Presented at Orbit Correction and Analysis Workshop,
BNL, Upton, New York, December 1-3, 1993.

Conf-931254--8

BNL-61251

CLOSED ORBIT related problems: CORRECTION, FEEDBACK and ANALYSIS [1]

Eva S. Bozoki

1 Introduction

Orbit correction -moving the orbit to a desired orbit, orbit stability -keeping the orbit on the desired orbit using feedback to filter out unwanted noise, and orbit analysis -to learn more about the model of the machine, are strongly interrelated. They are three facets of the same problem.

The better we know the model of the machine, the better predictions we can make on the behavior of the machine (inverse modeling) and the more accurately we can control the machine. On the other hand, one of the tools to learn more about the machine (modeling) is to study and analyze the orbit response to 'kicks'.

2 Orbit correction

With smaller and smaller vertical beamsizes ($\leq 50\mu$) in the new generation of Synchrotron Radiation facilities and with the advent of small angle and crab crossing colliding beams in the B and ϕ factories, there is a need for **very accurate** ($\leq 10\mu$) orbit control.

2.1 Approaches

There are different approaches to the orbit correction. They are not all exclusive to each other, some methods can fall into more then one categories.

Model based

The corrective step is calculated based on model predictions. It needs a very good knowledge of the model of the machine (including model-magnet calibration – see Section 4). Since these models are almost all linear, orbit correction needs several iterative steps. All early methods and many of the more recent ones ([1]-[14], [16], [26]-[27]) fall into this category.

Expert Systems

They consist of a knowledge base (data) and a separate reasoning mechanism (inference machine). The knowledge base contains the model of the machine and rules what the most expert accelerator physicist/engineer/operator would apply. Thus an Expert System represents a **model based** system. Expert Systems are capable of treating large amount of information in an empirical way.

There were attempts in mid and late 80's by CERN -using commercial ES packages, primarily for fault diagnosis [15]b, but in [27]h it was concluded, that a simple algorithmic process is more efficient for the analysis and correction of closed orbits.

Basically, this was the approach taken by SLAC in collaboration with the Stanford Knowledge Systems Laboratory ([15]c, [27]b,c) A Fortran coded in-house 'expert', Able was incorporated into the system for finding errors and misalignments in the machines and for controling the machines.

[1] Work performed under the auspices of the U.S. Dept. of Energy under contract no. DE-AC02-76CH00016.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

On the other hand, the design of an ES with limited scope was reported by LLNL [15]a.

Adaptive Systems

It is a **model based** approach with the added property of being able to learn from the difference of the predicted and actual response of the machine. These closed loop (adaptive) systems can automatically compensate for modeling unencertainties, are able to adopt to changes in the dynamics of the process they control and also for external disturbances as in [27]d.

Such systems can be used to minimize the orbit, as for example in [27]d, or more recently [16] the kicks are calculated by minimizing a general 'merit function'. In either case, the model is updated after each step. The 'merit function' in question is a function of the measured orbit, corrector strength, weight factors associated with the orbit monitors (to provide means for stricter orbit tolerance on selected ones), weight factors for the correctors which accounts for their limiting strength and the initial Response Matrix (RM). At each step, the minimizing algorithm also recalculates and updates the RM (see also in Section 2.4). Since in case [16] the model in included in the RM, if measured and not model-calculated RM is used as starting point this method can also be considered as a **black box** approach.

Black box

These methods assume no knowledge of the model of the machine, the machine is treated as a 'black box'. In [18] an algorithm adjusts the parameters of the black box (these parameters are the kicks) which will minimize a 'merit function', and this algorithm is acting as a control. The 'merit function' here is a function of the measured orbit and the used corrector strength (see also in Section 2.4).

When used with a measured RM instead of a model calculated one, all methods based only on the RM also falls into this category of approach (see eg. [7]-[10] [16]).

Neural Networks

In the last couple of years the use of a very powerful mathematical tool, Neural Networks, is being explored [17] for orbit correction. This is the first method which is designed to handle nonlinear problems and, as a consequence, it is able to treat the horizontal and vertical orbits in a coupled way. Furthermore, when used to 'feed back' on the error in its corrected orbit (adaptive Nnet), then it is able to follow drifts in the machine. Additional advantage is that an Nnet hardware can be used to implement the orbit correction in a feedback system.

