NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Data Mining-Aided Crystal Engineering for the Design of Transparent Conducting Oxides

Preprint

Changwon Suh, Kwiseon Kim, Joseph J. Berry, Jinsuk Lee, and Wesley B. Jones

To be presented at the Materials Research Society Fall Meeting
Boston, Massachusetts
November 29-December 3, 2010

Conference Paper
NREL/CP-2C00-50079
December 2010

Contract No. DE-AC36-08GO28308

NOTICE

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information

P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Data Mining-Aided Crystal Engineering for the Design of Transparent Conducting Oxides

Changwon Suh, Kwiseon Kim, Joseph J. Berry, Jinsuk Lee, and Wesley B. Jones

National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401, U.S.A.

ABSTRACT

The purpose of this paper is to accelerate the pace of material discovery processes by systematically visualizing the huge search space that conventionally needs to be explored. To this end, we demonstrate not only the use of empirical- or crystal chemistry-based physical intuition for decision-making, but also to utilize knowledge-based data mining methodologies in the context of finding p-type delafossite transparent conducting oxides (TCOs). We report on examples using high-dimensional visualizations such as radial visualization combined with machine learning algorithms such as k-nearest neighbor algorithm (k-NN) to better define and visualize the search space (i.e. structure maps) of functional materials design. The vital role of search space generated from these approaches is discussed in the context of crystal chemistry of delafossite crystal structure.

INTRODUCTION

Crystal structure of materials is closely linked with its final property [1]. In this regard, the ability to understand structural factors governing desired properties is critical in better designing functional materials. However, one of the current challenges in finding functional materials arises from the lack of tools to explore the huge search space. A good example is the discovery process for advanced TCOs due to the extremely huge search space from the many possible combinations from the periodic table to meet the TCO requirements.

There are two main approaches to handle huge search space. One is a combinatorial high-throughput synthesis and materials informatics to synthesize and interpret composition spreads, respectively [2, 3]. The other approach is to directly define the search space for TCOs such as structure mappings based on the concept of crystal chemistry [4]. While the former have been modernized successfully, the latter is still considered as a classical tool for identifying search space of materials. An example of the latter for designing new TCO includes identifying the role of the cations by Shannon *et al* in the 1970's to the phase stability, chemical bonding, and transport properties [5, 6]. The starting point of their approach was a classical bivariate structure field map consisting of ionic radii of A and B sites of the delafossite ABO_2 (ex. A=Cu, Ag, Pd, Pt; B=Co, Cr, Fe, Ga, In) structure [4], which were successfully revisited by Marquardt *et al.* later for exploring p-type TCOs [7]. This delafossite structure map was again noteworthy to the TCO community in the 2000s because only a few p-type TCOs such as Cu_2O have been developed so far [8], and there is much that still needs to be explored. Nevertheless, the approaches of the latter are inconclusive to elucidate interrelationships between structural factors and electrical/optical properties because two structural parameters used in the classical structure maps are usually not enough to fully delineate relationships of structural factors and performance of TCOs and therefore there is a need to systematically explore their inherent inter-complexities between multivariate structural factors.

APPROACHES

To solve the addressed issues here for oxide discovery as a TCO application, we revisit classical bivariate ABO_2 structure maps and modify them as modernized multivariate search spaces for ultimately finding potential p-type TCOs. We aim at fundamentally changing the conventional processes for TCO design to rational approaches by mainly focusing on a way to reduce the search space and simultaneously explore design routes in the space.

Structural aspects of ABO_2 delafossites for p-type TCOs

The delafossite structures have several advantages as the candidates for p-type TCO. For instance, tetrahedral coordinations of oxide ions reduce the non-bonding characteristics of the oxide ions, which lead to the delocalizations of the holes at the valence band edges. This layered structure enhances the bandgap enlargement and the low coordination number of the A ions is more effective to introduce comparable energy levels of Ad^{10} to those of O2p [9, 10].

