieved through the approximate values of the first perturbation harmonics. In the method proposed by G.Guignard and its later improvements, the «probability» of having errors in a given area of the accelerator circumference is found. Finally, in the Warren and Channell's suggestion, the values of all the errors are calculated, applying a special measurement procedure.

After the perturbations have been assessed, corrections with proper values and positions are applied in order to compensate the orbit deviations.

In the methods with orbit deviation compensation a quality criterion is created. It can be either a function or a functional, from the orbit deviations and correctors strengths. Then the minimum of the quality criterion is chosen. The single-step, the multi-step, and the dynamical methods are included in the group.

In the single-step methods, the orbit is considered only at the points where BPMs are situated and is characterized by the state vector $\mathbf{x} \equiv (x_1, x_2, \ldots < x_n)$, x_i being orbit deviations. A goal function

$$q = Q(\mathbf{x}, \mathbf{B}_c) \tag{6.1}$$

is created.

In (6.1), $\mathbf{B}_c$ is a vector whose components are corrector strengths. They are determined so that q has a minimum.

In the multi-step methods, the whole accelerator circumference is divided into separate areas. At first, the orbit over the first area is corrected. After that we go to the next area taking into consideration the results from the correction in the former step and so on.

The dynamical methods of correction treat the orbit as a whole curve $\eta(\varphi)$ rather than just orbit deviations in BPMs $\eta_i = \eta(\varphi_i)$. The quality criterion is a functional

$$I = \int_0^{2\pi} Q(\eta(\varphi), B_c(\varphi)d\varphi. \tag{6.2}$$

The condition for I to have a minimum gives us the chosen corrector strengths.

7. CORRECTION METHODS
WITH COMPENSATION OF THE PERTURBATIONS

7.1. Harmonic Correction Method. In the harmonic correction method, the orbit Fourier spectrum

$$\eta(\varphi) = \frac{u_0}{2} + \sum_{k=1}^{\infty} (u_k \cos k\,\varphi + v_k \sin k\,\varphi) \tag{7.1.1}$$

is calculated at first.

As we know, the orbit deviations are only in the points where BPMs are situated and not on the whole curve $\eta(\varphi)$; the methods of the applied harmonic analysis have to be used. Measuring the orbit with $2N$ BPMs we can determine only the approximate values of the first N orbit Fourier harmonics. In the case

of uniformly situated monitors the simplest way is to approximate first orbit harmonics with the so-called Bessel coefficients [5]

$$u_0 \approx U_0 = \frac{1}{N} \sum_{i=1}^{2N} \eta_i,$$

$$u_k \approx U_k = \frac{1}{N} \sum_{i=1}^{2N} \eta_i \cos k\, \varphi_i, \quad (k = 1,2,...,N),$$

$$v_k \approx V_k = \frac{1}{N} \sum_{i=1}^{2N} \eta_i \sin k\, \varphi_i, \quad (k = 1,2,...,(N-1)). \tag{7.1.2}$$

In the general case of non-equidistant monitors the rule of trapezoids for integral approximation gives

$$u_k \approx \frac{1}{\pi} \sum_{i=1}^{N} \eta_i \cos k\, \phi_i \left(\frac{\phi_{i+1} - \phi_{i-1}}{2} \right),$$

$$v_k \approx \frac{1}{\pi} \sum_{i=1}^{N} \eta_i \sin k\, \phi_i \left(\frac{\phi_{i+1} - \phi_{i-1}}{2} \right),$$

$$\phi_0 = \phi_N - 2\pi, \qquad \phi_{N+1} = \phi_1 + 2\pi. \tag{7.1.3}$$

In this general case the LSQ criterion of approximation

$$\sum_{i=1}^{p} \left[\eta_i - \sum_{k=0}^{n} c_k \, \psi_k(\phi_i) \right]^2 \rightarrow \text{Min}, \tag{7.1.4}$$

where c_k are the Fourier coefficients and ψ_k is the system of trigonometrical functions, gives

$$\begin{bmatrix} u_0 \\ U_1 \\ \cdot \\ \cdot \\ \cdot \\ u_N \\ v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_{N-1} \end{bmatrix} = (S^T S)^{-1} S^T \begin{bmatrix} \eta_1 \\ \eta_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \eta_{2N} \end{bmatrix} \tag{7.1.5}$$

where

$$
S = \begin{bmatrix}
1/2 & \cos\phi_1 & \cos 2\phi_1 & .. & \cos N\phi_1 & \sin\phi_1 & .. & \sin(N-1)\phi_1 \\
1/2 & \cos\phi_2 & \cos 2\phi_2 & .. & \cos N\phi_2 & \sin\phi_2 & .. & \sin(N-1)\phi_2 \\
. & . & . & .. & . & . & .. & . \\
1/2 & \cos\phi_{2N} & \cos 2\phi_{2N} & .. & \cos N\phi_{2N} & \sin\phi_{2N} & .. & \sin(N-1)\phi_{2N}
\end{bmatrix}. \tag{7.1.6}
$$

Knowing the orbit harmonics, the perturbation harmonics can be calculated using (3.14).

In the harmonic correction method, the fields in $2N$ correcting dipoles are chosen so that the first N perturbations harmonics to be canceled [6—8].

Keeping in mind that the correcting dipoles are short ($\Delta\varphi_c \ll 2\pi$) and transforming integrals to sums, we are to solve therefore the following system of equations

$$
\frac{1}{\pi} \sum_{i=1}^{2N} f_i^c \Delta\varphi_i^c = U_0,
$$

$$
\frac{1}{\pi} \sum_{i=1}^{2N} f_i^c \cos k\,\varphi_i \Delta\varphi_i = \frac{Q^2 - k^2}{Q^2} U_k,
$$

$$
\frac{1}{\pi} \sum_{i=1}^{2N} f_i^c \sin k\,\varphi_i \Delta\varphi_i = \frac{Q^2 - k^2}{Q^2} V_k. \tag{7.1.7}
$$

Then the fields in the correcting dipoles B_i^c, $i = 1,2,...,2N$ can be calculated from the generalized perturbations using (3.6), (2.14), (2.15).

7.2. G.Guignard's Method. It is very important for the biggest error sources to be found. Then we can check more carefully the corresponding elements and the practice shows that in many cases after removing this very rough imperfection the maximum orbit deviation is reduced to sufficiently small values. As in an error-free area of the accelerator circumference, the RHS of equation (3.5) is equal to zero. The orbit in this error-free area is

$$
\eta = A \cos Q\varphi + B \sin Q\varphi + C. \tag{7.2.1}
$$

The orbit (7.2.1) should be matched to the whole orbit in both ends of the area under consideration. From here the constants A, B, C are calculated.

If now we assume that there are errors in the above area, then the smooth solution (7.2.1) will not be valid any more. In the points with errors, η' will undergo kicks.

Obviously the divergence of the real measured orbit from the smooth model (8.2.1) will be a measure for the perturbation strengths in the area.

According to G.Guignard [9] we shall define the «probability» for the existence of perturbations in an area of the accelerator as

$$\Psi = \frac{1}{n} \sum_{k=p}^{p+n-1} [\sqrt{\beta_k} \, (A \cos Q\varphi_k + B \sin Q\varphi_k + c) - x_k]^2. \qquad (7.2.2)$$

In (7.2.2), p is the number of the first BPM and $n \geq 4$ is the whole number of BPMs included in the area.

The constant A, B, C in (7.2.2) defines the non-perturbated orbit. We shall determine their values so that the non-perturbated orbit will lie as close as possible to the readings x_k in the BPMs (in LSQ sense). In other words we calculate A, B, C from the condition $\Psi \to$ min.

In Guignard's method, we calculate the «probability» Ψ for successive areas of the accelerator ring. The areas with big Ψ values are possible sources of errors.

7.3. Warren and Channell's Suggestion. As this has been already mentioned in general the number of BPMs, N, is less than the numbers of dipoles — M and quadrupoles — L, $N < M + L$. Therefore we are not able to solve the system of equations (3.9) with respect to the errors. In Ref.10 Warren and Channell suggest enlarging this system adding some new equations. For this purpose, new measurements of the orbit have to be done changing synchrotron parameters and keeping the errors unchanged. A new set of orbit measurements can be carried out changing the quadrupole strengths (i.e., changing Q) or changing the signs of the quadrupole gradients (for example reducing FODO structure into DOFO). The enlarged system of equations is solved by the LSQ method.

8. LOCAL CORRECTION METHODS

8.1. Beam-Bump Method. At some places (injection, extraction, and others) a very high degree of orbit correction is necessary. From the other hand, at some places of the accelerator ring the orbit deviation may be much greater than the average one due to strong local stray fields, strong local imperfections or ground movement. These require the development of methods for local correction.

Such local correction method is the beam-bump method, suggested by Collins in the sixties [11—14].