2.2 Algorithms

Most orbit correction algorithms used in circular machines are variations on three 'themes': the harmonic and the least-square algorithms and the superimposition of local bumps. There are also some extra 'twists' which modify these methods [[7]-[10]]. (We disregard 'beamthreading', used mostly in linear machines and transport lines, e.g. [12]b, [14]).

Harmonic

In the harmonic methods [1]-[3], [22]a, the harmonic content of the orbits are determined and selected harmonics are corrected. The corrector strengths are calculated, assuming that the harmonic composition of the dipole field 'errors' reflects that of the orbit.

These methods are very sensitive to the positioning of the correctors and monitors. They yield good results only when those elements are close to being $\pi/2$ phase apart. They use many correctors around the machine and the individual kicks are relatively small. The sources of inaccuracies in

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

this method are that (*i*) the harmonic content of the orbit is calculated from measurements at a finite number of locations, (*ii*) correcting for the most important few - usually lower - harmonics introduces higher harmonics into the corrected orbit and (*iii*) some orbits contains high harmonics, which is not easy to correct.

Least square

In this method, kicks which correct the orbit are calculated by solving a set of linear equations:

$$\vec{X} = A\vec{\Theta} \quad . \tag{1}$$

where the components of $\vec{X}$ are the measured orbits at the N_m monitors, the components of $\vec{\Theta}$ are the kicks from the N_c correctors, and A is the $N_m x N_c$ Response Matrix.

In the standard least-square methods the residual rms orbit around the machine is minimized:

$$|\vec{r}| = |\vec{X} + A\vec{\Theta}| \tag{2}$$

These methods yield very good results in finding and compensating for a small number of dipole 'errors' in the ring. They use a few correctors only but the kicks tend to be much larger then in case of the harmonic methods. Also, when the number of correctors is increased, the correctors start to 'fight' each other, resulting in very large kicks.

The most commonly used least-square algorithm, Micado [4] is based on the unitary and orthogonal Householder transformation and pivoting. It decomposes the Response Matrix into an $N_m x N_m$ upper diagonal matrix:

$$H_n ... H_1 A = \begin{pmatrix} U_o \\ 0 \end{pmatrix} = U \tag{3}$$

where (H_i, j=1,N) are the Housholder transformation matrices. When asked to use N correctors out of the available total of N_c, the algorithm is choosing the 'most effective' N correctors. [2]

Compared with an 'unconstrained' least-square algorithm [6], which tries all possible combination and accepts the one resulting in the absolute minimum of the residual rms orbit, Micado will result in slightly smaller kicks. As a consequence, sometimes, Micado will point to an equivalent 'most effective' corrector instead of the actual error location in the machine. In addition, the redundancy of correctors in the machine causes the matrix to be not well behaved. Thus, reducing the accuracy of the solution.

Regularization

The 'fighting' of the correctors in the standard least-square methods are caused by the ill-posedness of eq. (1) when A is singular or almost singular leading to large Θ's for small X's. In [8] standard mathematical methods can be used to stabilize the problem. In [8]a the Tikhonov regularization method [9] is used. Instead of eq. (1) it solves the equation:

$$(A^T A + \alpha I)\Theta = A^T \vec{X} \tag{4}$$

where I is the unit matrix and α is stabilizer. The method is very sensitive to the choice of α, which depends on the degree of ill- posedness and on the magnitude of the A_{ij} elements

[2] At each transformation step, the column with the largest pivot is exchanged with the current column.

Eigenvector decomposition

This method finds the normalized eigensolutions of eq. (1) and calculates the $\vec{\Theta}$ kick vector as a superposition of the normalized eigenvectors [7].

$$\vec{\Theta} = \sum_{j=1}^{N_c} c_j \vec{\theta}_j \quad . \tag{5}$$

where the c_j coefficients are calculated from the eigen-decomposition of the measured orbit, $c_j = \vec{X} \cdot \hat{x}_j$, and $\vec{\theta}_j$, $\hat{x}_j$ are the normalized eigenvectors.

It can be shown, that this $\vec{\Theta}$ is the minimum kick vector and that the uncorrectable part of the orbit is

$$\epsilon^2 = \frac{1}{N_m}(\|\vec{X}\|^2 - \|A\sum_{j=1}^{N_c} c_j \hat{x}_j\|^2). \tag{6}$$

SVD

Using the technique of Singular Value Decomposition Method [11], the Response Matrix is written as

$$A = UWV^T \tag{7}$$

where U and V are unitary matrices and W is an $N_m x N_c$ diagonal matrix with non-negative elements. This algorithm reconfigures the monitors and correctors into transformed ones, each transformed monitor being coupled to max one transformed corrector with coupling strengths, W, which determines the efficiency of the orbit correction [10].

$$\vec{X}_t = U^T \vec{X} \quad , \qquad \vec{\Theta}_t = V^T \vec{\Theta} \quad , \qquad and \qquad \vec{X}_t = W \vec{\Theta}_t \tag{8}$$

The method allows for estimating the limits of the orbit correction as well as for the optimization of the corrector strengths.

