SLAC-PUB-6472
May 1994
(N)

ACCELERATOR CONTROLS AND MODELING*

JEFF CORBETT◊ and **CLEMENS WERMELSKIRCHEN**◊
Stanford Synchrotron Radiation Laboratory, Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94309

9.0 Introduction

The first part of this chapter provides an overview of the general requirements for modern synchrotron light source control systems. This description covers different components, architectures, and aspects of the operator interface. In addition, features of the computer infrastructure, on-line communication facilities, and front-end interfaces are described. As the control system is a central part of any accelerator, it interacts with many of the hardware components described in other chapters, and provides exchange, storage, manipulation, and display of data. The control system also contains hardware calibration constants needed to run machine analysis and control programs. The second half of this chapter provides an overview of accelerator models with particular attention to accelerator control. The model is an important part of the control system because it helps physicists adjust the operating conditions. To understand the role of the model, from its inception in the accelerator design phase to its application for machine control, we first consider the components of the model and the interface to the control system. We next derive physical parameters from the model: transport matrices, beta functions, closed orbit perturbations, and the synchrotron integrals. Many of these parameters can be measured experimentally. Procedures to determine calibration constants that make the model agree with experimental measurements are outlined. The final section concentrates on model based machine control.

9.1 Control System Overview

A general function of the accelerator control system is to establish coordination between all hardware components, so that the goal of the accelerator, to control a charged particle beam over a certain period of time, is accomplished. Specifically, the goal of an accelerator may be to produce electrons (or ions) in a beam with short pulse duration, or a beam stored for a long period of time, but the basic requirements of the control system and its components are similar for all accelerators.

Control systems have always been adapted to the specific needs of each individual accelerator. But over the years, the technologies used for the control systems (and so the architecture) have undergone significant changes. In fact, the differences in the control systems are usually driven more by the development of technology and component cost (mainly computer and microprocessor development) than by the scientific goals of the accelerator.

---

* Work supported by the Office of Basic Energy Sciences, Department of Energy contract DE-AC03-76SF00515.
◊ Author e-mail: corbett@ssrl01.slac.stanford.edu and FAX: (415) 926-4055, and
author e-mail: wermelsk@vms.gmd.de and FAX: +49 2241 14 3002.

The following sections will discuss the typical layers of a synchrotron light source control system from the hardware level to the operator interface, and touch on more sophisticated controls. At the same time, this development provides an overview of the historical evolution of accelerator control systems.

9.2 Control System Basics

Other chapters of this book describe the hardware components needed to build and operate a synchrotron light source. Here, we will show how these components are connected through the control system.

From the point of view of the control system, each hardware component consists of a set of control signals, either to control the device or to monitor the device status. In the very simple case, for instance to set the field strength of a magnet, the control system provides a low-level control voltage to the power supply, which in turn drives current through the magnet. At the same time, an electronic sensor detects the current going through that magnet and passes this information back to the control system computer. The first *control systems* consisted of potentiometers (knobs) and ammeters to perform the task of controlling devices like power supplies. As accelerators grew larger, the number of components increased and the components were spread over a larger area, so other control system techniques had to be developed.

The simple example of magnet power supply control shows all the main characteristics of any modern control system. Assume, for instance, the operator wants to adjust certain parameters of the beam. For this, the operator sets the magnet strengths, and the new power supply currents are calculated from calibration constants. Each power supply is adjusted by setting the control voltage through the control system. There is in principle no difference if the control is from local knobs or from the computer interface. The operator then checks the correct operation of the power supply by measuring the supply current. Again, there is no difference if it is through local meters or from values displayed by a computer. The power supply may also have switches and status indicators. This binary information (on/off) is part of the control system, and can be controlled or monitored.

This simple example illustrates the following basic functions required of any accelerator control system:

- it consists of everything between the operator and the hardware that is needed to control and monitor each device
- it gives the operator access to the accelerator from the control room or any office (via computer network connection)
- it transfers information from all devices, potentially spread over a large physical area, to one or more locations where this information is needed
- it provides analog set point values
- it accesses analog readback values and
- it sets digital or binary values, and processes single bit information

Modern synchrotron radiation sources include more complicated devices than magnet power supplies, but each device is similar from the point of view of the control system. The radio-frequency (RF) drive system, for example, when looked at as a black box, is controlled by exactly the same kind of electrical interface signals. The difference is that the analog set point values control the RF power applied to the cavity. For the control system, it is the same task as before; i.e., provide set point values, monitor power supply values, or set and monitor digital control signals.

The tasks of the control system defined so far are simple, in the sense that the control signals are constant in time (DC). Accelerators also include components that are controlled with time correlations between each other. The injection system, for instance, might include a linac, booster synchrotron, and injection and ejection hardware. In this case, the control system must synchronize time-related signals so that these components operate appropriately during each acceleration cycle.

For these applications, depending on the type of hardware device to be controlled, the control system provides trigger signals for well defined events (linac, septa, kicker), and provides analog set point values that evolve in time (waveforms). For a booster synchrotron, both the RF and magnet currents require time evolving waveforms during each acceleration cycle. Time evolving signals are also acquired to detect beam position, component instabilities, and feedback system signals.

9.3 General Control Systems

During the evolution of accelerators and their control systems over the last 50 years, the control of the accelerators has become more and more complex. In place of knobs, meters, and oscilloscopes, there are now front-end interfaces, computer controls and work stations. The basic task of the control system is still to allow operators to interact with the accelerator, in particular to set and monitor the machine parameters. But in the context of modern control systems with object-oriented databases, the meaning of each machine parameter can now be more general.

Today, each accelerator control system is built around a specific computer architecture which serves as the system infrastructure. The selection of infrastructure has become increasingly important as computer speed increases and costs decrease. The enhanced computing power allows much more sophisticated handling of machine parameters, more complex on-line calculations, more accurate simulations of machine physics, and more user friendly operator interfaces.

Accelerators also have been built to reach higher particle energies. This has spread the hardware components over a larger area. Fortunately, local area computer networks have evolved, and these high-speed networks can be incorporated into the accelerator control system. The function of the local area networks is to connect distributed processors to central processing computers and distributed operator consoles (work stations).

A modern control system is structured into hardware and software layers that manipulate data on different levels of abstraction. The bottom layer interacts with the electrical signals where the processors have to implement real-time control. The top layer provides the human interface where operators can control the accelerator. The layers in between maintain the machine parameter database, and provide data collection, data distribution, networking, and monitoring.

As the available computing power has increased, more sophisticated application programs can be used for machine control. Theoretical calculations and accelerator models that where formerly used only in the machine design phase are now used for machine commissioning and model calibration.[1] Today's more sophisticated control systems can model the state of the accelerator and the physics of beam-hardware interactions. This begins with calibration factors that convert hardware signals to physical beam parameters. The modern database may also include information for expert systems or artificial intelligence.[2]

9.4 Control System Architectures

Synchrotron light sources differ in size and in requirements for physical layout of the control system. In some cases, such as high-energy rings APS,[3] ESRF,[4] or SPring–8,[5] the control systems must be distributed over a large area. Smaller storage rings can have a more centralized control system. Almost independent of these constraints, the control system hardware components are arranged in a hierarchical structure as indicated in Fig. 1. The actual implementation of the control system will differ depending on the size and operational goals of the machine. As the scope of this book is restricted to synchrotron light-source accelerators, the following discussion will focus on control system architectures for such machines.

The levels of functionality shown in Fig. 1 can be concentrated in a single or a small number of computers. For small accelerators, there might be one computer with no network at all. The more control intelligence (i.e., computers or microprocessors) that is involved, the more the aspects of physically distributed processing capability and computer networking become important. Although the layers shown in the Fig. 1 give the principal overview of the control system building blocks, the meaning or even necessity of the different blocks can vary over a wide range for different control systems.

The control system for a synchrotron light source can also be divided into the control system for the injector and the control system for the light source (storage ring). The degree of separation between these two systems is arbitrary, but

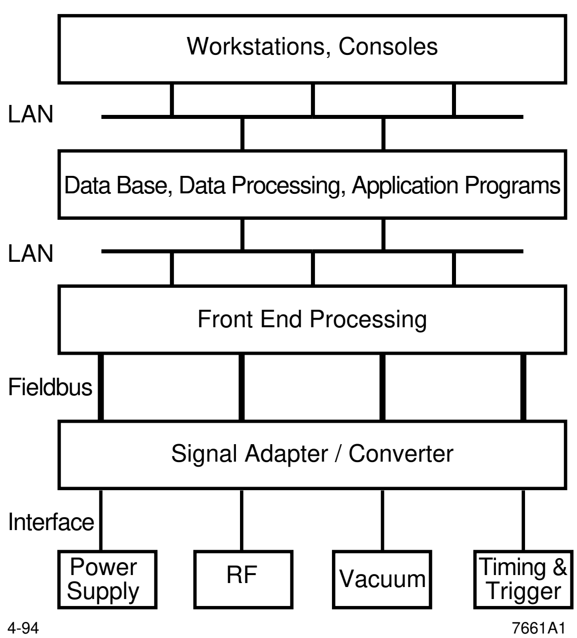

Fig. 1. Schematic of control system hierarchy for hardware components.

there must be some communication between the systems.

In the past, each accelerator had a home-made control system that was adapted to the needs of the individual machine. Components of modern control systems are frequently based on more standardized commercial products. The processors, as well as the networking infrastructure, are often off-the-shelf products, and there are efforts to standardize machine software.[6] There is also commercial software available that can be customized to specific needs.[7,8]

In general, work stations are widely used for the operator interface. They provide high-quality graphic displays, and take over the job of visualizing the machine parameters and handling operator input.