Database of ABO_2 compounds

The search space includes selection of elements for multinary metal oxides with possible structures such as delafossites (ABO_2) as well as available compositions for alloys based on literature survey. Following the scheme of Marquardt *et al.*[7], A and B are systematically mutated by coordination classes of ABO_2 compounds (i.e. $A^{VI}B_{1/2}{}^{VI}B_{1/2}{}^{VI}O_2{}^{VI}$, $A^{IV}B_{1/2}{}^{IV}B_{1/2}{}^{IV}O_2{}^{IV}$, $A^{VIII}B_{1/2}{}^{IV}B_{1/2}{}^{IV}O_2{}^{VI}$, and $A^{II}B_{1/2}{}^{VI}B_{1/2}{}^{VI}O_2{}^{IV}$) and the delafossites in ABO_2 compounds include $AB_{1/2}{}^{3+}B_{1/2}{}^{3+}O_2$, $AB_{1/2}{}^{2+}B_{1/2}{}^{4+}O_2$, $AB_{2/3}{}^{2+}B_{1/3}{}^{5+}O_2$, and $A(B'^{3+}, B''^{3+})O_2$. Note that B site cation is marked with B_1 and B_2 to effectively include all the complex delafossites into the database. The order of B_1 and B_2 is not critical because we include all the mutations such that $B_1 = B_{2/3}{}^{2+}$ and $B_2 = B_{1/3}{}^{5+}$ as well as $B_1 = B_{1/3}{}^{5+}$ and $B_2 = B_{2/3}{}^{2+}$ in $AB_{2/3}{}^{2+}B_{1/3}{}^{5+}$, for example. While we mainly used an inorganic material database, the so-called atomwork of National Institute for Materials Science in Japan, for structural information, TCO related electrical and/or optical properties were also collected from the literature [7, 9, 11-37].

Figure-of-merit of ABO_2 structure maps

With two (sometimes three) carefully chosen physical factors, structure maps can be regarded as the first step of the materials design processes because they ensure that each compound is spatially located by its structure type [1, 38]. In this regard, it is possible to search stable phases of hypothetical materials. The figure-of-merit of structure maps can be defined as the ability to separate different structure types. However, it is a formidable task to choose appropriate physical factors to meet the above requirements of structure maps. With the aid of data mining techniques, there has been a study for developing multivariate structure maps without any *a priori* assumption of which two parameters are to be selected [38]. However, the approach of principal component analysis (PCA) that was used in that study can only be used when our data sets have no missing data points since it treats data as a matrix for eigenvalue decomposition. Instead, in our approach we demonstrate other data mining aided approaches such as high-dimensional visualization that are more robust to missing data.

High-dimensional visualization for multivariate structure maps

The next stage of the data-driven TCO discovery process is to map out a set of multi-dimensional points onto low dimensional space using high-dimensional visualization. In this way, the multivariate search space provides more possibility to identify various governing structural factors that determines performance of TCO, the degree of the relative impact of factors, and interdependency between factors are extracted.

1) Radial visualization as high-dimensional data representation

As a high-dimensional visualization tool, we utilize a radial visualization in a way that the multiple variables are equally laid out on the circumference of the circle (Figure 1(a)). This visualization uses the concept of Hooke's law such that a data point for a sample sits in a circle and it is connected with virtual springs. The spring constant k_i is the scaled data value in each variable. The position of each data point is assigned at the equilibrium position where the sum of spring forces is zero. The location of each data point is assigned at the equilibrium position where the sum of spring forces is zero. The mathematical foundations and some features of radial visualization can be found in literature [39, 40].

2) k-nearest neighbor (k-NN) algorithm to optimize radial visualization

While the advantage of radial visualization over any other dimension reduction techniques such as PCA is the direct use of original data sets, the display of radial visualization highly depends on the layout of the variables. Moreover, when we show I variables of total M variables, the possible projections of I variables are M!/((M-I)!I!) and each selection of I variables produce different radial projections of (I-1)!/2 [41, 42]. In our approach, we use k-NN algorithm to evaluate usefulness of radial projections created by changing the order of multiple variables as Leban *et al.* suggested [41]. As shown in Figure 1(b), k-NN algorithm searches the patterns for the k-training samples that are closest to the unknown samples in high-dimensional space. The unknown sample is assigned the most common class (i.e. structure types in this study) among its k-nearest neighbors. To this end, we calculate the accuracy of k-NN on positional information generated from radial visualization. The accuracy is estimated using the leave-one-out cross validation methods such that each data is classified in terms of structure types while other samples participate in the prediction of structure types. The computational details are beyond the scope of this paper but the reader is referred to the literature for more information [41-44].

RESULTS AND DISCUSIONS

Generated ABO_2 structure maps by radial visualization are shown in Figure 2. The top-left projection may be most useful for assigning structure types of hypothetical ABO_3 compounds since it most clearly separates ABO_2 compounds with different structures. Figure 3(a) is an example of a radial structure map consisting of five structural factors. From the locations of each data point, delafossites have higher values of M_A. Most known delafossite TCO are Cu- and Ag-based which have relatively high Mendeleev numbers of 72 and 71, respectively. It should be noted that it is always possible to unexpectedly uncover the roles of any structural factors with this approach.

