Particle Accelerators, 1986, Vol. 19, pp. 93–98
0031-2460/86/1904–0093/$15.00/0

Printed in the United States of America

ROUNDOFF ERROR IN PARTICLE-ORBIT CALCULATIONS

M. E. JOHNSON
Fermi National Accelerator Laboratory, Batavia, IL 60540†
and
A. J. SLAUGHTER
Yale University, New Haven, CT 06511

(Received March 7, 1985)

This paper analyzes the errors due to roundoff in kick-type accelerator tracking codes. We show that the error in the angle of the phase-space vector goes as m^2 and the error in magnitude goes as m, where m is the number of turns that a particle has been tracked. The latter is true only for small values of the kick term. This error is due entirely to the fact that because of roundoff errors, the determinant of the quadrupole transport matrix is not unity. This is demonstrated with a simple kick model of the Tevatron.

I. INTRODUCTION

The design of the Superconducting Super Collider (SSC) will require extensive particle tracking to check the design of the accelerator. Since the accelerator is so large, these calculations will require a very large number of numeric operations. This paper discusses the roundoff errors in these calculations. The analysis is done for the so-called "kick" codes, i.e., the dipoles and quadrupoles are described by transport matrices, and all higher multipoles are treated in the thin-lens approximation as a kick at one or more points in the magnet.

A recent paper by Wilhelm and Lohrmann[1] gives extensive experimental measurements of the calculational errors in modeling the HERA accelerator using the RACETRACK program. Their data show that the error in position goes as m for a linear model (dipoles and quadrupoles only) and as m^2 for the nonlinear machine where m is the number of turns. They explain the error dependence for the linear machine by noting that the determinant of the transport matrices for quadrupoles is not exactly 1 because of the finite precision of a computer.

Transport through a quadrupole magnet is just a rotation in phase space. Using complex variables, the input coordinate, $z(=x+iy)$, is multipled by $\exp(i\theta)$ where θ is the betatron phase advance of the quadruple:

$$z = z_0 e^{i\theta}. \tag{1}$$

† Operated by Universities Research Assoc., Inc., under contract with the U.S. Department of Energy.

Because of the finite length of a floating-point number in a computer:

$$|e^{i\theta}| \neq 1. \tag{2}$$

This difference from unity can be represented by adding a small real number, a, to the rotation angle. The value of z after passing through n quadrupoles of the same type is then

$$z = z_0 e^{in\theta + na}. \tag{3}$$

Expanding $\exp(na)$ to first order in a gives

$$\Delta z = z_0 nae^{in\theta}; \qquad \Delta r = r_0 na, \tag{4}$$

where r is the radius of the circle. Thus the error in radius grows linearly with n. For single precision on the Cyber 875, $|a|$ is about 1×10^{-15}. The exact value of a depends of course on the transport matrix and, as shown in Ref. 1, can be positive or negative.

There is, however, no explanation in Ref. 1 for the m^2 dependence of the nonlinear machine. In the following discussion we show that this dependence is also due to Eq. (2).

II. ANALYSIS OF THE NONLINEAR MACHINE

Qualitatively, the errors can be understood as follows. Just as in the linear machine, there will be an error in radius that goes linearly with n because the transport matrices are nonunitary [Eq. (4)]. There is an additional error whose source can be seen from Fig. 1. The dotted line in the figure shows a particle trajectory with no errors. The solid line shows the trajectory with errors for the case where the transport-matrix error shrinks the radius. The calculated trajectory will have a displacement less than the real one. The trajectory will then pass through the thin lens at a smaller x distance than the real particle and so will have a small kick than the real one because the field is proportional to x^2 or higher powers.

This can also be seen from the circle diagram (Fig. 2). A kick from a nonlinear component appears as an instantaneous change in angle (y coordinate) and no change in x, i.e., a vertical jump in the circle diagram. Because of the nonunitarity of the transport matrix, this vertical jump starts at a different x coordinate than one where this error was not present. Since the strength of the

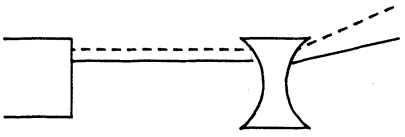

FIGURE 1 The dotted line is the orbit of a particle with no roundoff error. The solid line is the orbit with errors. In this example the determinant of the transport matrix is less than 1, so the particle's displacement from the axis is reduced. Consequently, the kick from the nonlinear element is also reduced.