9.4.1 Front-End Systems

The front-end system can be characterized as the *low level* part of the control system. The front-end system interfaces, or connects, physical signals derived from electrical hardware components to the computer. In practice, no signals are wired directly into the main control room; rather the electrical signals are converted into a digital form that can be transferred via data communication lines. For hardware control, signal values are transferred in digital form and converted into electrical signals. Digital signals have the added benefits that they can be transferred over long distances much easier than analog signals, they are less sensitive to electromagnetic noise, and data transmission errors can be detected. Therefore, the front-end control system components that convert between analog and digital signals (ADC and DAC) can be located close to the hardware components. This also reduces the cost of cables, since a single digital communication line can control many devices.

Some accelerator components, especially pulsed devices like kickers, produce high-energy electromagnetic noise in the surrounding area. In this case, the front-end interface electronics must be designed to isolate the rest of the control system from the induced noise. This can be done by using signal transformers and optical isolation circuits.

For economic and maintenance reasons, front-end systems are set up in a modular design. Standard modules that can be easily arranged and replaced within a housing crate are used to connect the control system to the electrical interfaces of the hardware component. The CAMAC standard, for example, is widely used for this purpose. It has a well established protocol, and modules for many different purposes are commercially available. Traditionally, the CAMAC modules were used primarily for data conversion (buffers, ADCs, DACs), but an increasing number of modules now have their own processing capability and the functionality of front-end systems. Although CAMAC systems originated from the time of low integration level in microelectronics, they have evolved to include microprocessors and minicomputers.

Beginning in the 1980s, when high-power single-chip microprocessors became available, a new front-end standard was created. This system, the VME standard, began with modules that had powerful microprocessors (Motorola 68000) and a data bus structure that was designed for high-speed communication between the modules. By using additional modules that provide data converters and drivers, VME systems are now widely used for front-end purposes in accelerator control systems.[9] Because of the more CPU–oriented approach of the VME design, a field bus layer was added to the front-end system. This allows each VME processor to handle several hardware components through inexpensive digital communication lines like RS232, GPIB, or MIL–STD–1553B, either connected directly or through an additional interface.

Similar to the VME standard, there are other developments like Multibus and Bitbus. These options provide mechanical and electrical standards that allow the system designer to combine commercial modules, interface modules, and specialized home-made modules.[10-13]

9.4.2 Centralized Control Systems

Centralized control systems are characterized by a database that contains all the machine parameters and the control system software on a single host computer. Since central control systems with a single computer are limited to the resources of the host processor, they are not practical for larger accelerators. Where a centralized control system is adequate, it has the advantage that software management and data organization are easier than with distributed systems.

For small control systems, the central host may be the only processor in the whole control system. In this case, the central host not only has to maintain the database, but also has to perform the task of data acquisition and provide the operator interface.

To increase the performance of a centralized system, the task of data acquisition can be delegated to locally intelligent front-end processors. The central host is then no longer responsible for reading individual data values out of each front-end interface, such as a CAMAC module. Instead, data is transferred in blocks to and from front-end microprocessors, such as intelligent crate controllers or VME crates.

9.4.3 Distributed Control System

Whenever the processor power or data storage capability of the central computer are not sufficient, the control system can be spread over several computers. This may involve distribution of software applications, or splitting the machine parameter database into a distributed database.

A distributed control system consists of a hierarchy of processor layers. They either share a completely distributed database where an application program can be run on any processor, or they are arranged as subsystems where each subsystem is responsible for a certain group of machine components. As in Ref. 14, there can be subsystems for the cryogenics, vacuum, injector, RF, beam transport, and beam switchyard, etc. Each subsystem may also have its own front-end system, communicate with a dedicated front-end through a special subnet, or the subsystems may share the front-end systems by exchanging data through an underlying communication network.

Perhaps the most extreme example of a distributed control system comes from the LEP collider at CERN. This machine has a circumference of 28 kilometers and a huge number of components to control, so the control system is distributed over a large area with a hierarchy of processors, and networks, and a distributed database.[15,16] Similar systems, such as reflective memory for fast orbit feedback, can be used for large, high-energy light sources.

9.5 Networks

The components of accelerator control systems are linked by a network to transmit data. A wide variety of technologies are used for these networks. Depending on the quantity and speed of the data transmission, either coaxial cable or fiber optic connections are used. Specialized links are used on the lower network layers. These must be adapted to the real-time needs of the front-end processors or the communication needs of the control system, especially for timing purposes.

Ethernet, the first high speed local area network standard (10 Megabits/sec) is widely used at this time. Because Ethernet offers CSMA-CD (carrier sense multiple access with collision detect), every station connected to the Ethernet can send data whenever no other station is transmitting. Where collisions occur, data packets must be retransmitted, which can lead to unpredictable transmission delays. Where guaranteed network bandwidth is an issue, token-ring architectures are now in use.

The use of standardized networks within a control system directly couples the control system with the computer network within the whole laboratory. This is often very convenient, but care has to be taken to prevent unauthorized manipulation of the accelerator.

9.6 Application Programs

The control system has to perform a series of tasks to keep the accelerator operational. Each of these tasks is controlled by software packages, commonly referred to as application programs. Examples of application programs include software to monitor beam current, vacuum conditions, power supply settings, and programs for orbit control. The overall design of the control system software is very important since the control system must provide mechanisms for all application programs to access machine parameters. The application programs should also be designed independent of the underlying hardware architecture. This arrangement can be accomplished by providing a well defined software interface through the control system kernel. A well designed kernel makes testing of new application programs much easier, and keeps application programs small, separated, and dedicated to specific purposes. Modern software development techniques and tools can be used to keep the software manageable, and to meet software and quality control standards. This improves overall system reliability.

In the following sections, a selection of basic tasks required of any control system is described. The functions of some higher level application programs will be mentioned. It should be kept in mind that the development and maintenance of application programs is an ongoing task during the lifetime of an accelerator. Research and development in the area of running accelerators under automated control is still progressing.

9.6.1 Machine Parameter Maintenance

The control system maintains a complete database with the status of all accelerator components. Either a centralized or a decentralized database is used to perform this task. The database software can be a commercially available product such as ORACLE,[16] it can be a relational database,[17] or it can be individually designed software.

It is often necessary to maintain data for several machine operating conditions, and consequently, the complete set of machine parameters or a subset of parameters has to be saved or restored from external files. The collective set of hardware set points is often referred to the machine configuration. These configurations can be maintained at three parallel levels:

- hardware set points from file archives
- present hardware set points
- present hardware read back values

If the hardware set points are adjusted, the read back values should change to verify the result. Any new configuration can be archived to the computer disk and retrieved at a later date.

9.6.2 Operator Interface

The most visible part of the control system is the operator interface; i.e., the consoles and displays. Most systems use work stations now, but control systems based on personal computers can also be found. For ease of machine control, it is important to have a high-resolution ergonomic display. The X-Windows graphics standard has been adopted, or is being planned for future use, for many synchrotron light source control systems.

Figure 2 shows an example of a graphical presentation for the machine status and operator interaction points for the injector synchrotron at SSRL. This control menu displays all parameters in the transport line between the linac and the booster, and allows the operator to manipulate their values. In practice, a high resolution color control panel gives a user friendly overview of the complex hardware configurations.

9.6.3 Monitoring, Alarms, Logging, and Protocol

Monitoring of the accelerator status is done at several levels of the control system. The front-end system, for example, monitors the correct behavior and function of each component. Monitoring is also done at the higher levels of the control system to check the function of the complete machine; for example, the beam position or beam intensity.

Whenever a device-monitoring application detects a parameter value out of range, an alarm is generated and transferred throughout the system. These alarms can have different levels of significance. Some may indicate emergencies that lead to automatic shutdown of the machine, while others alert the machine operators of a potential hardware problem.

Alarms and operator actions can also be logged and displayed on consoles. Logging systems provide on-line information about the actual status of the accelerator.

For long-term diagnostics, and possibly for accounting purposes, records of significant machine parameters have to be kept in a protocol file. Typical parameters in a protocol file include beam intensity, beam line status, and accelerator startup and shutdown information.

9.6.4 Automatic Procedures

Besides attending to the passive accelerator control tasks described so far, control systems are taking over responsibility for the active operation of the accelerator.

