

 Open access • Journal Article • DOI:10.1063/1.2755487

Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis. — [Source link](#)

Christian J. Long, Jason R. Hattrick-Simpers, Makoto Murakami, R. C. Srivastava ...+3 more authors

Institutions: University of Maryland, College Park

Published on: 31 Jul 2007 - Review of Scientific Instruments (American Institute of Physics)

Topics: Ternary operation and Ternary plot

Related papers:

- [On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets](#)
- [Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization.](#)
- [High-throughput determination of structural phase diagram and constituent phases using GRENDL](#)
- [Constraint reasoning and Kernel clustering for pattern decomposition with scaling](#)
- [Commentary: The Materials Project: A materials genome approach to accelerating materials innovation](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/rapid-structural-mapping-of-ternary-metallic-alloy-systems-1dcckaf25>

2007

Rapid structural Mapping of Ternary Metallic Alloy Systems using the Combinatorial Approach and Cluster Analysis

C. J. Long

Jason R. Hattrick-Simpers

University of South Carolina - Columbia, simpers@cec.sc.edu


M. Murakami

R. C. Srivastava

I. Takeuchi

See next page for additional authors

Follow this and additional works at: https://scholarcommons.sc.edu/eche_facpub

 Part of the [Engineering Physics Commons](#), [Optics Commons](#), and the [Other Chemical Engineering Commons](#)

Publication Info

Published in *Review of Scientific Instruments*, Volume 78, Issue 7, 2007, pages #072217-.

©Review of Scientific Instruments 2007, AIP (American Institute of Physics).

Long, C. J., Hattrick-Simpers, J., Murakami, M., Srivastava, R. C., Takeuchi, I., Karen, V. L., & Li, X. (July 2007). Rapid structural Mapping of Ternary Metallic Alloy Systems using the Combinatorial Approach and Cluster Analysis. *Review of Scientific Instruments*, 78 (7), #072217. <http://dx.doi.org/10.1063/1.2755487>

This Article is brought to you by the Chemical Engineering, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Author(s)

C. J. Long, Jason R. Hattrick-Simpers, M. Murakami, R. C. Srivastava, I. Takeuchi, V. L. Karen, and X. Li

Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis

C. J. Long, J. Hatrick-Simpers, M. Murakami, R. C. Srivastava, I. Takeuchi, V. L. Karen, and X. Li

Citation: [Review of Scientific Instruments](#) **78**, 072217 (2007); doi: 10.1063/1.2755487

View online: <http://dx.doi.org/10.1063/1.2755487>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/rsi/78/7?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Structure and mechanical properties of Al-3Fe rapidly solidified alloy](#)

AIP Conf. Proc. **1315**, 645 (2011); 10.1063/1.3552521

[A Study of Al-Zn-Sn Alloy Sacrificial Anode Cathodic Protection Requirements for Structure Used In Seawater](#)

AIP Conf. Proc. **1202**, 181 (2010); 10.1063/1.3295594

[Preparation of ternary alloy libraries for high-throughput screening of material properties by means of thick film deposition and interdiffusion: Benefits and limitations](#)

J. Vac. Sci. Technol. A **22**, 1788 (2004); 10.1116/1.1692366

[Materials physics: A new contemporary undergraduate laboratory](#)

Am. J. Phys. **66**, 724 (1998); 10.1119/1.18930

[Single-crystal \$\text{Pb}\(\text{Zr}_x\text{Ti}_{1-x}\)\text{O}_3\$ thin films prepared by metal-organic chemical vapor deposition: Systematic compositional variation of electronic and optical properties](#)

J. Appl. Phys. **81**, 2349 (1997); 10.1063/1.364239



Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis

C. J. Long, J. Hatrick-Simpers, M. Murakami, R. C. Srivastava, and I. Takeuchi
Department of Materials Science and Engineering, University of Maryland, College Park, Maryland 20742
and Center for Superconductivity Research, University of Maryland, College Park, Maryland 20742

V. L. Karen and X. Li
National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8520

(Received 7 November 2006; accepted 28 December 2006; published online 31 July 2007)

We are developing a procedure for the quick identification of structural phases in thin film composition spread experiments which map large fractions of compositional phase diagrams of ternary metallic alloy systems. An in-house scanning x-ray microdiffractometer is used to obtain x-ray spectra from 273 different compositions on a single composition spread library. A cluster analysis software is then used to sort the spectra into groups in order to rapidly discover the distribution of phases on the ternary diagram. The most representative pattern of each group is then compared to a database of known structures to identify known phases. Using this method, the arduous analysis and classification of hundreds of spectra is reduced to a much shorter analysis of only a few spectra. © 2007 American Institute of Physics. [DOI: 10.1063/1.2755487]

