18-Data Bases and Communications

and the advent of the automated powder diffractometers in the 1970's, there has been a growing need for the publication of high quality powder data. Unfortunately, little, if any, of the original experimental data are available, thus a significant task which has faced the ICDD in recent years is the re-evaluation and re-doing of the original data.

Data base maintenance is an expensive and time-consuming proposition, and the ICDD has been especially creative in its attempts to provide data worthy of use with modern instrumentation. This paper reviews the more important of these procedures, and outlines some of the work of the ICDD committees in keeping a watchful eye on the ever-changing needs of the powder diffraction community.

MS-18.01.05 SYSTEMATICS IN MATERIALS DESIGN. By John R. Rodgers*, National Research Council of Canada, Ottawa, Ont. K1A 0S2, Canada, and Pierre Villars, Intermetallics Phases Databank, 6354 Vitznau, Switzerland.

The search for new materials which have new or greatly improved properties is an intrinsic part of materials science. One aspect of such a search is to identify relations, regularities, rules, models and laws within experimentally determined data. In the field of inorganic and intermetallic materials there are over 4,000 distinct types of ordering of atoms, called structure types. Since the crystal structures are a first indicator for the existence of specific intrinsic properties they should be the first element to be analysed. One approach to such studies is the use of structure maps and Quantum Structural Diagrams (QSD). Structure maps order the vast amount of structural information within 2- and 3-dimensional plots, using a Mendeleyevian philosophy, such that materials with a given structure type cluster together. Quantum Structural Diagrams (QSD) systematize the relationship between composition and structure type and employ atomic size difference, electronegativity difference and sum valence electrons as coordinates. We will show, with examples, how such maps are used in the search for superconducting materials, quasicrystals and permanent magnets. The use of other maps showing relationships between mechanical properties and structure will be reviewed. An attempt will be made to explain if violations of structure types lead us to the discovery of new, yet unknown, exotic materials.

MS-18.01.06 NIST MATERIALS SCIENCE DATABASES. Alan D. Mighell* and Vicky Lynn Karen, Reactor Radiation Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.

The NIST Crystal and Electron Diffraction Data Center builds a comprehensive database with chemical, physical, and crystallographic information on all types of well characterized substances. During the year, the database has been significantly augmented with respect to all categories of substances and now contains greater than 183,000 entries. From the central database, two distribution databases are produced: (1) NIST Crystal Data and (2) the NIST/Sandia/ICDD Electron Diffraction Database. In The Materials Science and Engineering Laboratory at NIST, two other databases are being built that have applications in the materials sciences: one contains information on phase diagrams and the other data on superconductors. The database on high-temperature superconductors will include materials specification and characterization information. It is being developed in a collaborative effort with the National Research Institute for Metals of Japan.

As a source of critically evaluated data, the databases can be used as an aid to research or as a basis for research. They play an invaluable role in materials design, characterization and phase identification. Distribution software for identification and lattice analysis is available from Vicky Lynn Karen. NIST Databases are made available to the scientific community through computer oriented modes of dissemination including: magnetic tape, on-line searching, CD-ROM, and scientific instruments.

MS-18.01.07 NEW TOOLS FOR SCIENTIFIC DATA EXCHANGE. N. Spadaccini, Department of Computer Science, University of Western Australia.

The proliferation of scientific data and related information has necessitated the development of automated procedures to speed archiving and retrieval. Many databases have invested years of effort to developing fast access algorithms designed specifically for their own format. Attempts to pool information across database systems is hampered by incompatible formats and access software.

Agreement on a single database format is not practical, however the adoption of a standardised exchange format (e.g. George et al, Protein Seq Data Anal 1, 27-39, 1987) with more straight forward conversion routines to most other database formats has been widely suggested. Many of these exchange formats are specific to particular applications or disciplines. They are generally fixed or pre-defined free formats, both of which require significant prior knowledge to facilitate data access and are inflexible to evolving data requirements.

Hall (JCICS 31, 326-333, 1991) has reported a flexible, self defining file format known as STAR. The file can be easily extended to incorporate new data storage requirements as they evolve without violating the original file structure. The data can be automatically validated against either universally defined data dictionaries (e.g. the CIF dictionary) or can be extended to user defined dictionaries. The crystallographic community has adopted the CIF format, a restricted implementation of STAR, as its basis for data exchange and retrieval. This paper reviews the underlying advantages of the STAR approach and elaborates on its structure.

Spadaccini, Hall & Hall (*Star_Base* Users Manual, 1993) have developed a query language known as *Star_Base* to facilitate data access from STAR files. *Star_Base* provides a number of options including the extraction of data based on conditional statements over a variety of user defined scopes. The language makes it possible to access data as required by the user with little prior knowledge of the file structure. Possible future extensions of *Star_Base* to a full application level language will be presented.

PS-18.01.08 PCPDFWIN - SEARCH/RETRIEVE PROGRAM FOR THE ICDD POWDER DIFFRACTION DATABASE ON CD -ROM. By R. Jenkins, F. Needham*, International Centre for Diffraction Data, Swarthmore, PA, USA; R. Garvey, North Dakota State University, USA; S. Lin, Nankai University, China.