Superposition of local bumps

Local bumps with a specified beam displacement and/or angle at a specified location can be created using three correctors [12], [18], [23]b or four correctors [5]c-d, [13], [22]e without effecting the orbit outside the correctors. This provides precise control of the beam orbit at SR sourcepoints (bending magnet or undulator) or interaction points of two beams, with a corrective bump confined to a limited section of the machine. One can achieve global orbit correction by superimposing such local bumps around the ring.

The $\vec{\Theta} = \{\Theta_j, \; j = 1,4\}$ vector of corrector strengths, yielding x_o and x'_o displacements and angle at the bump location, can be calculated from the set of (four) equations:

$$A_p \vec{\Theta} = \vec{x} \qquad and \qquad A'_p \vec{\Theta} = \vec{x}' \tag{9}$$

where $\vec{x} = \{x_o, \;\; x_4 = 0\}$, $\vec{x}' = \{x'_o, \;\; x'_4 = 0\}$, A_p is the partial Response Matrix (between the selected correctors on one hand and the bump location and the last corrector location on the other) and A'_p is its derivative. $x_4 = x'_4 = 0$ assures the locality of the bump. The three corrector bump is treated similarly, with the equation for either x_o or x'_o left out.

2.3 Corrector strengths reduction

Eigensolutions of eq. (1) with corresponding small eigenvalues generate negligible changes in the orbit [7]. One can take advantage of this to reduce a $\vec{\Theta}$ kick vector calculated by any other orbit correction method and still stay within the tolerance for orbit correction. The reduced vector is

$$\vec{\Theta}' = \vec{\Theta} - \sum_{j=1}^{J} r_j \hat{\theta}_j \quad . \tag{10}$$

where $r_j = \vec{\Theta} \cdot \hat{\theta}_j$ and $\hat{\theta}_j$ is the j-th eigen kickvector.

An other method, 'corrector ironing', introduced in [10]c, [21] finds equivalent set of $\vec{\Theta}'$ which satisfies

$$\sum A_{ij}\vec{\Theta}' = \sum A_{ij}\vec{\Theta} \tag{11}$$

with the added condition of $\sum \vec{\Theta}'^2 = min.$

2.4 Optimization

The objective is to minimize the deviation of the orbit from a desired orbit, the price one pays is the size of the 'kicks' (corrector current). In addition, one have to consider practical limitation in the size of the kicks and provide for adjustable accuracy of correction at the different locations of the ring (at source points and insertion devices for the lightsources and at interaction regions of colliding beams, the orbit tolerance is much smaller then in other parts of the ring). Consequently, it is better to optimize a 'merit function' instead of just the rms orbit.

In [18] this function is

$$f = \sum |X_i|^\alpha + \sum |\Theta_j|^\beta \tag{12}$$

where α and β are chosen to prevent large excursions from the minimum.

In [16], [19]a the choice and optimization of a 'merit function' is based on a principle after [19]b:

$$f = (\vec{X} + A\vec{X})^T P(\vec{X} + A\vec{\Theta}) + \gamma\vec{\Theta}^T\vec{\Theta} \tag{13}$$

where P is a diagonal matrix of the positive weight factors associated with the orbit monitors and γ is a positive weight associated with the limit of total corrector current.

Weight factors associated with the orbit monitors can be included in the function to be optimized -as in the above cases, or in the calculation of x_{rms} itself -as in [20]a. In some cases, the adjustable tolerance is achieved by using a combination of global correction with larger tolerance followed by tighter controlled local corrections (for example in [19] and [20]b-c).

3 Feedback

In addition to being able to correct the orbit with a few μ accuracy, there is a need for **high stability**of the orbit. In other words, continuous orbit correction is needed to suppress unwanted noise (of mechanical or electrical origin) in the 1 - 100 Hz bandwidth. There are a number of analog [22] and digital [23] feedback systems already working or under design.