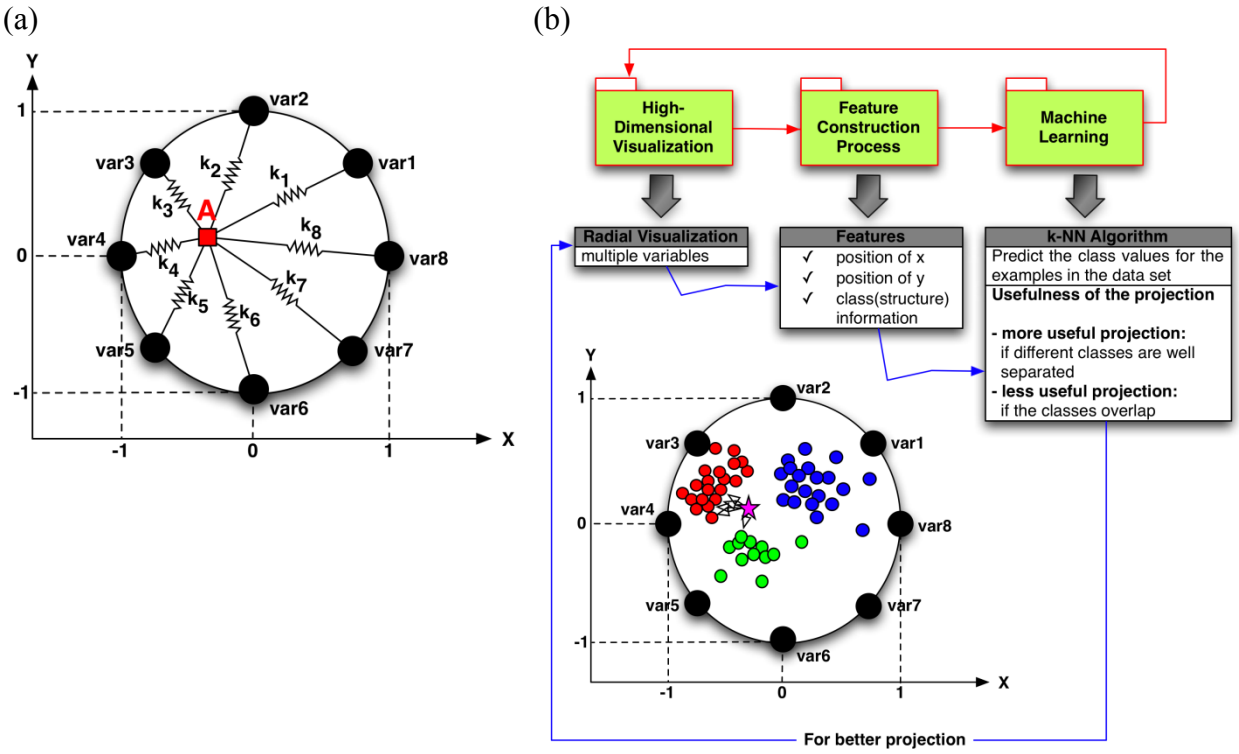

Figure 1. (a) A schematic of radial visualization for representing a point having 8 variables. (b) A procedure for finding interesting projections via k-NN algorithm in radial visualization. Note that the structure type of each sample (ex. marked with star-shaped) is assigned in a given order of variables for radial visualization. Here, when k=5, the four nearest neighbors are class of "red", while one is class of "green'. Therefore, the star-marked sample is assigned as "red" class.