INTRODUCTION

The combinatorial approach has been used to discover new materials phases as well as perform rapid mapping of composition-structure-property relationships in complex materials systems.¹⁻³ Using thin film composition spread libraries, large fractions of compositional phase diagrams can be mapped out with a high density of data points on a single wafer.⁴⁻⁷ Mapping phase diagrams is central to obtaining comprehensive pictures of materials systems, and mapping active physical properties as a function of composition is an integral part of understanding the underlying physical mechanism of the properties.⁷⁻⁹ Thin film materials can often display properties with deviation from bulk samples, but it has been shown in many systems that one can indeed obtain compositional trends which closely resemble or mirror those of bulk counterparts.^{7,10}

We are in the process of developing techniques to analyze a large number of x-ray spectra from thin film composition spreads in order to map out structural phase diagrams. Previously, we have discussed techniques for rapidly managing and visualizing the spectrum data taken from composition spread wafers using software written in MATLAB.¹¹ We showed that effective visualization techniques can be used to rapidly capture the essential features of structural variation across composition spreads.¹¹ In particular, we have demonstrated that, in the Ni-Mn-Al system, one can obtain insight about physical properties by combining visualization techniques and simple linear correlation analysis of the diffraction peak information and the physical properties.¹¹ We have used these techniques for an in-house Bruker tool (D8 DISCOVER) as well as for synchrotron microdiffraction.¹² The careful analysis platform provided by the visualization allowed us to prove the geometric nonlinear theory of martensite for the first time as well as to discover novel composi-

tions of shape memory alloys.¹² This software has been made available to the community and is beginning to enjoy widespread use.¹³ This work was largely motivated by the fact that there had been no existing software which allowed us to readily handle and analyze large volumes of diffraction data in an efficient way, a situation which is beginning to change.

In the present work, we focus on integrating cluster analysis using the POLYSNAP software¹⁴ into our rapid analysis platform. POLYSNAP was developed primarily for use in the pharmaceutical industry. To the best of our knowledge, it has never been used in an attempt to map structures across compositional phase diagrams of complex inorganic compounds. The ultimate goal of this effort is to develop a rapid method for obtaining a comprehensive and accurate mapping of the phase and structure distribution across composition spreads of rich and complex materials systems containing previously unknown materials phases.

The present work is also part of our effort to interface and integrate data analysis of combinatorial experimental data with the crystallographic databases available at the National Institute of Standards and Technology (NIST). As an example system, we looked at a region of the Fe-Ga-Pd ternary system.

EXPERIMENT

Our interest in the Fe-Ga-Pd ternary system stems from the fact that the Fe-Ga and Fe-Pd binary phase diagrams contain compositions with unusual magnetic actuator properties. Fe-Ga is a well known material system exhibiting large magnetostriction for Ga content between 20 and 30 at. %. The origin of this property is attributed to the complexity of the Fe-Ga binary phase diagram in this region.¹⁵ Fe₇₀Pd₃₀ is a ferromagnetic shape memory alloy¹⁶ (FSMA) whose martensitic transition is associated with a magnetic

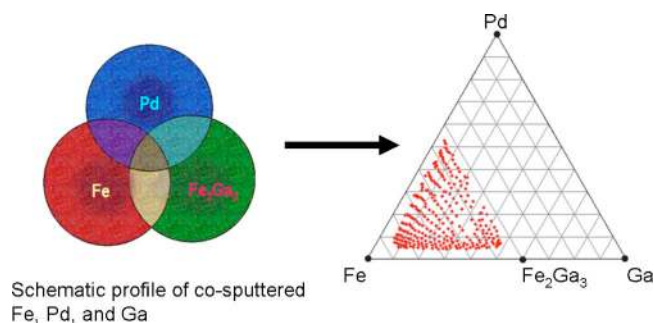


FIG. 1. (Color online) Schematic of a thin film composition library made using a three gun cosputtering system. The three targets used in the deposition were Fe, Pd, and Fe_2Ga_3 . At left a schematic of the deposition profile is displayed and at right the region of the phase diagram for which XRD information was obtained from the composition spread library is displayed.

field induced strain of about 10 000 ppm.¹⁷ As interesting as these two systems are, no work has yet been done in alloying them. Fortunately, Ga and Pd both form disordered crystals when they are substituted into the Fe lattice. This means that they could possibly be substituted into the Fe lattice without disturbing the original crystal structure.