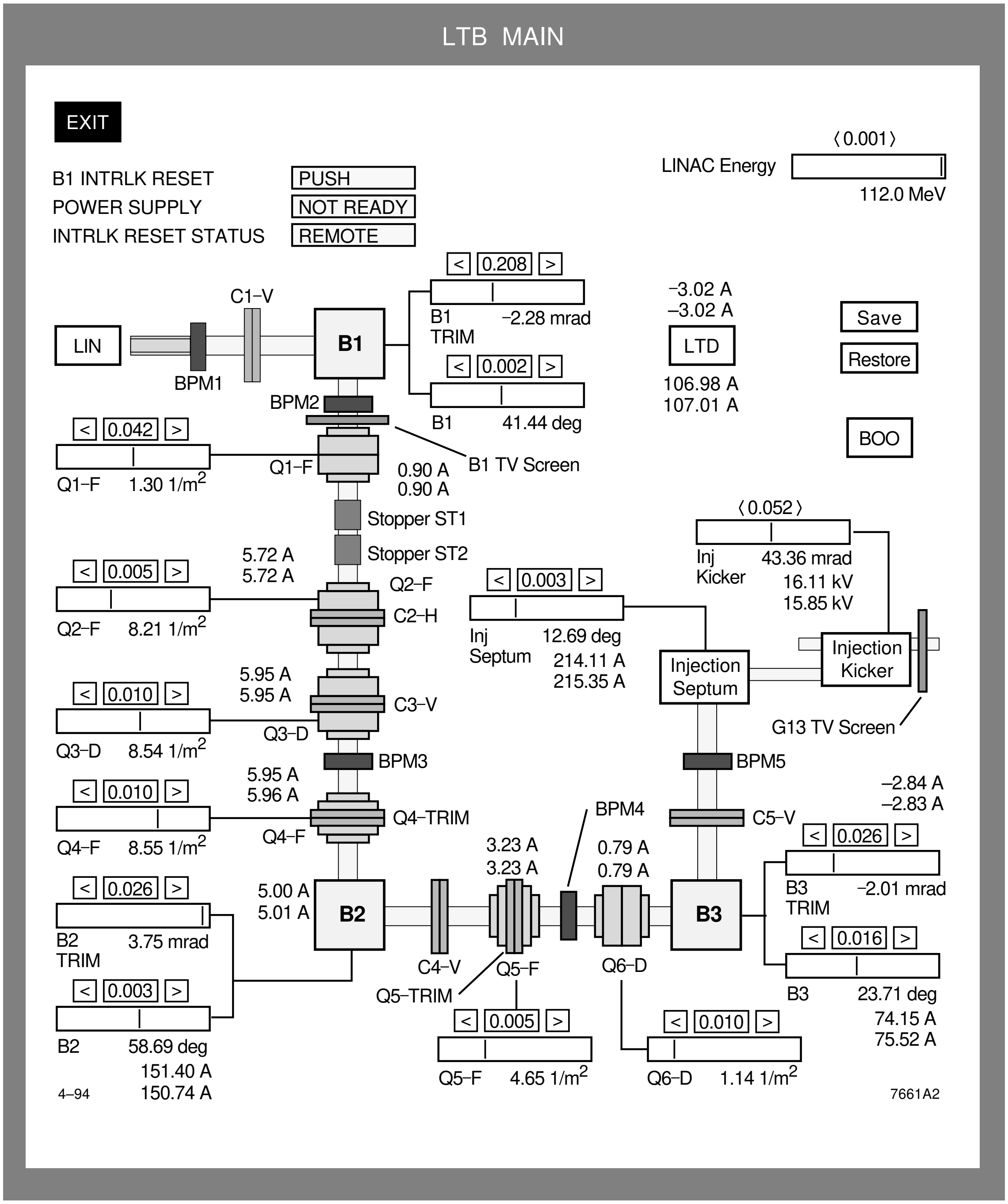

Fig. 2. SSRL injector linac-to-booster transport line.

As a first step, the control system does not just passively control machine parameters under operator command, but also executes automated procedures by itself. This is helpful during a machine startup or shutdown sequence, where all equipment must be turned on or off in a controlled sequence. Correlations between devices and parameters have to be maintained during these procedures to minimize commissioning time, and to protect against equipment damage. This becomes more important as accelerators become increasingly more complicated, to ensure reliable operation with minimum downtime.

Feedback systems (discussed in Chapter 13) are also becoming integrated into the control systems. Fast feedback systems can be implemented directly through the control system, while slow feedback systems can be implemented as software applications. For feedback systems of all kinds, there is an ongoing effort to increase the performance of the accelerator by using faster equipment.

9.7 Accelerator Modeling Overview

In Matt Sands' landmark paper,[18] he made the remark that with storage rings we must pay "particular concern for their performance as instruments for research... ." The recent explosion in the number of storage rings, with ever increasing performance standards, reinforces this notion. If we adopt this view of the storage ring as a scientific instrument, it becomes obvious that the performance of a synchrotron light source depends critically on the accuracy of the model. Machine control programs, for instance, use the model to manipulate the beam orbit and to adjust the electron beam properties. These requirements alone make the model an important part of any accelerator.

In the design phase, computer simulations are used to construct the *ideal* accelerator model. The goal is to produce a machine with electron beam properties that conform to stringent performance specifications, including photon beam brightness and tolerances to error. As a result, the accelerator design phase is a complicated process that seeks stable solutions to inherently nonlinear constraints. The finished design (the accelerator model) requires many iterations between accelerator physicists, engineers, and the synchrotron radiation user community.

Once the design is complete, the magnets, radio-frequency system, and support systems are constructed and installed in the accelerator tunnel. In the construction process, each component is carefully calibrated (see Chapters. 5, 6, and 7) and the calibration data are entered into an on-line version of the model that will be used for machine control.

However, even with great care in the manufacture of each accelerator component, the electron beam properties often differ from the values predicted by the model. To reconcile the difference, accelerator physicists carefully measure the beam properties and compare the results to the model. Large discrepancies between the measurements and the model often indicate problems that can be fixed at the hardware level. Small errors are accounted for by adjusting calibration factors in the model. The new *calibrated* model is then more suitable for machine control.

9.8 Model Components

Accelerator modeling originated as the art of finding solutions for charged particle trajectories under the influence of electromagnetic fields. It has since evolved into a broad field that relies on sophisticated computer programs. In this chapter, we focus on accelerator models for machine control—in particular, the linear field model. This simplification reduces the basic set of model components to:

- drift sections (no magnets)
- bend and corrector magnets

- quadrupole magnets (focusing)
- beam position monitors (bpm)

The next level of model sophistication usually includes nonlinear elements such as sextupole magnets for chromatic correction, and the RF acceleration system needed to restore energy lost from the beam.

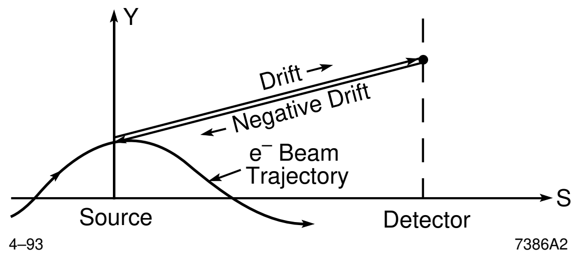

Fig. 3. Construction of a model component for photon beam lines.

Another important model element particular to synchrotron light sources is the photon beam position monitor. Conventional accelerator codes do not explicitly include photon beam position monitors because the magnetic fields acting on the electron beam do not deflect the photon beam. We can, however, use the element sequence (drift)–(bpm)–(negative-drift) to model the photon beam position. As indicated in Fig. 3, the length of the drift, (and the negative-drift) is the distance from the photon beam source point to the photon bpm. (The *negative-drift* simply returns the beam trajectory to the original location of the photon beam source point.) Using this element, it is possible to include the photon beam position in the accelerator model.

The physical configuration of magnets in a synchrotron light source is often referred to as the lattice. Storage ring lattices are usually built up from a repeating pattern of bend and focusing magnets, collectively referred to as a cell. For many applications, each cell has the special property that the electron beam envelope is the same at both ends of the cell. In a light source lattice, the magnet cells are typically designed to be achromatic (no energy dispersion at the ends of the cells), so that insertion devices installed between the cells are driven by a dispersion-free beam. These devices produce the high quality photon beams characteristic of third-generation synchrotron radiation sources (see Chapter 14).

One of the first achromatic cell configurations, the DBA (double-bend achromat), was developed by R. Chasman and G. K. Green for the NSLS.[19] The DBA contains a strong horizontally focusing quadrupole in the center of the cell to drive the dispersion to zero at the ends. A variant of the DBA, the TBA (triple-bend achromat) is based on similar principles. The TBA configuration forms the unit cell for the ALS (Advanced Light Source).[12] For a more complete discussion of lattice design considerations, see Chapter 2.

Returning now to the individual elements of the model, note that each element can be described with an almost arbitrary degree of mathematical complexity. For instance, both the central field and the fringe fields at the edge of the magnet can be represented by high-order multipole expansions. The most popular magnet model for machine control, however, is the simple linear field approximation, based on the first-order expansion of the equations of motion.[20] For machine control, the linear optics model is fast, and is accurate enough for most applications.

The most widely used format for the accelerator model file presently follows the MAD convention.[21] MAD will make all the standard linear optics calculations and produce a table of the accelerator optical functions suitable for post-processing. The HARMON package in MAD[22] also provides a means to compute sextupole strengths for on-line chromaticity correction. The effect of sextupole fields on the beam envelope and on the position of the beam centroid are often neglected in model-based control applications.

More precise accelerator simulation codes tend to concentrate on higher moments of the magnetic field, edge field effects, geometric corrections to the equations of motion for a small bending radius, and the impact of synchrotron oscillations. As discussed in Chapter 2, much of the emphasis on the design of storage rings for synchrotron radiation (and therefore the modeling codes) has been centered around characterizing the dynamic aperture. Codes like TRANSPORT[23] and Marylie[24] perform a Taylor series expansion of the equations of motion through each element, and simulate beam propagation via higher order transport matrices. These codes can generate a high-order power series expansion around the closed orbit to obtain a *one-turn map* of the phase space for single particle motion. Other programs *slice* each element into thin differential strips, and perform *symplectic integration* to find the particle motion.[25] TRACY and TRACY–II contain a PASCAL compiler with an accelerator physics *library* that allows the user to program an arbitrary sequence of calculations.[26]

Many other modeling programs exist for a wide range of applications. For a compendium of accelerator modeling programs, including a description of their applications, see Ref. 27. Given the available range of high-order programs, it is particularly important to recognize that the choice of program is an integral part of the accelerator model. Depending on the application, one should always bear in mind that the model is an approximation to the actual accelerator hardware, and that very high order models have a limited range of validity for machine control. Even where the range of validity is accurate to higher orders, one may wish to trade model accuracy for computation speed and ease of use.

9.9 Accelerator Model Interface

As described in the previous sections of this chapter, the on-line model is closely connected to the accelerator control system. One function of the on-line model is to monitor and control the electron beam orbit. In this case, the control system converts voltage signals from beam position monitors (bpms) into orbit position data by way of calibration factors. The model then calculates a set of dipole kicks to adjust the orbit, and a different set of calibration factors are used to convert these kicks to power supply voltages. The entire process takes place from a graphical interface connected to the control system.

