IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

# High-throughput determination of structural phase diagram and constituent phases using GRENDEL

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Nanotechnology 26 444002

(http://iopscience.iop.org/0957-4484/26/44/444002)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 129.6.223.84
This content was downloaded on 03/06/2017 at 20:38

Please note that terms and conditions apply.

You may also be interested in:

Segmentation of dynamic PET images with kinetic spectral clustering
S Mouysset, H Zbib, S Stute et al.

Mapping continuous phase diagrams of complex material systems
Young K Yoo and X-D Xiang

Community structure detection in complex networks with partial background information
Zhong-Yuan Zhang

Construction of crystal structure prototype database: methods and applications
Chuanxun Su, Jian Lv, Quan Li et al.

FDG parametric images by clustering and principal components
Y Kimura, M Senda and N M Alpert

Deconvolution of petroleum mixtures using mid-FTIR analysis and non-negative matrix factorization
George Livanos, Michalis Zervakis, Nikos Pasadakis et al.

An isometric muscle force estimation framework based on a high-density surface EMG array and an NMF algorithm
Chengjun Huang, Xiang Chen, Shuai Cao et al.

Effect of Ni0.5Zn0.5Fe2O4(NZFO) layer thickness on the magneto-electric properties of BaTiO3(BTO)-NZFO composite bilayer thin films
Sanjay Kumar Upadhyay, V Raghavendra Reddy, Ajay Gupta et al.

Ferroelectric, ferromagnetic and magnetoelectric properties of BiFeO3:Ba films
Meiya Li, Min Ning, Yungui Ma et al.

OPEN ACCESS
IOP Publishing
Nanotechnology

Nanotechnology **26** (2015) 444002 (9pp)
doi:10.1088/0957-4484/26/44/444002

# High-throughput determination of structural phase diagram and constituent phases using GRENDEL

**A G Kusne[1], D Keller[2], A Anderson[2], A Zaban[2] and I Takeuchi[3]**

[1] Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[2] Department of Chemistry, Center for Nanotechnology and Advanced Materials, Bar-Ilan University, 5290002 Ramat Gan, Israel
[3] Materials Science and Engineering Department, University of Maryland, College Park, MD 20742, USA

E-mail: aaron.kusne@nist.gov

Received 25 June 2015, revised 8 September 2015
Accepted for publication 16 September 2015
Published 15 October 2015



**Abstract**

Advances in high-throughput materials fabrication and characterization techniques have resulted in faster rates of data collection and rapidly growing volumes of experimental data. To convert this mass of information into actionable knowledge of material process–structure–property relationships requires high-throughput data analysis techniques. This work explores the use of the Graph-based endmember extraction and labeling (GRENDEL) algorithm as a high-throughput method for analyzing structural data from combinatorial libraries, specifically, to determine phase diagrams and constituent phases from both x-ray diffraction and Raman spectral data. The GRENDEL algorithm utilizes a set of physical constraints to optimize results and provides a framework by which additional physics-based constraints can be easily incorporated. GRENDEL also permits the integration of database data as shown by the use of critically evaluated data from the Inorganic Crystal Structure Database in the x-ray diffraction data analysis. Also the Sunburst radial tree map is demonstrated as a tool to visualize material structure–property relationships found through graph based analysis.

S Online supplementary data available from stacks.iop.org/NANO/26/444002/mmedia

Keywords: materials informatics, phase diagram, high throughput materials science

(Some figures may appear in colour only in the online journal)

## 1. Introduction

The last few decades have seen rapid development in experimental and theoretical tools for fabricating, simulating, and characterizing materials systems, bringing the dream of rapid advanced materials discovery closer to reality. For example, high-throughput thermodynamics and electronic-structure calculations have resulted in multiple large databases of materials properties predictions [1–5]. Meanwhile, advancements in combinatorial library synthesis and characterization allow for rapid analysis of thousands of potential functional materials in a matter of hours [6–8]. However, while both computational and experimental data are being collected at faster and faster rates, and are being compiled into various databases, there is a clear lack of high-throughput data analysis tools to unify the information of these databases and convert them into actionable knowledge in the pursuit of advanced materials. The development of such data analysis tools is a key objective of the White House Materials Genome Initiative [9].

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

 © 2015 IOP Publishing Ltd Printed in the UK

This work focuses on an algorithm for converting composition and structure data from combinatorial libraries and material structure databases into potential composition phase diagrams and potential constituent phases. The results can then be used to better understand the material process–structure–properties relationship for the material system under investigation. The high speed of the algorithm also allows for on-the-fly analysis—providing results to the experimentalist during data collection and permitting a more thorough investigation of the library, e.g. a higher resolution structure investigation of potential structure phase change regions. Results can also be used to seed more advanced and computationally expensive analysis methods. The algorithms described here are unsupervised methods, and the results must therefore be verified by an expert. Hence, phase diagrams and constituent phases identified by the algorithm should be considered 'potential' until critically evaluated by an expert.

In the past, a constraint reasoning based method was developed to identify both a composition phase diagram and constituent phases for combinatorial libraries [10]. However, this method relies on a collection of computationally expensive operations, such as peak detection and the dynamic time warping metric, resulting in solution times of tens to hundreds of hours for a typical system. A human input guided version of the algorithm was also developed [11], resulting in significantly reduced computation time. In this paper we present a high speed, low computational cost method for determining composition phase diagrams and constituent phases while also requiring minimal user input during data analysis.

Previous investigations into high speed, low computational cost algorithms have included clustering methods such as hierarchical cluster analysis and mean shift theory to rapidly identify composition phase diagrams from combinatorial libraries [12, 13]. The non-negative matrix factorization (NMF) method was used to identify potential constituent phases [14]. In either case, the algorithms have been used to evaluate either the phase diagram or the constituent phases, despite the fact that the two types of data analyses are interdependent—knowledge of one can be used to improve the evaluation of the other.