Natural thin film composition spreads of the Fe–Ga–Pd system were deposited at room temperature using an ultrahigh-vacuum three gun magnetron cosputtering system with a base pressure of 10^{-9} Torr (10^{-7} Pa) on 3 in. (76.2 mm) diameter (100) oriented Si wafers. The details of the synthesis procedure can be found in Refs. 7 and 12. The samples were then postannealed at 650 °C for 2 h in our sputtering chamber. The base pressure during annealing was 10^{-8} Torr. The total processing time (i.e., deposition and heat treatment) of a composition spread library is roughly 3 h before it is ready for rapid characterization. After the deposition, the composition of the wafer spread is immediately determined via wavelength dispersive spectroscopy (WDS) in at. %. This measurement can determine the percentage element contained at each point on the wafer accurately. Figure 1 shows the schematic procedure for the synthesis of a ternary composition spread which covers the relevant part of the phase diagram.

X-ray microdiffraction (XRD) of the fabricated films was performed using the ω -scan mode of a D8 DISCOVER for combinatorial screening (Bruker-AXS). It is equipped with a GADD two dimensional detector which captures data for a fixed range of 2θ and ω at once. The composition spread wafer contained 535 individual 1.75×1.75 mm² squares with continuously changing composition. However, XRD was performed on only 273 of the 535 squares due to time constraints. In order to scan the 2θ range of interest (20° – 90°), microdiffraction was performed in three frames for each square. We use an x-ray beam spot size of 1 mm diameter. Because we must scan the entire library for the same frame before moving to the next frame, the entire spread library must be scanned three times to cover the 2θ range. Once this is accomplished, the microdiffraction data are in the form of two-dimensional (2D) detector images. The raw detector images are then compiled and integrated to obtain the 2θ angles and peak intensities using the D8 GADDS program and a script to automate the process.

Since there is some extraneous information in the XRD spectra (e.g., substrate peaks and background signal) some preprocessing was done on the data before they were analyzed. In particular, background subtraction, cropping, and normalization were performed. Background subtraction was done by fitting and subtracting a piecewise polynomial from the data on a spectrum by spectrum basis (i.e., the background determination of one spectrum is not affected by that of any other). After background subtraction, the full measured 2θ range was cropped down to the minimum range such that all of the detected XRD peaks from all the samples (but not from the substrate) were contained in the spectra. For the Fe–Ga–Pd system, we found that this range was from 25° to 55° . The spectra were then imported into the POLYSNAP program for clustering analysis. When the spectra are imported into the program, they are normalized such that the largest intensity in each spectrum is unity.

CLUSTERING ANALYSIS

The process of sorting the spectra into discrete groups consists of deciding on a similarity metric, calculating and visualizing the similarity among all of the spectra, and then, based on the clustering of the spectra in the visualization scheme, assigning them into some number of distinct groups using a dendrogram. This clustering process was performed exclusively in the POLYSNAP software. Once the groups are decided upon, we then export the information about which spectra fall into which groups from POLYSNAP into MATLAB and, using the composition information obtained via WDS, draw a ternary diagram. Looking at the distribution of groups on the ternary diagram is akin to looking at a phase diagram, since the groups are based entirely on structure information.

There are many choices one must make along the way to creating this phase diagram. First, one must choose the metric by which the similarity between two spectra is determined. For our case, we chose to use the Pearson correlation coefficient. For two spectra, x and y , with means \bar{x} and \bar{y} , the Pearson correlation coefficient takes the form

$$C_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}$$

The values of C can range from -1 to 1 , with a value of 1 indicating identical spectra, a value of 0 indicating spectra which have no correlation, and a value of -1 implying that the spectra are anticorrelated. Anticorrelation means that where one spectrum has large values, the other has small values. In the context of comparing spectra produced by XRD, existence/absence of a diffraction peak is indicated by the high/low value of the diffraction intensity. Therefore, when one spectrum has a perfect match of every diffraction peak with another spectrum, that would give the highest correlation coefficient of 1 ; the more different the peak positions and/or intensities are, the less the correlation coefficient would be. This interpretation motivates the definition of a distance matrix $D = (1 - C)/2$. This matrix represents the dis-

similarity among the spectra. A distance of zero between two spectra implies that they are identical, while a distance of 1 implies spectra which are completely different. As long as our choice of similarity metric is a good one, the distance matrix should contain all of the information needed to group the spectra.