The idea of the beam-bump method consists in a local orbit correction by means of three correcting dipoles (Fig.3). The orbit deviation is compensated in a BPM or pick-up electrodes (PUE) situated near the middle corrector, as long as out of the area occupied by the correctors the orbit is kept unchanged. Therefore we have the conditions
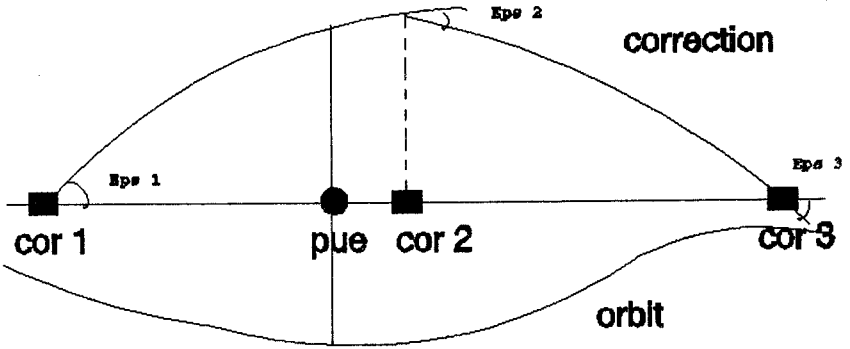
Fig.3. Beam-bump correction

$$x(\varphi_{PUE}) = - x_m,$$

$$x(\varphi \langle \varphi_1, \varphi \rangle \varphi_3) \equiv x'(\varphi \langle \varphi_1, \varphi \rangle \varphi_3) \equiv 0. \tag{8.1.1}$$

From (8.1.1) and (3.9) the following system of three equations about corrector strength can be deduced

$$\sin \mu_{12} \sqrt{\beta_2} \, \varepsilon_2 + \sin \mu_{12} \sqrt{\beta_3} \, \varepsilon_3 = 0,$$

$$\sqrt{\beta_1} \, \varepsilon_1 + \cos \mu_{12} \sqrt{\beta_2} \, \varepsilon_2 + \cos \mu_{13} \sqrt{\beta_3} \, \varepsilon_3 = 0,$$

$$(\text{ctg } \pi Q \cos \mu_{1pue} + \sin \mu_{1pue}) \sqrt{\beta_1} \varepsilon_1 +$$

$$+ (\text{ctg } \pi Q \cos \mu_{pue2} + \sin \mu_{pue2}) \sqrt{\beta_2} \varepsilon_2 +$$

$$+ ((\text{ctg } \pi Q \cos (\mu_{pue2} + \mu_{23}) +$$

$$+ \sin (\mu_{pue2} + \mu_{23})) \sqrt{\beta_3} \varepsilon_3 = \frac{2x_m}{\sqrt{\beta_{pue}}}, \tag{8.1.2}$$

where

$$\varepsilon = \frac{B_c \Delta s}{B\rho} \tag{8.1.3}$$

is the bump (kick) in the correcting dipole,

$$\mu_{12} = \int_{s1}^{s2} \frac{ds}{\beta(s)} = Q(\varphi_2 - \varphi_1) \tag{8.1.4}$$

is the betatron phase advance.

For the speciaL case of a BPM situated very close to the central corrector, i.e., when $\mu_{pue2} \approx 0$, the system (8.1.2) is reduced to

$$\varepsilon_1 = -\frac{\eta_m}{\sqrt{\beta_1}\,\sin\mu_{12}}, \qquad \varepsilon_2 = \frac{\sin(\mu_{12}+\mu_{23})}{\sin\mu_{12}\,\sin\mu_{23}}\,\frac{\eta_m}{\sqrt{\beta_2}\,\sin\pi Q},$$

$$\varepsilon_3 = -\frac{\eta_m}{\sqrt{\beta_3}\,\sin\mu_{23}}. \tag{8.1.5}$$

In the completely symmetrical case $\mu_{12}=\mu_{23}-\mu$, $\mu_{pue}=0$

$$\sqrt{\beta_1}\,\varepsilon_1 = \sqrt{\beta_2}\,\varepsilon_2 = \frac{x_m}{\sqrt{\beta_{pue}}\,\sin\mu},$$

$$\sqrt{\beta_2}\,\varepsilon_2 = -\frac{2x_m}{\sqrt{\beta_{pue}}}\,\mathrm{ctg}\,\mu. \tag{8.1.6}$$

Besides using the expression (3.9) about the closed orbit, we can obtain the system (9.1.2) in another way, by using the beam transport matrix between two arbitrary points s_1 and s_2 in Twiss' form

$$M_{12}=\begin{bmatrix} \sqrt{\dfrac{\beta_2}{\beta_1}}\,(\cos\mu_{12}+\alpha_1\sin\mu_{12}) & \sqrt{\beta_1\beta_2}\,\sin\mu_{12} \\[2mm] \dfrac{1}{\sqrt{\beta_1\beta_2}}\,(-\sin\mu_{12}-\alpha_1\alpha_2\sin\mu_{12}+(\alpha_1-\alpha_2)\cos\mu_{12} & \sqrt{\dfrac{\beta_1}{\beta_2}}\,(\cos\mu_{12}-\alpha_2\sin\mu_{12}) \end{bmatrix}. \tag{8.1.7}$$

Let M^1 be the transport matrix from CD_1 to PUE; M^2 — from CD_1 to CD_2 and M^3 — from CD_2 to CD_3. Then from (9.1.1) it follows that

$$\varepsilon_1 = \frac{x_m}{m_{12}^1}, \qquad \varepsilon_2 = -\frac{m_{11}^3 m_{12}^2}{m_{12}^3}\,\varepsilon - m_{22}^2 \varepsilon_1,$$

$$\varepsilon_3 = -m_{12}^2 m_{21}^3 \varepsilon_1 - m_{22}^3(m_{22}^3\varepsilon_1^2 + \varepsilon_2). \tag{8.1.8}$$

The optimum phase distance between the correctors in the triple is $\mu \approx \dfrac{\pi}{2}$, when $\varepsilon_2 \approx 0$.

8.2. Generalizations of the Beam-Bump Correction Method.
As a rule, the magnetic structure of storage rings and colliders is irregular. However, if the phase distance between BPM and CD_2 is big (Fig.3), the beam-bump method

correcting the orbit in the monitor closely to zero will increase the orbit devia-
tion in CD_2. This happens when correctors and monitors are situated non-regu-
larly (because of the lack of place) or when the betatron phase changes quite
quickly (in the areas with small β). That is why some improvements of the
beam-bump method have been put forward. Another reason for such impro-
vements is the fact that in the beam-bump method $\varepsilon_2 \sim 1/\sin\mu$ and in the case
when $\mu \approx k \cdot \pi$; $k = 1,2, \ldots$ quite large corrector kicks will be necessary for a full
correction to be fulfilled.

In Ref.15, groups of N monitors and $K = N + 2$ correctors are used. The
BPMs are considered to be situated close to the corresponding internal cor-
rectors.

The kicks of the internal correctors are determined by

$$R\varepsilon_{int} = -\, \mathbf{x}_{pue}, \qquad (8.2.1)$$

where the matrix R is

$$R_{ij} = \frac{\sqrt{\beta_i \beta_j}\ \sin_{jk} \sin_{1i}}{\sin \mu_{1k}}. \qquad (8.2.2)$$

In (8.2.2), μ_{1i} is the phase advance between the first corrector and i-th
internal corrector; μ_{jk} is the phase advance between the j-th internal corrector
and the last corrector; μ_{1k} is the phase advance for the whole group of
correctors.

Eq. (8.2.1) is a generalization of eq. (8.1.2) for the case of $(N+2)$ cor-
rectors and N BPMs.

The two end kicks are determined in a way to keep the orbit out of the
group unchanged

$$\sqrt{\beta_1}\ \varepsilon_1 = \frac{x_{pue1}}{\sqrt{\beta_{pue1}}\ \sin \mu_{1pue1}}, \qquad (8.2.3)$$

$$\sqrt{\beta_k}\ \varepsilon_k = \frac{x_{pueN}}{\sqrt{\beta_{pueN}}\ \sin \mu_{pueNk}}. \qquad (8.2.4)$$

In the computer program PETROC (an improved CERN variant of the
HERA's PETROS) another beam-bump improvement is done [16,17]. One can
use K correctors, k being 0, 3, 4, 5. Between the correctors $N \le 2k$ beam position
monitors are situated. No more than two BPMs can be placed between two
correctors. Only three correctors are active (with non-zero kicks) — the first,
the second, and the last. Let ε be the kick in the first active corrector.

In order for the orbit not to be changed, the two other active kicks have to be εr_2 and εr_k where

$$r_2 = -\sqrt{\frac{\beta_1}{\beta_2}}\,\frac{\sin \mu_{1k}}{\sin \mu_{2k}}, \tag{8.2.5}$$

$$r_K = -\sqrt{\frac{\beta_1}{\beta_k}}\,\frac{\sin \mu_{12}}{\sin \mu_{2k}}. \tag{8.2.6}$$

The kick ε in the first corrector is determined by the minimum of the function

$$q = \sum_{\text{BPMs}} (x_i + x_{\text{pue}i})^2 + w^2 \beta_1 \beta_2 \sin^2\mu_{12}\varepsilon^2. \tag{8.2.7}$$

In (8.2.7), $x_{\text{pue}i}$ is the measured orbit deviation and x_i is the orbit deviations due to correctors; w is a weight. The second member in (8.2.7) limits the corrector strengths.

9. METHODS WITH ORBIT DEVIATION COMPENSATION

9.1. Iterative Beam-Bump Method. If we move consequently along the accelerator circumference, the local beam-bump method can be used as a method for general correction. Each corrector works once as a first corrector in the correctors triple (Fig.3), once as a second and once as a third corrector.

In this straightforward improvement, a little problem is hidden. In the beam-bump method, the correctors of the triple are tuned to cancel the orbit in the monitor which is situated as a rule near the middle corrector. Correctors affect the orbit locally, i.e., the impact of the correctors is equal to zero out of the triple. But as we pointed out above, in the iterative beam-bump the successive triples overlap a little (they have two common correctors). Hence, each corrector triple will cause a small orbit distortion in the next (belonging to the next triple) beam position monitor. The last correctors triple consisting of correctors $(k-1)$, k and 1 will reduce the orbit deviation in the k-th monitor to zero, but will distort the orbit a little in the first monitor.