Also, in the case of NMF constituent phase determination, application of NMF to an entire combinatorial library dataset assumes that each and every constituent phase exists throughout the combinatorial library to varying degrees. This is not true for systems in thermodynamic equilibrium, where Gibbs phase rule limits the number of constituent phases in each phase region. For example a phase mixture region for a ternary system at fixed pressure is limited to a maximum of three constituent phases. These issues can be resolved by utilizing the clustering results in identifying constituent phases. Additionally, the results of NMF can be highly unstable from run-to-run. This is true even for a single dataset, due to NMF's reliance on a random number generator to seed the matrix factorization [15].

Some issues have also arisen due to the use of clustering algorithms that only analyze structure data to determine phase diagrams. These algorithms do not require that clusters, and therefore phase regions, are cohesive and well connected in composition space. An example can be seen in the application of the hierarchical cluster analysis to x-ray diffraction data in figure 5 of [12], where a point associated with one phase region (yellow) is surrounded by points belonging to another phase region (red). This may be due to a few causes. The invasive point may have been miss-clustered due to data artifacts such as noise or background. Alternatively, it may indicate the existence of a true structure change that should be recognized and translated to the phase diagram. These issues are resolved by utilizing a graph based method that allows the user to ensure connectivity in composition space.

In this work, a matrix factorization method is combined with a clustering method to ensure that each constituent phase set of $M$ members only exists over a limited range of the composition phase diagram—the local phase region determined by the clustering algorithm, thus resolving the inappropriate assumption made by NMF. The two methods are iteratively run in alternation, and are guided by an objective function. In each iteration, knowledge of potential constituent phases (matrix factorization) is used to improve identification of potential phase regions (clustering) and vice versa. Thus, by combining the clustering and matrix factorization methods, the results of each are enhanced, resulting in an improved composition phase diagram and set of constituent phases, as defined by the objective function. Also, the combined method does not rely on a random number generator for matrix factorization and therefor provides stable, repeatable results.

As mentioned, the combination of the clustering and matrix factorization methods ensures that the analysis results obey the physical constraint that constituent phases exist over restricted regions of the composition phase diagram. Additional data and physical constraints were imposed through the selection of the clustering method, the matrix factorization method, and the objective function: (1) a NMF method was selected to ensure that the identified constituent phases are described by non-negative structure data (e.g. strictly positive x-ray diffraction or Raman spectra) and that each sample is described by a positive combination of constituent phases. (2) The objective function was chosen to ensure that the constituent phases are similar to the original structure data through a volume constraint on the vectors describing the constituent phases. (3) A graphical model based clustering method was chosen to ensure that the clusters, and the related phase regions, are cohesive and well-connected in the composition space. The method selected allows the user to control the relative influence of the structure and composition data in determining the phase diagram. A discussion of ongoing research into including additional physical constraints can be found in the discussion section.

## 2. Graph-based endmember extraction and labeling (GRENDEL) algorithm

The analysis method used here combines three algorithms. First, spectral clustering is used to generate an initial, 'seed' composition phase diagram. Then the graphical-model-based graph cut algorithm is used to control cluster connectivity and

a NMF method is used to identify constituent phases for each cluster. The graph cut and NMF methods are part of a modified version of the GRENDEL [16] (GRENDEL) algorithm, which was originally developed to analyze hyperspectral satellite images. The original GRENDEL does not call for a non-negative constraint on the constituent phases ('endmembers') during matrix factorization. The incorporation of the non-negative constituent phase constraint as well as other modifications are described in the supplementary information. The graph cut and NMF methods are run iteratively in a process to minimize an objective function. The objective function is given as:

$$J(\boldsymbol{E}, \boldsymbol{P}, \boldsymbol{U}) = \sum_{i=1}^{C}\left(\sum_{j=1}^{N} u_{ij}\left(\boldsymbol{x}_j - \boldsymbol{E}_i\boldsymbol{p}_{ij}\right)^{\mathrm{T}}\left(\boldsymbol{x}_j - \boldsymbol{E}_i\boldsymbol{p}_{ij}\right) + \alpha\sum_{k=1}^{M-1}\sum_{l=k+1}^{M}(\boldsymbol{e}_{ik} - \boldsymbol{e}_{il})^{\mathrm{T}}(\boldsymbol{e}_{ik} - \boldsymbol{e}_{il})\right). \quad (1)$$

Here $i$ and $j$ index cluster number and sample number, respectively; $C$ is the total number of clusters; $M$ is the number of endmembers per cluster; $N$ is the number of samples; $\boldsymbol{x}_j$ is the structure data for sample $j$; $\boldsymbol{E}_i$ is the set of cluster dependent constituent phases for cluster $i$; $\boldsymbol{p}_{ij}$ are the mixture proportions of each endmember set for each sample; $\alpha$ is a coefficient for balancing the importance of the first and second summations in the objective function; and $u_{ij}$ defines the cluster membership for each sample $j$. For this implementation, $u_{ij}$ is a scalar that is one if sample $j$ belongs to cluster $i$ and zero otherwise, with each sample belonging to one and only one cluster. Also, in this implementation, each cluster has $M$ endmembers that are not shared between clusters. Both the use of 'soft' cluster membership values (ranging between zero and one) and the sharing of endmembers across cluster boundaries are intended features of a future implementation.

Minimizing the first summation with a non-negative constraint factorizes the sample structure spectra ($\boldsymbol{x}_j$) into the combination of cluster dependent constituent phases (non-negative endmembers $\boldsymbol{E}_i$) and abundances ($\boldsymbol{p}_{ij}$). Minimizing the second summation ensures that the volume described by the constituent phases (endmember vectors) is minimized, and that the constituent phases look as much like the original set of structure data as possible. A further discussion of GRENDEL, including the objective function, the optimization method, and integration of the graph cut method can be found in [16] and [17]. The graph cut method is discussed in [18–20]. A discussion of the implementation used here can be found in the supplementary information.

