Design of Electroceramic Materials Using Artificial Neural Networks and Multiobjective Evolutionary Algorithms

D. J. Scott, S. Manos, and P. V. Coveney*

Centre for Computational Science, Department of Chemistry, University College London, Christopher Ingold Laboratories, 20 Gordon Street, London WC1H 0AJ, U.K.

Received July 25, 2007

We describe the computational design of electroceramic materials with optimal permittivity for application as electronic components. Given the difficulty of large-scale manufacture and characterization of these materials, including the theoretical prediction of their materials properties by conventional means, our approach is based on a recently established database containing composition and property information for a wide range of ceramic compounds. The electroceramic materials composition-function relationship is encapsulated by an artificial neural network which is used as one of the objectives in a multiobjective evolutionary algorithm. Evolutionary algorithms are stochastic optimization techniques which we employ to search for optimal materials based on chemical composition. The other objectives optimized include the reliability of the neural network prediction and the overall electrostatic charge of the material. The evolutionary algorithm searches for materials which simultaneously have high relative permittivity, minimum overall charge, and good prediction reliability. We find that we are able to predict a range of new electroceramic materials with varying degrees of reliability. In some cases the materials are similar to those contained in the database; in others, completely new materials are predicted.

1. INTRODUCTION

Ceramic materials have wide ranging applications, and there is a continuing demand for materials which exhibit specific properties for particular purposes.[1] The composition-structure-property relationships found in ceramic materials are extremely complicated.[2] Furthermore, ceramic materials cover an almost limitless range of possible compositions exhibiting a large range of different properties, making the discovery of new materials by conventional methods extremely difficult. Traditionally, exploration of new compositions is often carried out incrementally, in close proximity to already well-known materials and is usually slow and costly. The move to combinatorial methods, first proposed by Hanak,[3] has accelerated the synthesis and analysis of new materials, permitting a wider range of compounds to be studied and resulting in large databases of materials property data.[4] Such databases are ideal for the application of "Baconian" methods[5] in which statistical inference is used to make scientific predictions, in direct contrast to the conventional "Popperian"[6] scientific method. While Baconian techniques begin with experimental data and attempt to make predictions using statistical inference, Popperian methods commence by adducing a scientific theory which is then tested by experiment. Popperian scientific theory is falsifiable; that is, experimental data in conflict with theoretical predictions may invalidate the theory, which then needs to be adjusted or rejected. Baconian methods are based on induction and depend solely on the quality and the quantity of the data. The Popperian approach is the standard one adopted in the physical sciences.[7,8] Although undoubtedly powerful, its domain of success is tightly circumscribed: in practice, it is often very hard to predict ab initio the properties of new materials using such deductive methods.

Baconian methods offer a potentially more efficient means of predicting new materials properties, especially if they are able to draw on the large data sets obtained through modern combinatorial techniques. Even when large materials data sets are taken into account, there remains a vast and largely unexplored compositional search space for which an exhaustive systematic investigation is virtually impossible. The purpose of the present paper is to describe an inductive approach to electroceramic materials design that draws on a recently created materials database, an artificial neural network (ANN) that maps the composition-property relationship, and an evolutionary algorithm to search for novel electroceramic materials. The approach yields materials compositions suitable for initiating conventional laboratory work and further combinatorial searches.

Here, we use a Baconian approach which employs statistical techniques applied to materials property databases in order to develop electroceramic material compositions. A ceramic materials database and informatics system[4,9] has been developed with data gleaned from the literature and produced by the Functional Oxide Discovery (FOXD) project.[10] An ANN is applied to data within the database and is used to encapsulate the composition-function relationship,[11] which is then inverted using an evolutionary algorithm (EA)[12] to provide the materials composition predictions.

EAs have wide ranging applications. They have been used for combinatorial library design[13,14] as well as inverse quantitative structure−property relationship (QSPR) problems in biomolecular science.[15] Biomolecular science also contains examples of the use of EAs for the inversion of

* Corresponding author phone: +44 20 7679 4560; fax: +44 20 7679 7463; e-mail: p.v.coveney@ucl.ac.uk.

10.1021/ci700269r CCC: $40.75
Published on Web 01/25/2008

artificial neural networks.[16] EA inversion of ANN predictors has also been used in capacitor design.[17] Additionally, EAs have been used to train ANNs to predict the mechanical properties of ceramics.[18] However, to our knowledge, the use of EAs to invert ANNs is novel for the design of electroceramics. Although there has been some work in the field of catalyst design,[19] the EA inversion of ANNs is rare in the field of functional ceramic materials.

Our paper is structured as follows. Section 2 provides a more detailed overview of electroceramics and their design. Section 3 discusses the database used to store literature data and the ANN which encapsulates the ceramic composition-function relationships; it also gives an overview of the multiobjective EA, including the representation, constraints, and objectives used. The three objectives and the implementation of the multiobjective EA are discussed in section 4. Our results are presented in section 5 and are discussed in section 6. Section 7 contains our conclusions and a consideration of future research directions.

2. THE DESIGN OF ELECTROCERAMIC MATERIALS

The study of ceramic materials is a complex subject due to both the large range of materials available and the varied properties exhibited.[2,20] Research in the field is driven by the many applications of these materials and covers a wide range of scales from their crystalline atomic lattice structure to the properties of entire devices. For decades, many scientists have attempted to develop models in order to generalize our understanding of materials systems and make predictions of their properties. Although a modern Popperian technique would be based on fundamental physiochemical principles, such an approach is in fact remarkably difficult to sustain for sufficiently complex multivariate systems including many materials.[7,8] The complexity of ceramics, due to their polycrystalline nature, multicomponent compositions, multiple phases, structural defects, and mobile ionic species which all affect their functional properties, further hinders our ability to develop Popperian models. The Baconian alternative to Popperian methods, used extensively in bioinformatics, including the pharmaceutical industry,[21] is much less mature in materials design;[22] nevertheless, this paper describes a Baconian technique for predicting new electroceramic materials.

Devices containing ceramic materials find applications in an increasing number of fields; our work concentrates on electroceramic materials which include dielectric and conductive ceramics.[4,11] Here, we focus on dielectric ceramics which are extensively used in microwave and telecommunications applications. The continuing growth of mobile telecommunications has sustained an interest in novel ceramics for use as dielectric resonators at microwave frequencies (1−20 GHz) and new materials with ever more specific properties are constantly required. Many useful dielectric resonator materials are perovskite oxides with the general formula ABO_3, where A and B are rare earth/alkaline earth ions and transition-metal cations, respectively. The perovskite structure, shown in Figure 1, is very versatile; doping both the A- and B-sites with similar metallic elements produces many possible compounds with widely varying properties. The addition of even a small amount of dopant species can have a major effect on the properties of the material. The ideal properties of a dielectric resonator are a sufficiently high relative permittivity ϵ_r (dimensionless) to allow miniaturization of the component ($\epsilon_r > 10$) and a high 'Q' factor at microwave frequencies to improve frequency selectivity ($Q > 5000$). The quality factor, Q, is given by the inverse of the dissipation factor $Q = 1/\tan(\delta)$ where δ is the loss angle, the phase shift between the voltage and current when an AC field is applied to a dielectric material.[23]

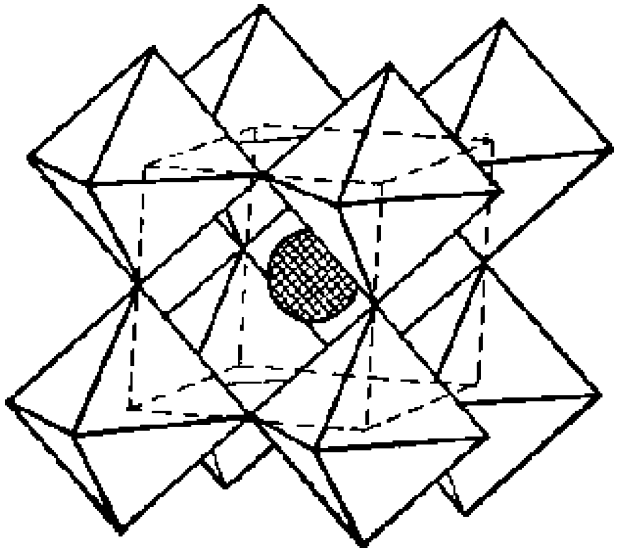

Figure 1. Basic perovskite structure of $CaTiO_3$ with the Ca^{2+} ion in the center of the cell, Ti^{4+} ions on the corner lattice sites, and O^{2-} ions on the center of each edge.[24]

We have created a publicly available database[4] containing composition and property information on dielectric ceramic materials with data culled from the literature. This database also contains results from combinatorial experiments performed on the London University Search Instrument (LUSI),[25] a high-throughput combinatorial robot based on ink-jet printing technology. The database can be analyzed using data mining algorithms which model the composition-property relationships inter alia.