The difficulty now lies in trying to understand the relationships among all of the spectra, embodied by the distance matrix D . To visualize D , each matrix element is interpreted as the distance between two spectra in some Euclidian space. The problem is that this space may have dimensionality as large as $n-1$, where n is the number of spectra. To see this, first take two points, S_i and S_j , each representing a spectrum. Place them into a Euclidian space such that they are the distance D_{ij} apart. This is a one dimensional space. Now take a third spectrum, S_k and place it however far it is from the first two (D_{ki} and D_{kj} , respectively). The three points form a triangle in two dimensional space. Repeating this process again with another spectrum produces a triangular pyramid in three space. At this point, we cannot necessarily fit another point into this space and satisfy the demands that it be the appropriate distance from each other point and that we only use three dimensions. However, we could try to put it in the best possible location such that its distance from each of the other points is the best possible approximation to the actual distances listed in the matrix D . This discards some of the similarity information. However, it gives us a way to visualize the correlation matrix in ordinary three dimensional space. The question now is how, using three dimensions, can one come up with the best possible approximation of the distribution of these points? This is a classic problem in reducing the dimensionality of a data set. To reduce the dimensionality of a distribution of points in a many dimensional space, we can use principal component analysis (PCA) or we can use metric multidimensional data scaling (MMDS). The mechanics of how to perform these techniques are well documented^{18,19} and we will not discuss them here. In this article, we use MMDS to find the best possible (three dimensional) approximation to the $(n-1)$ dimensional distribution of points embodied by the distance matrix. For the Fe–Ga–Pd system, a plot of the distribution of points, embodied by the distance matrix and approximated using MMDS, appears in Fig. 2(a). A three dimensional (3D) animation of the plot is also available at Ref. 20.

Looking at the distribution of points in the MMDS plot, one could start to visually divide the spectra into discrete groups. However, this process would be somewhat subjective. A more rigorous mathematical method for deciding which spectra to group together is to use a dendrogram.²¹ An example of a dendrogram for the Fe–Ga–Pd system is shown in Fig. 2(b). Along the x axis, each discrete point represents a spectrum. From each spectrum leads a vertical line. At some height, these lines all connect. The height at which they connect is determined by the similarity between the spectra. If some numbers of spectra are already connected, then the height represents the similarity between groups of spectra. To get an intuitive idea of how this dendrogram is made, first take the two most similar spectra and put them in a group together. This is done by connecting them on the dendrogram

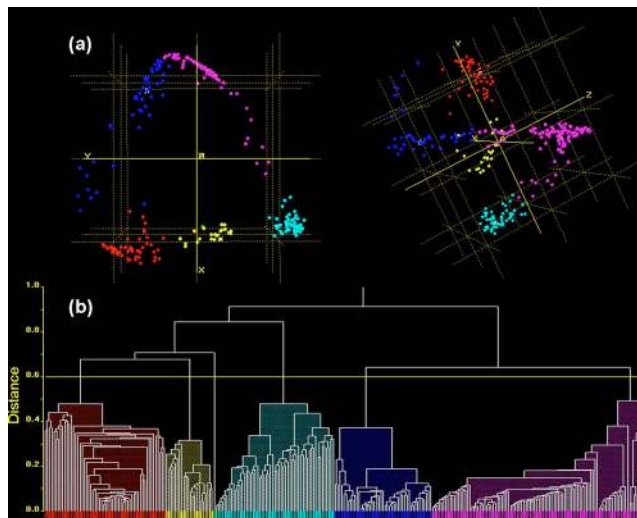


FIG. 2. (Color online) (a) Two different views of a 3D approximation of the distribution of points defined by the distance matrix, as approximated using metric multidimensional data scaling (MMDS). For this system, the MMDS plot is a very good approximation to the distance matrix, accounting for about 93% of the information in the matrix. Each point in the MMDS plots corresponds to a spectrum. Points which are close together correspond to spectra which are similar to each other, as measured using the Pearson correlation coefficient. Color is used to identify groups, and groups are defined using the dendrogram. (b) A dendrogram based on the distance matrix. Points on the x axis correspond to spectra. The height at which the lines leading from the spectra are connected indicates their similarity. In this dendrogram, the cut level is set at 0.6, creating five groups.