To avoid this trouble, one more turn of successive beam-bump corrections along the accelerator will be necessary.

9.2. Least Squares Method. The method can be characterized as a single-step correction method.

The orbit is considered only in the points where beam position monitors are situated and is characterized by the state vector η^c corresponding to the corrections:

$$\eta = \eta^{(B+g)} + \eta^c = \eta^{(B+g)} + A\delta^c. \tag{9.2.1}$$

To write (9.2.1) we have used the formula (3.9), δ^c being the generalized corrections (3.11) and the fact that η^{B+g} are known from the BPMs readings.

In the least squares method (LSQ) the following goal function is minimized [8,18—20]

$$q = \sum_{i=1}^{N} \eta_i^2 = \eta^T\eta \rightarrow \min. \tag{9.2.2}$$

It is important to note that as a rule the number of correctors, k, is less than that of detectors, n, i.e., $K < N$. It is possible to be demonstrated that the minimum of (9.2.2) occurs when corrector strengths are

$$\delta_{opt}^c = -(A^T A)^{-1} A^T \eta^{(B+g)}. \tag{9.2.3}$$

The minimum value of q (the so-called residual sum of squares) is equal to

$$q_{\min} = \eta^{(B+g)T} \eta^{(B+g)T} A\delta_{opt}^c. \tag{9.2.4}$$

9.3. Fast Realization of the LSQ Correction. A fast realization of the LSQ correction method is proposed in Ref.15. The idea is at each iteration to optimize the orbit using only one corrector. Having tried in this way all available correctors, we choose for further use this one for which the residual sum of squares of orbit deviations is the smallest.

At the 0-th iteration we begin with the orbit measured by BPMs

$$\eta^{(0)} = \eta^{(B+g)}. \tag{9.3.1}$$

Let $\eta^{(n-1)}$ be the orbit having been optimized at the $(n-1)$-th iteration. As at each step at the n-th iteration, we will use only single correctors, therefore

$$\eta_j^{(n)} = \eta^{(n-1)} + \delta_j^{(n)} a_j, \tag{9.3.2}$$

where $\delta_j^{(n)}$ is the generalized strength of the j-th corrector at n-th iteration; a_j is the j-th column of the matrix (3.10).

Following the LSQ algorithm we will minimize

$$g_j^{(n)} = \min_{\delta_j^{(n)}} (\eta_j^{nT}\eta_j^{(n)}). \tag{9.3.3}$$

Thus at the n-th iteration, we will obtain k values $q_j^{(n)}$, $j = 1,2,...,k$, one for each corrector. From these values $q_j^{(n)}$ we will choose the smallest one. The corresponding corrector will be the optimum corrector at the n-th iteration. Its

optimum strength $\delta_{j(n)}$ will be the correction used at the n-th iteration. Therefore our goal function is

$$q_j^{(n)} \xrightarrow[j]{} \min. \qquad (9.3.4)$$

In Ref.15 it is proved that this iterative procedure is convergent and that the corrector strengths found at different steps are added.

9.4. Algorithm MICADO. The MICADO algorithm for closed orbit correction was put forward by B.Autin and Y.Marti in Ref.21 and is realized in many accelerator design programs, for example in MAD [22]. This algorithm can be seen as an improvement of the LSQ method.

MICADO is an iterative algorithm. At the first iteration only single correctors are used. The goal function is as in the LSQ method

$$g_j^{(1)} = \min_{\delta j} (\eta_j^{(1)T}\eta_j^{(1)}), \qquad (9.4.1)$$

where

$$\eta_j^{(1)} = \eta^{(B+g)} + \alpha\delta_j^{(1)}, \qquad (9.4.2)$$

$\eta^{(B+g)}$ being the vector of BPM readings and $\delta_j^{(1)}$ being the vector of the generalized corrections (3.11). As at the first iteration we use only single correctors:

$$\delta_j^{(1)} = (0,..., \delta_j^{(1)},..., 0), \qquad (9.4.3)$$

therefore, at the first iteration we have to solve k equations with one unknown value — $\delta_j^{(1)}$, $(j = 1, 2,..., k)$.

Next we find out the number j^* of that corrector among all k correctors for which $q_j^{(1)}$ (9.4.1) has the smallest value — $q_{j^*}^{(1)}$.

It is j^*-th corrector that will be used at the second iteration together with each one of the rest $(k-1)$ correctors. We minimize again the square function

$$q_j^{(2)} = \min_{\delta j} (\eta_j^{(2)T}\eta_j^{(2)}). \qquad (9.4.4)$$

But now the corrections vector $\delta_j^{(2)}$ is

$$\delta_j^{(2)} = (0,..., \delta_{j^*},... \delta j,... 0), \qquad (9.4.5)$$

i.e., has two non-zero components: one, δ_{j^*}, is the strength of the j^*-th corrector (remember that this number was found and fixed at the previous step) and another one — the strength of any other corrector. Thus at the second iteration we have to solve $(k-1)$ systems of two equations with two unknown values — δ_{j^*} and δ_j.

Following the first iteration idea we will search now for the number j^{**} of that additional corrector for which $q_j^{(2)}$ has the smallest value. So that after the second iteration we will have fixed numbers j^* and j^{**} of two optimum correctors.

Iteration procedures of this kind go ahead, adding one new corrrector at every step, until the sum of the corrected orbit deviations becomes less than a given small value ε. Furthermore the method finds the smallest number of correctors satisfying this condition.

9.5. Hereward–Bacconier's Method. In this correction method the following goal function is introduced [23—26]

$$q = \gamma \sum_{i=1}^{N} \eta_i^2 + (1 - \gamma) \sum_{j=1}^{K} \delta_j^{c2} \to \min. \tag{9.5.1}$$

In (9.5.1), γ is a parameter, $0 < \gamma < 1$.

The first sum in (9.5.1) minimizes the orbit deviations in the BPMs, the second one restricts the strengths of the correctors, limiting in this way the influence of the high harmonics of the correction. We saw that from the readings in N BPMs the amplitudes of the first $N/2$ orbit harmonics can be calculated. These $N/2$ orbit harmonics are compensated by the corresponding harmonics of the correction. The higher harmonics of the correction remain uncompensated. As they distort the orbit additionally it is useful to restrict their influence.

The optimum value of the parameter γ is determined either experimentally or by computer simulations.

One can write

$$q = \gamma \sum_{i=1}^{N} \left(\sum_{j=1}^{K} a_{ij} \delta_j^c + \eta_i^{(B+g)} \right)^2 + (1 - \gamma) \sum_{j=1}^{K} \delta_j^{c2}. \tag{9.5.2}$$

The necessary condition for q to have a minimum is

$$\frac{\partial q}{\partial \delta_j^2} = 2\gamma \sum_{i=1}^{N} a_{ij} \sum_{p=1}^{K} (a_{ip} \delta_p^c + \eta_i^{(B+g)}) + 2(1 - \gamma) \delta_p^c = 0,$$

$$j = 1, 2, ..., k. \tag{9.5.3}$$

This can be written in matrix form

$$(\gamma A^T A + (1 - \gamma)E) \delta^c = - \gamma A^T \eta^{(B+g)}. \tag{9.5.4}$$

9.6. LSQ with Singular Value Decomposition. According to the linear algebra [27] any $(N \times K)$ matrix A with $N >= K$ can be represented as a product of three matrices:

$$A = U \cdot W \cdot V^T, \qquad (9.6.1)$$

where U is $(N \times K)$ unitary matrix $(U^T \cdot U = U \cdot U^T = E)$, W is a $(K \times K)$ diagonal matrix with positive or zero diagonal elements ($w_i >= 0$, called singular values) and V is a $(K \times K)$ unitary matrix $(V^T \cdot V = V \cdot V^T = E)$. (1) is known as SVD decomposition of the matrix A.

If the matrix A is square (equal numbers of monitors and correctors) and some w_i' s are equal to zero, A will be singular; in case some of the w_i' s are nonzero but very small, A will be illconditioned.

If $N > K$ (which is usual for accelerators), the system of equations for the corrector strengths

$$A \cdot \delta_{BC} = -\eta \qquad (9.6.2)$$

will be overdetermined.

Even in this two cases the SVD decomposition of A provides a «reasonable» solution of (2) (in LSQ sense).

It is proven [27] that the vector

$$\delta^{BC} = -V \cdot [\text{diag}\,(1/W_i)] \cdot U^T \cdot \eta, \qquad (9.6.3)$$

where for all w_i's equal to zero or very small, the corresponding diagonal elements $(1/w_i)$ in the second matrix of the matrix product (9.6.3) are replaced by zero ($\infty \to 0$!), satisfies

$$|A \cdot \delta^{BC} + \eta|^2 \to \text{Min}. \qquad (9.6.4)$$

If $U_{(i)}$ are the columns of the matrix U (N-vectors) and $V_{(i)}$ are the columns of the matrix V (K-vectors), then

$$\delta^{BC} = -\sum_i \left(\frac{U_{(i)}\eta}{w_i} \right) V_{(i)}, \qquad (9.6.5)$$

i.e., the corrector strengths are linear combination of the vectors $V_{(i)}$ with coefficients equal to the dot product of the vectors $U_{(i)}$ with the measured orbit η weighted by the singular values w_i.

In order to reduce the corrector strengths one can replace $(1/w_i)$ by zero not only for zero (or very small) w_i but when $w_i < \varepsilon$, where ε is chosen in a way to avoid the power supplies saturation [28].