3. MATERIALS DESIGN ALGORITHM

Our approach to the discovery of new materials involves a two-step process. Initially, we develop a "forward" prediction model which specifies the mapping from chemical composition to materials properties. In a second stage, this mapping is inverted[26] to determine the chemical compositions that produce specified materials properties. We use an artificial neural network (ANN) to determine the composition-properties relationship.[27] An ANN is a learning algorithm which uses existing data to develop a forward prediction model, capable of estimating output value(s) from a set of inputs, provided that sufficient data are available. Although ANNs are extremely powerful and can model very complicated data relationships, no analytical method exists for inverting an ANN in order to determine the inputs required to obtain a specific output. This kind of inverse problem can be solved using an evolutionary algorithm which exploits analogies with biological evolution to develop solutions. In this section we describe both the neural networks and evolutionary algorithms which are combined to search for new ceramic materials.

3.1. Artificial Neural Networks. Statistical analysis techniques can be divided into two domains: data modeling and algorithmic modeling.[28] Unlike standard statistical regression techniques, which fall into the data modeling field, ANNs are algorithmic modeling techniques, making no prior assumption about input−output relationships, a major advantage in their application to complex systems. Artificial neural networks can be used to develop functional approximations to data with almost limitless application[29,30] and are comprised of individual processing units, which are arranged into layers that are interconnected to give a powerful computational system. An individual unit consists of weighted inputs, a *combination* function, an *activation* function, and one output. The outputs of one layer are connected to the inputs of the next layer to form the network topology. The kind of activation function, the network architecture, and the training algorithm strongly influence the network performance.

We now outline the multilayer perceptron (MLP) network we have developed using the back-propagation algorithm. The details of the construction and performance of this ANN have been recently published[11] so we only summarize that work here. In an MLP network, the individual processing units are known as *perceptrons* which are usually arranged into three layers: input, hidden, and output. The number of hidden neurons is determined by the complexity of the problem and is often obtained by trial and error, although evolutionary computing techniques such as genetic algorithms[31] have been used to determine optimal network architecture. The training algorithm consists of two steps. In the first, the forward processing is performed, giving an output prediction based on the input values. In the second, the network weights are adjusted so that the prediction becomes closer to the actual output values contained in the training data.

Once the weights have been adjusted, this process is carried out for each of the training data records and is repeated many times for the whole data set until the training is complete. A validation data set is used to monitor the process and provides an estimate of the general prediction error of the network. Training is halted when the error begins to increase which indicates that the network is learning specific features of the training data set rather than learning the generic data relationships.

Once training is complete, the network is assessed for its general predictive ability using a test data set. A previously unused test data set is applied to the network, and the predicted outputs are compared with the actual results. The root relative squared error between the predictions and the real outputs is given by

$$\epsilon_{\mathrm{RRS}} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - t_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t})^2}} \qquad (1)$$

where y is the output predicted by the network, t is the actual output, N is the number of records in the data set, and $\bar{t}$ is the mean of the experimentally measured outputs. The root relative squared error is independent of the size of the test data set and provides an estimation of the performance of the ANN relative to the performance of a simple "mean predictor". The error of the test data set predictions is known as the generalization error.

We used cross-validation to produce the most reliable estimate of the generalization error of the ANN. A single network training carries with it the possibility that the predictive performance is due to chance selection of training, validation, and test data sets. Cross-validation[32] is a technique which attempts maximal use of the available data to obtain a good estimate of the generalization error. Typically, 10-fold cross-validation[27,33] is used which operates by dividing the data set into 10 subsets. In our work, one of these subsets was withheld from the data set, and the remaining data were used to train and validate the ANN. The withheld data set was used to test the network by calculating the root relative squared (RRS) error between the ANN predictions and the actual results. The process was repeated with each of the ten subsets being withheld in turn. Once complete, the data set was randomized, and the cross-validation was performed again. The data set was randomized ten times in total; 10-fold cross-validation was performed each time, reducing the possibility of the ANN performance being due to a particular data set selection. Cross-validation allows maximal use of the data set for training while withholding a separate data set which is used to determine the generalization error. The overall estimation of the general performance of the network is given by the mean value of the generalization error for each of the cross-validation subsets.

3.2. Multiobjective Evolutionary Algorithms. Real-world design problems typically involve multiple, and often conflicting, objectives.[34] Alongside increasing interest in applying evolutionary algorithms (EA) to complex systems, there has been ongoing research into multiobjective evolutionary algorithms. Evolutionary algorithms are a stochastic optimization technique inspired by biological evolution which use artificial equivalents of individuals, populations, mutations, and the concept of "survival of the fittest" to evolve optimal solutions to a problem. Evolutionary algorithms are particularly powerful since they make no assumptions about the underlying problem landscape and require no knowledge of the function gradient, making them very well suited to ANN inversion problems. They are also advantageous over gradient descent techniques because they are less likely to become trapped in local minima of the target function. Although evolutionary algorithms have a wide scope, including evolutionary and genetic programming, the most popular evolutionary algorithm is the genetic algorithm (GA) of Holland.[35]

3.2.1. Genetic Algorithms. In GAs, individual solutions are described by an array of numbers which represent the genes of the individual and form the input to a "fitness" or "objective" function. The fitness function determines the best individuals from within a population of putative solutions which are selected for recombination or "crossover". Crossover is the exchange of genetic information between two individuals resulting in one or more "offspring" and is reminiscent of sexual reproduction in living organisms. A random, low-probability adjustment to each of the genes is also included and is used to introduce new genetic material into the population. Known, as "mutation", this process also has its equivalent in biological evolution. The mutations are

the cause of the stochastic nature of the search and help prevent the algorithm becoming trapped in local minima. A favorable interchange/mutation produces an individual solution closer to the optimum of the target function; a poorer interchange/mutation results in a less optimal individual. Repeated iterations of the selection and crossover processes result in an improvement in the collective fitness of the population. The process is halted when no further improvement in the best individual occurs. There are many textbooks which describe the operation and implementation of GAs[36,37] and so only a summary is provided here.

3.2.2. Representation. In its classic form, an individual solution can be represented as an array of binary numbers which are concatenated to form a genotype. The crossover and mutation operations are then trivially performed on the complete string-crossover by selecting a crossover point and exchanging the bits on one side of the point between two parent strings, and mutation by randomly selecting a location for the mutation to occur and then "bit flipping" the element at that location with a random probability. A real-valued GA is used when the genotype can be represented in terms of real values. Real-valued GAs use probability distributions to perform the mutation operations and the popular simulated binary crossover (SBX)[38] to perform recombination, although other algorithms are available. Mutation and crossover are performed depending on a random probability parameter, and the "strength" of the operation is defined by a probability distribution index. SBX uses a probability distribution to generate offspring. Based on the search features of single point crossover used in binary coded algorithms, SBX attempts to generate "children" near to the parents. During the initial stages of the optimization, the population is spread, and the children are diverse, resulting in a coarse grained search. As the optimization progresses, the population converges, resulting in clustering of the children, and a fine grained search emerges.

3.2.3. Multiple Objectives. Although a simple optimization problem may involve only a single objective, many situations require simultaneous optimization of multiple objectives. Often in real world design problems the objectives are conflicting and trade-offs exist between them; as the fitness of one objective improves, the fitness of another is reduced. Two main approaches for dealing with multiple objectives exist. First, the objectives can be combined into a single fitness value. This presents two possible problems: the inherent trade-offs would result in an overall suboptimal solution, and the combination of objectives often requires the selection of appropriate coefficients to avoid skewing toward a suboptimal single solution. Second, we can use a full multiobjective optimization approach. In contrast with single-objective optimization, owing to the presence of trade-offs, no "single best" solution exists. Multiobjective EA techniques are well suited to this problem since they operate on a population and result in a group of solutions, each satisfying the objectives to varying degrees. Hand-selection of individuals from the final population provides final candidate solutions. In the instance where the objectives are simultaneously attainable, the population reduces to a single point. Otherwise, a trade-off surface containing "nondominated" solutions results. A particular solution is said to be nondominated if there exists no other individual in the population which is more optimal in **all** objectives. Formally, when minimizing all M objectives, with objective values f_i, design **a** dominates design **b** if

$$f_i(\mathbf{a}) \leq f_i(\mathbf{b}),\ i = 1, \ldots, M \quad \text{and}$$
$$\exists i \in (1, \ldots, M),\ f_i(\mathbf{a}) < f_i(\mathbf{b}) \quad (2)$$

A group of nondominated solutions is known as a nondominated set or "Pareto-set". For a particular population, the first nondominated set is given a "rank" of zero, and the entire population of solutions can be further ranked by temporarily ignoring the first nondominated set and calculating the nondominated set of the remaining solutions. This process, which can be repeated until the entire population is categorized, is used during the selection process to determine suitable parents for crossover. Within each nondominated set, it is desirable that the solutions remain well spread along the "Pareto-front", the continuous line passing through all of the points in the Pareto-set. This can be accomplished through "diversity preservation" algorithms which order solutions within a nondominated set such that the diversity of the solutions is maintained even when few solutions are selected.

