Computational Materials Science 150 (2018) 381–389

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Approach and algorithm for generating appropriate doped structures for high-throughput materials screening

Mingming Zhang, Xiaoyu Yang*

Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Inequivalent doped structures
High-throughput screening
First principle simulations
MatCloud

ABSTRACT

As the extensive use of first-principles Density Functional Theory (DFT) simulations, using DFT for high-throughput screening to predict the desirable doped structures that are physically stable with optimal properties becomes common. Usually, the challenge of running doping calculation is how to obtain inequivalent doped structures as input for DFT simulations to find the desirable doped structure(s). The current practice of substitutional doping is mainly based on experience to use one or more dopant atoms to replace target atoms to be substituted. Using this manual approach to produce all inequivalent doped structures based on expertise is tedious, and the results are usually incomplete. To address this need, we propose a "doping-filtering" collaboratively working approach and develop associated high-throughput computational algorithms to obtain inequivalent doped structures for substitutional doping-based high-throughput screening effectively. A computational time benchmark matrix table of using this approach to obtain inequivalent doped structures for different doping concentrations is also given. The algorithm is integrated into a high-throughput computational material infrastructure named MatCloud. It has been demonstrated in the study case of doping Ni, Co, Ti and Sc into Zr_2Fe that the approach and algorithm are effective in reducing the computational time in obtaining inequivalent doped structures.

1. Introduction

In material science, doping is an important way of material modification [1]. With the widespread use of the first-principles Density Functional Theory (DFT) [2] calculations in the field of material design in recent years [3–6], more and more material researchers start to use the computer for simulating materials doping process to explore the enhanced materials properties. The purpose of doping is to find out the best doping concentration and substitutional positions of a doped structure that presents stable structures or optimal properties. For this purpose, all the possible doped structures need to be obtained firstly according to the given doping concentration and then select out the desirable ones. In other words, scientists need to produce all possible doped structures, calculate properties of each doped structure and then find out which doping concentration and substitutional positions make doped structures the best.

Currently, the substitution position in a doped structure is often determined by the experimental refinement data and the researcher's personal experience. In most DFT calculation based doping research, researchers usually choose limited sites in crystal cell for substitutional doping [7]. They use Material Studio or other tools to complete the doping process, and then export the doped structures into a file for DFT calculation. This approach is restricted to a limited number of doped structures and certain experiment. Using high-throughput calculation in substitutional doping can help analyze a large number of samples and screening the best properties [8–10]. High-throughput screening is an effective approach to discover new materials, but the main challenge is how to obtain appropriate doped structures effectively.

Literature review shows that there is little attention paid to doping calculation modeling. Some doping modeling was implemented by using mcsqs or gensqs in ATAT package [11–13], but they are only for binary alloy. In most cases, a convenient approach is required to help to generate those doped structures without equivalent ones.

In order to use high-throughput calculation technologies to address the substitutional doping calculation issues, we propose a general substitutional doping calculation model as follows: for a given crystal structure, using different dopant elements $X(x_1,x_2, \ldots x_i)$ to replace different target element species $Y(y_1,y_2,\ldots y_i)$ that contains certain number of atoms Z $(z_1,z_2,\ldots z_i)$ that involves certain series of sites U $(u_1,u_2,\ldots u_i)$ to reach different doping concentrations $V(v_1,v_2,\ldots v_i)$. Note

* Corresponding author at: Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: kxy@cnic.cn (X. Yang).

https://doi.org/10.1016/j.commatsci.2018.04.017
Received 23 October 2017; Received in revised form 20 March 2018; Accepted 8 April 2018
Available online 21 April 2018
0927-0256/

one kind of target element can be substituted by either single dopant atom or multiple different dopant atoms at the same time. The high-throughput calculations can happen at "loop through" of X, Y, Z, U and V. In order to reach a certain concentration, an appropriate size of supercell need to be created. For a series of different doping concentrations, usually, a minimal size of supercell for each concentration need to be created. In our paper, we mainly investigate the case of using one dopant atom (X is fixed) to replace one target species (Y is fixed) that contains specific number of atoms that involves a series possible sits to reach one concentration (V is fixed).

One challenge in substitutional doping calculation is to remove equivalent structures. Substitutional doping of one dopant atom into a cell can result in a large number of possible doped structures generated. In these doped structures, there are a lot of equivalent structures which need to be removed. Two crystal structures being equivalent means: by some rotation or shift symmetry operations, one structure can transform to another structure which has same properties. For example, Substitutional doping into a 16-atom Al cell, the number of possible doped structures is 12870, but the number of inequivalent doped structures is 151 after removing equivalent structures. In order to reduce computational time, we need to remove all the equivalent structures for the subsequent high-throughput screening using DFT.