The LSQ correction method with SVD decomposition of the response matrix A is coded in the computer program for orbit simulation and correction ORBIT [29] and has been used in the X-ray ring NSLS and in SPEAR [30].

9.7. Eigenvectors Correction. In the LSQ correction method we have to invert the matrix $(A^T A)$ — (9.2.3). $A^T A$ is square $(K \times K)$ symmetric real matrix, i.e., hermitian matrix, i.e., there exists a transformation $\Lambda = T^{-1}(A^T A)T$ to a square diagonal matrix Λ. The diagonal elements λ_i of the matrix Λ are the eigenvalues of the matrix $(A^T A)$, and the transformation matrix $T = (\mathbf{X}_1, \mathbf{X}_2,..., \mathbf{X}_K)$ is formed by the eigenvalues $\mathbf{X}_i$ of the matrix $(A^T A)$.

It follows from this transformation and from (9.2.3.) that [8]

$$\delta_i^{BC} = -\sum_{p=1}^{K} \sum_{q=1}^{K} \sum_{r=1}^{K} \frac{X_{p,i} X_{p,q}}{\lambda_p} A_{rq} \eta_r. \qquad (9.7.1)$$

Let us represent the corrector vector δ in the basis $\mathbf{X}_l$, $l = 1,..., K$

$$\delta = \sum_l C_l \mathbf{X}_l. \qquad (9.7.2)$$

Then we have for the residual sum of squares (9.2.4)

$$q_{min} = \delta^T \delta + \sum_l \lambda_l C_l \delta^T \mathbf{X}_l. \qquad (9.7.3)$$

The eigenvectors with small eigenvalues will have no significant contribution to the correction and may be neglected thus reducing the corrector strengths.

9.8. Dynamical Correction Algorithm. Algorithms described above strive to compensate orbit deviations in the points where BPMs are situated by means of some number of orbit correctors. As a result of the orbit correction, the corrected orbit will have approximately zero deviations in the BPMs. However, between the monitors the corrected orbit will have non-zero deviations. During the computer simulation we noticed that in some particular error distributions the deviations of the corrected orbit in some points are out of control. In general it is important to correct the orbit over the whole accelerator ring and not only in the BPMs.

Emphasizing this fact, we will choose the criterion of correction quality as a functional [31]

$$q = \frac{1}{2\pi} \int_0^{2\pi} \eta^2(\varphi) \, d\varphi. \qquad (9.8.1)$$

But the orbit $\eta(\varphi)$ is a random function of the generalized azimuth. This requires to improve the criterion (9.8.1) taking the mathematical expectation of the functional

$$q = M \left(\frac{1}{2\pi} \int_0^{2\pi} \eta^2(\varphi)d\,\varphi \right), \tag{9.8.2}$$

(9.8.2) is the final form of the correction quality criterion in the DINAM algorithm.

One can write

$$\frac{1}{\pi} \int_0^{2\pi} \eta^2(\varphi)\delta\varphi = \frac{u_0}{2} + \sum_{k+1}^{\infty} (u_k^2 + v_k^2) =$$

$$= \left[\frac{u_0^{(B+g)^2}}{2} + \sum_{k=1}^{\infty} \left(u_k^{(b+g)^2} + v_k^{(B+g)^2} \right) \right] +$$

$$+ \left[u_0^{(B+g)}u_0^{Bc} + \sum_{k=1}^{\infty} \left(2u_k^{(B+g)}u_k^{Bc} + 2v_k^{(B+g)}v_k^{Bc} \right) \right] +$$

$$+ \left[\frac{u_0^{Bc^2}}{2} + \sum_{k=1}^{\infty} \left(u_k^{Bc^2} + v_k^{Bc^2} \right) \right] = \Sigma_1 + \Sigma_2 + \Sigma_3. \tag{9.8.3}$$

In (9.8.3), $u_{k1}V_k$ are the Fourier coefficients of the orbit, $u_k^{(B+g)}$ and $v_k^{(B+g)}$ being those caused by the dipole and the quadrupole errors and u_k^{Bc} and v_k^{Bc} being those caused by the correctors.

Let us have $2N$ Beam position monitors, M-dipoles, L-quadrupoles, and K-correcting dipoles. As a result of the orbit measurement by BPMs, we shall know the orbit distortions in $2N$ points. However, the total number of the perturbations $(M+L)$ is, as a rule, much larger than the number of BPMs. Therefore, we cannot solve (3.9) about the perturbations. From this system of equations only $2N$ perturbations can be expressed by the readings in BPMs and other perturbations. In other words, we have a case of correction with uncertainty.

Let η_i, $i = 1,2,..., 2N$ be the orbit distortions in BPMs. Then we can write:

$$\eta_i = \frac{u_0}{2} + \sum_{k=1}^{\infty} (u_k \cos k\,\varphi_i + v_k \sin k\,\varphi_{i_{(i=1,2,...,2n)}}). \tag{9.8.4}$$

From these $2N$ equations, the first N Fourier Coefficients $u_0, u_1, ..., u_N, ..., v_{(n-1)}$ can be expressed by the orbit distortions η_i and higher orbit harmonics. Assuming that BPMs are placed uniformly along the accelerator ring, the following relations can be deduced from (9.8.4)

$$u_0 = U_0 - 2 \sum_{j=1}^{\infty} u_{2Nj} = U_0 - S_0,$$

$$u_k = U_k - \sum_{j=1}^{\infty} (u_{2Nj-k} + u_{2Nj+k}) = U_k - S_k,$$

$$v_k = V_k - \sum_{j=1}^{\infty} (-v_{2Nj-k} + v_{2Nj+k}) = V_k - R_k, \qquad (9.8.5)$$

where

$$U_0 = \frac{1}{N} \sum_{i=1}^{2N} \eta_i,$$

$$U_k = \frac{1}{N} \sum_{i=1}^{2N} \eta_i \cos k\,\varphi_i; \qquad (k = 1, 2, ..., N),$$

$$V_k = \frac{1}{N} \sum_{i=1}^{2N} \eta_i \sin k\,\varphi_i; \qquad (k = 1, 2, ..., (n-1)) \qquad (9.8.6)$$

are Bessel's coefficients.

Making use of the relations (9.8.5) one can write for the sums in (9.8.3):

$$\Sigma_1 = \left[\frac{U_0^2}{2} + \sum_{k=1}^{n-1} (U_k^2 + V_k^2) + \frac{U_N^2}{2} \right] +$$

$$+ \left[-U_0 S_0 + \sum_{k=1}^{N-1} (-2U_k S_k - 2V_k R_k) - U_N S_N \right] +$$

$$+ \left[\frac{S_0^2}{2} + \sum_{k=1}^{N-1} (S_k^2 + R_k^2) + S_N^2 - \frac{U_N^2}{4} + V_N^{(B+g)^2} + \right.$$

$$\left. + \sum_{k=N+1}^{\infty} \left(u_k^{(B+2)^2} + v_k^{(B+g)^2} \right) \right] = \Sigma_a + \Sigma_b + \Sigma_c, \qquad (9.8.7)$$

$$\sum_2 = \left[U_0 u_0^{Bc} + 2 \sum_{k=1}^{n-1} (U_k u_k^{Bc} + V_k v_k^{Bc}) + U_N u_N^{Bc} \right] +$$

$$+ \left[2v_N^{B+g} v_N^{Bc} - u_0^{Bc} S_0 - 2 \sum_{k=1}^{n-1} (u_k^{Bc} S_k + v_k^{Bc} R_k) - 2u_N^{Bc} S_N \right] +$$

$$+ \left[2 \sum_{k=N+1}^{\infty} (u_k^{(B+g)} u_k^{Bc} + b_k^{(B+g)} v_k^{Bc} \right] = \sum_d + \sum_e + \sum_f. \qquad (9.8.8)$$

Using the fact that the higher orbit harmonics are statistically independent and have a zero mathematical expectation we obtain that

$$M(\Sigma_d) = M(\Sigma_e) = M(\Sigma_f) = 0. \qquad (9.8.9)$$

Using the Fourier expansion (3.13) and the reactions (3.14) between the orbit and the perturbation harmonics one can also obtain that

$$M(\Sigma_c) = 4 \sum_{k=N+1}^{\infty} \left(\frac{Q^2}{Q^2 - k^2} \right)^2 D, \qquad (9.8.10)$$

where D is the variance of the perturbation harmonics

$$D = D(f_k) = D(g_k). \qquad (9.8.11)$$

Because the correcting dipoles are short compared to the accelerator circumference, one can reduce integrals in the expressions for correctors Fourier harmonics to sums and write

$$\sum_d = \sum_{p=1}^{2N} \sum_{q-1}^{K} A_{pq} \eta_p \delta_q^{Bc}, \qquad (9.8.12)$$

where

$$A_{pq} = \frac{1}{N} \left(c_0 + 2 \sum_{k=1}^{n-1} c_k \cos k(\varphi_p - \varphi_q) + c_N \cos N\varphi_p \cos N\varphi_q \right).$$

$$c_k = \frac{2Q \sin \pi Q}{\pi^2 (Q^2 - k^2)}. \qquad (9.8.13)$$

In the same way one can prove that

$$\sum_3 = \sum_{q=1}^{k} \sum_{r=1}^{k} B_{qr} \delta_q^{Bc} \delta_r^{Bc}, \qquad (9.8.14)$$

where

$$B_{qr} = \frac{c_0^2}{2} + \sum_{k=1}^{\infty} c_k^2 \cos K(\varphi_q - \varphi_r) +$$

$$= \cos Q \left| \varphi_q - \varphi_r \right| + \frac{\sin \pi Q}{\pi Q} \cos (\pi - \left| \varphi_q - \varphi_r \right|) Q -$$

$$- \frac{\left| \varphi_q - \varphi_r \right| \sin \pi Q}{\pi} \sin (\pi - \left| \varphi_q - \varphi_r \right|) Q. \qquad (9.8.15)$$

Finally in the applied harmonic analysis it is proved that

$$\sum_a = \frac{1}{N} \sum_{i=1}^{2N} \eta_i^2. \qquad (9.8.16)$$

Summarizing all the above results for the quality criterion (9.8.2) we get

$$2q = \frac{1}{N} \sum_{i=1}^{2N} \eta_i^2 + \sum_{p=1}^{2N} \sum_{q=1}^{k} A_{pq} \eta_p \delta_q^{Bc} +$$

$$+ \sum_{q=1}^{k} \sum_{r=1}^{k} B_{qr} \delta_q^{Bc} \delta_r^{Bc} + 4 \sum_{k=n+1}^{\infty} \left(\frac{Q^2}{Q^2 - k^2} \right) D. \qquad (9.8.17)$$

The coefficients A_{pq} and B_{qr} are given by formulae (9.8.13) and (9.8.15), respectively. In these formulae, φ_p are the azimuths of the BPMs and φ_q and φ_r are the azimuths of the correcting dipoles.