3.2.4. Constraints. In addition to the dominance relationships which are used to determine which individuals are selected for recombination, constraints, which determine the legality of solutions can be applied to the genotype. Solutions may be illegal for several reasons, possibly due to real world constraints or the genotype representation. Constraints can cause a significant problem for GAs since the mutation and crossover operators will in general result in individuals which are not permitted by the constraints. Several approaches to solving this problem exist. These include using genotype representations which do not permit illegal solutions, defining crossover and mutation operators which preserve the legality of solutions, adding penalty terms to the fitness function for illegal solutions and adding a selection penalty to solutions which are not permitted. Constraints can be defined as "hard" or "soft". Redefined genotype representations and crossover and mutation operators result in the hard application of constraints; the GA will never find a solution which violates a constraint. Fitness and selection penalties are known as soft constraints because it is possible, though unlikely, that a solution which violates the constraints may be found.

3.2.5. Genetic Algorithm Software. A number of public domain and General Public License[39] GA codes are available from various research groups. The Non-Dominated Sorting Genetic Algorithm II (NSGA-II)[40] is one such example. The NSGA-II algorithm's strength lies in its elitist selection strategies in selection for survival and selection for breeding. "Elitism" is a technique which has been found to enhance the convergence of multiobjective EAs[41] and operates by retaining a group of optimal solutions between generations, thus reducing the risk that good genetic information might be lost by chance. The NSGA-II algorithm uses a constraint-dominance relationship to determine the selection order of solutions. Individuals are first selected on the basis of constraint validity and then for their nondominance as determined by their objective values. In this way, legal solutions always have a better nondomination rank than illegal solutions. In combination with diversity preservation algorithms, the NSGA-II's constraint-dominance selection strategy ensures that legal solutions which are spread along

the Pareto-front are most likely to be selected to create the next generation.

3.2.6. Genetic Algorithm Parameters and Operation. Several parameters control the crossover and mutation operations. p_c is the probability that crossover occurs between two variables, while η_c is the width of the probability distribution function used in SBX and can be thought of as the "strength" of the crossover operation. Similarly, p_m is the probability of a mutation occurring to each variable, and η_m is the width of the probability distribution function used.

The algorithm operates as follows.

1. A random population is created.

2. The objective functions are evaluated, and the population is sorted based on the nondomination of the individuals.

3. Elitism is introduced by combining a previous population, if available, with the current population and selecting the optimal solutions to form a population for crossover.

4. Selection, crossover, and mutation are performed to generate a new population.

5. The objective function evaluation, population combination, crossover, and mutation are repeated for a number of generations.

The resulting population of the GA should find solutions which are close to the true Pareto-front and are also well distributed across the multiple objectives.

3.3. Integration of Neural Network and Genetic Algorithms. In previous work, GAs have been used for training ANNs[42] and to determine the optimum number of hidden nodes for an ANN.[31] Although examples exist in other fields,[17,43] the use of a GA to "invert" an ANN is, to our knowledge, novel to the field of electroceramic materials design. This combination was mentioned in passing more than 10 years ago in unpublished work in the design of cementitious materials.[12]

Although a well trained ANN is capable of accurately predicting an output corresponding to a set of inputs, the converse is not true. That is, we cannot determine the inputs required to produce a specified output, making ANNs inefficient for solving optimization problems. A "trial and error" method is computationally inefficient, especially when dealing with high dimensional search spaces.

4. GENETIC ALGORITHM IMPLEMENTATION

This section describes the implementation of the "forward" ANN composition-property predictor which is then inverted using a GA. First, we use the ANN described in section 3.1 to develop a system which provides permittivity predictions from composition information.[11] By inverting the permittivity predictor with a genetic algorithm, we can then design materials with specific properties. In this case, we search for materials which exhibit high permittivity. However, since the ANN provides permittivity predictions for any material containing the permitted elements with no regard for the likely accuracy or the stoichiometry of the prediction, we include two further objectives for the optimization. The reliability of permittivity predictions and stoichiometry calculation are used along with the actual permittivity prediction as the three objectives. This section describes the implementation of the objectives, along with the constraints imposed on the solutions. The section concludes by discussing the performance of the algorithm.

4.1. Objective 1: Artificial Neural Network Permittivity Predictor. The first GA objective is the prediction of the relative permittivity of the material. From our materials database,[4] comprising $N = 700$ records of ceramic materials which contain composition, manufacturing, and property data, an ANN has been developed which is capable of predicting the relative permittivity ϵ_r of a material from its composition. The input data consist of the quantity of ions present in the material as provided by the chemical formula. The 700 materials in the database are comprised of 52 different elements (Ag, Al, B, Ba, Bi, Ca, Cd, Ce, Co, Cr, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Ho, In, La, Li, M, Mg, Mn, Mo, Na, Nb, Nd, Ni, O, P, Pr, Sb, Sc, Si, Sm, Sn, Sr, T, Ta, Tb, Te, Ti, Tm, V, W, Y, Yb, Zn, Zr) which determine the number of inputs to the ANN. Oxygen is a ubiquitous element, being present in all materials. Barium, calcium, niobium, and titanium are present in >200 compounds, while tantalum is present in 150. The remaining elements are present in <100 compounds. The mean number of elements per compound is 4.2.

The output of the ANN is the prediction of the permittivity for the requested composition. The materials in the database contain relative permittivities (dimensionless) from 1.7 to 100.0 with a mean of 35.8 and a standard deviation of 22.2. The data set used to train the ANN, which consists of data extracted from the literature, also contains data pertaining to the sintering conditions for the sample. Sintering temperature is recorded for approximately 65%, and the sintering time is available for only 15% of the records in the data set. While processing conditions can have a large effect on the properties of ceramic materials,[2] their inclusion in the ANN would result in a reduction in the number of records available, likely reducing the ANN's performance. Consequently, only the sample's compositional information, that is, the individual quantities of each element, are used as inputs to the ANN.

Ceramic material formulas are commonly scaled for ease of notation. Thus, for example, $Ba_{0.2}Sr_{0.8}TiO_3$ is denoted as $BaSr_4Ti_5O_{15}$. Although these materials are chemically identical, they would be considered different compounds by an ANN. To eliminate this problem, all of the materials are normalized relative to the oxygen content. Using this convention, the material above is expressed as $Ba_{0.07}Sr_{0.27}Ti_{0.33}O$, thus ensuring that all materials, regardless of notation, are treated consistently.

Principal component analysis was used to reduce the input data from the 52 elements found in the data set to 16 dimensions by removing 2% of the variance. The 16 compressed dimensions are used as inputs to the ANN and represent linear combinations of the 52 elements and are not (necessarily) physically meaningful. The best performing ANN was a 3 layer perceptron network,[11] consisting of 16 node input and hidden layers together with a single output node providing the permittivity prediction. No momentum terms were required since training was very fast, the fastest requiring 261 and the slowest 1754 generations before early stopping halted the training process. As discussed in section 3.1, 10-fold cross-validation was used to train multiple networks to ensure an accurate estimation of the generalization error. The 700 records were divided as follows: 70 were used for testing and 315 each for the training and validation data sets. Of the 100 networks trained, the mean $\epsilon_{RRS} = 0.76$