A closer look illustrates how the calibration factors are integrated into the control system. The example in Section 9.2 outlined the measurement and control of magnet power supplies. For model purposes, a set of calibration factors is used to convert each power supply read back value to a magnetic field strength. This requires at least two signal conversion stages: first, the voltage of a current transducer is converted to a value for amperes flowing through the magnet windings; then, a second set of calibration factors convert the amperes to magnetic field strength. This conversion stage can be complicated when the magnet cores have nonlinear magnetization characteristics. To parameterize the behavior of the iron, a set of polynomial coefficients can be numerically fit to data points measured in the laboratory. These polynomials convert the power supply read back values to field strength. Further details on the magnet measurement and calibration process can be found in Chapter 7.

Once the magnet calibration factors are in place, the model can be used to estimate the beam parameters. For on-line applications, a *skeleton* file containing the lattice geometry and effective length of the magnets forms the basis for the model. The control system loads magnet strengths into this file and the model program calculates the beam parameters. Using a similar procedure, the accelerator model can simulate longitudinal beam dynamics based on read back values from the radio-frequency system.

To provide the experimentalist with more flexibility, it is common to add an extra calibration factor for each lattice component in the database. For instance, the quadrupole strengths can be multiplied by the extra factors before running the on-line model. Proceeding in the other direction, quadrupole strengths calculated by the model are divided by the extra calibration factors before conversion to power supply currents. Although these calibration factors do not strictly conform to the polynomial formalism, they allow the experimentalist to adjust the model to conform with measurements. Section 9.11 of this chapter outlines techniques that can be used to determine these factors experimentally.

9.10 Model-Based Calculations

Model-based machine control relies primarily on calculations of the beam transport matrices and β-functions. The beam transport matrices describe single particle motion in terms of phase-space coordinates as the beam travels through a transport line or storage ring. In this chapter, we restrict the motion to first-order oscillations in the plane transverse to the beam direction. Closely related to the transport matrices are the response matrices that describe the motion of the closed orbit caused by dipole kicks in a storage ring.

The β-functions also describe single particle motion (and closed orbit perturbations), but the β-functions have a much wider range of applications. In particular, the β-functions are used for resonance calculations, to yield statistical properties of the beam, and to parameterize the beam envelope. It is important to note that both the transport matrices and the β-functions are nonlinear functions of the model parameters, and their ability to predict machine behavior depends on the accuracy of the model.

9.10.1 Transport and Response Matrices

Transport matrices are used in accelerator physics to propagate the phase-space coordinates of the beam from one position to the next. In general, the phase space can be six-dimensional, $(x, x', y, y', \Delta p/p, \Delta s)$, where each coordinate is evaluated relative to an ideal reference orbit.

In a synchrotron, the transverse oscillations come about as the result of periodic quadrupole focusing. These are the betatron oscillations. Similar to a harmonic oscillator, betatron oscillations can be represented on a phase-space diagram where the coordinates indicate the displacement and the angle of the motion. The transport matrices map the change in coordinates between any two positions in the accelerator. To first order, we can write

$$(x, x')_2 = \mathbf{R}(x, x')_1 \ , \tag{9.1}$$

where the matrix $\mathbf{R}$ is a function of the lattice parameters between position 1 and position 2.[28] The elements of $\mathbf{R}$ have two equivalent representations, either in terms of the lattice parameters or in terms of the β-functions. Accelerator modeling programs often calculate $\mathbf{R}$ across each element and multiply matrices to find the cumulative effect. To model linear particle dynamics, including synchrotron motion and coupling effects, a 6×6 dimensional transport matrix is used. Nonlinear terms can be added by expanding the equations of motion to higher order, and extending the column vector of phase-space coordinates.[23,28,29] In the accelerator literature, the components of the 2×2 transport matrix are often denoted

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \ . \tag{9.2}$$

The focusing effect of a quadrupole, for example, enters through the R_{21} term.

For a storage ring, we can form a matrix of the R_{12} elements that connect each corrector magnet kick to the orbit displacement at each bpm. This *response matrix* can be either calculated by the model, or measured directly. For precise orbit control applications, such as fast orbit feedback, the direct measurement may be preferred for maximum precision. It is not always convenient, however, to interrupt operation of the machine to measure the response matrix when the machine optics is changed (e.g., insertion device adjustment). To update the response matrix, either a model-based calculation is needed, or the feedback system must have an adaptive feature that can *learn* the new response matrix.

9.10.2 Beta Functions

Beta functions were first used by D. Kerst to describe transverse particle oscillations in the early betatron machines at General Electric and the University of Wisconsin.[30] Hence, the name *beta functions*. The theory of β-functions was then formalized by E. D. Courant and H. S. Synder for alternating gradient accelerators.[31] Since that time, β-functions have become a standard tool for accelerator analysis. A development of β-function theory, including a wide range of applications to storage rings, was recently compiled by Wiedemann.[20]

The β-functions are extremely useful for accelerator analysis because together with the dispersion function they form a compact representation of the electron beam parameters. They also provide a powerful tool to estimate the effect of field perturbations on the particle beam envelope, the closed orbit, and the impact of resonances on beam dynamics. Mathematically, the β-functions are a solution for single particle motion in the accelerator structure. Physically, the β-functions predict the amplitude and phase of transverse particle oscillations relative to the equilibrium orbit. They also parameterize the phase space ellipse of the particle beam distribution function as the beam circulates around the accelerator. It is important to note, however, that the β-functions are derived quantities, and do not themselves represent the accelerator model.

To see how the β-functions are derived from the model, we first consider the equation of motion for an on-momentum particle oscillating with respect to the closed orbit (see Chapter 2, Section 3):

$$x'' + K(s)x = 0 \ . \tag{9.3}$$

The function K(s) in Eq. (9.3) represents the quadrupole focusing action as a function of position 's' along the ideal orbit. For on-line model applications, K(s) is determined directly from power supply current measurements. Looking ahead to Section 9.11, one of the great challenges of experimental accelerator physics is to determine the exact value of the focusing function K(s) in the synchrotron!

To within a constant phase factor, the two linearly independent solutions to Eq. (9.3) can be combined to yield an expression for the particle position in terms of the β-function, $\beta(s)$, and the phase advance of the oscillation, $\phi(s)$,

$$x(s) = x_o\sqrt{\beta(s)}\ \cos[\phi(s) + \phi_o] \ , \tag{9.4}$$

where x_o and ϕ_o are the initial oscillation amplitude and phase of the oscillation. In this representation, it is clear that the function $\sqrt{\beta(s)}$ gives the amplitude modulation of the motion, and $\phi(s)$ is the phase advance for the motion. Furthermore, it is clear that for a system of many particles, each with random oscillation phase ϕ_o, the function $\sqrt{\beta(s)}$ is proportional to the size of the beam envelop at each position 's' (neglecting dispersion effects[20]).

Two parameters closely related to the β-function are $\alpha = -\beta'/2$ (derivative) and $\gamma = (1+\alpha^2)/\beta$. Together, the set of functions $\{\alpha, \beta, \gamma\}$ are often referred to as the Twiss parameters. The quantity $\varepsilon = \gamma\ x^2 + 2\alpha x x' + \beta x'^2$ (the celebrated Courant-Snyder invariant[31]) is a constant of the motion for lossless betatron oscillations. For a given distribution of betatron oscillation amplitudes, the invariant quantity 'ε' is used to characterize the emittance of the beam. See also Chapter 2 and Section 9.10.5 below.

To calculate the β-functions from the model, we can use the fact that the beam envelope repeats each turn and the invariance of the Courant-Snyder constant to yield a similarity transformation for propagation of the Twiss parameters:

$$\sigma(s_2) = \mathbf{R}\sigma(s_1)\mathbf{R}^T \ , \tag{9.5}$$

where the beam sigma matrix is given by

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix} \tag{9.6}$$

In this equation, **R** is the same phase-space transport matrix used in Section 9.10.1 to propagate single particle trajectories. Replacing **R** with the transport matrix for one complete revolution around the accelerator, and rearranging terms, we have

$$\{\beta_o, \alpha_o, \gamma_o\} = \mathbf{S}\{\beta_o, \alpha_o, \gamma_o\} \ , \tag{9.7}$$

where the elements of the 3×3 matrix **S** are functions of the storage ring model parameters (quadrupoles, bends, drifts, etc.).[32] Equation (9.7) can be solved for the initial

Twiss parameters, $\{\beta_o, \alpha_o, \gamma_o\}$, at one point in the storage ring. The Twiss parameters in the remainder of the ring are then found by propagating these values via the similarity transformation of Eq. (9.5). The x and y β-functions as calculated by the COMFORT program for SPEAR are indicated in Fig. 4. The phase function,

$$\phi(s) = -\int_o^s \frac{ds}{\beta(s)} \tag{9.8}$$

is found by direct integration, and plotted in Fig. 5 for one-half of the SPEAR ring.

9.10.3 Closed Orbit Perturbations

It is well known that a section of the closed orbit in a storage ring can be analyzed just like a transport line. To find the beam coordinates, we specify the initial conditions (x_o, x_o'), and propagate the trajectory with transport matrices through each element. The difference between transport lines and storage rings, however, is that in a storage ring only one initial condition of the orbit (x_o, x_o') repeats after each revolution. We call this trajectory the closed orbit. A betatron oscillation is motion with respect to the closed orbit that does not repeat each turn (for non-integer tunes).