The general method is diagramed in figure 1. Composition and structure data from the combinatorial library (a) is first collected. For the case of x-ray diffraction analysis, relevant known materials in the ICSD (b) are manually identified, their composition and structure data are imported, and the structure data is automatically converted into simulated diffraction data as described in [13]. The combined structure data (c) is then analyzed using the spectral clustering method [21] with the cosine metric to generate an initial composition phase diagram (d) and the composition data is tessellated to generate a composition space graph (e) where each sample is a vertex and the edges connect neighboring samples. The data, along with the graph and the seed phase diagram are then fed into GRENDEL which determines a composition phase diagram (f) and constituent phases (g). figure 2 provides a flow chart for the data process, showing how each variable is updated through the analysis method, beginning with the use of spectral clustering to generate the initial composition phase diagram described by cluster membership matrix $\boldsymbol{U}$ and tessellation to create the composition space graph described by similarity matrix $\boldsymbol{S}$.

## 3. Results

The analysis was run on three datasets for the three thin-film composition spreads of Fe–Ga–Pd [12], (Bi,Sm)(Sc,Fe)$O_3$ [22], and Fe–Nb–O [23]. For the Fe–Ga–Pd dataset, experimental x-ray diffraction and composition data were used as well as composition and structural data for the known constituent binary phases from the ICSD. The ICSD structural data was converted to simulated diffraction patterns as described in [13]. The (Bi,Sm)(Sc,Fe)$O_3$ dataset is composed of composition and Raman spectra. The Fe–Nb–O system is described by a combinatorial library coordinate system and Raman spectra, as composition data was not available. All analysis was performed in MATLAB on a dual core i5-2467M 1.6 GHz laptop with 4 GB of RAM[4].

The Fe–Ga–Pd dataset was previously investigated using hierarchical cluster analysis and mean shift theory based clustering as well as NMF. The results found by the GRENDEL method are shown in figure 3 and are fairly similar to those reported for the previous algorithms [12–14], while also providing stable, repeatable results for the constituent phase determination. Computation time was 42 s. figure 3(a) shows the initial clustering result given by spectral clustering, with the experimental data shown as circles and the ICSD simulated data shown as squares. The stray red and green points show the previously mentioned issue of cluster connectivity. Figure 3(b) shows the graph generated by tessellating the composition of the samples.

Figure 3(c) provides the results of GRENDEL using the parameters listed in the supplementary information. GRENDEL identifies the stray red and green points as belonging to the blue phase region. Figure 3(d) shows the two most dominant constituent phases for each phase region, which are labeled with the same color as their respective phase regions. Of particular interest is the identification of an ICSD-like diffraction pattern for the second constituent phase of the light blue phase region. The diffraction pattern is almost identical to the ICSD-calculated diffraction pattern for $Fe_{0.7}Ga_{0.3}$. The

[4] Certain commercial equipment, instruments, or materials are identified in this report in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