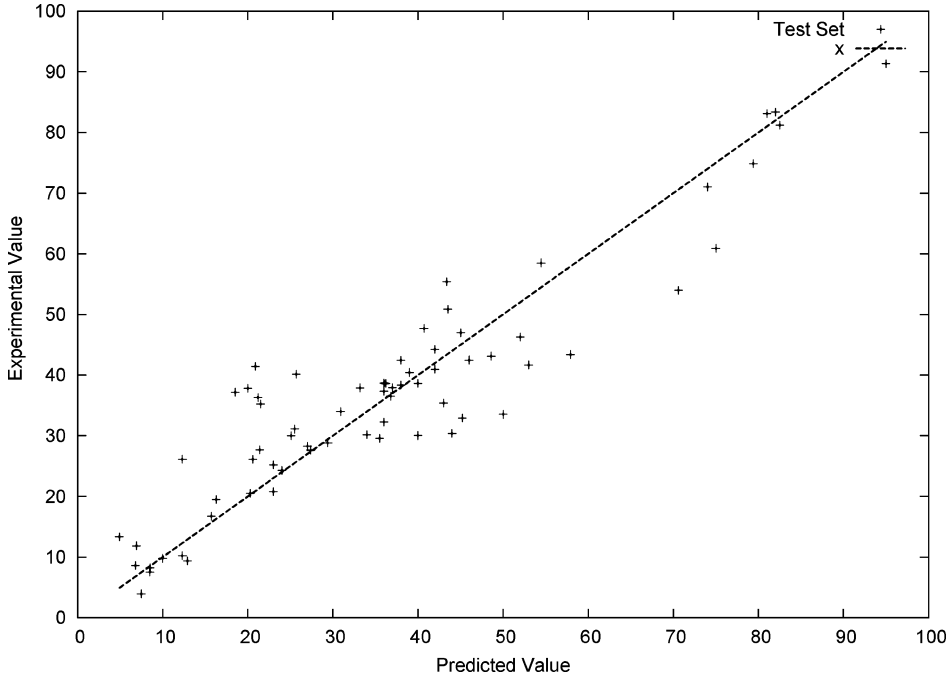

Figure 2. The performance of the back-propagation MLP neural network used to predict the permittivity of the test data set. An ideal straight line with intercept 0 and slope 1 is also shown. The RRS error of the predictions is 0.71.

with a standard deviation of 0.03 and the network selected for this work has an $\epsilon_{RRS} = 0.71$. An RRS error of 1 means that the ANN performs as well as a simple "mean value" predictor; an RRS error of 0 means that the ANN predicts the values in the test data set perfectly. An RRS error of 0.71 therefore means that the ANN predicts 29% better than the simple mean value predictor. The root-mean-square (RMS) difference between the predicted and experimentally measured values for the ANN is 16.0. When compared with the mean value of the permittivities in the data set which is 35.8, we see that the ANN is capable of predicting permittivity values within 50% of the experimentally measured value. Figure 2 illustrates the ANNs prediction accuracy compared with experimental results. An RRS error of 0.71 and RMS prediction accuracy within 50% are reasonable considering the range of materials available in the ANN training data. Additionally, we emphasize that this is a "screening" technique, and the results obtained are used to provide directions for new research. Although more accurate predictions are always desirable, we wish to consider the widest range of materials possible. Hence, the ANN should be sufficiently accurate to determine new material compositions for high throughput manufacture by LUSI.

Members of the FOXD project have access to a Web-based interface to this neural network system,[4] allowing users to obtain permittivity predictions of ceramic materials. A Web-based interface to the database is also available, permitting access to the literature data used for ANN training. The database also contains data from samples produced by LUSI which will be used to improve the performance of future models. Although LUSI is able to rapidly produce large numbers of samples, materials properties testing is not performed in situ but rather at individual project partner sites. The analysis cannot, therefore, be performed at the same speed that LUSI can produce samples, and only a limited number have thus far been fully characterized.

The ANN's predictions are more likely to be accurate when attempting predictions for materials similar to those found in the training data set, and so we have also included a reliability index to assess the accuracy of the ANN predictions. This is described in the following subsection.

4.2. Objective 2: Reliability Index for Network Predictions. The second of the GA objective addresses the "reliability" of the predictions produced by the ANN. The data set used to train the ANN consists of clusters of ceramic compounds that correspond to the types of ceramics that are of current interest to researchers (for example, barium strontium titanate (BST) system).[44] Additionally, particular elements, such as oxygen and titanium, occur more frequently in the database, hence predictions made using these combinations of elements and materials which are similar to those found in the database will be more accurate. This feature is encapsulated via a "reliability index" which assesses the reliability of predictions made using the ANN.[11] The algorithm operates by comparing the input material with the "average material" within the ANN training data set to give a distance vector **R**. Specifically, the algorithm compares the proportions of each element in the input with the mean and standard deviation of the elements in the training data set. The overall reliability is given by the magnitude of the distance vector

$$|\mathbf{R}| = \sqrt{\sum_{i=1}^{N} \left(\frac{x_i - \overline{e_i}}{\sigma_i} \right)^2} \qquad (3)$$

where x_i is the amount of the ion present in the ith material, and $\overline{e_i}$ and σ_i are the mean and standard deviation of the amount of the same element in the ANN training data set, respectively. N is the number of elements present, which is 52 in this case.

The reliability index provides a measure of the distance of the entered material from the average material in the data

set. For any two materials, that with the lower $|\mathbf{R}|$ is likely to be more reliably predicted. A reliability of zero indicates that the quantity of each element present is equal to the mean quantity of that element in the database and the prediction is likely to be reliable. However, the material may not exist in the database since the elements may not be present in the particular combination entered. Nevertheless, the reliability index provides a valuable assessment of the likely accuracy of the prediction and forms the second objective of the GA. When the reliability index is used in combination with the first objective, the ANN permittivity prediction, the GA will search for materials which exhibit high permittivity while remaining "close" to the materials present in the training data set, thus increasing the likelihood that the ANN prediction is accurate.

Although these objectives may produce some excellent theoretical solutions, such a GA contains no information about the physical constraints on the compounds. A real material must be electrically neutral if it is to be manufactured. The third objective directs the search toward electrically neutral materials.

4.3. Objective 3: Excess Charge Calculation. Stoichiometric compounds can be represented using a ratio of well-defined natural numbers. If the quantities of each element, when multiplied by the oxidation state of the element, sum to zero, then the material is electrically neutral, as required for a stable ceramic compound. A compound which contains an excess or deficiency of one or more elements due to defects in the crystal lattice is said to be nonstoichiometric. Although the perovskite crystal structure is very versatile and can tolerate a degree of nonstoichiometry, each defect decreases the stability of the crystal: there is a limit to the amount of nonstoichiometry which can be tolerated before a compound becomes unstable,[45] and therefore stoichiometric or near-stoichiometric material designs are required to be reliably manufacturable. The development of stoichiometric materials is accomplished by the addition of a third objective to the GA which is the minimization of the overall electrical charge carried by the compound.

Since elements can have multiple oxidation states, a charge calculation is performed for each combination, and the one which provides the minimum excess charge is taken to be the excess charge of the compound. Additionally, some materials contain elements in more than one oxidation state. Such materials are less common than materials in which all of the element is in the same oxidation state, and here we do not consider these materials. The presence of elements in multiple oxidation states can also cause electrical conduction, diminishing the dielectric properties. In the charge calculation formula, all of the element is assumed to be in the same oxidation state. The excess charge calculation forms the third objective of the GA: compounds with a lower excess charge are selected in preference to those with a higher excess charge during the GA selection process.

The 52 elements present in the data set, on average, provide two oxidation states which would result in $2^{52} \approx 4.5 \times 10^{15}$ combinations to evaluate, which would take an unfeasibly long time to perform. Since we are only interested in materials which contain four or fewer elements, the excess charge calculation is only performed for materials which contain ten or fewer elements. Thus, the excess charge objective begins to contribute to the search only once the compound has been reduced to a reasonable number of different elements. For materials with more than 10 different elements, the excess charge objective is fixed to a value of 10.

4.4. Genetic Algorithm Implementation. The GA code used in this paper is the Non-Dominated Sorting Genetic Algorithm II (NSGA-II).[40] We use a real representation, a vector of real values which represent the different elements available for materials design. The database used to train the ANN contains 52 different elements, and, therefore, the ANN can accept 52 different elements at the input. Among the 52 input elements are several which are unsuitable for materials design, and so we remove these from the GA's genotype. Recently introduced legislation[46] prevents the use of lead and cadmium in materials research, and so these elements are not present in the genotype. Hydrogen and fluorine are valid inputs for the ANN, since they are present in the training data set; however, we do not plan to use these elements in manufacture, and so they are also absent from the genotype. Finally oxygen is present in all ceramics and has a fixed quantity in the resulting material designs. As explained in section 4.1, the material formulas are normalized with respect to the oxygen content; this means that oxygen can be removed from the genotype since it is a constant quantity in the materials. The resulting genotype consists of a vector which contains 47 elements: 52 are required for the ANN input, while 5 have fixed quantities and so are not present in the GA. When calculating the value of the ANN objective function, the fixed quantities are inserted into the genotype to ensure the correct form of the ANN input vector. Lead, cadmium, hydrogen, and fluorine are entered with zero contribution, and oxygen is inserted with a contribution of one.

4.5. Constraints and Objectives. The GA attempts to optimize three objectives simultaneously:

1. Maximization of the relative permittivity: The relative permittivity ϵ_r as predicted by the neural network is maximized.

2. Minimization of the reliability index: The reliability index, which provides an assessment of the accuracy of the ANN prediction, is minimized to identify reliably predicted materials.

