재료공학 석사학위논문

Data mining for next-generation high-κ candidates;

automation of band gap and permittivity calculation

차세대 high-κ 물질 탐색을 위한 데이터 마이닝;

밴드갭 및 유전율 계산 자동화

2013 년 8 월

서울대학교 대학원

공과대학 재료공학부

임강훈

Data mining for next-generation high-κ candidates; automation of band gap and permittivity calculation

차세대 high-κ 물질 탐색을 위한 데이터 마이닝; 밴드갭 및 유전율 계산 자동화

지도교수 : 한승우

이 논문을 재료공학 석사학위논문으로 제출함

2013 년 5 월

서울대학교 대학원

공과대학 재료공학부

임강훈

임강훈의 석사학위논문을 인준함

2013 년 6 월

위 원 장 황 철 성 (인)

부위원장 조 경 재 (인)

위 원 한 승 우 (인)

Abstract

Data mining for next-generation high-κ candidates; automation of band gap and permittivity calculation

Kanghoon Yim

Materials Sciences and Engineering

The Graduate School

Seoul National University

Making a perfect handbook for desired properties of all existing materials is always a dream work for materials scientists. Remarkable advances in computing power and first-principles techniques now present a good opportunity for building a vast theoretical database of material properties. Fast high-throughput *ab initio* calculations with good accuracy can be achieved by aligning proper computational methods with a sophisticated automatic procedure. By minimizing human interventions, even though every

procedural step should be carefully verified and tested in advance, massive data of material properties could be obtained within a reasonable computational cost. The materials map can be obtained from the new database and it would enable us to identify a material with unexpected property that can provide a breakthrough in various applications. As our first attempt to make such a materials map, we developed an automation code for computing band gap and static dielectric constant of various oxides. The code automatically generates input structure data from ICSD[1], and VASP[2] is used as core engine for the *ab initio* calculations. Human interventions are minimized by fully automated procedures and reliability of the computed property was carefully examined in the first stage. We employ HSE06 and LDA functional for calculating band gap and dielectric constant respectively, as they are known to produce good results for each property and a special k-points searching method is suggested for finding band edge positions. Various optimization procedures precede the core computations, for instance, filtering redundant data from ICSD and optimizing computational parameters to reduce the cost without compromising the accuracy. As the result, we manage to calculate properties of more than 1,000 oxides and we

identified some interesting materials with unexpected properties of large band gap and high dielectric constant.

Keywords : data mining, high-k, band gap, permittivity, first-principles calculation

Student Number : 2011-23320

Contents

Abstract

Contents

List of table and figures

1. Introduction

1.1 New era of materials design ····· 1

1.2 Materials design for high-k dielectrics ····· 2

2. Computational theory and methods

2.1 DFT and beyond DFT

2.1.1 Density functional theory ····· 6

2.1.2 Beyond DFT ····· 10

2.2 Method for structure relaxation ····· 12

2.3 Method for band gap ····· 13

2.4 Method for dielectric constant ····· 16

3. Automation algorithms ····· 18

3.1 Screening structures ····· 19

3.2 Structure relaxation ····· 21

3.3 Calculating band gap

3.3.1 Searching band edge positions from symmetric k-points ····· 24

3.3.2 HSE gap using GGA structure····· 30

3.4 Calculating dielectric constant ····· 33

4. Result and discussion

4.1 Performance of the automation code ····· 37

4.2. Materials map

4.2.1 Overview of materials map ····· 39

4.2.2 New candidate for high-k dielectrics ····· 41

4.3 Correlate relationship of calculated variables ····· 45

5. Conclusion ····· 50

References ····· 52

List of tables and figures

Figure 1. Static dielectric constant versus band gap of candidate oxides for high-K dielectric ····· 4

Figure 2. The energy gaps of various materials are compared between experiment and theory ····· 15

Figure 3. Main strategy of automation for searching new high-k candidates ····· 18

Figure 4. k-point test scheme ····· 22

Figure 5. Schematic picture of a dispersive conduction band around the band edge position ····· 25

Figure 6. Calculated band gap of Al2O3 (ICSD#: 173014) by by mesh search and line search ····· 29

Figure 7. Calculated band gap by automation using three different method versus experiment data ····· 32