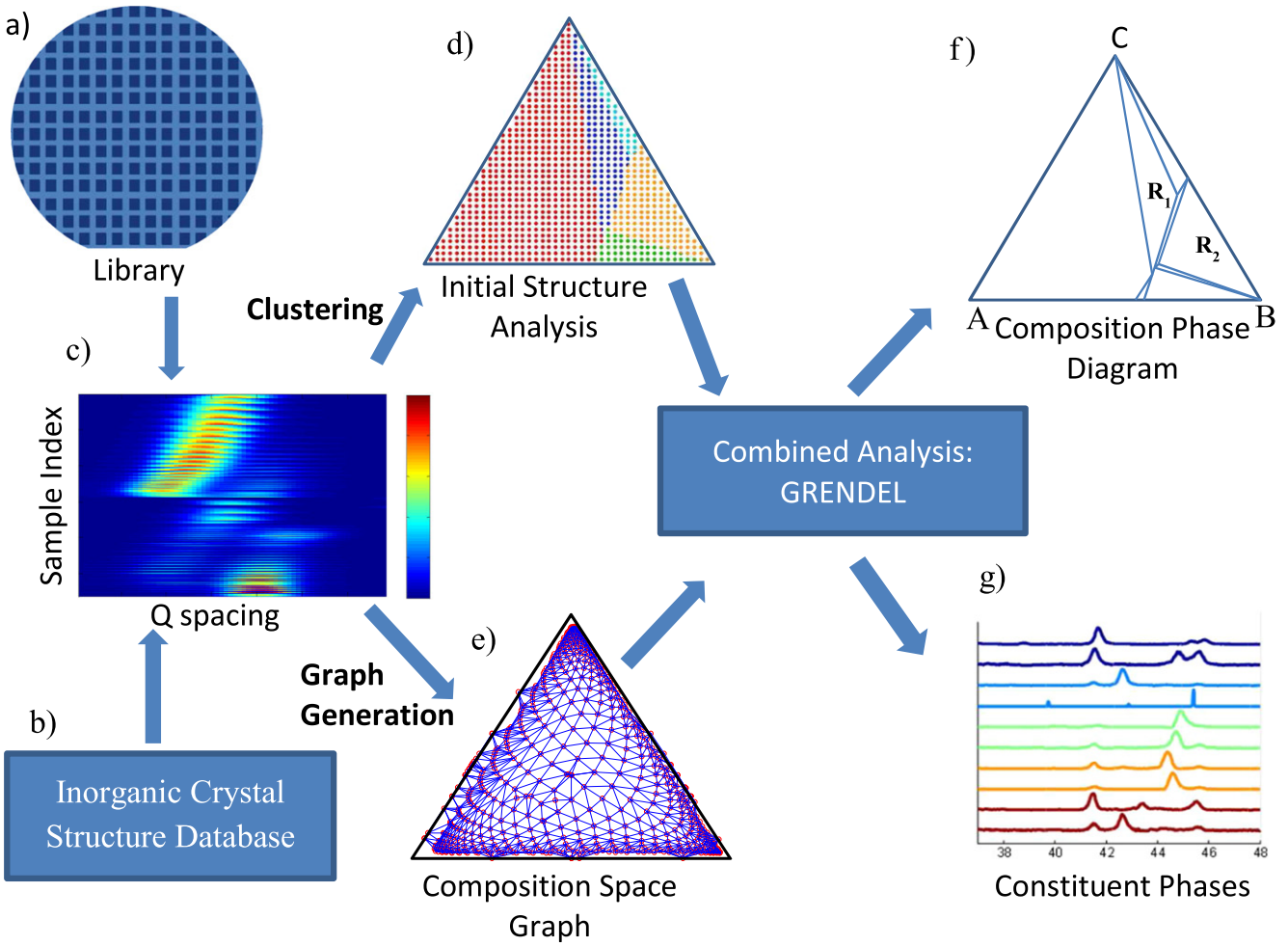



**Figure 1.** Diagram of data flow. Composition and x-ray diffraction are collected from the composition spread combinatorial library (a) and database (b) and then compiled (c). The x-ray diffraction data is clustered to provide an initial composition phase diagram (d) and the composition data is tessellated to generate a composition space graph (e). GRENDEL is then used to determine a potential phase diagram (f) and potential constituent phases (g).

$Fe_{0.7}Ga_{0.3}$ simulated sample is indicated by an arrow in figure 3(c).

As discussed in the literature, hierarchical clustering and NMF do not take into account the possibility of peak shifting, which causes a split cluster in the Fe heavy region [13]. This is also an issue with the current implementation of GRENDEL as can be seen by the two clusters (green and orange) in the Fe rich region of the composition phase diagram. The four constituent phases identified for these two regions can all be recognized as the BCC Fe structure, with composition-dependent shifts to the (110) diffraction peak. Work is ongoing to establish whether the use of a shift resilient metric would remove this limitation.

Analysis results for the (Bi,Sm)(Sc,Fe)$O_3$ system are shown in figure 4. In this study, the pseudo-ternary thin-film spread was fabricated by pulsed laser deposition. The spread was designed such that along one direction, there was a continuous substitution of the A-site (Sm substituted for Bi in perovskite $BiFeO_3$) up to an amount-of-substitution fraction of 20%. Along the other direction, a continuous B-site substitution of Fe with Sc up to an amount-of-substitution fraction of 20% was explored: the ternary maps $Bi_{1-x}Sm_xFe_{1-y}Sc_yO_3$ with $x$ and $y$ up to 0.2. This analysis took 7 s of computation time. Figure 4(a) gives the phase diagram results when only applying spectral clustering. The lack of a composition space cluster connectivity constraint can be seen in the scatter of the blue and green points (indicated by blue and green diamonds), mirroring the results found when using hierarchical cluster analysis [22]. The composition graph for the system is shown in figure 4(b) and the phase diagram results and constituent phase results are given in figures 4(c) and (d) respectively. Both are in good agreement with the results of [22]. Peaks seen in the spectra figure 4(d) at around 150 $cm^{-1}$ and 175 $cm^{-1}$ are associated with A1 and A2 modes of the R3c structure of the $BiFeO_3$ parent compound, respectively. Separate x-ray and electron microscopy studies on selected composition samples from this system have been carried out, and the clustering results here are consistent with the fact that in the light blue region, there is an antiferrodistortive phase [24]. The compositions marked in red correspond to a region where there is a co-existence of a ferroelectric and an antiferroelectric phase [25].