The strengths of the correcting dipoles are determined by the condition for a minimum of q to occur

$$2 \sum_{p=1}^{k} B_{sp} \delta_p^{Bc} = - \sum_{p=1}^{2N} A_{ps} \eta_p, \qquad (s = 1,2, ..., k). \qquad (9.8.18)$$

Introducing matrices $A = \{A_{ij}\}$ and $B = \{B_{ij}\}$ and the vectors of the corrections δ^{Bc} and BPMs readings η, eq. (9.8.18) can be written in the following matrix form

$$2B \delta^{Bc} = -A^T \eta. \tag{9.8.19}$$

Let us introduce the matrix

$$R = -\frac{1}{2} B^{-1} A^T. \tag{9.8.20}$$

The matrix R depends only on the azimuths of the BPMs and on the correctors and for the given accelerator it can be calculated prior to the correction. Then the required strengths of the correcting dipoles will be determined by the matrix expression

$$\delta^{Bc} = R \cdot \eta. \tag{9.8.21}$$

So the algorithm is relatively fast. The computer simulations showed that it works reliably and is free from the undesirable effects mentioned in the beginning of this part.

10. ORBIT CORRECTION
WITH NEURAL NETWORKS

Artificial Neural Networks (ANNs) [32] with their ability to tune themselves according to the output errors are well suited for on-line orbit correction. The relationship between corrector kicks and orbit displacements in BPMs is in general non-linear due to the presence of non-linear elements in the machine. The orbit fluctuates during the time. Both these circumstances are well treated by ANNs which have features for solving non-linear problems and self-teaching.

As it is not possible to go in more detail in this paper we will restrict ourselves only to the main principles of the ANNs correction.

An ANN consists of neutrons arranged in layers and directed and weighted connections between them (Fig.4).

In the forward propagation of the signals (from input to output) each neutron processes has input signals s_j and produces an output signal s_i according to the relations

$$S_i = f\left(\sum_{j=1} T_{ij} S_j \right), \tag{10.1}$$

where T_{ij} are the synapse weights.

The function $f(x)$ in (10.1) (the so-called action function) is usually a step-function thus if the weighted sum of the input signals exceeds the threshold for the neuron, the neuron «fires» and produces output signal.

In the backward propagation of the signals first of all the output errors ε_1 of the ANN are calculated

$$\varepsilon_1 = z_i - s_i, \tag{10.2}$$

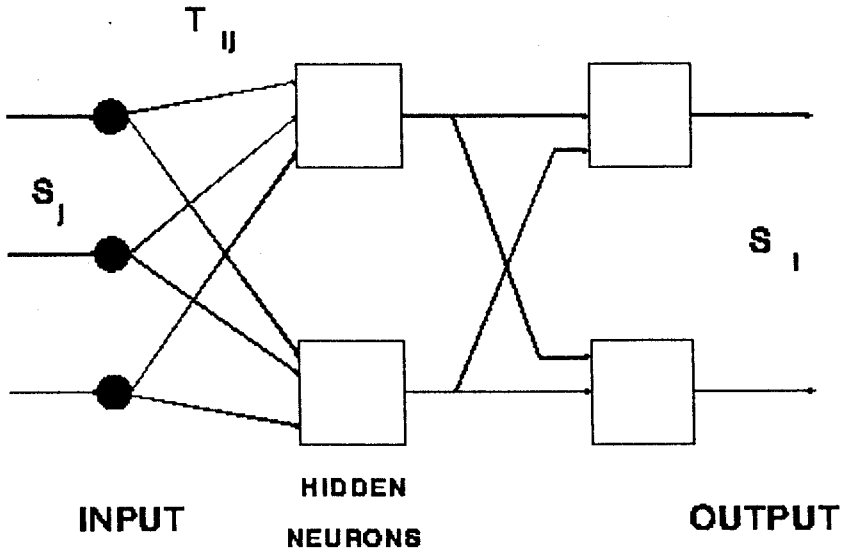where z_i are the «ideal» output signals.

Fig.4. Artificial Neural Network

Let B_{ij} be the blames — the degree of responsibility that each input signal has to the output error

$$B_{ij} = \frac{T_{ij} S_j}{\sum\limits_{j=i} T_{ij} S_j} .$$ (10.3)

Finally the weights are corrected according to

$$T_{ij}^{\text{new}} = T_{ij}^{\text{old}} + k B_{ij} \varepsilon_i,$$ (10.4)

k being a coefficient.

Before to be ready for use an ANN must be trained. To do this we apply a set of input signals for which the «ideal» outputs are known, record the output errors and tune the weights according to (10.4). After many training cycles the desired accuracy of the output can be reached.

The use of ANNs for orbit correction is quite straightforward. The input signals are the measured orbit displacements X in BPMs and the output signals are the corrector kicks ε. There is one input and one output neuron for each BPM and corrector.

In the training stage the applied input signals X are determined by

$$X = A\varepsilon,$$ (10.5)

where A is the response matrix whose elements can be calculated by the machine model or be measured, which is more accurate, and ε are random kicks.

After the training is over and the desired accuracy of the output is reached ANN can be used for on-line orbit correction.

During the machine operation a continuous process of retraining (fine tuning) goes on. The detected orbit errors are feed back and the weights are recalculated-adoptive correction.

ANNs have been used for orbit correction in the NSLS VUV and X-ray storage rings [33,34].

Neural networks have been simulated by means of SNNS computer simulator [35]. A three-layer shortcut connected network (all neurons are connected to each other) and Quickprop training strategy have chosen best results. After 2200 training cycles ANN has been able to correct the orbit to 44 μm maximum deviation.

11. EXPERT SYSTEMS

Knowledge based expert systems differ from conventional computer programs in their intensive use of intuitive and empirical rules which together with facts about the task form the so-called knowledge base [36]. The control strategy (the order in which rules are applied) is determined by the inference engine (Fig.5).
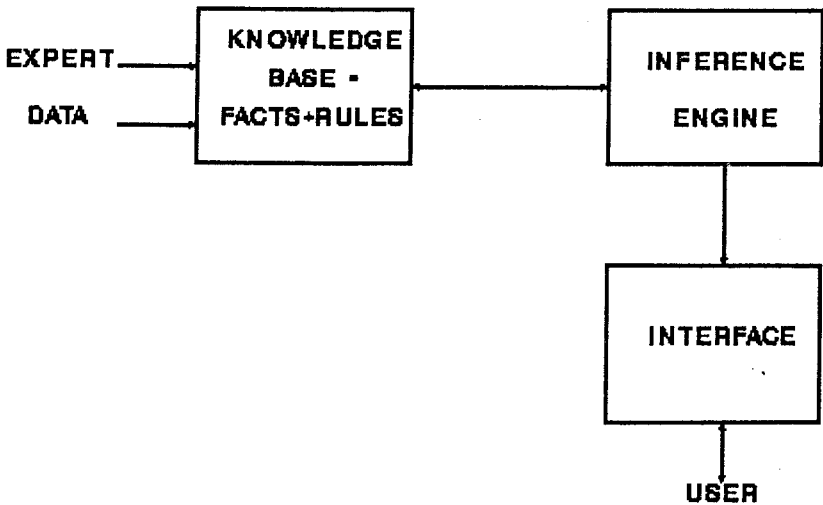

Fig.5. Block diagram of a knowledge based expert system

The expect system approach to closed orbit correction is developing in CERN by D.Brandt and A.Verdier [37,38].

The expert system is based on the fitting method of G.Guignard (chapter 7.2) of searching for field and alignment errors. It can predict and estimate the location of large field defects and to check the correctness of the BPM reading. A convergence test is used as a sensitive mean of detecting whether all errors have been identified.

The main program is written in PROLOG; while the numerical subroutines, in PASCAL.

The expert system has been successfully tested on EAP and LEAR in CERN.

12. ORBIT CORRECTION FEEDBACK SYSTEMS

In synchrotron radiation sources (SRS) we need not only closed orbits with small distortions but also high stable orbits. Orbit stability is a crucial point in achieving low emittance electron beams and therefore high brightness of the photon beams. Orbit correction must be applied dynamically which eliminates fluctuations produced by ground vibrations and magnet power ripples.