Figure 8. Calculated static dielectric constant with two XC functional versus experiment ····· 35

Figure 9. Selected atom A for AO binary compounds ····· 38

Figure 10. Selected atom A and B for ABO binary compounds ··· 38

Figure 11. Materals map of band gap and static dielectric constant (as of May 2013) ····· 40

Figure 12. New predicted phase transition of rocksalt BeO ····· 44

Figure 13. band gap versus electronic contribution of dielectric constant ····· 47

Figure 14. band gap versus ionic contribution of dielectric constant ····· 47

Figure 15. Correlation between Born effective charge and static dielectric constant (fitted line also plotted) ····· 48

Figure 16. Corrlation of Bader charge and static dielectric constant (a) for metal ion (b) for oxygen ····· 49

Table 1. Comparison for number of k-points and calculated band gap by mesh search method and line search method ····· 30

Table 2. Calculated Band gap by automation code reference calculatation data and experimet data ····· 31

Table 3. Calculated static dielectric constant with two XC functional and experimental data ····· 34

Table 4. Percentage of each automated process computing cost ····· 39

Table 5. New candidate materials for high-k dielectrics ····· 41

Table 6. Reported phase and calculated properties of BeO ····· 43

Table 7. Comparison for calculation result of tetrahedral BeO ····· 44

Chapter 1. Introduction

1.1 New era of the material design : data mining by high-throughput calculations

In most electronic devices, highly developed processing and manufacturing techniques derived almost top-notch of the consisting material's performance. To make more advanced devices, there must be an accompanying materials design. The most robust and direct approach for a new materials design is searching target properties from large number of candidate materials as many as possible. After the new massive property database is constructed, it may be possible to find some new promising materials or examine the correlational relationships between measurable variables of the vast candidate pool. Using this so-called "data mining" method, controlling the desired property by material selection would be enabled. To actualize such a dreamy data mining in materials science, three conditions need to be satisfied. First, a high-throughput estimation or prediction should be possible. Second, the candidate material structures could be collected from large and reliable structure database. And last, estimation or prediction of the

properties should be performed with good enough accuracy for comparing each other. But these conditions are almost impossible to be achieved by experimental approaches.

Recently, remarkable advances in computing power and first-principles techniques present a good opportunity for building a vast theoretical database of material properties. Fast high-throughput calculations with high accuracy can be achieved by aligning proper computational methods into a sophisticated automatic procedure. By minimizing human interventions, even though every procedural step should be carefully verified and tested in advance, massive data of material properties could be obtained within a reasonable computational cost. Besides, nowadays, there are large and reliable databases of materials structure which is easily accessible via internet, such as ICSD(Inorganic Crystal Structure Database)[1]. In this thesis, I'll introduce our data mining strategy by a full-automated code adopting VASP[2] as the core engine of DFT calculations and provide recently accumulated property data base will be also provided.

1.2 Materials design for a high-k dielectrics

The remarkable improvement of microelectronic device is one of the most evolutionary achievements in modern science and engineering. Until recently, the silicon-based semiconductor technology has been taking a leading role for the performance enhancement of MOS devices, especially by scaling down the device's dimensions. But as the scale of MOS approach to a few nanometers, the performance of the device cannot be enhanced by scaling down anymore as degrading factors start to behave dominantly[3],[4]. At this physical lower limit, a new materials design is the only key to break through the blind lane of prior technology. Introducing high-κ material for the gate dielectric is a representative case.

For decades, silicon dioxide has been used as a gate oxide material. As the size of transistor has decreased, the thickness of the silicon dioxide gate dielectric has steadily decreased to increase the gate capacitance and thereby drive current, raising device performance. But as the thickness of gate dielectric decreases to less than 2 nm, leakage currents due to tunneling increase drastically and it causes serious problem to devices with high power consumption and reduced device reliability.

To solve this problem, replacing the silicon dioxide gate dielectric

with a high－κ material allows increased gate capacitance without the leakage effects. High dielectric constant with large band gap is the centric properties to be a high－k material for gate dielectric application.