3. Minimization of the overall charge: The overall charge of the compounds searched is minimized, resulting in manufacturable designs of stoichiometric or near-stoichiometric compositions.

Figure 3 shows the (normalized) minimum and maximum values of the quantities of the elements present in the database and gives an indication of the range of each element present. Since ceramic material formulas are often scaled for notational convenience, a consistent representation of the materials is ensured by normalizing the elemental quantity of each compound with respect to the oxygen content. The constraints on the 47 metal ions in the genotype were set to have a minimum of zero and a maximum of one.

The number of elements n_e present in the material is also constrained. Ceramic compound compositions typically consist of six or fewer elements; here, we set a constraint that the GA must obtain results which consist of four elements. This number was chosen in consultation with ceramicists for ease of manufacture. The smallest nonzero element contribution to a material in the database is 0.0095

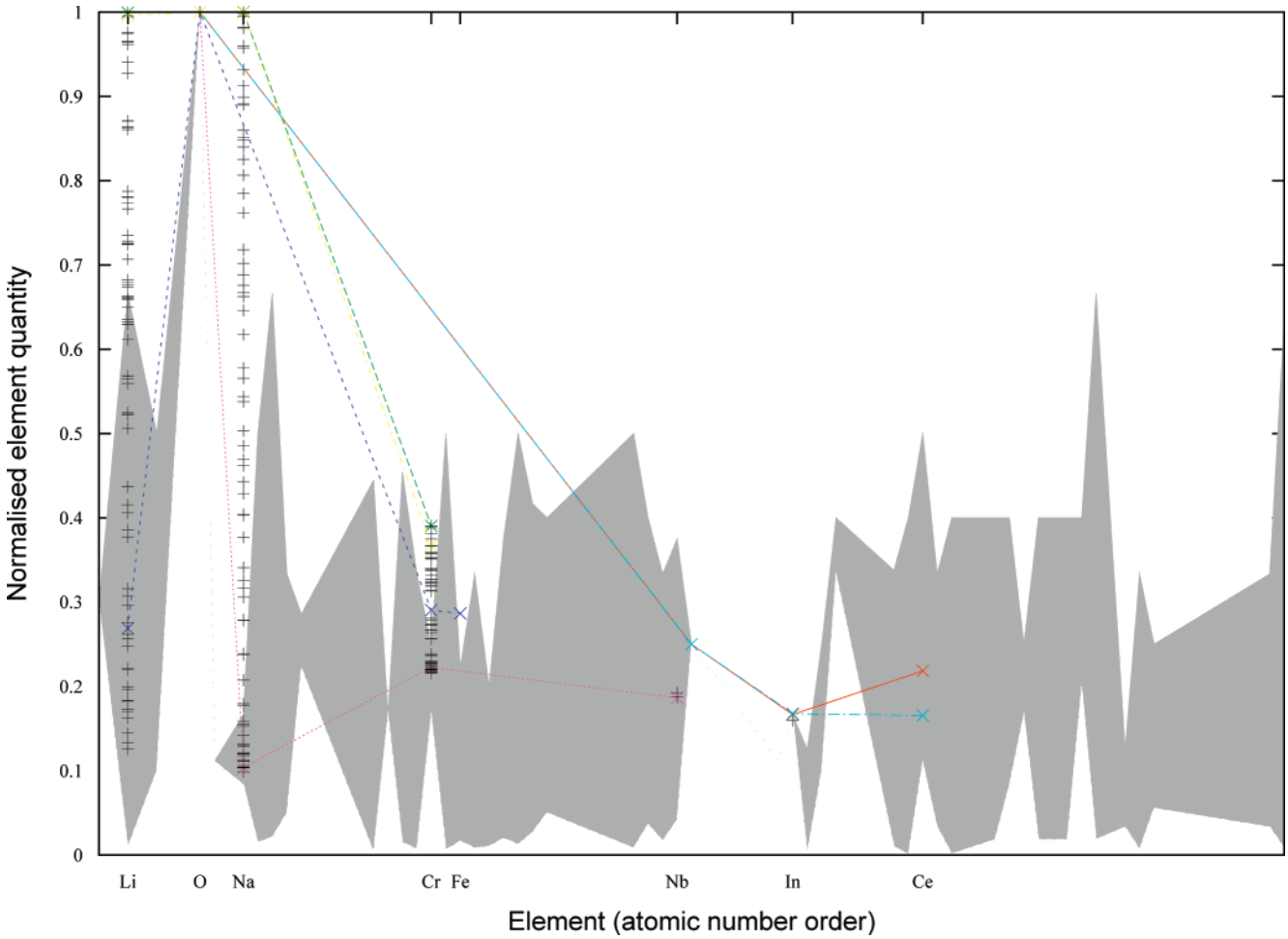

Figure 3. FOXD database statistics and GA results. The shaded area illustrates the range of quantity of each element found in the ceramic materials database. The points show the quantities of each element present in the resulting GA population. The results from the extremes of the final population, shown in Table 1, are highlighted with connecting lines. The quantities of each element within the compounds in the final population and within each material in the database have been normalized with respect to the quantity of oxygen present in each material. Chromium, lithium, and sodium are the most commonly occurring elements in the final population although iron, indium, cerium, niobium, and molybdenum are also present in a number of predicted materials.

(normalized), and so 0.001 would be a reasonable choice to determine the presence of an element. This is a very stringent constraint, and reliable convergence could not be obtained even when running the algorithm for 50 000 generations. Furthermore, the LUSI system which is intended to produce the resulting material predictions can only reliably produce compositions with precision 1−3%[47] for the sample sizes that we are examining. Therefore, we choose 0.01 (1%) as a tolerance value to determine the presence of an element. The number of elements is evaluated by counting the number in the genotype with composition values greater than a threshold of 0.01, elements with a contribution ≤0.01 being ignored. The database contains 10 materials with a contribution of less than 1%, so we are not eliminating a significant region of the search space by choosing this threshold.

The constraints are implemented during the selection process. Designs are selected based on their feasibility (lack of constraint violation) and objective values. For two designs **a** and **b** with number of elements n_e(**a**) and n_e(**b**):

1. If **a** and **b** are both feasible (n_e (**a**) ≤4 and n_e (**b**) ≤4), then **a** dominates **b** in the usual Pareto-optimal sense (eq 2), otherwise

2. If **a** is feasible (n_e (**a**) ≤4) and **b** is not (n_e (**b**) >4), **a** dominates **b**, otherwise

3. If neither **a** nor **b** is feasible (n_e (**a**) >4 and n_e (**b**) >4), if n_e (**a**) <n_e (**b**), then **a** dominates **b**.

In this way, designs are first selected for their feasibility and then for their objective value. A feasible design will always dominate an infeasible design regardless of the objective values.

The resulting 4 elements are combined with the fixed oxygen contribution and scaled by a factor of 3 to obtain a material composition. Thus, for example, if the GA produces a result which contains $Ba_{0.1}Ca_{0.1}Sr_{0.13}Ti_{0.33}$, the resulting material is obtained by adding the O_1 contribution and scaling by 3: $Ba_{0.3}Ca_{0.3}Sr_{0.4}Ti_1O_3$. In future research, we plan to relax the 4 element constraint, thereby permitting materials with a greater number of elements to be explored.

4.6. Running the Evolutionary Algorithm. Deb's code,[40] written in C, was used to develop the GA. The only modifications made were the code additions required to calculate the objectives. The GA was run using a randomly generated starting population of size 100. The initial population contained 100 different materials containing a contribution from each of the 47 elements in the genotype which satisfied the constraints, i.e., the contribution from each element was a randomly generated number between zero and one. A mutation probability rate of $p_m = 0.025 \approx 1/47$ and