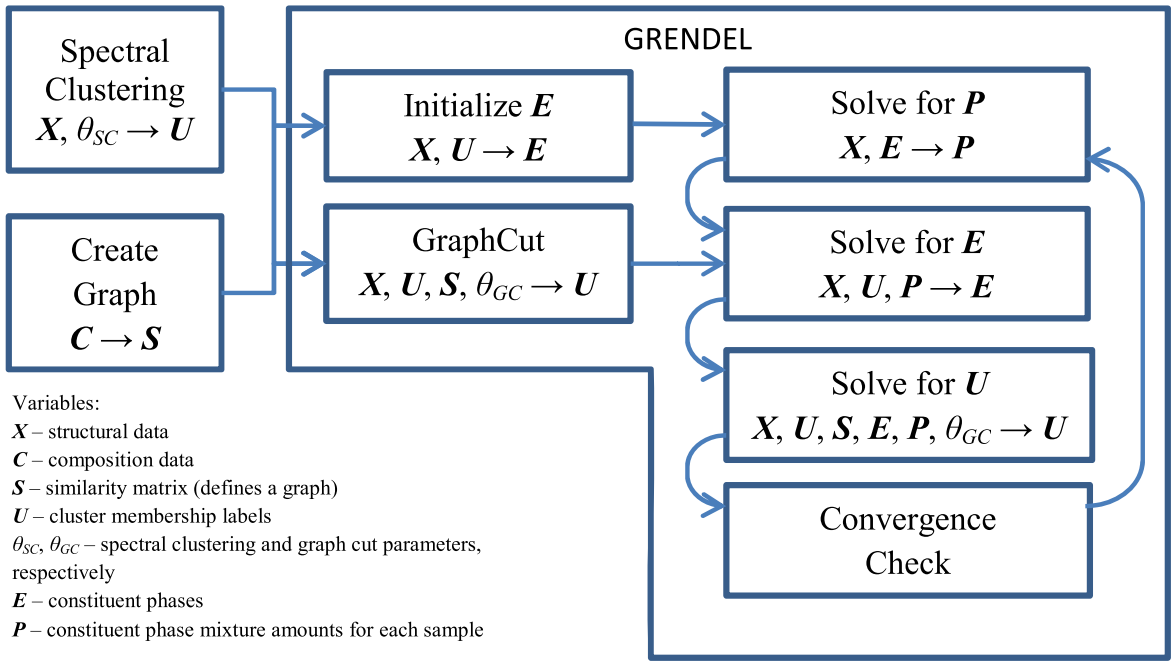



**Figure 2.** Data flow chart for the initial spectral clustering and graph determination through GRENDEL.

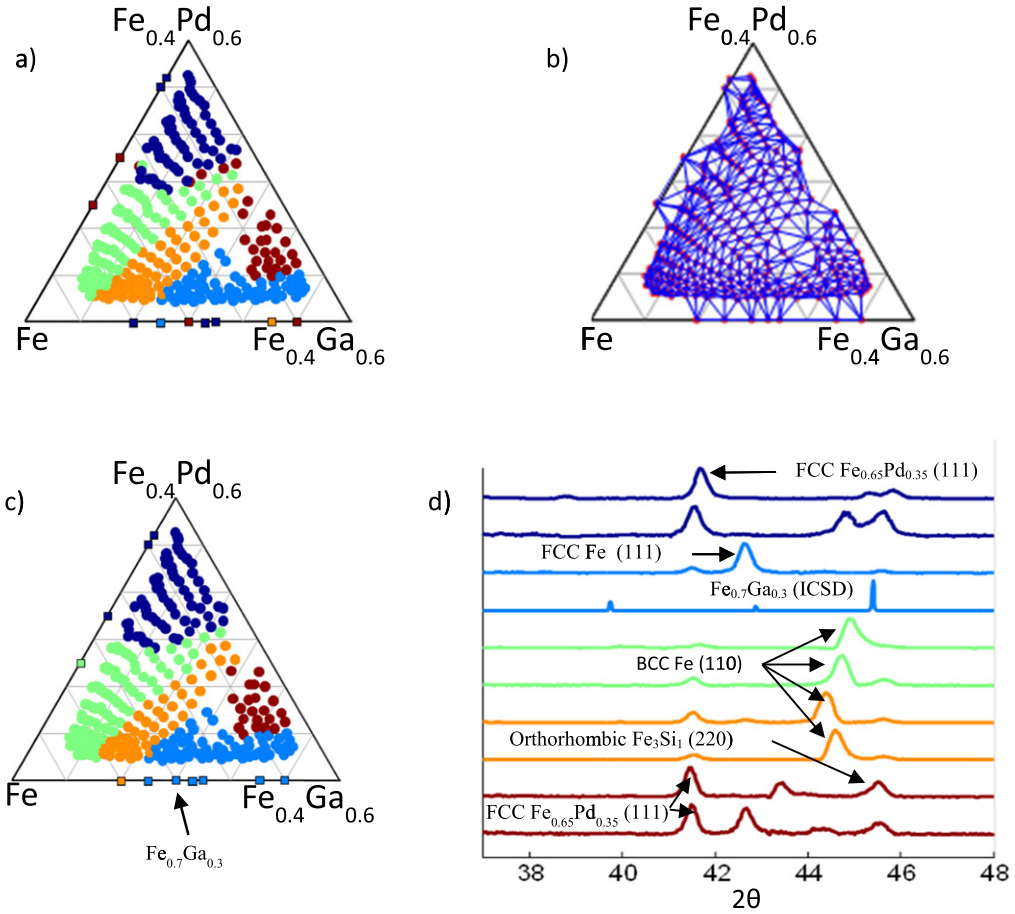



**Figure 3.** Analysis results for Fe–Ga–Pd composition spread: (a) spectral clustering results, (b) tessellated composition graph, (c) GRENDEL phase diagram, (d) GRENDEL constituent phases and their reflections.