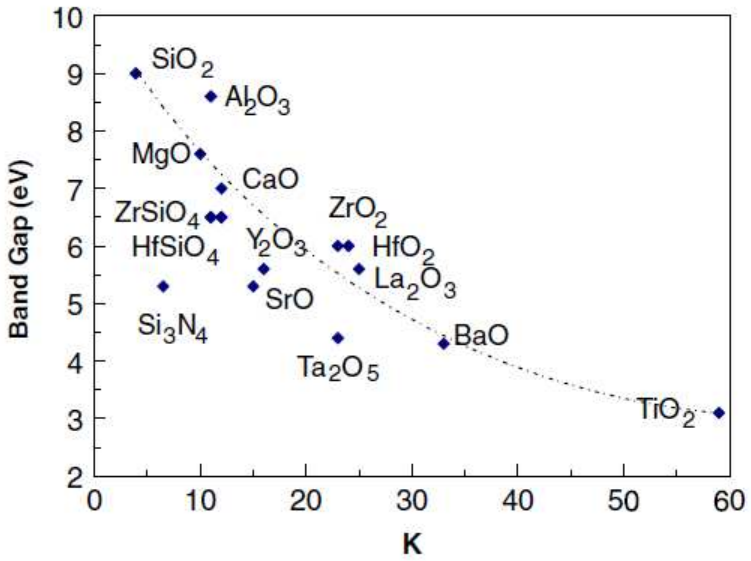

Figure 1. Static dielectric constant versus band gap of candidate oxides for high-K dielectric (Experiment) [5],[6]

Unfortunately, like many other nature of materials, it is hard to find a dreamy material which has simultaneously large band gap and large dielectric constant. As shown in Fig 1 [5],[6], experimentally measured bulk properties of known candidate materials for high－k dielectrics show that band gap and static dielectric constant have the

“trade-off” relation. But this relation is a substantial trend rather than a strict correlation. Many other factors could be involved and correlated in each property. And huge numbers of materials still are not yet searched for this specific application.

So, as the first goal of data mining, We developed the full-automation code for calculating band gap and static dielectric constant for finding new high-k dielectrics. Using the automation code, We managed to calculate more than 1,000 binary and ternary crystal oxides from ICSD and we generated “materials map” from the accumulated material property database until recently.

In section 2, I’ll briefly introduce the computational theories and methods for calculating those properties from collected structures. In section 3, The algorithm of our automation code is described and in section 4, I’ll provide currently updated “materials map” and discuss correlating relationship of obtained results.

Chapter 2. Computational theory and method

2.1 DFT and beyond DFT

2.1.1 Density functional theory

Density functional theory(DFT) has been most powerful first-principles calculation method to predict physical properties of various system since Hohenburg and Kohn theorem[7] was introduced in 1964. By Hohenburg and Kohn theorem, the external potential can be a unique functional of ground state charge density $\rho(r)$. If the ground state density is known, Hamiltonian $\mathbf{H}$, ground state wave function ψ, and ground state energy can be defined as a unique functional of $\rho(r)$.

The total energy E of the system is define as

$$E = \int v(r)\rho(r)dr + F(\rho(r)) \qquad (2.1)$$

Where $v(r)$ is the external potential, $\rho(r)$ is the charge density, and $F(\rho(r))$ is the universal functional. The variational principle

proves that the total energy given by Eq.(2.1) is minimum when $\rho(r)$ corresponds to the true charge density in the ground state.

To calculate practical electronic structures of the real material, the Hamiltonian must be defined. The Hamiltonian of interacting electrons in the ionic potential is shown in Eq.(2.2)

$$-\sum_{i=1}^{N}\frac{\hbar^2}{2m}\nabla_i^2-\sum_{i,I=1}^{N}\frac{Z_I e^2}{\left|\overrightarrow{R_I}-\overrightarrow{r_i}\right|}+\sum_{i>j}^{N}\frac{e^2}{\left|\overrightarrow{r_i}-\overrightarrow{r_j}\right|}\nabla_i^2 \qquad (2.2)$$

The first term is kinetic energy, the second term is ion-electron interaction, and the third term is electron-electron interaction. But it is practically impossible to solve the Schrodinger equation with the Hamiltonian given by Eq.(2.2) because of the extremely high degree of freedom.