with a horizontal line at a height corresponding to the distance between them, as described in the distance matrix. In order to proceed further, we next determine the similarity between this group of spectra and all of the other spectra. The way in which this similarity is defined is called the linkage method. One of the ways to decide the similarity between a group of spectra and another spectrum (or between two groups of spectra) is to compare the average spectrum of all the members of the group to the other spectrum. This is called the group average linkage method, and it was the method used to produce the dendrograms in Fig. 2 and Fig. 3. The process of grouping together the most similar spectra (or groups of spectra) is then repeated, each time linking less and less similar spectra at higher and higher levels on the dendrogram and each time reducing the number of groups by 1. Eventually, at the top of the dendrogram, there is only one vertical line. If one were to stop making groups at some threshold similarity level, then one would be left with some number of groups. When talking about a dendrogram, this threshold is called the cut level. By adjusting the cut level, one can adjust the number of groups. If each of these groups is assigned a color, then it is possible to look again at the MMDS plot, with the points colored as they fall into the different groups. All of this color coding is handled automatically by POLYSNAP. Fig. 2 and Fig. 3 are color coded in this way.

If the data fall into well separated clusters in the MMDS plot, then there should be some level in the dendrogram where there is a big step in the linkage height. That is, at some point, there will be a link whose height is clearly larger than all of the links below it. In Fig. 2, we see just such a

step. This is where we place our cut level. This creates three more or less spherical clusters (red, yellow, and cyan) and two “boomerang” shaped clusters. Let us first discuss the spherical clusters. Since points which are close together on the MMDS plot correspond to similar spectra, a spherical group should correspond to a single phase. This simple logic is the reason that we chose the group average linkage method, which prefers to create spherical clusters. However, not all of the groups in the MMDS plot are spherical; it also contains some arclike distributions of points. An arc of points represents spectra which are all related to each other, but which are undergoing some systematic change as a function of position along the arc. For example, this may correspond to a set of samples across which a peak slowly shifts (i.e., a change in a lattice angle) or it may correspond to a set of samples with a slowly varying mixture of phases. The presence of these nonspherical clusters implies that the group average linkage method may not be the best choice. There does exist a linkage method which is less sensitive to the shape of the group, called the single link method. In this linkage method, when creating the dendrogram, the distance between two groups of spectra is defined to be the distance between the two spectra (one from each group) with the minimum distance between them. This method preserves the arc as a single group. Unfortunately, this linkage method is also very sensitive to variations in the signal to noise level among the spectra. This sensitivity to the noise level results in many small groups being formed, which makes this method less fit from a clustering standpoint. We have also tried several other linkage methods, namely, the weighted group average method (a.k.a. weighted pair-group method with averaging, or WPGMA), the centroid method (a.k.a. unweighted pair-group method using centroids, or UPGMC), the median method (a.k.a. weighted pair-group method using centroids, or WPGMC), the complete link method (farthest neighbor), and the ward method (minimum total pairwise squared distance), all of which gave similar results to the group average linkage method. A new linkage method which is a combination of the group average and single link methods would be ideal. However, given the software at our disposal, we were not able to find a linkage method which would allow for nonspherical clusters and also allow for variations in the signal to noise ratio. As an alternative to developing a new linkage method, we have modified our dendrogram by hand to create three arclike groups out of the two boomeranglike groups in Fig. 2. The new dendrogram and the resulting distribution of groups in the MMDS plot are shown in Figs. 3(a) and 3(b), respectively. To see if our clustering technique is reasonable, we follow a path along the distribution [see the blue line in Fig. 3(b)] and plot the XRD data as a function of position along this line [Fig. 3(c)]. This allows us to see the relationship between the groups which we have created and the actual XRD data. It is clear when looking at the XRD data that the yellow, red, cyan, and magenta groups correspond to unique patterns, while the green and blue groups correspond to transitions between one pattern to another. Furthermore, the cyan group in Fig. 3 corresponds to spectra which are all of the same phase, but with a peak which is continuously shifting.