For a well aligned lattice, the on-momentum beam passes through the magnetic axis of each element. In this case, the initial conditions of the closed orbit are simply (x_o, x_o')=(0, 0). If a dipole field perturbation is introduced, at the point of the kick the new orbit satisfies the closure condition

$$M \begin{pmatrix} x_o \\ x'_o \end{pmatrix} + \begin{pmatrix} 0 \\ \theta \end{pmatrix} = \begin{pmatrix} x_o \\ x'_o \end{pmatrix} , \tag{9.9}$$

where **M** is the one-turn transport matrix, and θ is the magnitude of the dipole kick. The position and angle of the closed orbit at the

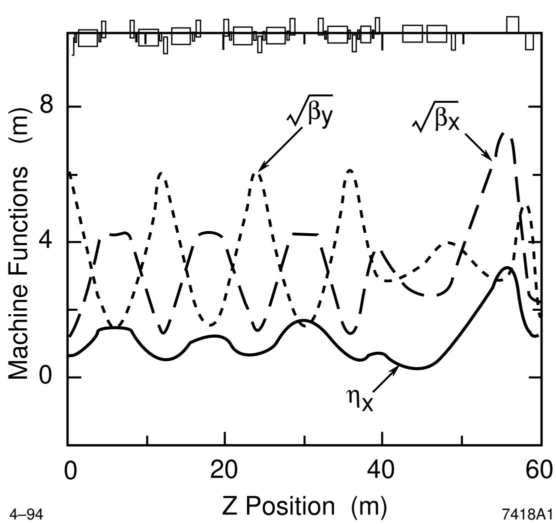

Fig. 4. β-functions for the SPEAR light source lattice.

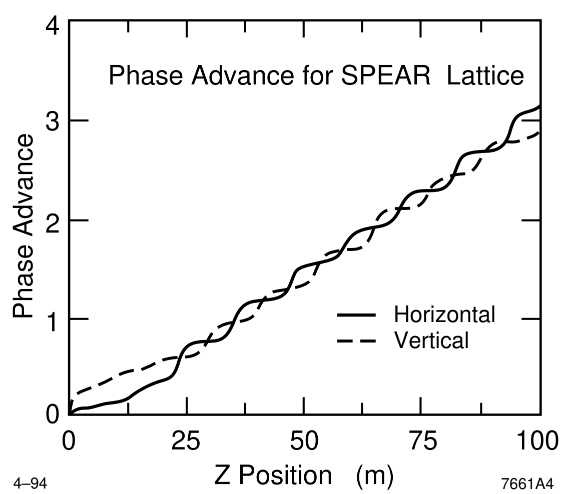

Fig. 5. Phase advance functions for the SPEAR lattice.

point of deflection are

$$(x_o, x'_o) = (1-\mathbf{M})^{-1}\ (0,\theta)\ , \qquad (9.10)$$

At any other point in the ring, the coordinates of the closed orbit perturbation are found by propagating the orbit via the transport matrix, $(x, x') = \mathbf{R}\ (x_o, x_o')$. If more than one dipole kick is present, the closed orbit is a superposition of orbit perturbations.

As indicated in Fig. 6, the orbit deflection at the location of the corrector magnet in a storage ring cannot be distinguished from a corrector kick in a transport line. In the storage ring, however, the coordinates (x_o, x'_o) of the closed orbit satisfy the closure condition, Eq. (9.9).

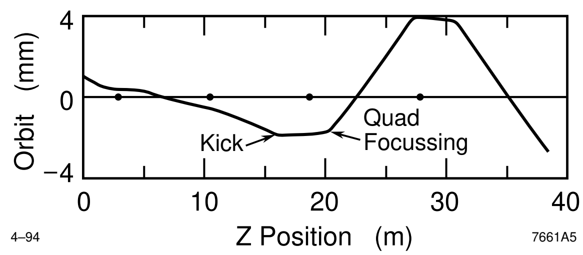

Fig. 6. Local closed orbit deflection from a single dipole kick.

No discussion of closed orbit perturbations is complete without mention of the β-function representation. In this context, the response matrix elements relating a kick $\Delta\theta_i$ at point 'i' to the beam displacement Δx_j at point '*j*' are written

$$\Delta x_j \ = \ \Delta\theta_i\ \frac{\sqrt{\beta_i\beta_j}}{2\sin\pi\nu}\ \bullet\ \cos\nu\ (\pi - |\phi_j - \phi_i|)\ , \qquad (9.11)$$

where the phase factors advance from 0 to 2π, and ν is the tune of the accelerator (phase advance per revolution). The absolute value of the phase argument permits the point of observation to be either ahead or behind the kick.

Although the β-function representation of the response matrix is equivalent to the response matrix formulation of Section 9.10.1, it has several features that illuminate the physical behavior of orbit motion. First, from Eq. (9.11), it is clear that dipole kicks located at points where the β-function is large will induce relatively large closed orbit perturbations. For this reason, strong quadrupoles located in regions of *large beta* must be well aligned. Likewise, bpms located at points of large beta are the most efficient way to detect orbit perturbations. Moreover, Eq. (9.11) tells us that integer tunes are unstable, and indicates how to select a corrector with the appropriate phase advance to cancel an orbit perturbation (which set of correctors will produce a local orbit *bump*).

9.10.4 Dispersion

The dispersion function tells us the deviation of the closed orbit for off-momentum particles from the (nominal) closed orbit defined for particles at the synchronous energy (Chapter 2). The dispersion function can either be evaluated from the model or measured directly (radio-frequency variation). To calculate the dispersion function from the model, we first extend the transport matrix formalism to include momentum dependence, and then solve for the off-momentum closed orbit, often called the η-function. In the horizontal plane, the transport matrix that maps the η-function for one turn around the storage ring becomes,

$$\mathbf{M} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ 0 & 0 & 1 \end{pmatrix} , \tag{9.12}$$

where the new elements R_{13} and R_{23} give the dependence of horizontal position and angle on energy. Analogous to Eq. (9.9), the initial coordinates (η, η') for the dispersion function satisfy the closure condition

$$\mathbf{M}(\eta,\eta',1) = (\eta,\eta',1) , \tag{9.13}$$

and the dispersion is evaluated at all other points by propagating the initial conditions with the appropriate 3x3 transport matrix computed from the model. Figure 4 includes a plot of the dispersion function for SPEAR. A more general discussion of the dispersion function evaluated in terms of integrals over the β-functions and phase advance is covered in reference.[20] For our purposes, it is important to note that independent of the method of calculation the model must be well calibrated for the predicted dispersion to agree with the measured value.

9.10.5 Synchtrotron Integrals

Up to this point, we have demonstrated how the accelerator model can be used to calculate the shape of the electron beam envelope (β-functions, dispersion) and the coordinates of the closed orbit. These calculations do not, of course, constitute a full representation of the electron beam. What is still missing are the particle distribution functions in both the longitudinal and transverse planes.

As described in Chapter 2, photon emission in the presence of dispersion creates a shift in the closed orbit, and consequently excitation of betatron oscillations. Photon emission also stimulates synchrotron oscillations in the longitudinal plane. Both of these effects are mediated by the action of the radio frequency drive system, with the result that the particle distribution functions relax to a stationary equilibrium value. Since the emission of energy quanta (photons) is statistically random, the Central Value Theorem of statistical mechanics tells us that the particle distribution functions are Gaussian.

In the plane transverse to the beam motion, the spread in oscillation amplitudes is characterized by the emittance of the beam. Recalling the Courant-Snyder invariant for betatron oscillations, $\varepsilon = \gamma x^2 + 2\alpha xx' + \beta x'^2$, the transverse particle distribution can be written

$$\rho(x,x') = \frac{1}{\sqrt{2\pi}\ \varepsilon} \exp -\left(\frac{\gamma x^2 + 2\alpha xx' + \beta x'^2}{\varepsilon} \right) . \tag{9.14}$$

In this case, the emittance 'ε' characterizes the distribution of oscillation amplitudes. Neglecting dispersion effects, in a given plane the rms size of the beam cross section is $\sqrt{\varepsilon\beta(s)}$, at point '$s$.' Low emittance (cold) electron beams imply dense distribution functions, leading to high brightness photon beams. The longitudinal distribution is also Gaussian, and can be characterized by an rms energy spread, σ_E.

How are the equilibrium emittance and energy spread determined for stationary stored beams? This question was answered by the synchrotron radiation integrals.[18] In short, the synchrotron radiation integrals summarize the effects of photon emission in the accelerator guide field to yield estimates for the following important parameters:

- energy loss per revolution
- equilibrium beam emittance
- energy spread
- damping times for transverse and longitudinal oscillations
- path length dependence on energy

In the linear approximation, the five primary synchrotron radiation integrals all contain at least one factor of the inverse bending radius (ρ^{-1}) in the argument of the integral (photon emission in dipole fields), with additional dependence on n (field index), η, β, and their derivatives, depending on the integral. To calculate the synchrotron radiation integrals, it has been demonstrated that only the β– and η–function values at the entrance of each magnetic element with finite bending radius are required.[33] This feature makes evaluation of the integrals fast and efficient for on-line modeling applications.

9.11 Model Calibration

In this section, we review experimental procedures that can be used to calibrate the linear optics model. These procedures are based on comparing beam measurements to the values predicted by the model. Discrepancies between the two sets of data are eliminated by adjusting parameters in the model.

Depending on the goals of the experimentalist, several approaches to model calibration are possible. It is easy, for example, to adjust the model so that it predicts the measured tunes. Finding a more accurate model requires more careful measurements, and sometimes complicated numerical analysis. In general, the goal of the model calibration procedure is to first adjust the model to accurately reflect the measured beam parameters, and then adjust the accelerator so that the beam parameters agree with their design values. For a review of diagnostic techniques used for synchrotron light sources, see Chapter 10 and references therein.