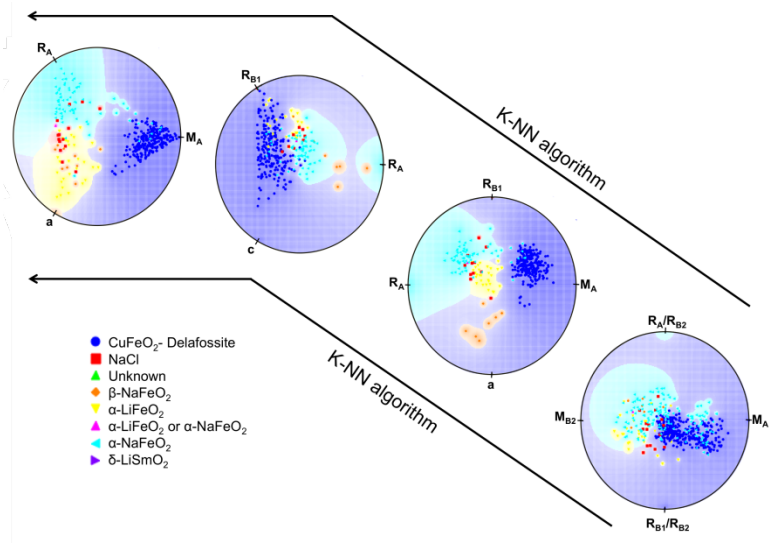

Figure 2. Exemplary ABO_2 structure maps generated by radial visualization and k-NN algorithms. The used notations in Figure 2 and Figure 3 are R_A and R_B: Shannon's ionic radius of A and B, a and c: lattice parameter of a- and c-axis, M_A, M_{B1}, and M_{B2}: Mendeleev sequential number of A and B, f_{B1}, f_{B2}: fraction of B, and V: volume of unit cell. B_1 and B_2 were devised to deal with complex delafossites in the database (see the section of database of ABO_2 compounds).

The ratio of ionic radii in A and B site (R_A/R_B) can be a good classifier to discriminate α-$NaFeO_2$ from $CuFeO_2$ delafossites. The reason is that in Figure 3(b) the structure of α-$NaFeO_2$ has high values of R_A/R_B while the structure of $CuFeO_2$ delafossites has low values. Note that

this relationship is known as a result of painful crystallochemical analysis of the structure with the criterion of $R_A/R_B>1$[36].

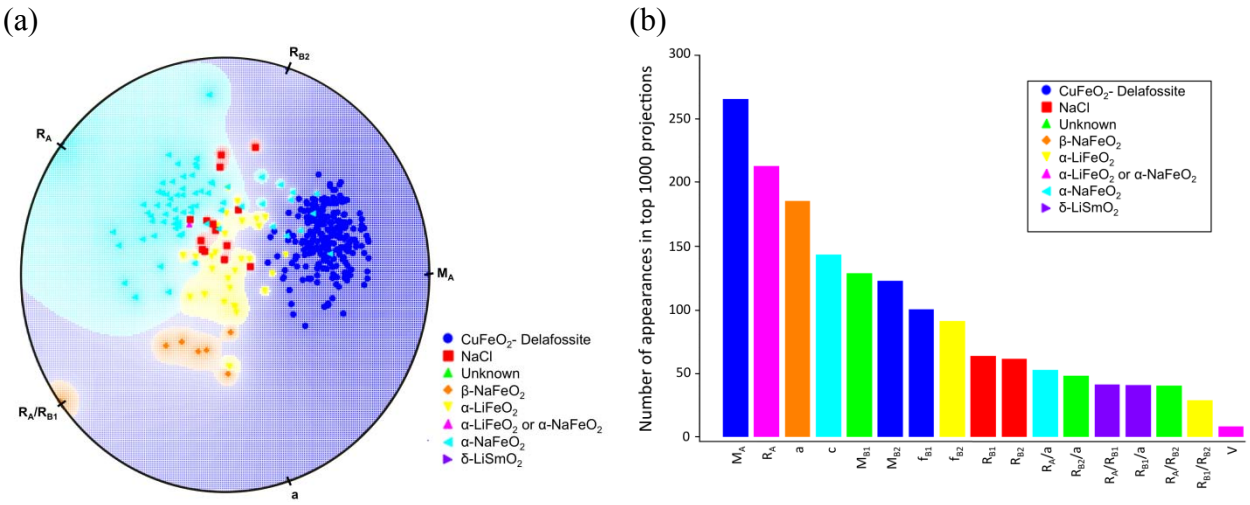

Figure 3. (a) An example of radial visualization-derived multivariate ABO_2 structure map. (b) Variable rankings of ABO_2 in terms of number of appearances in top 1000 radial projections.

The radial visualization can also provide variable ranking in a given data. For example, Figure 3(b) shows the number of appearances in top 1000 projections for each variable. From this, we identify that the main classifiers (i.e. variables to classify these structures) in a given ABO_2 formula are M_a, R_a, a, c etc. From the color of the figure, delafossites are more related to Pettifor's Mendeleev sequence numbers defined in literature [1], which were also confirmed in Figure 3(a).