The advantage of using a digital- as opposed to an analog-feedback is in its great flexibility. Everything is programmable and changeable. Any of the orbit correction algorithms, described in 2.2 can be used and it is easy to switch between different algorithm. The model does not have to be 'hardwired'. The effective bandwith can be changed.

Important problems, which will be discussed during this workshop are sampling rate, aliasing and phase delay. What sampling rate is needed to correct noise in a given bandwith and at the same time avoid aliasing in that bandwith. How can one achieve enough negative gain in the bandwith and still not exceed the limiting 135° phase delay.

4 Modeling information from orbits

4.1 Model and magnet calibration

All models represent simplification introducing deviation of the real machine from its mathematical model. All real machines represent a deviation from the ideal/designed machine. During commissioning and machine studies both, the machine and our understanding of the machine (model) have to be improved.

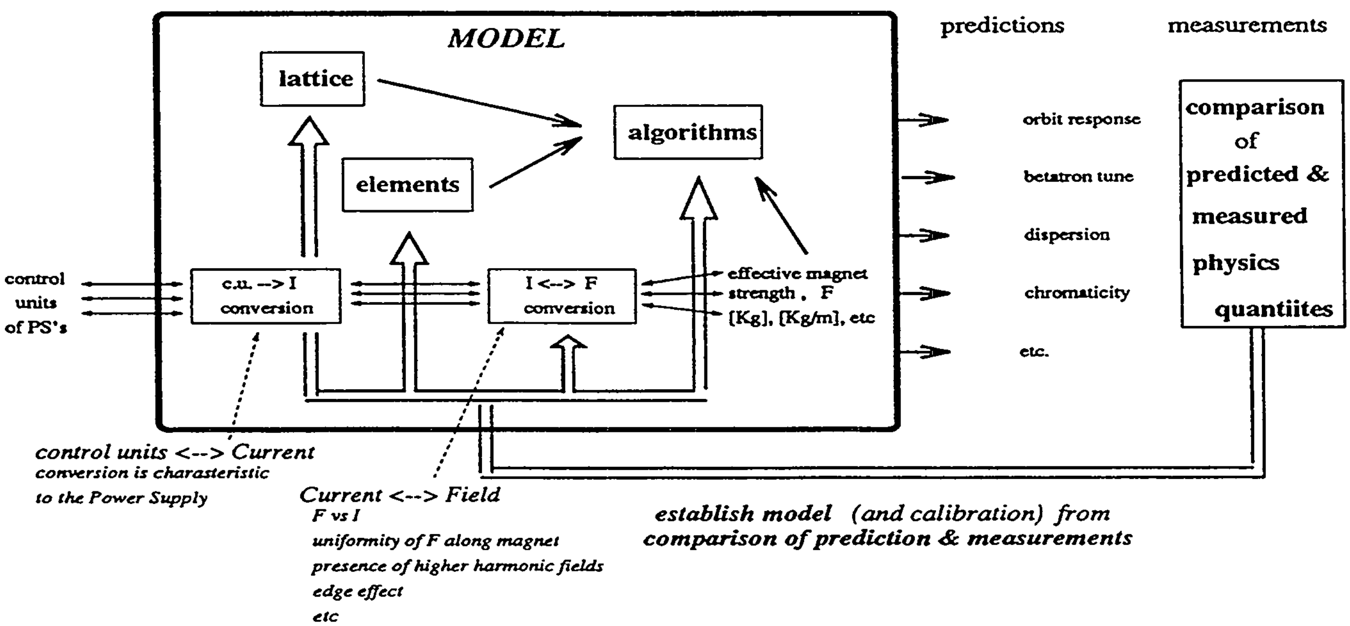

Figure 1: Interrelation between the elements of the Model

The accuracy of the model of a machine is limited by the approximations and errors made in describing the structure and the components of the accelerator and in the algorithms accounting for the beam behavior, and also by the accuracy of magnet calibrations[3] which are incorporated into the model. The interrelation of all the above modeling elements are shown on Fig. 4.1.

[3]The beam behavior depends on the actual fields it experiences as it passes through the magnets with all their nonuniformity, edge effects, higher harmonic content, etc. Beam particles passing through the magnets at different orbits will also feel slightly different fields. On the other hand, there is only one computer controlled set-

It is important to build a realistic model such that predictions on the beam behavior by the model matches the actual behavior of the machine with all its fabrication and installation errors.