9.11.1 Optics Calibration

One common diagnostic is to spectrum analyze transverse oscillations of the beam to determine the tunes of the lattice. If the measured value of the tunes are different from the model value, we can adjust the quadrupole strengths in the model to make the model agree with the measurement. The easiest way to find the new quadrupole strengths is to Taylor-expand the model tunes to first order in terms of the quadrupole strengths. Since the horizontal and vertically focusing quadrupoles dominate ν_x and ν_y, respectively, we can choose one QF and one QD quadrupole family for the Taylor expansion. The choice of quadrupole families is, to some extent, arbitrary. Note that although accelerator model calibration is an inherently nonlinear problem, the equations can often be solved by linear analysis if we make a first order Taylor expansion.

It is also possible to estimate β–function values at discrete locations in the lattice from measurements of the corrector-to-bpm response matrix, or from turn-by-turn oscillation measurements. Each entry of the corrector-to-bpm response matrix, for instance, has a β–function representation (neglecting RF effects)

$$\frac{\sqrt{\beta_i \beta_j}}{2\sin \pi\nu} \cdot \cos \nu(\pi - |\phi_j - \phi_i|) \ , \tag{9.11}$$

where 'i' and 'j' indicate the location of the kick and location of the bpm measurement, respectively. Initially, we have a set of model values $\sqrt{\beta_o}$ and ϕ_o at each corrector and each bpm. These values can be expanded to first order,

$$\sqrt{\beta} \ = \ \sqrt{\beta_o} + \Delta\sqrt{\beta} \ ,$$

$$\phi \ = \ \phi_0 + \Delta\phi \ , \tag{9.15}$$

and substituted into Eq. (9.11). Retaining only terms to first order, we obtain a linear matrix equation for the perturbed values of $\Delta\sqrt{\beta}$ and $\Delta\phi$. The tunes can also be varied as fitting parameters. Since the problem is inherently nonlinear, the linear fitting procedure may require several iterations for the solution to converge. Although knowledge of the β-functions at discrete locations can be useful for machine control, it is not equivalent to knowing the quadrupole field strengths.

Several algorithms exist for numerically fitting the model parameters (e.g., quadrupole strengths) so that the model predictions agree with the measured response matrix.[34] These algorithms typically compare columns of the model matrix (calculated orbit perturbations) to columns of the measured matrix (measured orbit perturbations). The advantages of using a response matrix for model calibration include the following:

- The beam motion probes the spatial field structure of each quadrupole
- The beam motion tests for bpm linearity
- The corrector kicks can be calibrated
- The large number of measurements provides good redundancy for numerically fitting the model

Moreover, since the response matrix measurements produce difference orbits, the unknown effect of bpm offset errors and dipole field kicks from misaligned quadrupoles are eliminated from the data. Clearly, the measured response matrix contains a wealth of information. To keep the response matrix measurement linear, sextupole magnets, and insertion devices should be turned off whenever possible.

Quadrupole field errors are difficult to analyze, however, because the response matrix has a nonlinear dependence on the field strengths. In this case, we can assign several columns of the measured response matrix as fitting constraints and search for the correct quadrupole strengths for the model with a nonlinear fitting program. The solution is well constrained if orbit perturbations measured in both the *'x'* and *'y'* planes are used.

As an alternative to nonlinear model calibration methods, it is straightforward to expand the difference between the model response matrix and the measured response matrix to first order in terms of the following parameters:[35]

- quadrupole field strengths
- corrector gain factors
- BPM gain factors

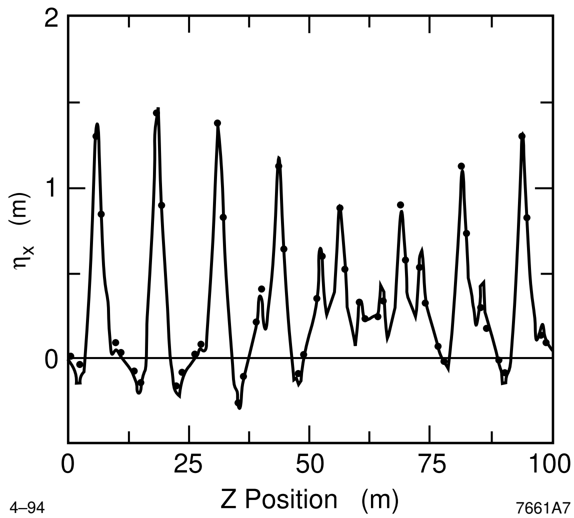

Fig. 7. Model prediction and measurement of dispersion function in NSLS x-ray ring.

If we include contributions from the dispersion orbit caused by the corrector kicks in the horizontal plane,[36] the linearized equations can be written in vector form:

$$\Delta\mathbf{C} = \frac{d\Delta\mathbf{C}}{d\mathbf{k}}\,\Delta\mathbf{k} + \frac{d\Delta\mathbf{C}}{d\theta}\,\Delta\theta + \frac{d\Delta\mathbf{C}}{d\mathbf{g}}\,\Delta\mathbf{g} + \frac{d\Delta\mathbf{C}}{d\delta}\,\Delta\delta \;, \qquad (9.16)$$

where ΔC is the difference between the model and measured response matrix, and the variables $\{\Delta\mathbf{k}, \Delta\theta, \Delta\mathbf{g}, \Delta\delta\}$ are the quadrupole gradient errors, corrector gain factors, bpm gain factors, and momentum shift ($\Delta P/P_o$) produced by each horizontal corrector kick, respectively. Solving for these values by linear matrix inversion helps to bring the model into agreement with the measurements. Before inverting the matrix, error bars can be added for the bpm read back noise to weight the individual equations.

A convincing test of this algorithm was carried out on the NSLS x–ray ring.[36] First, all sextupole magnets were turned off and the insertion devices were retracted to obtain a linear lattice, and the response matrix was measured several times to accumulate statistics. After several iterations of the fitting procedure, the new model agreed with the measured data to an accuracy of about 0.2%, or 2 microns for a 1 mm orbit perturbation. The *calibrated model* also predicted the measured tunes to within experimental accuracy. Hence, the calibrated computer model was consistent with the measured data.

Having established a linear optics model for the x-ray ring without sextupoles, the sextupoles were turned on and the response matrix remeasured. Since a beam passing off-axis through a sextupole magnet experiences a quadrupole field, this time the calibration variables were quadrupole field components at the sextupole magnet locations. The fitting again produced a statistically viable solution, in agreement with the measured response matrix, that indicated the beam was off-center in some of the sextupoles by several mm. This result agreed with bpm measurements of the closed orbit. As a cross check of the optics model with the sextupoles on, the new model was used to predict the dispersion function. As demonstrated in Fig. 7, the predicted dispersion agreed with the measured dispersion, including an error component

The result of the NSLS x-ray ring experiments was to produce a linear optics model that predicted:

- the measured response matrix
- individual quadrupole field strengths
- corrector calibration factors
- bpm calibration factors
- beam position in sextupole magnets
- tune (in both planes)
- the dispersion function

Given this pedigree for the calibrated model, it is reasonable to believe that it will predict the actual β-functions and synchrotron integrals to a high degree of accuracy. The model should also be sufficiently accurate to begin searching for alignment errors.

9.11.2 Beam Based Alignment

Magnet and bpm alignment errors are considerably more difficult to detect than focusing errors. In the first place, prior to searching for alignment errors, the linear optics model should be established. This reduces the number of variables in the problem. The dominant source of orbit perturbations are then dipole field components (e.g., quadrupole misalignments), but the orbit measurement can also be contaminated by bpm read back errors, and can contain a momentum error. Including these terms, the expression for the *measured* orbit can be written in vector notation as

$$\mathbf{x}_{bpm} + \Delta\mathbf{x}_{bpm} = \mathbf{C}\theta + \mathbf{Q}\Delta\mathbf{x}_q + \frac{\eta\Delta P}{P_o} \quad , \tag{9.17}$$

where **C** and **Q** are the response matrices, as viewed at the bpms, for the corrector kicks (θ) and quadrupole misalignments ($\Delta\mathbf{x}_q$), respectively. The bpm read back errors are contained in the column vector $\Delta\mathbf{x}_{bpm}$. The term $\eta\Delta P/P_o$ is the dispersion component of the orbit that appears if the beam is off-momentum. Note that although the corrector response matrix **C** and the dispersion function η can be measured, the response matrix **Q** for quadrupole misalignments must be calculated from the model.

In general, the solution for the exact electron beam orbit relative to the design orbit [$\mathbf{x}_{bpm}$ in Eq. (9.17)] is extremely difficult to determine because the number of unknown variables $\{\Delta\mathbf{x}_{bpm}, \Delta\mathbf{x}_q, \Delta P/P_o\}$ is greater than the set of measurements $\{\mathbf{x}_{bpm} + \Delta\mathbf{x}_{bpm}\}$ for a single optical configuration.

The under-constrained data analysis problem makes it difficult to distinguish dipole field errors from bpm offset errors. It is also difficult to distinguish between quadrupole misalignments and bend field errors. Furthermore, unlike the optics calibration procedure, deflecting the orbit does not yield new information (to first order) because dipole fields are spatially homogeneous.

To help solve these problems, the analysis of beam-based alignment data can be broken up into a sequence of piecewise solutions that agree with the model (the GOLD method[34]). An example taken from the SPEAR storage ring in shown in Fig. 8, where, a stray dipole kick from the injection transport line is found by propagating a fitted trajectory in the forward and backward directions past the location of the kick. An isolated bpm error is also indicated in the measured data. The closed orbit perturbation, with the dipole kick included in the model, is shown in Fig. 8c.