In 1965, Kohn and Sham suggested the effective potential V_{eff} as the solution to describe many-body system. They convert the many-body problem with complicated interactions to simple one particle problem by considering the effective potential. The effective potential is composed with three parts; 1) the ion-electron interaction V_{ion-e} (Ewald energy), 2) electron-electron coulomb

interaction V_{e-e} (Hartree energy), and 3) the exchange-correlation energy V_{xc}.

$$V_{eff}(r) = V_{ion-e} + V_{e-e} + V_{xc} \tag{2.3}$$

Using V_{eff}, the Schrodinger equation becomes simple one particle problem as Eq.(2.4)

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(\rho(\vec{r})) \right] \phi_i(\vec{r}) = \varepsilon_i \phi_i(\vec{r}) \tag{2.4}$$

Starting with a trial charge density and wave vectors from the atomic potentials of given system, the effective potential is calculated and then Schrodinger equation is reduced to single particle equation by the variational principle. Then the Fermi energy and new charge density is determined. This calculation processes is repeated recursively until the difference between starting charge density and new charge density becomes smaller than certain given criteria.

So, how to determine the effective potential is very important factor of DFT calculation. The exchange-correlation part of the effective potential V_{xc} is related to quantum mechanical properties of mnay-body system, and it can be generated by two different approximations. One is local density approximation(LDA) which is first introduced by Kohn and Sham[8]. The LDA scheme assumes that the exchange-correlation energy of the system is identical to that of the homogeneous electron gas. The exchange-correlation energy at r is regarded as the corresponding value of the electron-gas system with homogeneous charge density ρ everywhere.

$$E_{xc}^{LDA}[\rho(r)] = \int \varepsilon_{xc}^{LDA}(\rho(r))\rho(r)dr \qquad (2.5)$$

The other is generalized gradient approximation(GGA) first developed by Perdew[9] and Becke[10]. GGA is basically similar to LDA except that GGA contains the inhomogeneity of electron density considering the XC energy deviation from the uniform electron gas consumption.

$$E_{xc}^{GGA}[\rho(r)] = \int \varepsilon_{xc}^{GGA}(\rho(r)|\nabla\rho(r)|)\rho(r)dr \qquad (2.6)$$

In 1996, Perdew-Burke-Ernzerhof improved formal GGA scheme(PW91) and it is called PBE[11]. We use PBE for GGA functional in this thesis. GGA improved description of the binding energy of the molecules and the cohesive energy of solids compared to LDA[12].

2.1.2 Beyond DFT

For last decades, these DFT schemes widely and successfully solved numerous problems about structural, mechanical, vibrational, optical properties and etc. But, even though the DFT calculation broadly succeeds in various system and applications, both LDA and GGA intrinsically fail to predict realistic band energy level and they have serious problem of the band gap underestimation. There two reasons of these problems. One is omission of self-interaction energy and another is the absence of discontinuity in a derivative of exchange-correlation energy. Those problems can be solved by considering many-body theory such as GW approximation[13]. But GW approximation requires extremely expensive computing cost

compared to DFT and there is troublesome issue about selecting starting eigenfunctions and eigenvalues.

Confined to energy gap, the hybrid functional also yields remarkably improved result than conventional DFT. The hybrid functional is based on admixture of the bare exchange from Hartree-Fock(HF) and that from the DFT. The self-interaction error within DFT is largely canceled by HF method. We used HSE06 method proposed by Ref.[14] in this thesis because it is more efficient compared to former PBE0[15] method. In HSE06 method, HF exchange energy is only admixed within short-range(SR) part because the exchange energy is exponentially decaying as a function of the distance. Since the exchange energy in HF scheme is calculated by integrating the orbital directly, excluding long-range(LR) part makes a profit on performance. The resulting formula of this scheme is

$$E_{xc}^{HSE} = \frac{1}{4} E_x^{HF,SR}(\mu) + \frac{3}{4} E_x^{\mathrm{PBE},SR}(\mu) + E_x^{\mathrm{PBE,L}R}(\mu) + E_c^{PBE} \qquad (2.7)$$

Here the mixing coefficient α is 0.25 which indicate the ratio for the amount of exact exchange from HF. μ is the screening

parameter, the criteria for splitting SR and LR. In HSE06 μ is chosen by 0.2 $Å^{-1}$.