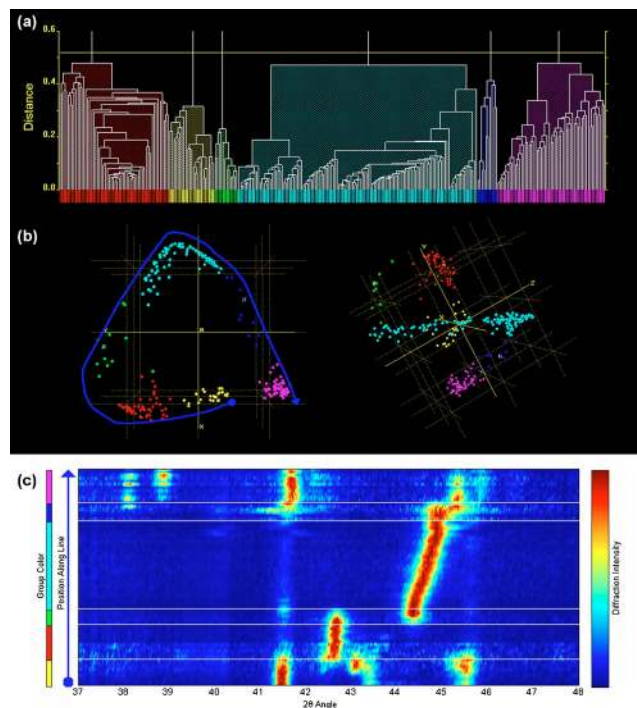


FIG. 3. (Color online) (a) A modified dendrogram where the “arms” of the arc are separated into separate groups. (b) Two views of a 3D MMDS plots showing the distribution of the groups created using the modified dendrogram. (c) The XRD data as a function of position along the blue line.

From this analysis, we can conclude that using the Pearson correlation coefficient as our similarity metric to create a distance matrix and visualizing that matrix using MMDS, we can identify clusters of spectra which correspond to unique patterns. In particular, spherical clusters in the MMDS plot correspond to a single pattern while an arc in the MMDS plot may correspond to a transition between two patterns or to a single pattern where one of the peaks is continuously shifting. Once the clustering analysis is completed, we can move on to trying to identify the actual phases represented by the clusters.

PHASE IDENTIFICATION

Identification of known phases was done by comparing the XRD spectrum of the most representative member (i.e., the member with the smallest distance to all other members) of a group of spectra to a set of reference spectra from the crystallographic databases available at NIST. In particular, we used the FIZ/NIST Inorganic Crystal Structure Database (ICSD)²² which contains more than 90 000 entries and the NIST Structural Database (NSD)²³ which contains more than 60 000 entries.²³ For the ternary systems we are investigating, we have found that there is significantly more information available for our purposes in these databases as compared to other data sources. The matching conditions we used were as follows: if the composition of the reference pattern was within (or very near) the composition space spanned by the group of XRD spectra produced by POLYSNAP and the peaks were separated by less than 0.4° , then we considered this to be a match.

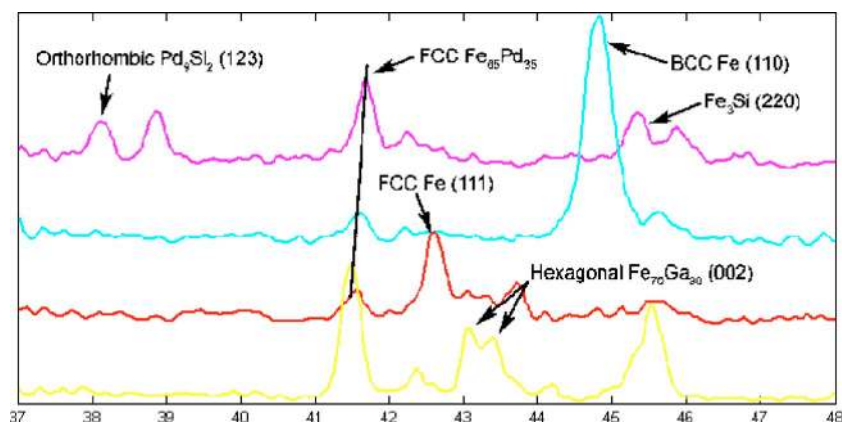


FIG. 4. (Color online) The most representative patterns in the yellow, red, cyan, and magenta groups from Fig. 3. The clustering approach reduces the problem of going through hundreds of spectra and identifying every peak in every pattern to just deciphering the meaning of these few patterns. The peaks are labeled if a possible match to known samples from the NIST crystallographic database was found.

Figure 4 shows a plot of the most representative members of the yellow, red, cyan, and magenta groups. The identified structures denote the basic crystal structures of each region. The green and blue groups are not displayed since they represent mixtures of the phases present in the other groups. We would like to note that in some cases, there was no match identified from the database.