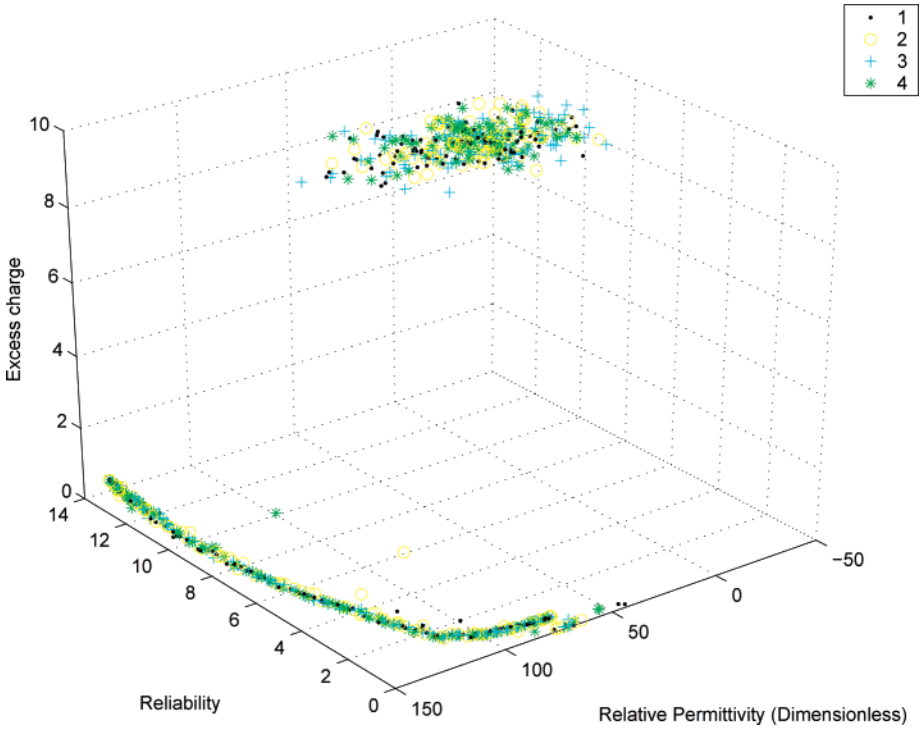

Figure 4. Three-dimensional nondominated set, showing the three objectives being simultaneously optimized. The figure shows the results of four different runs of the GA (dots, crosses, open circles, and asterisks) which are indicated in the legend and demonstrates that the resulting populations have very similar characteristics. As the GA progresses the population moves from the top of the figure, where the initial populations are shown, to the bottom of the figure which shows the final resulting populations. The figure contains negative permittivity predictions present within the initial set of solutions; these are physically meaningless but are due to extrapolation performed by the neural network predictor.

recombination probability of $p_c = 0.9$ were used.[38,40] Optimizations were performed with a range of values to determine the mutation strength and recombination strength indices. η_c and η_m values of 5, 10, and 20 were considered, and a value of 10 for both parameters was found to give consistent convergence with no measurable difference between final populations. The algorithm was executed for 5000 and 20 000 generations with 20 000 generations required for consistent convergence with a run-time of approximately 5 min on a 1.6 GHz PC.

5. RESULTS

Figure 3 shows the elemental compositions from the final GA population. In addition to oxygen, by far the most common elements are chromium, lithium, and sodium although iron, indium, cerium, niobium, and molybdenum are also present, albeit only in a small number of materials.

The results from four separate GA runs are shown in Figures 4 and 5. Figure 4 shows the evolution from the initial population of solutions to the nondominated sets in terms of the permittivity, reliability index, and excess charge objectives. The first is maximized, while the last two are minimized. The figure contains some negative values for the permittivity which are physically meaningless. These values occur within the initial population of randomized solutions, before the reliability and stoichiometry objectives are used to optimize the population toward realistic, manufacturable material compositions.

Figure 5 shows an enlarged view of the resulting populations; the trade-offs between all three objectives are visible. The figure effectively consists of three different sections. The left-hand side of the figure shows a trade-off between reliability and excess charge. Initially, the excess charge decreases as the reliability becomes worse; however, the excess charge eventually begins to increase again, indicating predicted compounds which have poor charge and reliability attributes.

The central section indicates a trade-off between permittivity and reliability with the excess charge remaining constant. Compounds with higher permittivities have a worse (higher) reliability index since these solutions correspond to compounds which are unlike most of those stored in the database and used to train the ANN.

Finally, on the right-hand side, the permittivity and excess charge trade-off, while the reliability remains constant. In general, the charge increases (gets worse) as the permittivity increases. However, various solutions in the nondominated set exhibit near-zero charge along with high permittivity values (ϵ_r 120−140).

Table 1 shows some of the compounds predicted within the final population. Table 2 lists both hand-selected materials from the final GA population (a) as well as similar materials residing in the database (b). The quantities of each element have been scaled by a factor of 3 to obtain the real chemical formula. The constituent elements are listed in alphabetical order and not in ABO_3 perovskite form. This is because the site occupation cannot be determined until the materials are manufactured and crystallographic analysis is used to determine the structure. Table 1 shows a selection of compounds with extreme objective values from the final population. Examination of these materials provides a good qualitative understanding of the trade-offs. Two of the results

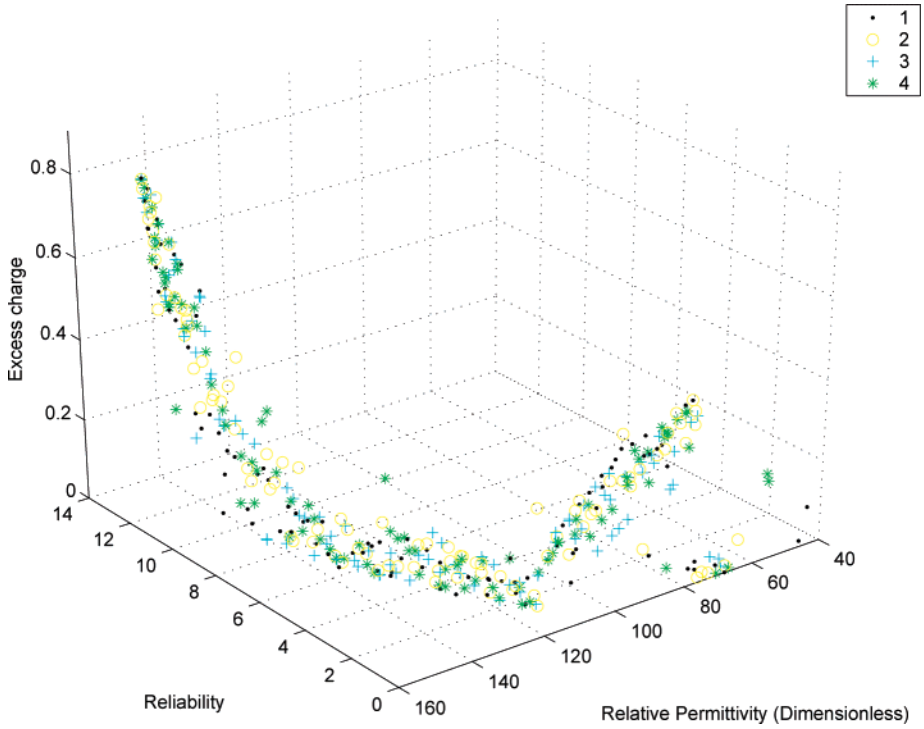

Figure 5. An enlarged view of Figure 4 containing four three-dimensional nondominated sets (dots, crosses, open circles, and asterisks) which are indicated in the legend and illustrating the three objectives being simultaneously optimized. The four resulting populations are all extremely similar, confirming that the final populations have very similar characteristics and contain similar materials. Due to the stochastic nature of the search method, the resulting populations are unlikely to be identical.

Table 1. Two Records from Each of the Objective Extremes of the Final Population[a]

compound	permittivity	reliability	excess charge
$Cr_{1.1}(II)Li_3(I)Na_3(I)O_3(-II)$	**146.17**	13.05	2.2
$Cr_{1.2}(II)Li_3(I)Na_3(I)O_3(-II)$	**146.24**	13.13	2.6
$Ce_{0.7}(III)In_{0.5}(III)Mo_{0.8}(III)O_3(-II)$	44.07	**0.00**	0.0
$Cr_{0.7}(IV)Na_{0.3}(I)Nb_{0.6}(V)O_3(-II)$	68.17	**0.01**	0.0
$Cr_{0.9}(III)Fe_{0.8}(III)Li(I)_{0.8}O_3(-II)$	98.83	1.99	**0.0**
$Ce_{0.5}(IV)In_{0.5}(III)Mo_{0.8}(III)O_3(-II)$	44.12	0.25	**0.0**

[a] Two have the highest permittivity values, two have the best reliability values, and two have the minimum excess charge. The tradeoffs between the objectives are evident.

display the highest predicted permittivity, two have the best reliability, and the remaining two contain the best charge attributes. Generally, these materials are optimal in one of the three objectives, and their remaining two attributes are poor. However, the fourth compound displays good reliability and minimal excess charge, while the permittivity is average. In another case, the fifth compound exhibits high permittivity, minimal excess charge, and poor reliability. The materials outlined in this table are the most unusual of the final population, containing some of the less common elements found in the predicted compounds.