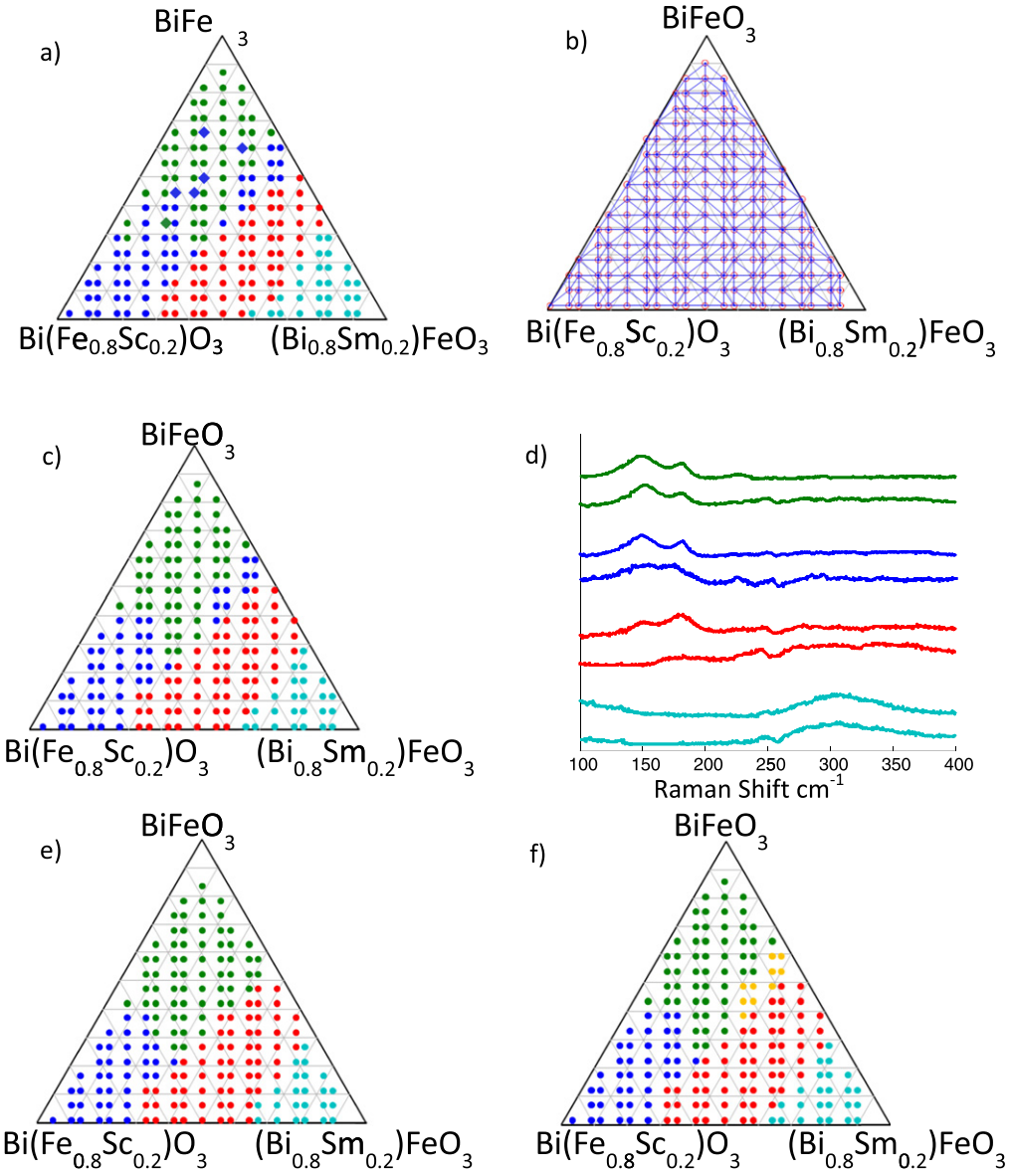



**Figure 4.** Analysis results for (Bi,Sm)(Sc,Fe)$O_3$ composition spread: (a) spectral clustering results with scattered cluster points indicated by diamond markers, (b) tessellated composition graph, (c) GRENDEL phase diagram, (d) GRENDEL constituent phases, (e) GRENDEL phase diagram results when requiring well-connected phase regions, (f) blue cluster of (c) is split giving the blue and yellow clusters.

Use of a moderate balance between the influence of composition and Raman data in clustering was used for these results, giving a split blue region with two distinct composition boundaries. The balance was obtained by a scalar weight applied to the data cost used in the graph cut clustering, as described in the supplementary information. Use of a stronger composition connectivity constraint in graph cut clustering (i.e. a reduced data cost weight) results in the absorption of the smaller blue cluster into the neighboring green phase region, as shown in figure 4(e). Also, the capability was added to identify a cluster that is not well-connected in composition space and relabel the parts as different clusters, as seen in figure 4(f). Here the blue cluster was identified to have two unconnected regions, which were then split into a blue and yellow cluster. This is useful when it is believed separated cluster points indicate the existence of an additional unique structure. In the future implementation of this algorithm, the data cost weight will be optionally determined through a model selection method such as the Bayesian Information Criteria. User control over the parameter will also remain an option as it allows the user to see the range of possible phase diagrams describable by the data under the constraints of the GRENDEL algorithm.

The Fe–Nb–O combinatorial library was analyzed using the library coordinates as a surrogate for the composition map, giving a sample coordinate based phase diagram. This library was fabricated by pulsed laser deposition. The results are shown in figure 5, with the initial spectral clustering results (Raman) given in figure 5(a), the library coordinate-based connectivity graph given in figure 5(b), the GRENDEL phase diagram shown in figure 5(c), and the constituent phases given in figure 5(d). The analysis took 26 s, and the results are in good agreement with those found in [23].