HSE06 calculation often result still a little underestimation of band gap energy compared to experiments. In those cases, adjusting the mixing coefficient α is usually conducted to give a closer description compared to the experimental result. At the point of computing cost, hybrid functional is still much more expensive than conventional DFT calculation, but much manageable than GW approximation.

2.2 Methods for Structure relaxation

Within a DFT scheme, the forces acting on each ion can be calculated at electronic ground state. Using this information, ionic relaxation including cell relaxation can be performed recursively to find minimum total energy state of the system. It is well-known that DFT calculation provides substantially good predictions of atomic structures of material. Most current XC functional give good structural agreement with experiment less than a few percent or and general tendencies of each functional is also well reported. LDA

functional gives mostly underestimated the lattice parameters, while GGA functional shows a little overestimating tendency usually with smaller error than LDA. In many cases, hybrid calculation predicts better lattice parameters than conventional DFT calculations but ionic relaxation of hybrid calculation consumes way much more computing cost than DFT.

2.3 Method for Band gap

Band gap is energy difference between minimum of conduction band energy and maximum of valence band energy($E_{CBM} - E_{VBM}$). To calculate accurate band gap energy, two difficulties have to be deal with. First, energy levels of conduction and valance band or at least their difference should be accurately predicted. Second, to get the E_{CBM} and E_{VBM}, exact band edge positions in k-space must be found.

As discussed in section 2.1, it is verified that eigenvalues of DFT have no precise physical meaning[8],[16]. To solve this flaw of DFT calculation, many-body effect should be considered, especially the self-interaction energy of electrons. Using GW approximation, qusiparticle band structure energies are obtained by replacing the

KS exchange−correlation potential with the electron self−energy $\sum$. But, as discussed in section 2.1.2, GW approximation requires much expensive computing cost even compared to hybrid calculation and have some difficulties to be implemented in high−throughput calculation. In case of our purpose, only the energy difference between E_{CBM} and E_{VBM} is important, so adoption of GW calculation is not necessary and definitely not proper for high efficient automation. The hybrid functional can be more rational solution to get improved band gap energies. Figure 2 shows calculated band gap of some insulators and oxides by LDA, GGA and HSE06 compared to experiments[17]. The figure shows that HSE06 shows best results among three first−principles calculations. In section 3.3.2, we proposed much efficient scheme to calculate band gap compared to strict hybrid functional scheme with minimizing loss of accuracy.

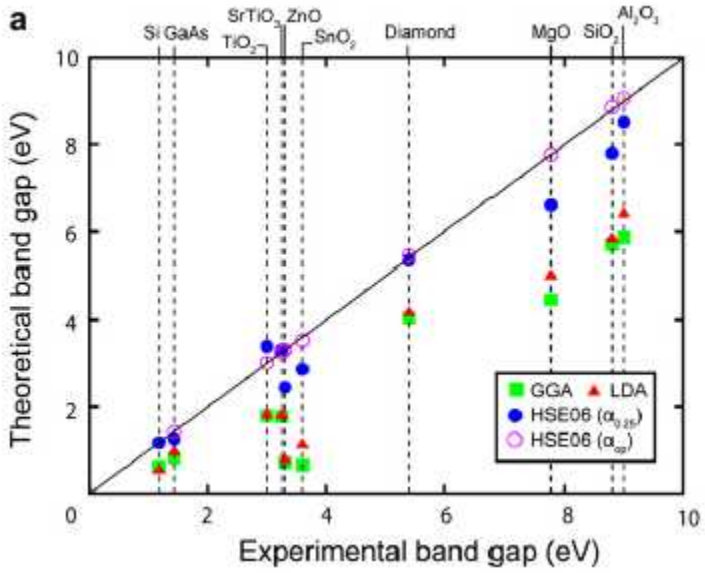

Figure 2. The energy gaps of various materials are compared between experiment (x coordinate) and theory (y coordinate)