Table 2 shows hand-selected materials from the final GA population (a) along with similar results from the database (b). The materials provided in Table 2(a) combine the best permittivity, reliability, and charge attributes. Since the excess charge calculation must be near zero for a compound to be manufacturable, the selected materials were first chosen to have extremely small excess charge. Then, materials with good reliability and high permittivity predictions were chosen. The permittivity and reliability of these materials are not as good as for the materials shown in Table 1; however, these results combine a high permittivity prediction with good reliability. These results illustrate one of the key benefits of the multiobjective evolutionary algorithm approach to materials design. "Reliable" materials are most similar to those found in the database, are likely to have accurate permittivity predictions, and therefore serve to

Table 2. (a) Human Selected Material Designs of Interest from the Optimized GA Population[a] and (b) Selection of Chromium, Lithium, and Sodium Containing Materials from the Database[b]

(a)

compound	permittivity	reliability	excess charge
$Cr_{0.8}(III)Li_2(I)Na_{1.6}(I)O_3(-II)$	135.63	6.41	0
$Cr_1(III)Li_{2.6}(I)Na_{0.5}(I)O_3(-II)$	112.69	3.4	0
$Cr_1(IV)Li_{1.6}(I)Na_{0.4}(I)O_3(-II)$	112.13	2.35	0.01
$Cr_{0.7}(V)Na_{0.5}(I)Nb_{0.6}(III)O_3(-II)$	73.85	0.71	0.01
$Cr_{0.7}(VI)Li_{1.3}(I)Na_{0.5}(I)O_3(-II)$	114.28	1.62	0.01
$Cr_{0.7}(V)Li_{1.5}(I)Na_{0.9}(I)O_3(-II)$	123.48	3.19	0.01

(b)

compound	permittivity
$CrNbO_4$	22[48]
$CrTaO_4$	9.7[48]
$Pb_{0.75}Ca_{0.25}(Cr_{0.5}Nb_{0.5})O_3$	48[49]
$Pb_{0.5}Ca_{0.5}(Cr_{0.5}Nb_{0.5})O_3$	43[49]
$Pb_{0.5}Ca_{0.5}Na_{0.25}Nb_{0.75}O_3$	72[50]
$Sm_{0.5}Na_{0.5}TiO_3$	80[51]
$LiNb_3O_8$	34[51]
$CaLi_{0.33}Nb_{0.66}O_3$	29.6[52]

[a] These materials have been hand-selected as possible candidates for manufacture. Materials with near-zero excess charge were selected to ensure that the compounds were near- or fully-stoichiometric; this set was further reduced by selecting materials with a good combination of high permittivity prediction and good reliability. [b] These materials can be compared to selected materials from the optimized GA population shown in Tables 1 and 2(a).

validate the technique of using a GA to invert the ANN. High permittivity materials are less reliably predicted and thus contain the most interesting materials, opening new research directions. Multiobjective evolutionary algorithms result in a population of solutions which can be hand-selected by domain experts to obtain candidates for manufacture.

6. DISCUSSION

The hand-selected results shown in Table 2(a) have been compared to records contained in the ceramic materials database. Table 2(b) displays materials from the database which contain chromium, the most prevalent element in the GA results. Lead is present in two of the materials but is not found in the GA results because it was eliminated from the genotype owing to safety legislation,[46] as previously discussed. Two of the materials from the database contain niobium in addition to chromium, and several compounds containing both elements are present in the optimized GA population; an example is shown in Table 2(a).

The permittivities of the database materials are not as high as those predicted for the GA results. However, one of the GA predictions, $Cr_{0.7}Na_{0.5}Nb_{0.6}O_3$, has a relative permittivity of 73.85, much closer to the database material $Pb_{0.75}Ca_{0.25}(Cr_{0.5}Nb_{0.5})O_3$, which has an experimentally measured permittivity of 48. The reliability index of the predicted material is also significantly lower than the other hand-selected materials, meaning that the permittivity prediction is likely to be accurate. By contrast, the predicted materials in Table 2(a) combine high permittivity with good reliability and are possible candidates for laboratory manufacture and measurement.

In a perovskite material, the element(s) on the A site are +2 ions and the element(s) on the B site are +4 ions; giving a neutral material when combined with three O^{2-} ions. Examination of the compounds shown in Table 1 reveals that none of the materials conform with the $A1_xA2_{1-x}B1_yB2_{1-y}O_3$ perovskite formula. However, the versatility of the perovskite structure means that it is very difficult to determine whether a material will crystallize in the perovskite structure prior to synthesis. Although not done here, we could impose further constraints on the GA to promote the selection of materials with this structure although this may prove to over constrain the discovery process. Additionally, the "Megaw tolerance"[53] compares the ionic radii of elements to determine the likelihood of perovskite structure formation and could be included as an additional constraint.

The charge calculation is currently performed using many possible oxidation states of the elements. Some oxidation states are more stable than others, so some of the compounds predicted by the GA may be chemically unstable. To alleviate this problem, we could in the future improve the reliability index algorithm by weighting the GA search space in favor of more stable compounds.

The quality factor, '*Q*', mentioned in section 2 is also an important property for dielectric resonators. The addition of '*Q*' factor prediction and optimization to the materials design algorithm presented here is a logical modification to the algorithm and is left as a subject for further research. With such a modification, we would be able to develop materials predictions which simultaneously optimize permittivity and '*Q*' factor properties.

7. CONCLUSIONS

We have shown in this paper that it is possible to design new materials using Baconian methods. Through combination of a neural network trained with data gleaned from the literature and an evolutionary algorithm we have developed a powerful tool for the materials design problem. Moreover, any number of constraints can be included in order to explore the compositional search space in arbitrary ways. Materials with a lower reliability index are similar to existing materials and may be useful for improvement of already well understood materials. Materials predicted with less reliability are unlike materials contained within the database; while the neural network predictions are likely to be less accurate, such materials compositions are a possible source of innovative designs.

Three objectives were used. Two pertain to physical properties of interest—the permittivity and the overall charge—while the reliability index provided an indication of the accuracy of the results found. The use of a multiobjective genetic algorithm resulted in a final population containing a nondominated set of potential designs which primarily conflict in permittivity and reliability. Human selection is used to identify compounds of modest permittivity, but very good reliability, along with new compounds exhibiting high permittivity, which are candidates for future manufacture and analysis. The development of more sophisticated constraints may help guide the evolutionary process to more practical designs. Of particular importance is the satisfaction of stoichiometric constraints; this is crucial not only here but also in the general class of problems where we are designing chemical compounds.

A full evaluation of the predictive capabilities of the technique presented in the current paper can only emerge from a combinatorial approach, such as that being pursued by the FOXD project using LUSI, in order to program the synthesis and testing of large numbers of proposed materials. Synthesis and characterization of the materials designs presented here "closes the loop" of the materials discovery cycle and represents work in progress at the present time. The resulting data can be used to improve the overall predictive performance of the model, thus permitting more accurate GA searches to commence. An ultimate aim is to be able to steer automated searches through the compositional search space to discover novel materials designs.

ACKNOWLEDGMENT

We are grateful to Simon Clifford for discussions and suggestions. We would also like to express our thanks to our colleagues in the FOXD project: Julian Evans, Shoufeng Yang, Lifeng Chen, and Yong Zhang at University College London and to John Kilner, Neil Alford, Sarah Fearn, Robert Pullar, and Jeremy Rossiny at Imperial College, London. Our research is funded by the EPSRC project "Discovery of New Functional Oxides by Combinatorial Methods" (GR/S85269/01).[10]

REFERENCES AND NOTES

(1) Kingery, W. D.; Bowen, H. K.; Uhlmann, D. R. *Introduction To Ceramics*; John Wiley and Sons Ltd.: 1976.
(2) Moulson, A. J.; Herbert, J. M. *Electroceramics*; John Wiley and Sons Ltd.: 2003.