One can compare measured and predicted machine parameters, like tune, dispersion, chromaticity, orbit, or changes in them as a response to small changes of the magnet strengths, like in quadrupoles, sextupoles, orbit correctors, etc. Analyzing the differences between the measured and predicted values, it is possible to find errors in the machine, some of which like miswired or grossly misaligned magnets can be corrected, and to improve on the model. Improvement of the model means modifying elements of the lattice - including small misalignments into the model rather then correcting them, modifying algorithms, and including model calibration part into the magnet strength conversion [25], [15]c.

4.2 Orbit as a modeling tool.

Orbit has the richest information on the actual model of the accelerator/storage ring. The change in the orbit as a response to a θ_j kick is given by eq. (1), and it can be expressed with the machine parameters; tune (ν), values of the betatron function (β) and betatron phase (ϕ) at the kick and at the observation point:

$$x_i = A_{ij}\theta_j = \frac{\sqrt{\beta_i\beta_i}}{2sin\pi\nu}\cos\nu\left(|\phi_i' - \phi_j| - \pi\right)\theta_j \tag{14}$$

The machine functions in turn, depend on the actual focussing field the beam feels in the magnets. The $\theta = \frac{\Delta B \cdot l}{B\rho}$ kick can come from many sources; a misaligned magnet, a nonzero orbit in a nonlinear element (e.g. sextupole) or a change in the strength of a magnet (corrector or quadrupole).

There many monitors distributed around the ring provide a large number of measured information, consequently, there is enough freedom in the system to find kicks from misaligned quadrupoles and sextupoles, orbit monitor offsets, etc. and to fit for focussing strength from quadrupoles and insertion devices, corrector strength, energy deviation, dispersion, coupling, monitor sensitivity.

4.3 Methods used

The common approach in [26]-[28] to allow in the model varying degree of misalignment of the main magnets, varying degree of coupling and variable focussing strengths in the quadrupoles (and insertion devices) and then fit the measured orbit (absolute or difference orbit) with the model calculated one in order to find the best value of the variable parameters. The model can then be updated with the findings.

In some cases, for transport lines or 1st turn of circular machines, the fitting procedure is progressing element by element to find the location and magnitude of kicks caused by misalignment and to find the best values for the initial position and angle of the beam and for the magnet strengths.

Other methods applied to hadron machines [30] will find best fitted values for betatron function, tune, coupling, chromatic behavior.

point/voltage/current of the power supply of a magnet, consequently the 'bench calibration' of the magnets alone can not account for the accurate conversion between the magnet's control unit and its field; the magnet and the model have to be 'calibrated' together yielding conversion algorithms for [θ] $mrad/c.u.$, [g] $\frac{kg}{m}/c.u.$, [B] $kg/c.u.$, etc.

An altogether different method, which performs fitting in a multiparameter space is the linear perturbation method [29]. In this case instead of analyzing one orbit at a time, the Response Matrix is analyzed. The model predicted RM is expanded to first order in the parameters and fitted to the measured one.

References

[1] A. Jackson, SLAC report No. SRS/NSS/75/103 (1975).

[2] B. Autin, P.J.Briant, CERM ISR-MA/71-36 (1971).
E. Close, M. Cornacchia, A.S. King and M.J. Lee, PEP Note-271 (1978).
E. Gianfelice and R. Giannini, CERN/PS 85-42 LEA (1985).
E. Bozoki, NSLS Tech.Note #459, (1992).

[3] D.J. Harding, and A.W.Riddiford, Proc. IEEE PAC, (1989).

[4] B. Autin and Y. Marti, CERN ISR-MA/73-17 (1973).

[5] A.S. King, M.J. Lee and P. Morton, IEEE Trans. Nucl. Sci., Vol. NS-20, (1973).
E. Close, M. Cornacchia, A.S. King and M.J. Lee, PEP Note-271 (1978).
E. Bozoki, BNL-31361, p.23-29, (1982). and CERN/PS/PSR/85-57 (1985).

[6] T. Risselada, CERN/PS(PSR) 87-90 (1987).

[7] E. Bozoki, A. Friedman, Proc. IEEE PAC, (1993).

[8] Y. Tang and S. Krinsky, Proc. IEEE PAC, (1993).
J. Bolduc and G. MacKenzier, IEEE Trans Nucl. Sci., NS-18, No.3, p.287, (1971).

[9] A.N. Tikhonov, Solution of Ill-Posed Problems, John Wiley & Sons, NY, (1977).