S. Park et al, Curr. Appl. Phys. 11, S337 (2011)

2.4 Method for Static Dielectric Constant

It is generally accepted that an accurate quantitative description for static and frequency-dependent dielectric response function requires the implementation of post-DFT schemes, such as GW[18] or the Bethe-Salpeter equation[19],[20], a qualitative agreement between theory and experiment can often be achieved on the level of DFT. Static dielectric matrix can be calculated by density functional perturbation theory(DFPT or linear response theory) which transform all summations over empty bands into linear equations[21]. Local field effect also can be included in the calculated matrix using Hartree approximation(RPA) or DFT level. Only the ion clamped static dielectric constant including local field effect must be compared to experiment. Relaxed-ion static dielectric tensor also band be calculated from the interatomic force constants using density perturbation theory. The total static dielectric constant can be obtained by these electronic and ionic contribution of dielectric constant.

$$\varepsilon_0 = \varepsilon_\infty + \varepsilon_{ionic} \tag{2.8}$$

Because the dielectric function is tensor property, averaged value of static dielectric constant like in eq.(2.9) can be compared with experimental value.

$$\varepsilon = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix}$$

$$\bar{\varepsilon} = \frac{1}{3}\left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right)$$

$$\overline{\varepsilon_0} = \overline{\varepsilon_\infty} + \overline{\varepsilon_{latt}} \qquad (2.9)$$

It is known that DFPT using LDA functional usually overestimate the dielectric constants by 5−20% and careful interpretation is required especially for a quantitative description of screening in small gap materials. But DFPT with LDA functional still gives good predictions particularly compared to DFPT with GGA.

Chapter 3. Automation Algorithm

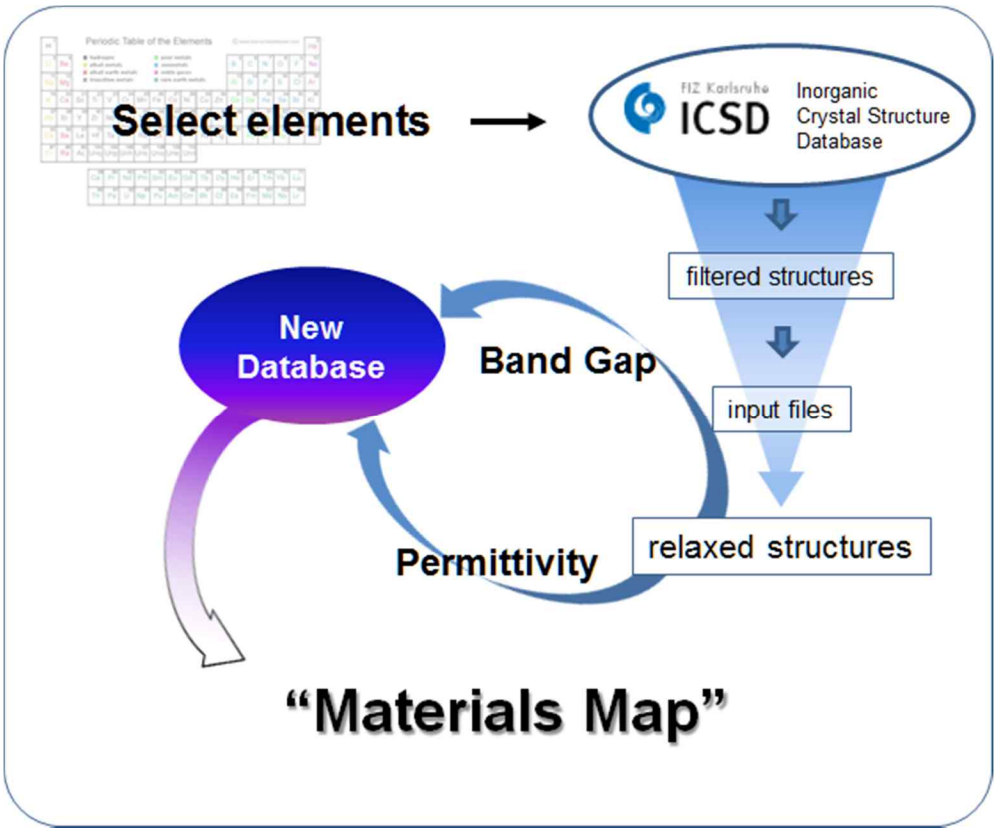

Figure 3. Main strategy of automation for searching new high-k candidates

Figure 3 shows brief scheme of our automation code for calculating band gap and static dielectric constant. The automation code minimize the human intervention, so the only thing to do by user is giving initial structure set of targeting materials by just selecting the

consisting atoms. After the atoms are selected, the code screen the non-crystalline structure and pick out the unique structure data within the excessive data which has same structures. After the starting structure set is extracted, input files for the first-principles calculation is automatically generated. We used VASP[2] as our core engine for the first-principles calculation. From the generated input files, structure relaxation, band gap calculation and static dielectric constant calculation are performed in sequence. After whole calculation of each structure is finished, the result data is collected to new property database. More detailed algorithms of the automation code are below sub-sections.