Putting the clustering, phase identification, and composition information together yields a map of the distribution of phases for the explored region of the phase diagram (Fig. 5). We also note that the blue and green groups, which we interpreted as transitions from one phase to another based on the plots in Fig. 3 do indeed fall along the borders of the other phases. Although the full ternary phase diagram of this system is not available for comparison, the projection of the identified distributions to the two binary (Fe–Ga and Fe–Pd) systems matches the known phase diagrams reasonably well. According to the published equilibrium phase diagrams, in the Fe–Ga system,²⁴ starting from the pure Fe end, the α -Fe phase persists up to about 80% Fe, beyond which various mixture regions containing the Fe–Ga $L1_2$ phase stretches up to about 50% Fe. The $L1_2$ phase has an fcc structure, which in our study was correctly identified as being isostructural to fcc Fe. In the Fe–Pd system,²⁵ a mixture of α -Fe and

Fe_{0.5}Pd_{0.5} (fcc) is known to extend from 100% Fe to about 50% Fe. It is expected that this region would predominantly “appear” as mainly α -Fe. In our study, we find that at approximately Fe_{0.65}Pd_{0.35}, the dominant phase switches from α -Fe to fcc Fe_{0.65}Pd_{0.35} which stretches beyond 50% Fe. Our analysis has identified this region (starting at the correct composition) as the fcc Fe_{0.65}Pd_{0.35}, which we believe is a quenched phase.

EXISTING PROBLEMS AND FUTURE WORK

The ultimate goal of our efforts is to reach a point where the analysis of hundreds of spectra automatically identifies all of the pure phases present in a system and quantifies the percent of each phase present in each sample. The work presented here represents only a first step towards this goal, as there are still several problems left to overcome. Some of these problems are inherent in the use of thin films, while others are a result of the analysis techniques.

One of the problems one faces when attempting to do structure identification of thin films is that it may simply not be possible to precisely identify all the lattice parameters, and thus, the exact structure of the material. In principle, in order to completely determine the lattice parameters, one must measure the intensity of all x-ray reflections, as in powder diffraction. The films under study here are often at least textured, if not sometimes even epitaxially grown, reducing the number of reflections to only the ones from the preferred orientations. It is also possible that the film may exhibit different preferred orientations at different sites, resulting in different sets of reflections for the same structure. As a partial solution to the problem of textured films, it is sometimes possible to obtain some additional information about textured samples by tilting the wafer. Other problems include the formation of “spurious” phases such as silicides, as observed here. There could also be formation of metastable phases which are unique to the film structures and not present in bulk form.

In addition to the difficulties of working with thin films, there are also weaknesses in the clustering techniques. One of the major problems with dividing spectra into discrete groups is that the groups do not always represent pure phases. This is certainly the case when there are mixtures of phases present. The classification of spectra into “transition groups” such as the blue and green groups in Fig. 3 is help-

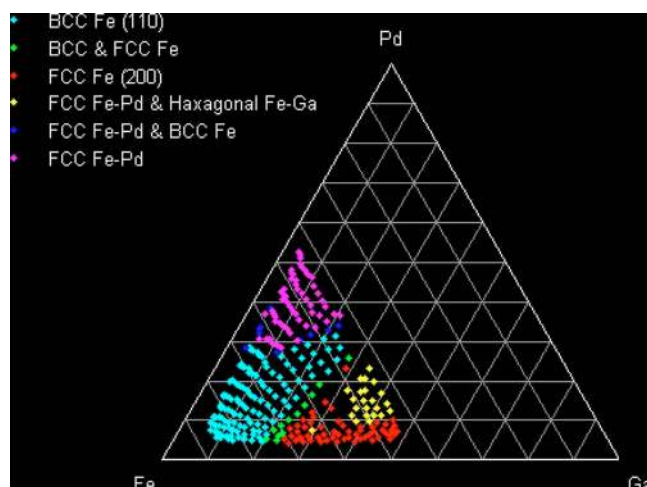


FIG. 5. (Color online) The distribution of diffraction patterns as produced using the groups of spectra produced by clustering analysis and comparison of most representative patterns from the clusters with a database of reference patterns.

ful, but if all of the samples were mixtures, then all groups would be transition groups, and no pure phases would be identified. It might also be the case that different patterns which correspond to different preferred orientations of the same structure might be identified as different structures. Often, these problems could be partially solved by manually scrutinizing individual peaks and/or by applying prior knowledge about the materials.