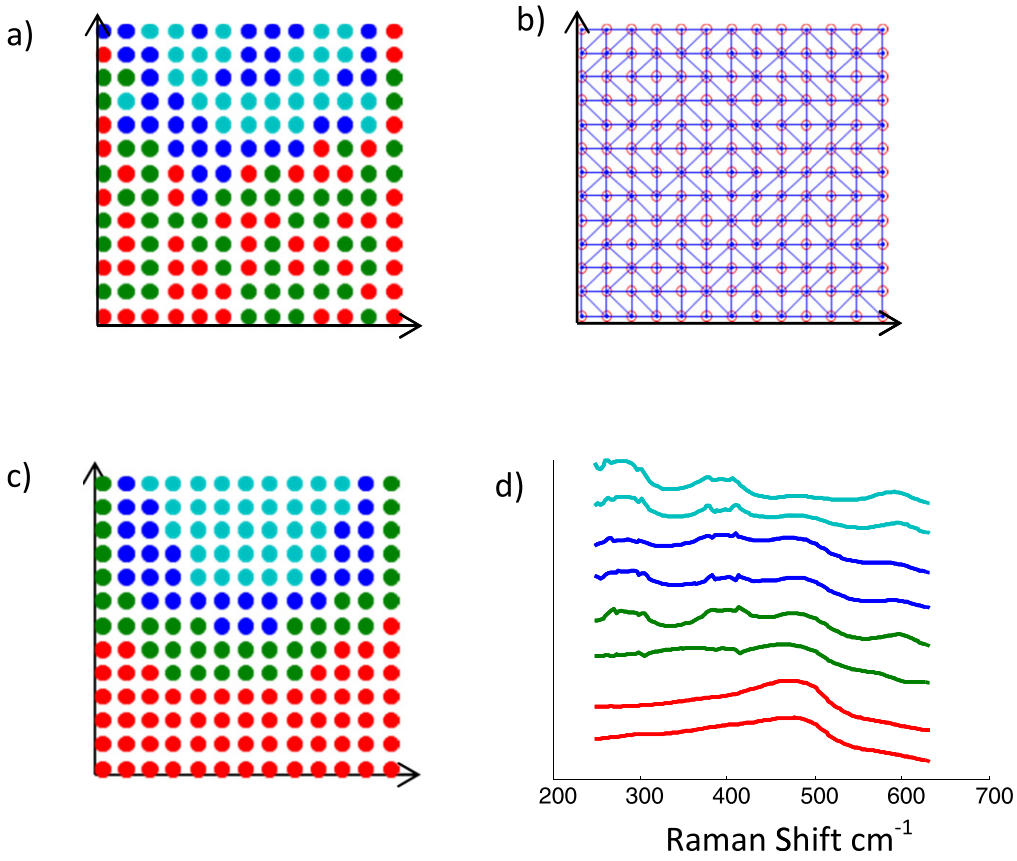



**Figure 5.** Analysis results for Fe–Nb–O library: (a) spectral clustering results, (b) tessellated composition graph, (c) GRENDEL phase diagram, (d) GRENDEL constituent phases.

## 4. Discussion

The phase diagram and constituent phases for three ternary or pseudo-ternary material systems have been analyzed using the high-throughput GRENDEL algorithm. The algorithm has been shown to be capable of integrating database data, i.e. critically evaluated data from the Inorganic Crystal Structure Database, and it has successfully analyzed both x-ray diffraction and Raman spectra for either a composition space or the combinatorial library space. While this study focused on ternary or pseudo-ternary material systems, GRENDEL may also provide benefits in analyzing binary, quaternary, and more complex material systems. However, significantly complex systems are likely to suffer from the 'curse of dimensionality' [15], requiring an exponential number of samples (scaling with the number of component materials) to determine the composition phase diagram. All analysis described in this paper was performed on a dual-core laptop and results for each dataset were obtained in under one minute.

The graphical model basis of GRENDEL allows for improved phase diagram determination as it allows the user to ensure that the clustering results that are used to identify the phase diagram are well connected in the composition space. GRENDEL also provides stable and improved results for the constituent phase identification due to its lack of requiring a random number seeding for matrix factorization. Furthermore, GRENDEL utilizes the information of each of the phase diagram and constituent phases to iteratively converge to optimized results.

As discussed in the introduction, GRENDEL has a set of built in physical constraints that ensure the resulting constituent phases are positive and similar to the given structure data and that the phase regions are well connected. Additional physical constraints can be added through modifications to the clustering algorithm, the matrix factorization algorithm, and the objective function. For example, work is underway to determine if a metric that permits feature shifts in the structure data can be used in computing structure similarity. This should provide improvements for both the phase region determination and constituent phase identification when dealing with peak shifts like the ones seen in the Fe–Ga–Pd system. Also, the Gibbs phase rule can be introduced through the addition of sparsity constraints in the objective function. The existence of multiplicative factors for each constraint (summation in the objective function) also allows the user to vary the impact of each constraint on the results or ignore them completely. Due to the versatility of the GRENDEL method, it may also be possible to incorporate such variances into an ensemble method.

It may be of interest to visualize the results of GRENDEL or other clustering analysis techniques in a manner that clearly represents the material structure–property relationship. The sunburst, a particular type of radial tree map, can be used

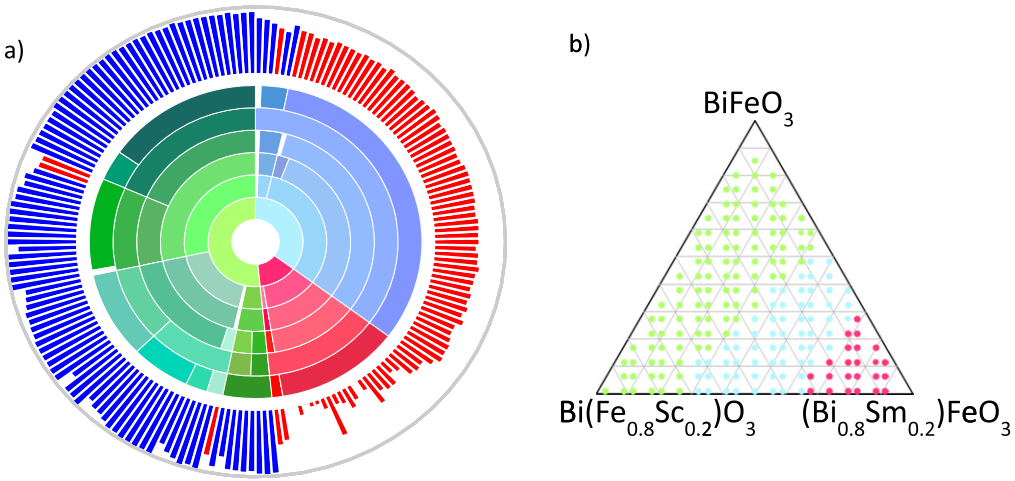



**Figure 6.** (a) A sunburst (radial tree map) diagram. Each of the inner rings shows a clustering result for the (Bi,Sm)(Sc,Fe)$O_3$ library Raman spectra. Each sample is represented by an angular direction and cluster membership is indicated by color label. The outer red-and-blue ring gives coercive electric field magnitude between zero and 491 kV $cm^{-1}$. Samples with 'open' ferroelectric hysteresis loops are indicated in blue. The sunburst visualizes the material structure–property correlation between the Raman spectra taken at each site and the ferroelectric hysteresis loops. (b) Samples from the (Bi,Sm)(Sc,Fe)$O_3$ library are labeled by cluster corresponding to the inner ring of (a).

to visualize the relationship between different structure clustering results and their relationship to functional properties. An example is shown in figure 6(a).