[10] Y.J. Chung, Proc. IEEE PAC, (1993).
Y. Chung, G. Decker, K. Evans, (APS)LS-213, (1992)
W.J. Corbett, B. Fong, M. Lee and V. Zieman, SLAC-PUB-6110(A), (1993).

[11] W.H. Press et all., Numerical Recepies, Cambridge Univ. Press, (1986).

[12] R.O. Hettel, IEEE Trans. Nucl. Sci. NS-30, p.2228, (1983).
R. Raja, A. Russel and C. Ankenbrandt, NIM A242, p. 15–22, (1985).
J. Milutinovic, Present Workshop

[13] A. Chao, PEP TechNote PTM-147, (1979).

[14] L. Dahl, Unilac report (199?)
K.A. Thompson et all., Proc IEEE PAC, (1989)

[15] D.L. Lager, Engineering & Computational Directorate Research, p.14, (1989).
E. Malandain, P. Skarek Proc. IEEE PAC, V.3, p.559, (1989).
M.J. Lee, Proc. Linear Accelerator Conference (1988). SLAC-PUB-4724

[16] Y. Cheng and C.S. Hsue, Proc. IEEE PAC, V.3, p.1704, (1991).

[17] J.E. Spencer, Proc IEEE PAC (1989).
D. Nguyen, M. Lee, R. Sass, H. Shoaee, SLAC-PUB-5503, (1991).
E. Bozoki and A. Friedman, Proc. Workshop on Simulation on Nnets, p.41, (1993).
J. Niederer, Present Workshop

[18] R. Carr, SLAC/SSRL-0001, (1993).

[19] P. Juan, J. Darpentigny, P.C. Martin, P. Nghiem, Proc. EPAC, V.1, p.729, (1990).
G. Guignard, CERN 70-24, (1970).

[20] E. Bozoki and A. Friedman, private communication
J. Safranek: private communication
A. Ropert, Present Workshop

[21] V. Ziemann, SLAC note, CN-393 (1992).
B. Fong, unpublished SPEAR note (199?).

[22] L.H. Yu, E. Bozoki, J. Galayda, S. Krinsky and G. Vignola, NIM A284, pp.268 (1989).
T. Katsura et all., Proc. IEEE PAC, V.1, p.538, (1989).
Y. Cheng, Proc. IEEE PAC (1993).
F.J. Decker, Proc. IEEE PAC (1993).
G. Portmann and J. Bengtsson, Proc. IEEE PAC, (1993).

[23] A. Friedman, E. Bozoki, O. Singh and J.D. Smith, Proc. IEEE PAC (1993).
N. Nakamura and T. Katsura, Proc. IEEE PAC (1993).
O. Singh, Present Workshop
V.M. Zhabitsky, I.L. Korenev, L.A. Yudin, JINR report, 1992
B.G. Martlew et all., Present Workshop.

[24] S.Y. Zhang Proc IEEE PAC (1989)
S.H. Ananian and R.H. Manoukian, Proc IEEE PAC (1992).

[25] E. Bozoki, Proc. of the Workshop on Modeling Based Accelerator Controls, p. 5, (1987).

[26] Y. Qian, E. Crosbie, L. Teng, Proc IEEE PAC, (1991)
J. Milutovic and A.G. Ruggiero, Proc IEEE PAC, (1991)

[27] M.J. Lee et all., Proc IEEE PAC, V.1, p.611, (1989).
M.J. Lee, S. Scottwater, Wkshp on Model-Based Accel. Control, BNL-40651, p.31, (1981).
S. Clearwater and M.J. Lee, Proc IEEE PAC, V.1, p.532, (1987).
M.J. Lee, SLAC-PUB-5700 (1991).
W.J. Corbett, M.J. Lee, Y. Zambre, Proc. EPAC (1992).
A. Luccio, NIM **A293**, p.460, (1990)
A. Luccio and E.H. Auerbach, Proc. IEEE PAC (1993).
A. Verdier and J.C. Chappelier, Proc. IEEE PAC (1993).

[28] J.L. Warren and P.J.Channell, Proc. PAC, V.2, p.2415 (1983).

[29] W.J. Corbett, M.J. Lee, V. Ziemann, Proc. IEEE PAC (1993). (Calif)
J. Safranek, Present Workshop

[30] S. Koscelniak and A. Iliev Present Workshop
Y. Shoji and C.J. Gardner Present Workshop
D. Li at all. Present Workshop
D. Trbojevic, S. Peggs and Jei Wei, Present Workshop
J.Y.Liu et all., Present Workshop

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.