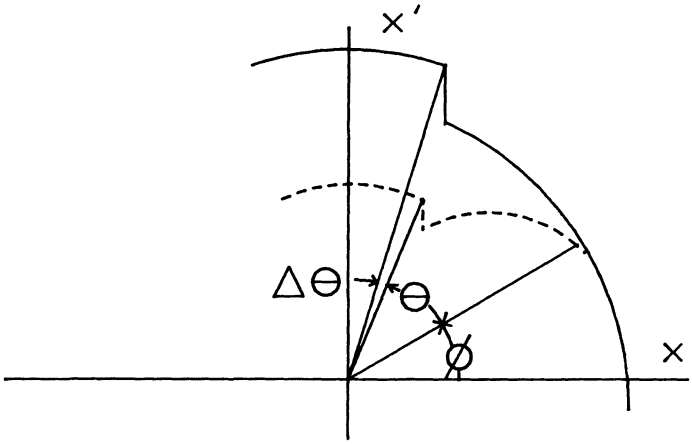

FIGURE 2 Circle diagram showing the effects of the nonlinear kick on a particle's trajectory in phase space. The initial phase-space coordinates are at r, ϕ. The transport matrix rotates it through an angle θ. The solid line is the phase-space trajectory with no roundoff error. The dotted line is an exaggerated trajectory including roundoff errors. $\Delta\theta$ is the angle error introduced by the kick term.

magnetic field is proportional to at least x^2, there is an additional error in z introduced by the kick. A more quantitative analysis shows that the error in radius, Δr, goes as n, while the error in phase, $\Delta\theta$, goes as n^2.

The nonlinear machine must have the linear error growth in r due to Eq. (2). Let us examine the effects of the kick term on the radius and the angle when the kick is applied at a radius that is slightly different from the no-roundoff one to try to develop a quantitative expression for the error amplification described in Fig. 1. This analysis assumes no error; we look only at the effect of the kick term applied to a two radii, r and $r + d$, to see if the position difference between them grows with n.

First, look at the effect of the kick term on the radius. From geometry the kick term increases the y displacement on one half of the circle and decreases the y dispalcement on the other half. Since the particle travels around the circle many, many times, the first-order average displacement will be zero provided the tune of the machine is not near a resonance and the strength of the kick term is small. Since the average radius does not change with n, the displacement between the two radii, r and $r + d$, is indpendent of n and equal to d. Note that we have ignored the effects of random roundoff errors which would cause this error to grow as $\sqrt{n}$. We have observed this behavior in simple models. As the strength of the kick term is increased, however, we observe a linear error growth in δr. We are currently investigating this effect.

The angle difference behaves differently. As can be seen in Fig. 2, the angle difference increases with each kick, so the difference should grow linearly with n. This can be demonstrated quantitatively as follows. A recursive relation for the nth position in terms of the $(n-1)$th term is:

$$z_n = z_{n-1}w + iG\ \text{real}^N\ (z_{n-1}w). \tag{5}$$

Here G is a constant that combines the strength of the multipole field and the terms that convert field strength and displacement to angle, N is the order of the multipole field ($N=2$ for sextupole), real takes the real part of the expression, and $w=\exp(i\theta)$. Note that the nonunitarity of the transport matrix has been ignored.

One can write the position after n quadrupoles to order G^2 as

$$z_n = z_0 w^n + iG \sum_{j=1}^{n} w^{n-j}\,\mathrm{real}^N(z_0 w^j) + G^2 3i \sum_{j=1}^{n-1}\sum_{k=1}^{j} w^{n-j-1}\,\mathrm{real}^{N-1}(z_0 w^{j+1})\,\mathrm{real}(iw^{j+1-k})\,\mathrm{real}^N(z_0 w^k). \tag{6}$$

Now, use Eq. (6) to compute the phase at radius r. Both the G and the G^2 term have sums over angle coordinates. From table of trigonometric sums we have

$$\sum_{k=0}^{n-1} \sin(x+ky) = \sin\left(x+\frac{n-1}{2}y\right)\sin\frac{ny}{2}\operatorname{cosec}\frac{y}{2} \tag{7}$$

and

$$\sum_{k=0}^{n-1} k\sin kx = \frac{\sin nx}{\sin^2\frac{x}{2}} - \frac{n\cos\frac{2n-1}{2}x}{2\sin\frac{x}{2}}. \tag{8}$$

From these formulas one can see that the sum of $\sin(kx)$ gives a function of $\sin(nx)$ and a sum of $k\sin(kx)$ gives a function of $n\sin(nx)$. That is, no higher powers of n are introduced. However,

$$\sum_{j=1}^{n} j = \frac{n^2+n}{2}, \tag{9}$$

which gives a higher power of n. Let us use the octupole ($N=3$) as an example. The sums in Eq. (6) generate many terms. Only the term with the highest power of n in each sum is kept. The phase at radius r is

$$\tan^{-1}\frac{128r\sin(n\theta+\phi)+48Gnr^3\cos(n\theta+\phi)-9G^2n^2r^5\sin(n\theta+\phi)}{128r\cos(n\theta+\phi)-48Gnr^3\sin(n\theta+\phi)-9G^2n^2r^5\cos(n\theta+\phi)}.$$