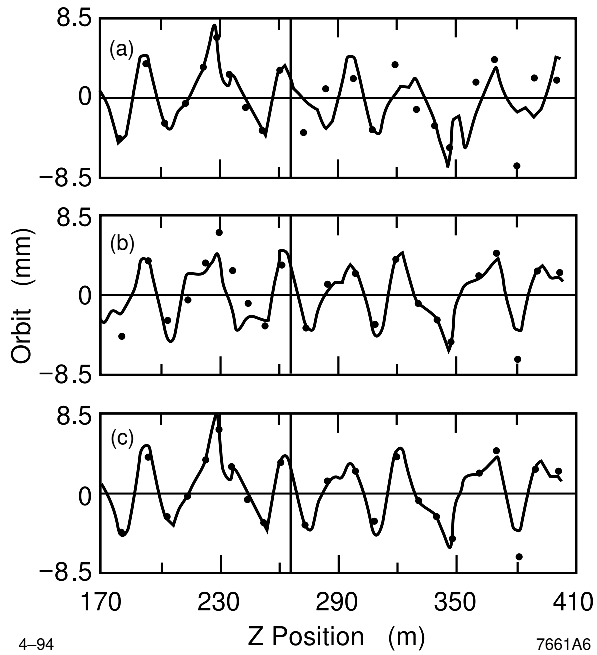

Fig. 8. Closed orbit trajectory, propagated (a) forward and (b) backward, over dipole kick error, and (c) closed orbit with dipole kick.

Another simple, yet informative, technique for analysis of the absolute orbit is to interpolate the beam position between adjacent bpm readings. In this case, we compute the angle at the first bpm to pass the trajectory through the second bpm. A quick scan of the result graphically illustrates potential bpm or quadrupole alignment errors.

An algorithm for analyzing sets of three contiguous bpm readings has been tested at BNL.[37] This technique found errors in the AGS Booster orbit that were attributed to improper electrical connections. An extension of the bpm triplet analysis was developed at CERN.[38] This technique relies on an analysis of the χ^2 (residue) generated by least-squares fitting the beam trajectory data for bpm triplets, quadruplets, quintuplets, and so on. The residue method is particularly suitable for large storage rings where the amount of data makes processing difficult—on LEP, for example, a localized field error was found that was corrected by realigning quadrupole magnets.

Alignment errors can also be found by simultaneous analysis of the orbit measured with several different optical configurations.[39] In principle, the bpm and quadrupole offset errors remain constant, independent of the optical configuration. By changing the quadrupole strengths, a set of closed orbit equations [Eq. (9.17)] with different response matrices (**C** and **Q**), and different dispersion functions can be produced. From the expanded set of equations, the quadrupole offsets, bpm offsets, and beam energy offset can be found. The key to this method, of course, is to generate a set of sufficiently different optical configurations to make the solution unique.

9.12 Model–Based Control

The discussion up to now has been directed toward calculation of beam parameters and model calibration. With these tools, we are now in a position to use the model for machine control. The goal, of course, is to guide the electron beam parameters to an optimum state for the photon beam users. This requires model-based calculations with reliable model-to-magnet calibration factors. Once the state of the accelerator is changed, the result should always be verified by appropriate diagnostic measurements.

9.12.1 *Orbit Control*

As an introduction to model based accelerator control, consider control of the closed orbit. In general, the goal of almost any controlled orbit change will fall into one of the following categories:

- orbit amplitude reduction
- local orbit bumps
- orbit adjustment to a previous reference

Often, the minimum orbit amplitude (as measured at the bpms) is not the optimum orbit for machine operation. For example, it is possible to include dispersion corrections in the orbit control program,[40] to adjust the position and angle of the photon beams, or to ask the orbit correction algorithm to minimize corrector strengths. Local orbit bumps are also used to fine tune the injection rate and to control coupling.

To manipulate the closed orbit, operators use dipole corrector magnets to deflect the beam, and beam position monitors detect the motion (either in the storage ring or on the photon beam lines). To first order, orbit adjustments require finding a solution to the linear matrix equation:

$$\Delta\mathbf{x} = \mathbf{C}\Delta\theta \ , \tag{9.18}$$

where $\mathbf{C}$ is the corrector-to-bpm response matrix, $\Delta\mathbf{x}$ is the desired orbit perturbation, and $\Delta\theta$ is a column vector of the corrector strengths.

Several common methods for ‘orbit control’ are outlined below. In each case, the mathematical calculation is based on either a computed or measured corrector-to-bpm response matrix. The solution to a set of linear equations (matrix inversion) yields a corrector pattern that will move the orbit toward the desired goal.

9.12.1.1 MICADO[41]

Historically, the MICADO package was one of the first orbit control algorithms. In progression, MICADO finds the single most effective corrector, the most effective additional corrector, and so on, to implement the orbit correction. The main advantage of MICADO is the ability to find the single most efficient corrector, which may point to the location of an error in the lattice.

9.12.1.2 Harmonic[42]

Every closed orbit perturbation can be decomposed into a Fourier series evaluated on the normalized, periodic phase interval 0 to 2π.[31] For a dipole perturbation, the closed orbit is in fact a periodic solution to a kick term driving the single particle equation of motion, Eq. (9.3). The harmonic spectrum of the orbit contains components in ratio

$$F(n) \propto \frac{1}{n^2 - \nu^2} \ , \tag{9.19}$$

where ν is the betatron tune and n is the harmonic wave number.

Since the spectra of orbit perturbations are dominated by harmonics near the tune, Fourier decomposition on a limited set of harmonics centered about the tune can produce an effective representation of the orbit. Conversely, to compensate components centered about the tune requires only small corrector amplitudes. These features make harmonic orbit correction a useful technique for orbit feedback systems.[42]

Complications with harmonic orbit analysis include:

- non-equal phase interval between bpms,
- Nyquist sampling limits,
- the need for accurate β-function values at the bpm locations

The latter restriction underscores the need for a calibrated optics model.

9.12.1.3 Eigenvector[43]

Given a linear set of equations, $\Delta \mathbf{x} = \mathbf{C}\,\Delta\theta$, we want to find solutions for a corrector pattern ($\Delta\theta$) that will produce the orbit perturbation, $\Delta\mathbf{x}$. One approach to finding $\Delta\theta$ is to solve this problem by the method of eigenvectors. In orbit correction applications, however, the response matrix $\mathbf{C}$ is not always square. For the general M×N dimensional matrix (M-bpms, N-correctors) there are three possibilities:

M = N (square matrix) In this case, for a full-rank matrix $\mathbf{C}$, there exists one unique solution, $\Delta\theta = \mathbf{C}^{-1}\Delta\mathbf{x}$. The inverse response matrix, $\mathbf{C}^{-1}$, can be found by eigenvector decomposition. Whenever a singular (not invertible) matrix is encountered, Singular Value Decomposition[44] can be applied (see below). The action of the eigenvectors is to project the orbit perturbation $\Delta\mathbf{x}$ onto the eigenbasis of the matrix $\mathbf{C}$. This projection yields the corrector pattern, $\Delta\theta$. Notice that orbit decomposition in terms of eigen-vectors of the response matrix is analogous to a Fourier series decomposition in terms of harmonic basis vectors. Depending on the application, the corrector pattern can be produced from all, or a subset, of the eigenvectors.

M > N (more bpms than correctors). This case is a linear least-squares problem where the number (or choice) of correctors cannot produce all orbit displacements that span the range of orbit perturbations, $\Delta\mathbf{x}$. The standard least squares solution procedure is to multiply Eq. (9.18) by $\mathbf{C}^{T}$ to make a square matrix: $\mathbf{C}^{T}\Delta\mathbf{x} = \mathbf{C}^{T}\,\mathbf{C}\Delta\theta$. The resulting corrector pattern is found by matrix inversion,

$$\Delta\theta = (\mathbf{C}^{T}\mathbf{C})^{-1}\ \mathbf{C}^{T}\Delta\mathbf{x} \quad . \tag{9.20}$$

In the language of linear algebra, the least-squares procedure projects the orbit perturbation onto a subspace of basis vectors spanned by the correctors (the columns of matrix $\mathbf{C}$). The process is analogous to finding the shortest path from a point to a line. Least-squares orbit control can be used to produce a smoothed orbit correction in the presence of bpm noise.

M < N (more correctors than bpms). In this case, the fitting problem is under-determined, that is, the matrix $\mathbf{C}$ is not full rank. Under these conditions, SVD will produce the 'psuedo-inverse' of the matrix, $\mathbf{C}^{-1}$. Since SVD minimizes the rms (root-mean-square) value of the solution vector $\Delta\theta$, this is where SVD makes its application to accelerator control particularly attractive.[45] [The minimum rms result is easy to demonstrate. Note that the solution vector, $\Delta\theta$, is a linear combination of the homogeneous solution $\Delta\theta_n$ (null vectors) and the particular solution $\Delta\theta_p$ (in the row space of **C**). Each null vector is a corrector pattern that does not move the beam at the bpms under observation. Since SVD can isolate the particular solution, the magnitude of the corrector kick vector is minimized.]

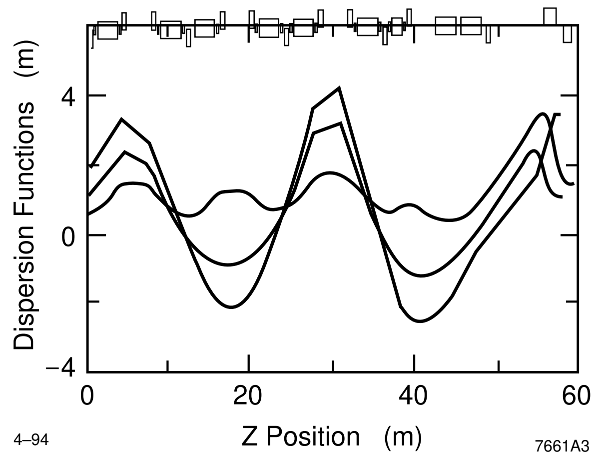

Fig. 9. Model calculation of dispersion function used to guide SPEAR lattice toward a low momentum-compaction lattice.