There are two important procedures for doping calculation: (i) Doping. Generating all possible doped structures for a given crystal structure and doping concentration, that is, to ensure we do not drop any possible structure in result set. (ii) Filtering. Removing a large number of equivalent structures in doping result set, that is, to ensure each doped structure in output set is unique, any other structure in this set is not equivalent to itself.

In this paper, we propose a "doping-filtering" collaboratively working approach to reduce computational time for obtaining inequivalent doped structures. The approach and algorithms are integrated into a high-throughput computational infrastructure, namely, MatCloud. This approach is proposed against the traditional "doping first, filtering second" approach (we call "doping-filtering" working separately approach), the idea of which is combining doping procedure and filtering procedure in a loop process. In comparison to "doping-filtering" working separately approach, it has been tested in three different test cases that this approach can effectively reduce the computational time in obtaining inequivalent doped structures.

The paper is structured as follows: Section 2 discusses the proposed approach and algorithms. Section 3 is results analysis. Section 4 is a case study of using the approach. Section 5 is discussions. Section 6 is conclusions.

2. Methods

2.1. Background of generating inequivalent doped structures for screening

Doping simply means adding some extra atoms into crystal cell to change properties of this crystal. The additional atom can be called dopant atom and the atom to be replaced can be called target atom. How many atoms to be replaced is determined by the required doping concentration. Usually, a supercell size needs to be determined to reach the doping concentration.

The goal of our algorithm is to generate all possible doped structures according to the given crystal structure and doping concentration. In most cases, the algorithm need choose some sites in the structure that contains target atom, then replace these target atom(s) with dopant atom to generate one possible doped structure. There are many replacement combinations we can choose, and there are many possible doped structures will be generated. For example, if we replace two Cl atoms with Sc atom, three possible doped structures can be as Fig. 1:

Theoretically, in the case of using one element species to replace another species element in the cell, for the given doping concentration, the number of total possible doped structures N is decided by two factors:

(1) The total number of target atoms in the initial structure, we denote as n.
(2) The doping concentration or the number of target atoms to be substituted by dopant atoms, we denote as m.

The total possible doped structures N can be calculated by following formula:

$$\mathrm{N}= \mathrm{C}_n^m = \frac{n!}{m!(n-m)!} \tag{1}$$

The possible doped structures generated by using dopant atom to replace the possible positions contain many equivalent structures, which need to be removed for the following subsequent screening process. The equivalent structures are expressed differently only in crystal data structure, but with the same properties. As mentioned previously, by some symmetry operation, one structure can transform to another, which have same properties. Removing equivalent structures can reduce the number of doped structures hence reduce the computational time. Throughout the paper, structures that are generated by replacing dopant atoms are called doped structures, and doped structures where equivalent structures are removed are called inequivalent doped structures.

2.2. The overview of the proposed doping approach for generating appropriate inequivalent doped structures

The proposed approach includes following 3 major procedures (see Fig. 2): (i) Supercell creation approach determination. (ii) Doping route determination. (iii) "Doping-Filtering" collaborative work to generate all possible doped structures. It contains doping process and filtering process. Doping process is responsible for generating all possible doped structures, whereas filtering process is to remove equivalent structure in the result that doping process generated. The proposed "doping-filtering" collaboratively working approach means the possible doped structures generating process and equivalent structures removal process work together in a loop process, rather than generating all possible doped structures first and removing all equivalent structures second. Also, the route for supercell creation and doping need to be considered.

(1) Supercell creation approach determination.

In most case, a supercell of appropriate size needs to be created to make sure the given doping concentration can be reached. We developed an algorithm that can help to automatically determine the appropriate supercell size for the given doping concentration. Take an example that doping a 4-atom Al at 5% doping concentration. In order to make sure that a minimum size of supercell to be created to reach the doping concentration, we need to replace 1 atom in the whole supercell which has X atoms. The number X can be obtained by 1/5% = 20, which means we need to create 5 times supercell that contains 20 atoms per cell.

Choosing an appropriate size of a supercell is crucial in doping calculation. According to formula 1, we know that bigger size of the cell demands more computational time and can produce more inequivalent doped structures.

It is suggested to avoid creating a big size supercell for a particular site occupancy (i.e. doping concentration) like 0.9%, etc. We recommend using normal site occupancy (or doping concentration) such as 5%, 10% for doping calculation and use the produced properties values to fit a model to derive associated properties at that particular doping concentration.

(2) Doping route determination.