In SRS orbit stability is improved by means of correction feedback system (Fig.6).

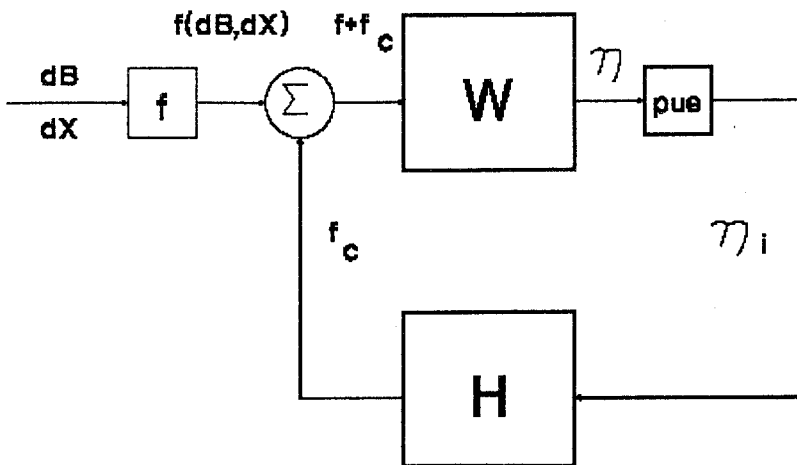In general the feedback systems are divided to local and global systems.

Fig.6. Orbit correction feedback system. $f(dB, dX)$ is the error function (3.6), W is the operator $1/Q^2 (d^2/d\phi^2 + Q^2)^{-1}$, H is the correction operator

In local feedback systems the orbit is locally corrected at the centre of an insertion device by three or four magnet bumps [39].

Global feedback systems can be analog [40] or digital [41].

In Ref.40 a feedback system based on Fourier analysis of the orbit is described. As it has been shown in chapter 7.1. the orbit Fourier coefficients can be expressed by the matrix equality, eq.(7.1.5). This means that a simple linear electronic network can be built to realize on-line Fourier analysis. The input voltages are proportional to the orbit displacements measured by BPMs and in real time the output voltages are proportional to the Fourier harmonics. After that the corrector strengths are adjusted to cancel these harmonics.

Digital feedback systems [41] allow one to avoid drift, offset and temperature problems typical for analog circuits. Orbit data is transferred in digital form by BPM processors distributed around the ring. The digital signal processors then calculate the corrector strengths and control the corrector power supplies.

Digital feedback systems are programmable and therefore more flexible in use.

In principle any orbit correction algorithm can be used in feedback systems.

13. OPTIMUM POSITIONING OF DIPOLE MAGNETS

A different approach to the closed orbit correction is the optimum positioning of the dipole magnets around the ring. According to this approach the dipole magnets are situated in their places on the ring not in an arbitrary way but following a special strategy.

In the stage of accelerator assembling, the field errors in dipoles are carefully measured and only after that they are installed according to a positioning algorithm.

This allows for the orbit distortion to be minimized [42,43].

Let the dipoles be installed around the ring according to the permutation

$$X = (\Delta B_{k_1}, \Delta B_{k_2}, ..., \Delta B_{k_M}),$$

$$k_i \varepsilon \{1, 2, ..., M\}, \quad k_i \neq k_j; \quad i \neq j. \tag{13.1}$$

If we choose as quality criterion the functional

$$q = \frac{1}{2\pi} \int_0^{2\pi} \eta^2(\phi) \, d\phi \rightarrow Min, \tag{13.2}$$

it can be shown [42] that the following sum must be minimized

$$\sum_{q=1}^{M} \sum_{r=1}^{M} B_{qr} \Delta B_{k_q} \Delta B_{k_r} \to \text{Min,} \tag{13.3}$$

where the coefficients B_{qr} are calculated by the machine model.

Let us introduce the X_{jk} integer variables

$$X_{ij} = \begin{cases} 1, & \text{if the } K\text{--th dipole is situated} \\ & \text{at the } j\text{--th position} \\ 0, & \text{otherwise,} \end{cases} \tag{13.4}$$

$$\sum_{k=1}^{M} X_{jk} = 1, \quad \text{i.e., one dipole at a place,}$$

$$\sum_{j=1}^{M} X_{jk} = 1, \quad \text{i.e., each dipole only at one place.} \tag{13.5}$$

From (13.3) the so-called «quadratic assignment problem» of the discrete programming follows:

«Let us have M dipoles, $k = 1, 2,..., M$ and M places $j = 1, 2,..., M$ around the ring. We look for such positioning of every dipole at only one place for which

$$\sum_{q=1}^{M} \sum_{r=1}^{M} \sum_{m=1}^{M} \sum_{n=1}^{M} [B_{qr} \Delta B_m \Delta B_n] X_{qm} X_{rn} \to \text{Min,} \tag{13.6}$$

where X_{jk} are M^2 integer variables (13.4) with constraints (13.5)».

Unfortunately in cases of large dimension of the task the quadratic assignment problem proves to be difficult one [42].

That is why in [42,43] new approaches to the problem making full use of its specific character have been developed.

The set of dipole errors (13.1) causes a combinatorial space of permutations P_B. The points in this space are all possible permutations X (13.1) and its power is $M!$.

Let us introduce a metric in the P_B space in the following way: the distance $r(X, Y)$ between the points X and Y is assumed to be equal to the minimum number of the transpositions (elementary or pair shifts) necessary to bring the point X to the point Y.

Here we will describe only one of the algorithms proposed in [42,43], namely the algorithm of controlled random search.

The algorithm uses the following goal function

$$Q = \max_i |x_i|, \tag{13.7}$$

where x_i is the orbit displacement in the i-th BPM.

The logical structure of the algorithm of controlled random search can be described in the following steps.

Controlled Random Search

Step 1. Choose an arbitrary initial arrangement of the dipoles X, i.e., an arbitrary initial point in the combinatorial space P_B. Draw a sphere S_0 centered in the initial point X_0 and having radius equal to R_0 ($R_0 < (M - 1)$ and its value is chosen by physical considerations). Choose the convergence parameter ε.

Step 2. Set $i = 1$.

Step 3. Choose a random point $X_i \varepsilon S_{(i-1)}$ using uniform probability distribution. Calculate $Q_i = Q(X_i)$.

Step 4. Check whether $Q_i < \varepsilon$. If yes, stop the iterations and exit the algorithm. If not go to step 5.

Step 5. Draw a sphere S_i centered in X_i and having radius R_i :

$$R_i = \delta R_0, \tag{13.8}$$

where

$$\delta = \frac{Q_i}{Q_0}. \tag{13.9}$$

Step 6. Set $i = i + 1$ and go back to step 3.

The computational experiments show that the CPU time necessary for the optimizing of a machine with M dipoles is proportional to $M!$.

Other algorithm for optimum positioning of dipoles can be found in [42,43].

14. FIRST TURN CORRECTION

In order for the particles in a circular accelerator to be able to perform hundred of thousands turns before to reach the final energy of the machine they must first of all pass the very first turn.

During the assembling and initial tuning of the accelerator much bigger errors than the random field and alignment errors mentioned above may occur. Sometimes they are caused by unpredictable mistakes and there are several such cases in the accelerator practice.

In the presence of big linear errors the centre of charge trajectory does not follow any more the orbit and can have very big deviations. Even more, the beam can hit somewhere that vacuum chamber not being able to make complete turn around the machine.

Launch errors in the injection system may also cause the beam loss or provided they are not so big-harmful coherent oscillations of the beam.

The situation is complicated by displacement errors in the BPMs and by the low resolution of the monitors for single-pass beam.

So we face the task of threading the beam entire turn around the accelerator ring.

Two different approaches are possible [44].

First we can try to thread the beam around the ring using the existing orbit correctors. Correcting algorithms have been developed for first turn treatment and some of them will be described in this paper.

In a different approach we can search for the sources of big errors — dipole magnets with big field errors of highly displaced quadrupoles. After finding of the candidates for error sources we should carefully examine the corresponding elements. This approach has been successfully applied for beam line steering and can also be used for the first turn treatment.

The first turn correction is closely related to the beam line steering. In fact before closing the first turn onto the second the magnetic structure of a circular accelerator can be looked at as a beam line.

The beam losses are especially dangerous in the superconducting machines causing the loss of the superconductivity and thus making the tuning process much longer.

In big synchrotrons the distorted orbit amplitude can reach values which are compatible with the vacuum chamber radius due mainly to the misalignments of the quadrupoles. This makes the first turn threading a very important task.

Having corrected the center of charge trajectory so that the beam goes entire turn around the ring we should close this trajectory, i.e., make the second turn (and all the following turns) to coincide with the first turn.

This chapter represents some algorithms for first turn steering.

14.1. Beam Threading Algorithm. The idea of the method is straightforward [45,46]. The steering system consists of small correcting dipoles and beam position monitors (BPM) situated near to them (Fig.7).

As BPMs electrostatic pickup electrodes are used.

Because the length of the correcting dipoles is small comparing to the accelerator circumference they can be considered to produce local orbit bumps.