the phase at radius $r+d$ is the same as in the above equation exept that r is replaced by $(r+d)$. The two equations are expanded in a Taylor series in G to second order and then subtracted. This gives an approximation of the phase error, $\delta\theta$, to order G^2. The result is

$$\delta\theta = \frac{3Gdrn}{4} \tag{10}$$

to first order in d. From this we see that the phase error grows linearly with n and depends on G and d to first order. The non-unitarity of the rotation matrix causes the radius difference d to grow linearly with n. Combining Eqs. (4) and (10) gives

$$\Delta\theta = \frac{3Gr^2n^2a}{4}. \tag{11}$$

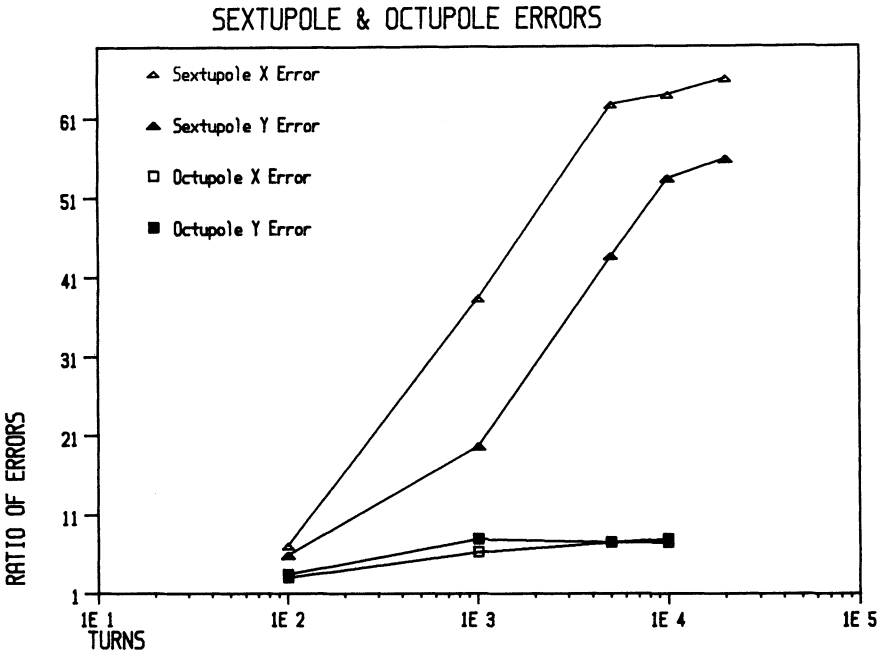

FIGURE 3 Error ratio for a factor-of-8 change in G for sextupoles and octupoles. The sextupole error depends on G^2 while the octupole error depends only on G.

Thus we see that the roundoff error in the quadrupole transfer matrix causes the total error to grow as n^2.

Calculating the phase error for the sextupole ($N = 2$) is complicated since there are a large number of terms in the sum. However, one can go back to Eq. (6) and include the exp (a) term. Expanding exp (a) as in Eq. (1) allows one to compute a series for Δz. If one evaluates the sum for the G term, one finds that it depends on n, not on n^2 (the same exercise for octupole does give an n^2 dependence). The G^2 term does give an n^2 dependence. From the structure of the equations one can show that all even multipoles will have no n^2 dependence for the G term while all odd multipoles will have an n^2 dependence for the G term.

The above predictions were checked with a simplified model of the Tevatron.

TABLE I

Position Error as Function of Number of Turns for Single-Precision and Double-Precision Matrix Multiples

Number of turns	Single-precision position error	Double-precision position error
1	5.9 E-13	3.3 E-26
10	1.0 E-11	2.7 E-25
100	6.5 E-10	3.2 E-25
1000	5.2 E-8	4.5 E-25
10 000	5.0 E-6	9.3 E-25

The erros were measured in the same manner as described in Ref. 1. That is, the particle was run forward for n turns and then backward for n turns. The difference between the initial and final value is the error. Figure 3 shows the error ratio for a factor-of-8 change in G for sextupole and for octupole. The G^2 dependence of the sextupole is clearly evident.

Our numerical results for single-precision floating point on the Cyber 875 (48-bit mantissa) are very similar to those of Ref. 1. Since the source of the error is the nonunitarity of the transfer matrix, we converted all the matrix multiplications to double precision (120–bit word length). All of the kick terms were left in single precision. This increased the computing time by about 50% over the all-single precision calculation. However, no effort was made to optimize the program. Table I shows a comparison of the single-precision and limited double-precision runs as a function of the number of turns in a simplified model of the Tevatron. For a 10 000-turn run the error was reduced by about 10^{-19}.

REFERENCE

1. P. Wilhelm and E. Lohrmann, DESA HERA Report 84/22, (1984).