(3) Hanak, J. J. The 'multiple sample concept' in materials research: synthesis, compositional analysis and testing of entire multicomponent systems. *J. Mater. Sci.* **1970**, *5*, 964−971.
(4) Scott, D. J.; Manos, S.; Coveney, P. V.; Rossiny, J. C. H.; Fearn, S.; Kilner, J. A.; Pullar, R. C.; Alford, N. McN.; Axelsson, A.-K.; Zhang, Y.; Chen, L.; Yang, S.; Evans, J. R. G.; Sebastian, M. T. Functional Ceramics Materials Database: An online resource for materials research. *J. Chem. Inf. Model.* **2008**, *48*, 449−455.
(5) Bacon, F. Novum Organum. In *The Philosophical Works of Francis Bacon*; Routeledge: London, 1905.
(6) Popper, K. R. *Conjectures and Refutations*; Routledge and Kegan Paul plc: New York, U.S.A., 1963.
(7) Evans, J. R. G.; Edirisinghe, M. J.; Coveney, P. V.; Eames, J. Combinatorial searches of inorganic materials using the ink-jet printer: science, philosophy and technology. *J. Eur. Ceram. Soc.* **2001**, *21*, 2291−2299.
(8) Coveney, P. V.; Fowler, P. W. Modelling biological complexity: a physical scientist's perspective. *J. R. Soc. Interface* **2005**, *2*, 267−280, 10.1098/rsif.2005.0045.
(9) Harvey, M. J.; Scott, D.; Coveney, P. V. An integrated instrument control and informatics system for combinatorial materials research. *J. Chem. Inf. Model.* **2005**, *46*, 1026−1033.
(10) *Functional Oxide Discovery*; EPSRC Grant: GR/S85269/01. http://www.foxd.org (accessed November 11, 2007).
(11) Scott, D. J.; Coveney, P. V.; Kilner, J. A.; Rossiny, J. C. H.; Alford, N. McN. Prediction of the functional properties of ceramic materials from composition using artificial neural networks. *J. Eur. Ceram. Soc.* **2007**, *27*, 4425−4435.
(12) Coveney, P. V.; Fletcher, P.; Hughes, T. L. Using Artificial Neural Networks to Predict the Quality and Performance of Oil-Field Cements. *AI Mag.* **1996**, *17*, 41−53.
(13) Gillet, V.; Khatib, W.; Willett, P.; Fleming, P.; Green, D. Combinatorial Library Design Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2002**, *42*, 375−385.
(14) Farrusseng, D.; Klanner, C.; Baumes, L.; Lengliz, M.; Schüth, F. Design Discovery Libraries for Solids Based on QSAR Models. *QSAR Comb. Sci.* **2005**, *24*, 78−92.
(15) Brown, N.; McKay, B.; Gasteiger, J. A novel workflow for the inverse QSPR problem using multiobjective optimization. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 333−341, 10.1007/s10822−006-9063−1.
(16) Burden, F. R.; Rosewarne, B. S.; Winkler, D. A. Predicting maximum bioactivity by effective inversion of neural networks using genetic algorithms. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 127−137.
(17) Yang, T.; Lin, H.-C.; Chen, M.-L. Metamodeling approach in solving the machine parameters optimization problem using neural network and genetic algorithms: A case study. *Rob. Comput.-Integr. Manuf.* **2006**, *22*, 322−331.
(18) Sudarsana Rao, H.; Ghorpade, V. G.; Mukherjee, A. A genetic algorithm based back propagation network for simulation of stress-strain response of ceramic-matrix-composites. *Comp. Struct.* **2006**, *84*, 330−339.
(19) Rodemerck, U.; Baerns, M.; Holena, M.; Wolf, D. Application of a genetic algorithm and a neural network for the discovery and optimization of new solid catalytic materials. *Appl. Surf. Sci.* **2004**, *223*, 168−174.
(20) Setter, N. Electroceramics: looking ahead. *J. Eur. Ceram. Soc.* **2001**, *21*, 1279−1293.
(21) Terfloth, L.; Gasteiger, J. Neural Networks and Genetic Algorithms in Drug Design. *Drug Discovery Today* **2001**, *6*, 102−108.
(22) Setter, N.; Waser, R. Electroceramic materials. *Acta Mater.* **2000**, *48*, 151−178.
(23) Wersing, W. Microwave Ceramics for Resonators and Filters. *Curr. Opin. Solid State Mater. Sci.* **1996**, *1*, 715−731.
(24) Randall, C. A.; Newnham, R. E.; Cross, L. E. History of the First Ferroelectric Oxide, $BaTiO_3$. *Web Source* 2004 (accessed November 11, 2007).
(25) *London University Search Instrument.* http://www.materials.qmul.ac.uk/research/facilities/lusi/index.php (accessed November 11, 2007).
(26) Caruthers, J. M.; Lauterbach, J. A.; Thomson, K. T.; Venkatasubramanian, V.; Snively, C. M.; Bhan, A.; Katare, S.; Oskarsdottir, G. Catalyst design: knowledge extraction from high-throughput experimentation. *J. Catal.* **2003**, *216*, 98−109.
(27) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: 1995.
(28) Breiman, L. Statistical Modeling: The Two Cultures. *Stat. Sci.* **2001**, *16*, 199−215.
(29) Basheer, I. A.; Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. *J. Micro. Meth.* **2000**, *43*, 3−31.
(30) Gotlib, V. A.; Sato, T.; Beltzer, A. I. Neural computing of effective properties of random composite materials. *Comp. Struct.* **2001**, *79*, 1−6.
(31) Alsultanny, Y. A.; Aqul, M. M. Pattern recognition using multilayer neural-genetic algorithm. *Neurocomputing* **2003**, *51*, 237−247.
(32) Mitchell, T. M. *Machine Learning*; McGraw-Hill: 1997.
(33) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. B* **1974**, *36*, 111−147.
(34) Coveney, P. V.; Highfield, R. *Frontiers of Complexity*; Ballentine Books: New York, U.S.A., 1995.
(35) Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, U.S.A., 1975.
(36) Goldberg, D. E. *Genetic Algorithms in Search, Optimisation and Machine Learning*; Addison Wesley Longman Inc.: 1989.
(37) Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*; World Scientific Publishing: Singapore, 1999.
(38) Deb, K.; Agrawal, R. B. Simulated Binary Crossover for Continuous Search Space. *Complex Syst.* **1995**, *9*, 115−148.
(39) *GPL*. http://www.gnu.org/copyleft/gpl.html (accessed November 11, 2007).
(40) Deb, K.; Agrawal, S.; Pratap, A.; Meyarivan, T. A Fast and Elitist multi-objective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Computation* (*IEEE-TEC*) **2002**, *6*, 182−197.
(41) Zitzler, E.; Deb, K.; Thiele, L. Comparison of Multiobjective Evolutionary Algorithms on Test Functions of Different Difficulty. *Proceedings of the 1999 Genetic and Evolutionary Computation Conference. Workshop Program*, Orlando, Florida, 1999; pp 121−122.
(42) van Rooij, A.; Jain, L. C.; Johnson, R. P. *Neural Network Training Using Genetic Algorithms*; World Scientific: River Edge, NJ, 1996.
(43) Kumar, A.; Hand, V. Feasibility of Using Neural Networks and Genetic Algorithms To Predict and Optimize Coated Paper and Board Brightness. *Ind. Eng. Chem. Res.* **2000**, *39*, 4956−4962.
(44) Ioachim, A.; Toacsan, M. I.; Banciu, M. G.; Nedelcu, L.; Dutu, A.; Antohe, S.; Berbecaru, C.; Georgescu, L.; Stoica, G.; Alexandru, H. V. Transitions of barium strontium titanate ferroelectric ceramics for different strontium content. *Thin Solid Films* **2007**, *515*, 6289−6293.
(45) Lee, S.; Woo, S. K.; Lee, K. S.; Kim, D. K. Mechanical properties and structural stability of perovskite-type, oxygen-permeable, dense membranes. *Desalination* **2006**, *193*, 236−243.
(46) European Parliament. *Directive 2002/95/EC of the European Parliament and of the Council of 27 January 2003 on the restriction of the use of certain hazardous substances in electrical and electronic equipment*; Technical Report; 2003.
(47) Zhang, Y.; Chen, L.; Yang, S.; Evans, J. R. G. Control of particle segregation during drying of ceramic suspension droplets. *J. Eur. Ceram. Soc.* **2007**, *27*, 2229−2235.
(48) Maeda, M.; Yamamura, T.; Ikeda, T. Dielectric Characteristics of Several Complex Oxide Ceramics at Microwave Frequencies. *Proceedings of the 6th Meeting of Ferroelectric Materials and Their Applications, Kyoto*, Department of Applied Physics Faculty of Engineering, Tohoku University, Sendai 980, Japan, 1987; pp 76−79.
(49) Kato, J.; Kagata, H.; Nishimoto, K. Dielectric Properties of (PbCa)(MeNb)O_3 at Microwave Frequencies. *Jpn. J. Appl. Phys.* **1992**, *31*, 3144−3147.
(50) Chen, X. M.; Lu, X. J. Characterization of $CaTiO_3$-modified $Pb(Mg_{1/3}Nb_{2/3})O_3$ dielectrics. *J. Appl. Phys.* **2000**, *87*, 2516−2519.
(51) Belous, A. G.; Ovchar, O. V. Temperature compensated microwave dielectrics based on lithium containing titanates. *J. Eur. Ceram. Soc.* **2003**, *23*, 2525−2528.
(52) Tong, J. X.; Zhang, Q. L.; Yang, H.; Zou, J. L. Low-temperature firing and microwave dielectric properties of $Ca[(Li_{0.33}Nb_{0.67})_{0.9}Ti_{0.1}]O_{3-\delta}$ ceramics with LiF addition. *Mater. Lett.* **2005**, *59*, 3252−3255.
(53) Megaw, H. D. *Ferroelectricity in Crystals*; Methuen & Co. Ltd.: London, 1957.

CI700269R