Let

$$\varepsilon_n = \frac{B_{cn} \cdot l}{B \cdot \rho} \qquad (14.1.1)$$

be the kick in the n-th corrector, where $B \cdot \rho$ denotes the beam rigidity.
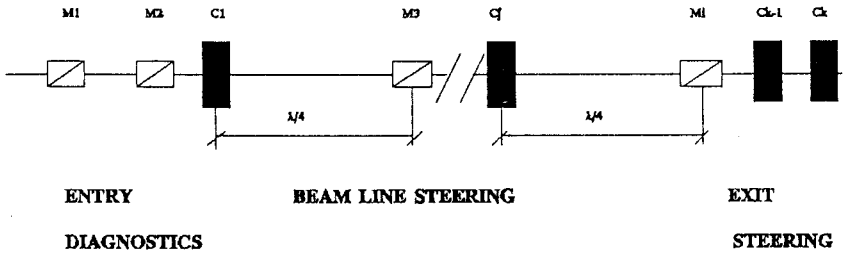
Fig.7. Beam line steering system

If M^{ji} is the transfer matrix between j-th corrector and i-th BPM, we can write

$$x_i = m_{12}^{ji} \cdot \varepsilon_j = \sqrt{\beta_j \beta_i} \, \sin \mu_{ji} \, \varepsilon_n, \qquad (14.1.2)$$

where x_i is the deviation in the i-th BPM and we have expressed the matrix elements by the Twiss parameters — amplitude function $\beta(s)$ and betatron phase advance $\mu(s)$:

$$\mu(s) = \int_0^s \frac{ds}{\beta(s)}. \qquad (14.1.3)$$

Applying the principle of superposition we can write

$$\begin{bmatrix} x_3 \\ x_4 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sqrt{\beta_1\beta_3}\sin\mu_{13} & 0 & \cdots & 0 \\ \sqrt{\beta_1\beta_4}\sin\mu_{14} & \sqrt{\beta_2\beta_4}\sin\mu_{24} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \sqrt{\beta_1\beta_n}\sin\mu_{1n} & \sqrt{\beta_2\beta_n}\sin\mu_{2n} & \cdots & \sqrt{\beta_{k-2}\beta_n}\sin\mu_{k-2n} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_{k-2} \end{bmatrix} \qquad (14.1.4)$$

The system of equations (14.1.4) can be solved with a trivial forward substitutiton.

It follows from (14.1.2) that the optimum phase distance between the corrector and the monitor in a corrector-monitor pair is $\pi/2$. In practice it is difficult for one to follow this requirement.

BPMs should be placed near to the points where β has maximum and where we can expect big centre of charge deviations. This improves the BPMs sensitivity.

In order to minimize corrector strengths the correcting dipoles also should be placed near to the points where β has maximum.

Whereas in the simplest FODO structure it is easy for one to follow these rules situating the horizontal BPMs and correctors near to F quadrupoles and

the vertical BPMs and correctors near to D quadrupoles, in more sophisticated magnetic structures with many elements it is almost impossible.

The beam line steering systems have as a rule two additional parts (Fig.7).

In the entrance of the line two pickups are used to measure the position and the slope of the injected beam. The credibility of the entry angle reconstruction however is greatly improved if the two pickups are placed in drift space [45].

At the exit of the line two correctors are used to match the centre of charge position and slope to ones in the following accelerator or target.

14.2. Least Squares Algorithm. The Least Squares (LSQ) approach was successfully applied not only as a method for general orbit correction but also for first turn treatment [47].

Let's have n BPMs and m correcting dipoles, $n \geq m$.

If $c_k, k = 1, 2, ..., m$ are some changes in the correcting currents, the theoretical changes in the center of charge trajectory x_i^c that they will cause generally speaking are

$$X_i^c = \sum_{k=1}^{m} \left(\frac{\partial x_i^c}{\partial c_k} \right) \cdot c_k = \sum_{k=1}^{m} F_{ik} \cdot c_k. \qquad (14.2.1)$$

In practice due to the error influence the measured centre of the charge position x_i^m will not coincide with the designed position x_i^d. Let $\Delta x_i = x_i^d - x_i^m$ denote the discrepancy between both.

The task of threading the beam around the entire circumference of the ring can be looked at as a least squares problem

$$S^2 = \sum_{i=1}^{n} (\Delta x_i - x_i^c)^2 \rightarrow \text{Min.} \qquad (14.2.2)$$

The corrector strengths which minimize the discrepancy between the measured and desired trajectories follow from the well-known solution of the LSQ problem

$$\mathbf{c} = (F^T \cdot F)^{-1} \cdot F^T \cdot \Delta \mathbf{x}. \qquad (14.2.3)$$

In the case of equal number of correctors and BPMs, when F is square matrix (2.2.3) becomes

$$\mathbf{c} = F^{-1} \cdot \Delta \mathbf{x}. \qquad (14.2.4)$$

For the first turn steering F is a triangular matrix (like (2.1.4)) as a BPM «sees» only those correctors which are placed downstream to it.

The LSQ method has been used in the Tevatron first turn correction [47]. In superconducting synchrotrons the orbit steering is of special importance because the beam losses cause the loss of the superconductivity. This means that the accelerator will have about one hour stop at every tuning step and is a serious problem. The automatization of the tuning by means of the LSQ algorithm has allowed for an orbit close to the designed orbit to be established after several iterations of the correction.

Another LSQ algorithm for first turn steering is described in [48].

The goal function is defined as

$$\Psi(\varepsilon_j^*) = \sum_{i=1}^{n} \left[(\eta_i^{PUE} - \eta_i^{des}) + \sum_{j=1}^{k} a_{ij} \varepsilon_j^* \right]^2 \rightarrow Min, \qquad (14.2.5)$$

where

$$\varepsilon_j^* = \sqrt{\beta_j^{COR}} \, \varepsilon_j \qquad (14.2.6)$$

are the «generalized» kicks.

In eq.(14.2.5), η^{des} denotes the desired centre of charge trajectory; n and k are the current numbers of used pickups and correctors. In the case when we lose the beam before the full turn to be carried out the current number of used pickups n is less than the total number of pickups N and the current number of switched on correctors k is less than the total number of available correctors K.

In practice there are limitations on the maximum allowed kicks in the correcting magnets

$$|\varepsilon_j| < \Delta. \qquad (14.2.7)$$

Thus we face a constrained optimization problem [49].

In order to solve this problem the penalty function method has been used.

The general idea of the method of the penalty functions is to reduce the constrained optimization problem to a series of unconstrained problems. To do this we will add to our goal function (14.2.5) the so-called penalty function $\alpha(\varepsilon_j^*)$ which is chosen in such a way that it will «punish» the function $\Psi(\varepsilon_j^*)$ if the constraints (14.2.7) are broken. In [48] the following penalty function is used

$$\alpha(\varepsilon_j^*) = \sum_{j=1}^{K} \max (0, \, \varepsilon_j^{*2} - \Delta^2). \qquad (14.2.8)$$

The following series of unconstrained optimization tasks is solved

$$\Psi'(\varepsilon_j^*, \mu_k) = \Psi(\varepsilon_j^*) + \mu_k \cdot \alpha(\varepsilon_j^*) \rightarrow Min, \qquad (14.2.9)$$

where $\mu_k > 0$, $k = 1, 2, 3...$ are parameters and

$$\lim_{k \to \infty} \mu_k = \infty. \tag{14.2.10}$$

14.3. Closed Bumps Algorithm. The classical method for global orbit correction by means of local orbit bumps can be applied also for first turn correction [49].

The local orbit bump is produced by three correcting dipoles (Fig.3).

Let n BPMs be places between the correctors.

Let «a» be the centre of charge deviation in the middle corrector. The height «a» can be used for the local orbit bump to be parametrized.

A kick ε_i centered in a point i produces changes in the centre of charge position and angle at another point s given by

$$\Delta x(s) = \sqrt{\beta_i \beta_s} \, \sin \mu_{is} \, \varepsilon_i,$$

$$\Delta x'(s) = \frac{\sqrt{\beta_i}}{\beta_s} (\cos \mu_{is} - \alpha_s \sin \mu_{is}) \, \varepsilon_i. \tag{14.3.1}$$

We will want to produce a local bump. This means that if the point s is situated anywhere outside the bump, the effects of the three correctors should compensate each other:

$$\sum_{i=1}^{3} \Delta x_i(s) = 0,$$

$$\sum_{i=1}^{3} \Delta x_i'(s) = 0. \tag{14.3.2}$$

It follows from (14.3.1), (14.3.2) that for s outside the bump

$$\sum_{i=1}^{3} \sqrt{\beta_i} \varepsilon_i \sin \mu_{is} = 0,$$

$$\sum_{i=1}^{3} \sqrt{\beta_i} \varepsilon_i \cos \mu_{is} = 0. \tag{14.3.3}$$

The kick in the first correcting dipole should produce a deviation equal to «a» in the middle corrector. It should be therefore

$$\varepsilon_1 = \frac{a}{\sqrt{\beta_1 \beta_2} \, \sin \mu_{12}}. \tag{14.3.4}$$

Given (14.3.4), (14.3.3) becomes a system of two equations with two unknowns ε_2 and ε_3.

The solution is

$$\varepsilon_2 = \frac{a \sin \mu_{12}}{\beta_2 \sin \mu_{12} \sin \mu_{23}} \qquad (14.3.5)$$

and

$$\varepsilon_3 = \frac{a}{\sqrt{\beta_2 \beta_3} \sin \mu_{23}} . \qquad (14.3.6)$$

In [50] the following goal function is introduced

$$G(a) = \sum_{i=1}^{n} w_i^{PUE}(x_i^m - x_i^d + x_i^c)^2 + \sum_{j=1}^{3} w_j^{COR} P_j(\varepsilon_j). \qquad (14.3.7)$$

In (14.3.7), x_i^m denotes again the measured deviation in the i-th BPM, x_i^d — the desired deviation. x_i^c is the deviation in the i-th BPM due to the three correctors and it can be easily calculated using the transfer matrices between the correctors and the monitor.

w^{PUE} is a weight associated with each monitor while w_j^c is a weight associated with each corrector.