The PCPDFWIN program uses Btree and Bitmap algorithms to search and retrieve the ICDD PDF2 (Powder Diffraction File) database on CD-ROM, which contains 150 MB of information. This information includes interplanar spacings, relative intensities, Miller indices, crystal system, space group symbol, axial lengths, interaxial angles and several chemical and physical properties. Among its many search parameters, the PCPDFWIN program includes Three Strongest Interplanar Spacings, Three Longest Interplanar Spacings, Cell Parameters, Cell Volume, and Density. In addition, it provides over 250 organic functional groups as search criteria for the organic substances in the database. A two

dimensional graphic structure display is also implemented for organic substances. All search criteria can be integrated with Boolean operators, including AND, OR and NOT, providing considerable flexibility in searches. For example, the user can search for substances containing only elements "Si" AND "O"; the user can also search for all substances containing "Si" OR "O".
The PCPDFWIN program has been developed as a Microsoft Windows 3.1 application. It runs on an Intel 386 (or above) based computer and requires at least 4 MB RAM, 6 MB hard disk space, a mouse, Microsoft Extensions and a PDF2 database on CD-ROM.

PS-18.01.09 SEQUENCE DEPENDENCE OF NON-PLANARITY OF PEPTIDES, AND EFFECT ON SECONDARY STRUCTURES. By N. Sukumar and Vasantha Pattabhi*, Department of Crystallography and Biophysics, University of Madras, Madras 600 025, India.

Non-planarity of peptides has been predicted from quantum chemical theory, and has also been observed frequently in crystal structures. A systematic study of this phenomenon is expected to give informations on the flexibility of peptide plane, and also help in model-building studies.

Class 48 of the Cambridge Crystallographic Data Base has been surveyed with this aim in mind. The analysis shows that deviations from planarity ($\Delta\omega$) range from 0° to 20° depending on the sequence. In general, deviations for Ser, Asn, Lys and His are very low in contrast with those for Phe, Pro and Leu. If Gly precedes or succeeds any residue, the deviation from planarity is significant. A linear correlation between molecular weight and torsion angle ω has been observed, indicating that deviations from planarity are involved to avoid steric interactions, and in some cases to enable hydrogen-bond formation. The analysis was extended to study the correlation between $\Delta\omega$ and the secondary structural features. The following observations have been made from the study. (i) The direction of the deviation, and the handedness of the helix seem to be correlated. (ii) In the case of peptides adopting bent conformation, the $\Delta\omega$ values are uniformly small. (iii) No such correlation was seen in the case of extended structures.

PS-18.01.10 COMPUTER-AIDED DESIGN OF NON-ISOMORPHIC HETEROCOMPOSITIONS. By A.N. Efimov, A.O. Lebedev*, Ioffe Physical-Technical Institute, St. Petersburg, Russia.

The correct choice of substrate compound and of its crystallographic orientation is of great importance for epitaxial growth of HTS. The following conditions concerning the substrate must be verified: compatibility of the substrate material with the physico-chemical conditions of growth and subsequent processing; absence of chemical interaction between substrate material and HTS; geometrical match with HTS-structure as a prerequisite for creating perfect single crystalline epitaxial layers (taking into account the temperature dependence of lattice parameters); physical properties needed for the concrete device application (for instance the small dielectric loss for HF- and UHF-range); low cost of starting materials and possibilities for industrial manufacturing of substrates. The optimal design of heterocompositions is nearly impossible without computers due to the number of compounds, and due to the number of requirements to be taken into account.
The present computer-aided design package for heterocompositions includes the following program units: an analysis of metric match based on the coincidence site lattice concept and a modified theoretico-numerical algorithm (Santoro A., Mighell A.D., Acta Crystallogr., 1973, A29, 169-175); an analysis of structure defects connected with the symmetry aspects of genetic relations between the substrate and layer structures; computer graphics for visualization of the interfacial region.
The data base contains the following information: crystal-geometrical parameters: structure type and unit-cell constants; dependence of these parameters on temperature including the temperatures and types of phase transformations; dielectric constants; peculiarities of phase diagrams; data source.

PS-18.01.11 A UNIVERSAL FILE FORMAT FOR POWDER DIFFRACTION DATA ARCHIVE AND INTERCHANGE: THE IUCR CRYSTALLOGRAPHIC INFORMATION FILE FORMAT. By B.H. Toby, Air Products & Chemicals Inc., 7201 Hamilton Blvd, Allentown, PA 18195, USA, J.I. Langford*, School of Physics & Space Research, University of Birmingham, Birmingham B15 2TT, UK and S.R. Hall, Crystallography Centre, University of WA, Nedlands 6009, Australia.

Powder diffraction data definitions have been devised for use with the Crystallographic Information File (CIF) facility for data storage and exchange. These will permit the ready exchange of raw and processed powder data generated by diffractometers of every type - commercial sealed-tube X-ray instruments, single-wavelength synchrotron and neutron systems and energy-dispersive or time-of-flight experiments.

CIF has been developed by the IUCr for the storage of crystallographic data which range from measured intensities to structural diagrams and text in a form suitable for publication (Hall et al., Acta Cryst. A47, 655-685). CIF is composed of ASCII characters which may be edited in the normal way or transmitted via international networks (eg E-mail or FTP). A CIF is completely self-descriptive, in that data items are identified by unique tags, known as data names, which may be defined locally or globally. Global data items (ie those used outside the local environment) are carefully defined in a CIF data-name directory that is also stored electronically in a CIF-like format.

Powder diffraction data are usually measured by means of a computer-controlled diffractometer, or obtained from film by a densitometer, and are stored in a data-file format specific to the instrument used. Typically, these data are then processed by using computer programs modified to use data with the local format. If the data are to be published or archived, the diffraction pattern will often be reduced to a table of peak position and intensity values. Apart from the requirement to standardise the data-processing software, the need for a global exchange of unprocessed powder diffraction data is increasing. This need arises because of the use of total-pattern methods for

(1) crystalline phase identification and quantification,

(2) ab initio structure determination,

(3) structure refinement,

(4) studies of microstructure and

(5) characterisation of materials with limited ordering (eg polymers and clays, where the diffraction pattern cannot adequately be represented by a set of peak parameters).