Each of the inner rings shows a clustering result for the Raman data from the (Bi,Sm)(Sc,Fe)$O_3$ library, with each sample represented by an angular direction. The most central ring shows the samples sorted into three clusters labeled green, blue, and magenta, corresponding to the composition diagram of figure 6(b). As the parameters of the clustering algorithm are varied, different clustering results are obtained, visualized by the concentric rings. Despite variation in clustering parameters, the majority of the samples in the blue and magenta clusters tend to remain in the same respective cluster, indicating that the sample structure similarities are stable across clustering results. Alternatively, the samples found in the green cluster tend to separate into three to five subclusters.

For this library, ferroelectric hysteresis loops were also measured at each point [22]. The outer red-and-blue radial ring displays the coercive electric field measured for each sample. Samples in the magenta cluster vary in coercive field from 9 kV $cm^{-1}$ to 316 kV $cm^{-1}$, samples in the blue cluster vary between 281 kV $cm^{-1}$ and 387 kV $cm^{-1}$, and samples in the green cluster generally showed 'open loops' indicating that these samples suffered from leakage current (and a coercive field cannot be assigned). Samples with 'open loops' are indicated in blue. It can be seen from the diagram that the clustered structures correspond to and are indicative of coercive field range. Thus, the Sunburst diagram provides a visualization of the materials structure–property relationship. In this case, the Sunburst shows that there is a direct correlation between Raman spectra taken at each site and the ferroelectric hysteresis loops.

## References


[1] Curtarolo S *et al* 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201

[2] Ortiz C, Eriksson O and Klintenberg M 2009 Data mining and accelarated electronic structure theory as a tool in the search for new functional materials *Comput. Mater. Sci.* **44** 1042–49

[3] Landis D D *et al* 2012 The computational materials repository *Comput. Sci. Eng.* **14** 51–7

[4] Mihalkovič M and Widom M 2004 Ab initio calculations of cohesive energies of Fe-based glass-forming alloys *Phys. Rev.* B **70** 144107

[5] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002

[6] Koinuma H and Takeuchi I 2004 Combinatorial solid-state chemistry of inorganic materials *Nat. Mater.* **3** 429–38

[7] Xiang X-D and Takeuchi I 2003 *Combinatorial Materials Synthesis* (Boca Raton, FL: CRC Press)

[8] Takeuchi I, van Dover R B and Koinuma H 2002 Combinatorial synthesis and evaluation of functional inorganic materials using thin-film techniques *MRS Bull.* **27** 301–8

[9] *Materials Genome Initiative for Global Competitiveness,* 2011 (National Science and Technology Council, White House) (https://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf)

[10] Le Bras R *et al* 2011 *Principles and Practice of Constraint Programming–CP 2011* (Berlin: Springer) pp 508–22

[11] Le Bras R, Bernstein R, Gomes C P, Selman B and van Dover R B 2013 Crowdsourcing backdoor identification for combinatorial optimization *Proc. 23rd Int. Joint Conf. Artificial Intelligence. IJCAI* vol 13

[12] Long C *et al* 2007 Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis *Rev. Sci. Instrum.* **78** 072217–072217

[13] Kusne A G *et al* 2014 On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets *Sci. Rep.* **4** 6367

[14] Long C, Bunker D, Li X, Karen V and Takeuchi I 2009 Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization *Rev. Sci. Instrum.* **80** 103902–103902
[15] Hastie T *et al* 2009 *The Elements of Statistical Learning* vol 2 (Berlin: Springer)
[16] Massoudifar P, Rangarajan A, Zare A and Gader P 2014 An integrated graph cuts segmentation and piece-wise convex unmixing approach for hyperspectral imaging *IEEE GRSS Workshop Hyperspectral Image Signal Processing: Evolution Remote Sensing* (*WHISPERS*) (http://cise.ufl.edu/~pegah/Whispers_Pegah.pdf)
[17] Zare A, Gader P, Bchir O and Frigui H 2013 Piecewise convex multiple-model endmember detection and spectral unmixing *IEEE Trans. Geosci. Remote Sens.* **51** 2853–62
[18] Boykov Y, Veksler O and Zabih R 2001 Fast approximate energy minimization via graph cuts *IEEE Trans. Pattern Anal. Mach. Intell.* **23** 1222–39
[19] Kolmogorov V and Zabih R 2002 *Computer Vision—ECCV 2002* ed A Heyden *et al* (Berlin: Springer) pp 65–81
[20] Boykov Y and Kolmogorov V 2004 An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 1124–37
[21] Luxburg U V 2007 A tutorial on spectral clustering *Stat. Comput.* **17** 395–416
[22] Kan D, Long C J, Steinmetz C, Lofland S E and Takeuchi I 2012 Combinatorial search of structural transitions: systematic investigation of morphotropic phase boundaries in chemically substituted $BiFeO_3$ *J. Mater. Res.* **1** 1–14
[23] Keller D A *et al* 2015 Utilizing pulsed laser deposition lateral inhomogeneity as a tool in combinatorial material science *ACS Comb. Sci.* **17** 209–16
[24] Cheng C-J *et al* 2009 Structural transitions and complex domain structures across a ferroelectric-to-antiferroelectric phase boundary in epitaxial Sm-doped $BiFeO_3$ thin films *Phys. Rev.* B **80** 014109
[25] Borisevich A Y *et al* 2012 Atomic-scale evolution of modulated phases at the ferroelectric–antiferroelectric morphotropic phase boundary controlled by flexoelectric interaction *Nat. Commun.* **3** 775