$P(\varepsilon)$ ia a penalty function which punishes the goal function if the kicks are too big.

In particular, the penalty function can be taken as

$$P(a) = \begin{cases} k|\varepsilon|, \text{ if } |\varepsilon| > \varepsilon_{max} \\ 0, \text{ otherwise} \end{cases} \qquad (14.3.8)$$

where $k > 0$ is a large constant.

$G(a)$ is a function of only one independent variable, a. It can be easily minimized using for example the simplex method.

In order for the entire circumference of the accelerator ring to be corrected the above algorithm should be applied iteratively. The ring is divided into overlapping closed bumps and the correction is repeatedly used taken into account the results from the previous steps.

The algorithm is proposed for SSC and has been largely used for SSC first turn and global closed orbit simulations.

14.4. Error Finding Methods. In the periods of machine assembling, initial tuning or upgrading relatively large errors can occur: launch errors in the position and slope of the center of charge of the injected beam, kick errors in the dipoles and misaligned quadrupoles and focus errors in the quadrupole gradients.

As the strengths of the correcting elements are limited it could be impossible to correct the centre of charge trajectory to the required extent. That is why a different approach to the beam steering has been developed [50,51].

In this approach we will rather search for the sources of big errors than try to correct them. After the places and magnitudes of such errors have been found one should carefully check the corresponding elements trying to uncover the physical bases of the errors.

The error-finding approach has been successfully used in SLAC for beam steering in the linear electron-positron collider SLC [50,51].

Two kinds of computer programs are used to determine the error candidates — modeling programs and error-simulating programs.

The modeling programs give the operator a mathematical model of the accelerator. They receive as input a full description of the accelerator elements (their location and strengths) and produce as output a file listing elements and corresponding transfer matrices and Twiss parameters.

Error-simulating programs calculate the effects on the beam parameters due to specific errors in the elements and generate simulated trajectories. To find the predicted trajectories this programs use the transfer matrices calculated by the modeling programs.

As a modeling program any lattice design program can be used.

For simplification the error-simulating programs describe errors introducing thin-lens elements in the lattice. For simulating of focusing errors thin-lens quadrupoles are inserted in the lattice quadrupoles and for simulating of kick errors thin-lens dipoles are inserted either in bending dipoles to describe field errors or in quadrupoles to describe alignment errors.

Two different kinds of error treatment exist:

A. Global Search. The method of global error search makes use of a powerful nonlinear optimization package. All the elements are suspected as possible sources of errors. Thin-lens elements describing errors are inserted in the elements. Then the optimization programs are used to find the settings of these thin-lens elements which will yield BPMs readings matching the measured trajectory. Those thin-lens elements which have nonzero strengths are the sites of possible errors.

B. Local Search. Only few of the elements are used as possible error sources. The operator has to make some guesses about the location of the errors and about the error-free regions. A highly-developed graphical interface can be very useful in this trial-and-error method. The operator display should be able to show a plot of the measured trajectory, desired trajectory and the difference

between both. It is clear that an error originates in the region where the diffe-rences grow up. After determining the error places the operator tries to adjust the strengths of the corresponding thin-lens elemenets so that the calculated trajectory lies close to the measured trajectory. Non-linear optimization programs are used again but due to the small number of independent variables the optimization problem is much more easy for solving.

Several expert systems have been developed to automate the use of the beam line correcting programs and to minimize the time necessary for line commissioning [52–54].

These are hybrid programs combining traditional expert systems with their capabilities for qualitative reasoning, modeling programs giving a mathematical model of the machine and optimization programs. Combining numerical algo-rithms with symbolic reasoning these hybrid expert systems systematically per-form the specific procedures that a human expert follows in order to correct the beam line.

First of all they look for error-free regions where the discrepancies between the predicted and the measured trajectories are small. An assumption that every subregion between two adjacent error-free regions is a possible location of errors and that there is only one error element within each subregion is made. Then the described above error-finding procedures are used in order for the existence of beam line errors to be uncovered. Frames are usually used for representation of the domain knowledge. LISP is the preferable language for the expert system implementation.

REFERENCES

1. **Kolomensky A.A., Lebedev A.N.** — Theory of Cyclic Accelerators. Amsterdam, North-Holland Publishing Co, 1966.
2. **Bruck H.** — Accelerateurs circulaires de particules. Paris, Press Universitaires de France, 1966.
3. **Strollin P.** — Preprint CERN, ISR-TH/68-4, 1968.
4. **Gluckstern R.** — Part. Accel., 1978, v.8, p.203.
5. **Hamming R.W.** — Introduction to Applied Numerical Analysis. New York, McGraw-Hill, 1971.
6. **Resegotti L.** — Preprint CERN, ISR-MAG/68-30, 1968.
7. **Autin B., Bryant R.J.** — Preprint CERN, ISR-MA/71-36, 1971.
8. **Ando A., Endo E.** — Preprint KEK, 75-4, 1975.
9. **Guignard G.** — Preprint CERN, ISR-BOM/80-21, 1980.

10. Waren J.L., Channell P.J. — IEEE Trans. Nucl. Sci., 1983, v.NS-30, No.4.
11. Averill P.J. — IEEE Trans. Nucl. Sci., 1965, v.NS-12, p.899.
12. Bovet C., Reich K.H. — Preprint CERN, SI/Int/Dl/69-9, 1969.
13. Holtey G. — Preprint CERN, Lab.2-DI-PA/Int. 73-3, 1973.
14. Pegss S. — Ph. D. Thesis, Cornell, 1981.
15. Burnod L., D'Amico E. — IEEE Trans. Nucl. Sci., 1983, vol.NS-30, No.4.
16. Guignard G., Marti Y. — Preprint CERN, ISR-BOM-TH/81-32, 1981.
17. Guignard G., Marti Y. — Preprint CERN, LEP-TH/83-50, 1983.
18. Bozoki E. — Preprint CERN, PS/PSR/85-57, 1985.
19. Martini M., Rinolfi L. — Proceedings of EPAC, Rome, 1988, p.842.
20. Riselada T. — Preprint CERN, PS/87-90, 1987.
21. Autin B., Marti Y. — Preprint CERN, ISR-MA/73-17, 1973.
22. Iselin Ch. F., Niederer J. — Preprint CERN, LEP-TH/87-37, 1987.
23. Bacconier Y. — Preprint CERN, 65-35, 1965.
24. Guignard G. — Preprint CERN, SI/Int.DI/70-1, 1970.
25. Guignard G. — Preprint CERN, SI/Int.DI/70-2, 1970.
26. Guignard G. — Preprint CERN, 70-24, 1970.
27. Press W. et al. — Numerical Recipes in C.Cambridge University Press, 1989.
28. Chung Y., Decker G., Evans K. — Proceedings of IEEE Part. Accel. Conf., Washington, 1993.
29. Dinev D. — Preprint KFA-Julich, Jul-2406, 1980.
30. Chung Y. et al. — Preprint of ANL, LS-213, 1992.
31. Dinev D., Vasilev P. — Bulg. J. Phys., 1985, v.12, No.5, p.480.
32. Simpson P.K. — Artificial Neural Systems. NY, Pergamon Press, 1989.
33. Bozoki E., Friedman A. — Proceedings of EPAC, London, 1994, p.1589.
34. Bozoki E., Friedman A. — AIP conference proceedings, 1994, v.315.
35. Zell A. et al. — SNNS User Manual. Report 3/13 of the University of Studtgart, 1993.
36. Levine R. I., Drang D.E., Edelson B. — AI and Expert Systems. McGraw-Hill Publishing, 1990.
37. Brandt D., Verdier A. — Proceedings of EPAC, Rome, 1988, p.654.
38. Brandt D., Varlot F., Verdier A. — Part. Accel., 1990, v.29, p.221.
39. Bocchetta C., Wrulich A. — Nucl. Instr. Meth., 1991, v.A300, p.223.
40. Yu L.H. et al. — Nucl. Instr. Meth., 1989, v.A284, p.268.
41. Chung Y. — Beam instrumentation workshop, Santa Fe, 1993.
42. Dinev D. — Nucl. Instr. Meth., 1985, v.A237, No.3, p.441.
43. Dinev D. et al. — J. Techn. Phys., 1986, v.56, No.6, p.1137.
44. Dinev D. — Preprint of KFA-Julich, Jul-2499, 1991.
45. Bryant P.J. — CERN Accelerator School, Gif-sur-Yvette, 1984, p.358.
46. Koutchouk J.P. — Preprint CERN, LEP-TH/89-2, 1989.
47. Raya R., Russell A., Aukenbrandt C. — Nucl. Instr. Meth., 1989, v.A242, p.15.
48. Dinev D. — Bulg. J. Phys., 1994, v.21, No.1.

49. **Paxson V., Peggs S., Schachinger L.** — Proceedings of EPAC, Rome, 1988, p.824.

50. **Sheppard J.C. et al.** — IEEE Trans. Nucl. Sci., 1985, v.NS-32, p.2180.

51. **Lee M. et al.** — IEEE Trans. Nucl. Sci., 1987, v.NS-34, p.536.

52. **Clearwater S.H., Lee M.J.** — IEEE Trans. Nucl. Sci., 1987, v.NS-34, p.532.

53. **Lee M.J., Kleban S.** — Proceedings of EPAC, Rome, 1988, p.767.

54. **Weygard D.P.** — IEEE Trans. Nucl. Sci., 1987, v.NS-34, p.564.