CONCLUSIONS

Our approaches have shown the value of high-visualization techniques and machine learning algorithms such as radial visualization and k-NN for developing multivariate structure maps of ABO_2 compounds, including delafossites. Our study can be applied any kinds of dataset. Apart from the visualization of search space in terms of structure mapping, we can also unravel complex structure-processing-property relationships of materials using the demonstrated concepts in the field of materials informatics by including structural factors, processing conditions, and properties. These relationships are useful to suggest hypothetical delafossites for p-type TCO applications. We are currently developing various physicochemical parameters for better multivariate TCO structure maps.

ACKNOWLEDGMENTS

The Alliance for Sustainable Energy, LLC (Alliance), is the Manager and Operator of the National Renewable Energy Laboratory. Employees of the Alliance for Sustainable Energy, LLC, under Contract No. DE-AC36-08GO28308 with the U. S. Dept. of Energy have authored this work. The United States Government retains and the publisher, by accepting the article for

publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

REFERENCES

1. D. G. Pettifor, in *Crystal Structures of Intermetallic Compounds*, edited by W. J.H. and R. L. Fleisher (John Wiley & Sons, New York, 1994) p. 195.
2. J. D. Perkins, J. A. del Cueto, J. L. Alleman, C. Warmsingh, B. M. Keyes, L. M. Gedvilas, P. A. Parilla, B. To, D. W. Readey and D. S. Ginley, Thin Solid Films **411**, 152 (2002).
3. C. Suh, C. W. Gorrie, J. D. Perkins, P. A. Graf and W. B. Jones, Acta Mater. **59**, 630 (2011).
4. R. Shannon and C. Prewitt, J. Inorg. Nucl. Chem. **32**, 1427 (1970).
5. R. D. Shannon, J. L. Gillson and R. J. Bouchard, J. Phys. Chem. Solids, **38**, 877 (1977).
6. D. Kammler, T. Mason and K. Poeppelmeier, J. Am. Ceram. Soc. **84**, 1004 (2001).
7. M. Marquardt, N. Ashmore and D. Cann, Thin Solid Films **496**, 146 (2006).
8. H. Mizoguchi, M. Hirano, S. Fujitsu and T. Takeuchi, K. Ueda, and H. Hosono, Appl. Phys. Lett. **80**, 1207 (2002).
9. A. Banerjee and K. Chattopadhyay, Prog. Cryst. Growth Ch. **50**, 52 (2005).
10. A. Buljan, M. Llunell, E. Ruiz and P. Alemany, Chem. Mater. **13**, 338 (2001).
11. A. Vegas and R. Isea, Acta Crystallogr. B **54**, 732 (1998).
12. N. Tsuboi, T. Hoshino, S. Kobayashi, K. Kato and F. Kaneko, Phys. Stat. Sol. (a) **203**, 2723 (2006).
13. J. Tate, M. Jayaraj, A. Draeseke, T. Ulbrich, A. Sleight, K. Vanaja, R. Nagarajan, J. Wager and R. Hoffman, Thin Solid Films **411**, 119 (2002).
14. W. C. Sheets, E. S. Stampler, M. I. Bertoni, M. Sasaki, T. J. Marks, T. O. Mason and K. R. Poeppelmeier, Inorg. Chem. **47**, 2696 (2008).
15. F. Sauvage and D. Munoz-Roja, K. R. Poeppelmeier, and N. Casañ -Pastor, J. Solid State Chem. **182**, 374 (2009).
16. P. W. Sadik, M. Ivill, V. Craciun and D. P. Norton, Thin Solid Films **517**, 3211 (2009).
17. T. Okuda, N. Jufuku, S. Hidaka and N. Terada, Phys. Rev. B **72**, 144403 (2005).
18. Y. Ono, K.-i. Satoh, T. Nozaki and T. Kajitani, Jpn. J. Appl. Phys. **46**, 1071 (2007).
19. T. Nozaki, K. Hayashi and T. Kajitani, J. Electron. Mater. **39**, 1798 (2010).
20. R. Nagarajan, A. Draeseke, A. Sleight and J. Tate, J. Appl. Phys. **89**, 8022 (2001).
21. R. Nagarajan, N. Duan, M. Jayaraj, J. Li, K. Vanaja, A. Yokochi, A. Draeseke, J. Tate and A. Sleight, Int. J. Inorg. Mater. **3**, 265 (2001).
22. N. Miyasaka, Y. Doi and Y. Hinatsu, J.Solid State Chem. **182**, 2104 (2009).
23. , F. Mathieu and P. Tailhades, Inorg. Chem. **48**, 6065 (2009).
24. R. Kykyneshi, B. Nielsen, J. Tate, J. Li and A. Sleight, J. Appl. Phys. **96**, 6188 (2004).
25. H. Kandpal and R. Seshadri, Solid State Sci. **4**, 1045 (2002).
26. R. Isea, A. Vegas and A. Ramos-Gallardo, Acta Crystallogr. B **54**, 35 (1998).
27. S. Götzendörfer and P. Löbmann, J.Sol-Gel Sci.Technol. DOI: 10.1007/s10971-010-2336-0 (2010).
28. N. Wongcharoen and T. Gaewdang, Phys. Proc. **2**, 101 (2009).