In the language of control theory, matrix inversion by SVD minimizes the cost function; in this case, the rms current in the corrector magnets. This feature of SVD is routinely applied to SPEAR to reduce the strength of the beam line steering correctors.[46] Here, the position of nine photon beams is held constant (modeled as per Section 9-8), and the SVD algorithm reduces the rms value of up to 30 corrector magnets. The reduced corrector strengths give more overhead for the local beam line steering servos and the global orbit feedback system. For global orbit feedback applications, SVD is useful because one can vary the number of eigenvectors needed for orbit representation (eigenvalue cut-off), and the corrector kicks are minimized at each feedback cycle.[43,47,48]

9.12.2 On-Line Optics Control

Another important function of the accelerator model is to control the optical functions. It is common practice, for instance, to use the model to calculate a new set of quadrupole strengths that will adjust the horizontal and vertical betatron tunes. The operator may also want to change the β-function values of the lattice to achieve different operating conditions. Examples of more complicated lattice modification include keeping the tunes constant while lowering the emittance, or lowering the momentum compaction. In each case, the new quadrupole strengths are calculated from the accelerator model.

The example of controlling the momentum compaction factor in SPEAR illustrates the main features of on-line optics control.[49] For this experiment, the operator wants to lower the momentum compaction by controlling the dispersion function while keeping the tunes and chromaticities constant. The evolution of the dispersion function through a progression of configuration changes is indicated in Fig. 9.

In order to guide the electron beam optics along a smooth path from the nominal lattice to the new lattice, an incremental series of small modifications is desirable. To change the machine optics in a controlled manner, the operator first needs a model for the present state of the accelerator. In addition, the operator needs accurate calibration factors that will predict the correct power supply settings to implement the desired change in magnet field strengths. Next, similar to the procedure used by off-line lattice designers, a new lattice is calculated, and the result is downloaded onto the machine hardware. At each step, measurements are made to diagnose the new lattice (tune, response matrix, etc.), and the results are compared to the model prediction.

References

1. The following papers contain examples of integrated control system/accelerator modeling packages: L. Catani et al., *Proc. IEEE 1991 PAC* (San Francisco, 1991); D. Dobrott et al., *ibid*.; M. J. Lee, *Proc. 1991 Int. Conf. on Accelerators and Large Experimental Physics Control Systems* (Tsukuba, Japan, 1991); H. P. Chang et al., *Proc. IEEE 1993 PAC* (Washington, D.C., 1993); H. Nishamura, *ibid*.
2. D. P. Weygand et al., *Proc. IEEE 1987 PAC* (Washington, D.C., 1987).
3. *7–GeV Advanced Photon Source: Conceptual Design Report* (Argonne National Laboratory, Illinois 60439, 1987), ANL–87–15.
4. European Sychrotron Radiation Facility, *ESRF Foundation Phase Report, Feb 1987.*
5. JAERI–RIKEN Spring–8 Project Team, *Spring-8 Project* (1991).
6. M. Knot, D. Gurd, S. Lewis, and M. Thout, *Proc. 1993 ICALEPCS* (Berlin, 1993); W. McDowell, M. Knott, and M. Kraimer, *Proc. IEEE 1993 PAC*, (Washington, D.C., 1993).
7. P. Clout et al., NIM **A293**, 456 (1990).
8. B. Ng et al., *Proc. IEEE 1991 PAC* (San Francisco, 1991), p. 1371.
9. J. P. Scott and D. E. Eisert, *Proc. of Europhysics Conf. on Control Systems for Experimental Physics (*Villars-sur-Ollon, Switzerland, 1987), p. 103.
10. N. D. Arnold et al., *Proc. IEEE 1991 PAC* (San Francisco, 1991).
11. G.J. Nawrocki et al., *ibid*.
12. *LBL 1–2 GeV Synchrotron Radiation Source, Conceptual Design Report*, (Lawrence Berkeley Laboratoy, Livermore, 1986); PUB–5172 Rev. (1986).
13. T. Russ and C. Sibley, NIM **A293**, 258 (1990).
14. R. Bork, *Proc. IEEE 1987 PAC* (Washington, D.C., 1987), p. 523.
15. Beetham and the SPS/LEP Accelerator Controls Group, *Proc. IEEE 1987 PAC* (Washington, D.C., 1987), p. 741.
16. P.G. Innocenti et al., NIM **A293**, 1 (1990).
17. C.O. Pak, NIM **A293,** 408 (1990).
18. M. Sands, *The Physics of Electron Storage Rings: An Introduction,* SLAC-121, 1970.
19. R. Chasman, G.K. Green, and E.M. Rowe, IEEE Trans. Nuc. Sci. **NS–22**, 1765 (1975).

20. Helmut Wiedemann, *Particle Accelerator Physics* (Springer Verlag, Berlin, 1993).

21. F. C. Isilin and J. Niederer, CERN/LEP–TH/88–38 (1988).

22. *A User's Guide to the Harmon Program* (CERN, 1982), LEP Note 420.

23. K. L. Brown et al., SLAC–91, Rev. 2, UC–28.

24. A. J. Dragt, et al., IEEE Trans. Nuc. Sci. **NS–32**, No. 5, 2311 (1985).

25. R. Ruth, IEEE Trans. Nuc. Sci. **NS–30**, No. 4, 2669 (1983).

26. TRACY evolved from the efforts of H. Nishamura and E. Forest, and was rewritten in a very compact and powerful form by J. Bengston. For an original reference, see H. Nishamura, LBL REPORT 25236, ESG–40 (1988).

27. Los Alamos Accelerator Code Group, *Computer Codes for Particle Accelerator Design and Analysis: A Compendium*, second ed., LA–UR–1766 (1990). For linear accelerators, see C.R. Eminhizer, ed., La Jolla Inst., AIP Conf. Proc. **177** (1988).

28. K. L. Brown, SLAC Report–75.

29. D. C. Carey, *The Optics of Charged Particle Beams* (Harwood Academic, New York, 1987).

30. D. W. Kerst and R. Serber, Phys. Rev. **60**, 53 (1941).

31. E. Courant and H. Snyder, Ann. Phys. **3**, 1 (1959).

32. M. D. Woodley et al., SLAC–PUB–3086 (1983).

33. R. H. Helm et al., *Proc. 1973 PAC* (San Francisco, 1973); SLAC–PUB–1193.

34. Examples of beam line error analysis codes based on comparing measured data to the model include: M. Lee et al., *Proc. Europhysics Conf. on Control Systems for Experimental Physics* (Villiars, Switzerland, 1987); SLAC–PUB–4411. M. J. Lee, Y. Zambre, and W. Corbett, *Proc. Int. Conf. Cum Workshop on Current Trends in Data Acquisition and Control of Accelerators* (Calcutta, India, 1991); SLAC–PUB–5701. W. J. Corbett, M. J. Lee, and Y. Zambre, *Proc. 3rd EPACS* (Berlin, 1992); SLAC–PUB–5776.

35. W. J. Corbett, M. J. Lee, and V. Ziemann, *Proc. IEEE 1993 PAC* (Washington, D.C., 1993); SLAC–PUB–6111.

36. J. Safranek and M. Lee, *Proc. of Orbit Correction and Analysis Workshop, Brookhaven Nat'l. Laboratory* (1993).

37. A. Luccio, *Proc. IEEE 1993 PAC* (Washington, D.C., 1993).

38. A. Verdier and J.C. Chappelier, *ibid.*

39. W.J. Corbett and V. Ziemann, ibid.; SLAC–PUB–6112.

40. Examples of simultaneous orbit and dispersion correction used for PEP-I: E. Close et al., IEEE Trans. Nuc. Sci, **NS–26**, No. 3, 3502 (1979). A. Chao et al., PEP–NOTE–318 (1979). M. H. R. Donald et al., *Proc. IEEE 1981 PAC* (Washington, D.C., 1981); SLAC–PUB–2666.

41. B. Autin and G. Marti, CERN–ISR–MA/73–17 (1973).

42. Harmonic representation of the closed orbit is developed in Ref. 31. Fast harmonic orbit control in synchrotron light sources was developed at NSLS: L. H. Yu et al., NIM **A284**, 268 (1984).

43. Eigenvector decomposition is widely used in theoretical and applied physics, and forms the basis for many orbit correction codes. For a recent treatment of applications to storage rings, see A. Friedmann and E. Bozoki, BNL–49527, submitted to NIM.

44. G. Strang, *Linear Algebra and It's Applications*, second ed. (Academic Press, Inc.); see also G. Strang, Amer. Math. Monthly, **100**, 9 (1993).

45. V. Ziemann, *SLAC Single Pass Collider Note* CN–393 (1992); see also E. Bozoki and A. Friedmann, *Proc. IEEE 1993 PAC* (Washington, D.C., 1993).

46. W. J. Corbett, B. Fong, M. J. Lee, and V. Ziemann, *ibid.*; SLAC–PUB–6110.

47. Y. Chung et al., *ibid.*, p. 2263.

48. W. J. Corbett and D. Keeley, *Proc. of Orbit Correction and Analysis Workshop* (Brookhaven National. Laboratory, 1993).

49. P. Tran et al., *Proc. IEEE 1993 PAC* (Washington, D.C., 1993), p. 173.