It is worth noting that POLYSNAP does have features to identify mixtures quantitatively given a database of pure phase patterns.^{26,27} We are currently working on integrating these procedures into our data analysis procedures. However, the problems of noisy spectra, preferred orientations, peak shifts within a single phase, and the limited number of known pure phases in the systems we are studying present problems which we are yet to overcome. Future work will also be focused on composition spread samples of other ternary systems.

ACKNOWLEDGMENTS

This work was supported by NSF DMR-0603644 (IMI Program), NSF DMR 0520471 (MRSEC at UMD), ONR N000140610530, and NIST.

¹H. Koinuma and I. Takeuchi, *Nat. Mater.* **3**, 429 (2004).

²*Combinatorial Materials Synthesis*, edited by I. Takeuchi and X.-D. Xiang (Dekker, New York, 2003).

³I. Takeuchi, R. B. van Dover, and H. Koinuma, *MRS Bull.* **27**, 301 (2002).

⁴T. Fukumura *et al.*, *Appl. Phys. Lett.* **77**, 3426 (2000).

⁵I. Takeuchi, W. Yang, K.-S. Chang, M. Aronova, T. Venkatesan, R. D.

Vispute, and L. A. Bendersky, *J. Appl. Phys.* **94**, 7336 (2003).

⁶R. B. van Dover, L. F. Schneemeyer, and R. M. Fleming, *Nature (London)* **392**, 162 (1998).

⁷I. Takeuchi *et al.*, *Nat. Mater.* **2**, 180 (2003).

⁸S.-W. Cheong and H. Y. Hwang, *Ferromagnetism vs. Charge/Orbital Ordering in Mixed-Valent Manganites: Colossal Magneto-Resistive Oxides*, edited by Y. Tokura (Gordon and Breach, Amsterdam, 2000), p. 237.

⁹A. Damascelli, Z. Hussain, and Z.-X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003).

¹⁰M. J. Turchinskaya, L. A. Bendersky, A. J. Shapiro, K. S. Chang, I. Takeuchi, and A. L. Roytburd, *J. Mater. Res.* **19**, 2546 (2004).

¹¹I. Takeuchi, C. J. Long, O. O. Famodu, M. Murakami, J. Hatrick-Simpers, and G. W. Rubloff, *Rev. Sci. Instrum.* **76**, 062223 (2005).

¹²J. Cui *et al.*, *Nat. Mater.* **5**, 286 (2006).

¹³For more information about the visualization software, contact I. Takeuchi at takeuchi@umd.edu

¹⁴G. Barr, W. Dong, and C. J. Gilmore, *J. Appl. Crystallogr.* **37**, 658 (2004).

¹⁵G. Petculescu, K. B. Hathaway, T. A. Lograsso, M. Wun-Fogle, and A. E. Clark, *J. Appl. Phys.* **97**, 10M315 (2005).

¹⁶M. Sugiyama, R. Oshima, and F. E. Fujita, *Mater. Trans., JIM* **25**, 585 (1984).

¹⁷T. Fukuda, T. Takeuchi, and T. Kakeshita, *Mater. Jpn.* **40**, 544 (2001).

¹⁸I. T. Jolliffe, *Principal Component Analysis* (Springer, New York, 1986).

¹⁹J. C. Gower, *Biometrika* **53**, 325 (1966).

²⁰<http://www.combi.umd.edu>

²¹A. D. Gordon, *Classification*, 2nd ed. (Chapman and Hall, London CRC, Boca Raton, FL, 1999).

²²A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, *Acta Crystallogr., Sect. B: Struct. Sci.* **B58**, 364 (2002).

²³NIST Standard Reference Data Program, Gaithersburg, MD 20899.

²⁴*Binary Alloy Phase Diagrams*, 2nd ed., edited by T. B. Massalki (ASM International, Ohio, 1990), p. 1064.

²⁵*Binary Alloy Phase Diagrams*, 2nd ed., edited by T. B. Massalki (ASM International, Ohio, 1990), p. 1751.

²⁶C. J. Gilmore, G. Barr, and J. Paisley, *J. Appl. Crystallogr.* **37**, 231 (2004).

²⁷G. Barr, W. Dong, and C. J. Gilmore, *J. Appl. Crystallogr.* **37**, 243 (2004).