29. J.-P. Doumerc, A. Ammar, A. Wichainchai, M. Pouchard and P. Hagenmuller, J. Phys. Chem. Sol. **48**, 37 (1987).
30. G. Dong, M. Zhang, T. Li and H. Yan, J. Electrochem. Soc. **157**, H127 (2010).
31. N. Ashmore and D. Cann, J. Mater. Sci. **40**, 3891 (2005).
32. R. Bywalez, S. Goetzendoerfer and P. Loebmann, J. Mater. Chem. **20**, 6562 (2010).
33. A. Buljan, P. Alemany and E. Ruiz, J. Phys. Chem. B **103**, 8060 (1999).
34. Y. Bessekhouad, Y. Gabes, A. Bouguelia and M. Trari, J. Mater. Sci. **42**, 6469 (2007).
35. K. El Ataoui, J. Doumerc, A. Ammar, P. Gravereau, L. Fournčs, A. Wattiaux and M. Pouchard, Sol. State. Sci. **5**, 1239 (2003).
36. B. V. Beznosikov and K. S. Aleksandrov, Institute of Physics, Siberian Division, Russ. Acad. Sci. Krasnoyarsk, Preprint No. 843F, 2007.
37. B. V. Beznosikov and K. S. Aleksandrov, J. Struct. Chem+. **50**, 102 (2009).
38. C. Suh and K. Rajan, Mats. Sci. Tech. **25**, 466 (2009).
39. C. Brunsdon, A. Fotheringham and M. Charlton, Technical Report Series, **43**, 55 (1998).
40. J. McCarthy, K. Marx, P. Hoffman, A. Gee, P. O'Neil, M. Ujwal and J. Hotchkiss, *Ann. NY. Acad. Sci.* **1020**, 239 (2004).
41. G. Leban, B. Zupan, G. Vidmar and I. Bratko, Data Min. Knowl. Disc. **13**, 119 (2006).
42. M. Mramor, G. Leban, J. Demsar and B. Zupan, Bioinformatics **23**, 2147 (2007).
43. J. Han and M. Kamber, *Data Mining - Concepts and Techniques*. (Morgan Kaufmann Publishers, San Francisco, 2006) p.348.
44. C. Kamath, *Scientific Data Mining - A practical Perspective*. (SIAM, Philadelphia) p. 185.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
December 2010	Conference Paper	

4. TITLE AND SUBTITLE
Data Mining-Aided Crystal Engineering for the Design of Transparent Conducting Oxides: Preprint

5a. CONTRACT NUMBER
DE-AC36-08GO28308

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
C. Suh, K. Kim, J.J. Berry, J. Lee, and W.B Jones

5d. PROJECT NUMBER
NREL/CP-2C00-50079

5e. TASK NUMBER
PVA9.2910

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401-3393

8. PERFORMING ORGANIZATION REPORT NUMBER
NREL/CP-2C00-50079

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
NREL

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 Words)*
The purpose of this paper is to accelerate the pace of material discovery processes by systematically visualizing the huge search space that conventionally needs to be explored. To this end, we demonstrate not only the use of empirical- or crystal chemistry-based physical intuition for decision-making, but also to utilize knowledge-based data mining methodologies in the context of finding p-type delafossite transparent conducting oxides (TCOs). We report on examples using high-dimensional visualizations such as radial visualization combined with machine learning algorithms such as k-nearest neighbor algorithm (k-NN) to better define and visualize the search space (i.e. structure maps) of functional materials design. The vital role of search space generated from these approaches is discussed in the context of crystal chemistry of delafossite crystal structure.

15. SUBJECT TERMS
materials information; data mining; structure-property relationships; TCO; structure mapping

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (*Include area code*)
Unclassified	Unclassified	Unclassified	UL		