3.1 Screening initial structure

To make a large database of material's properties, the first thing we need is the large database of structure. ICSD is one of the largest and most favored crystal structure database which contains more than 160,000 structure data including both experimentally searched and theoretically designed.

To make a good materials map, calculation of real existing structures should take precedence. So we first extract all experimentally reported structures from ICSD. After all experimental structures which consist of selected atoms are searched, the occupancies of each ion consisting the unit-cell are checked. If any ion has smaller occupancy than 1, the structure is rejected. Once only perfect crystal structures are gathered, the structures are arranged by atomic composition and space group. Within the set of same composition and space group, structures are lined up in order of lower experimental temperature, and then smaller lattice parameter. At first, the structure which is found at lowest temperature is selected as a unique structure of the set. Then each ratio of lattice parameters of other structures is checked. If the lattice ratio of the structure is same with the unique structure, that structure is rejected as redundant data. If the lattice ratio of a structure is significantly deviated from already selected unique structure, then the structure is also selected as another unique structure. The reason that the lowest temperature structure are selected is that DFT calculation describe the ground states of system at 0K, so it is easily expected that starting with the lowest temperature structure is the best choice for reducing the ionic

relaxation cost. In ICSD, in fact, temperature information is omitted in many data. In the case of no data has temperature information in the whole set, the structure with smallest lattice length is selected as the first unique structure of the set.

3.2 k-point test and structure relaxation

To obtain reliable and efficient calculation, cut-off energy test and k-point test should be preceded. These tests are simple but quite harassing, so we use recommended cut-off energy setting(high precision) by VASP. Unlike cut-off energy, k-point set cannot be determined explicitly without test calculations because the convergence of the selecting k-point set is highly dependent to the system. All calculation including ionic relaxation must be performed with most efficiency but without compromising accuracy, so k-point test should be carefully done. Usually k-point test is done by starting with smaller number of k-point mesh and increase the mesh size until calculated result satisfy the certain criteria, but in our code we adopted the opposite approach to avoid serious problem from

using insufficient size k-point mesh. This can be happened when a plateau is exist like shown in Fig. 4

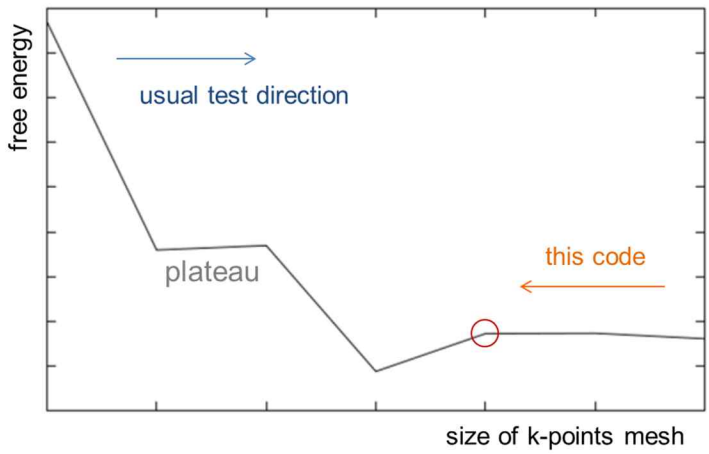

Figure 4. k-point test scheme

Criteria for k-point test is total energy difference < 5 meV/atom and inner pressure difference < 10 kB. There is KSPACING option in VASP which means the spacing between the nearest k-points (unit: $Å^{-1}$) to generate the k-point mesh automatically. Most metal systems show good convergence by using KSPACING = 0.2 $Å^{-1}$ and for insulators and semiconductors much smaller k-points are needed to get enough accuracy. Our code start the k-point test from certain KSPACING value and increase it until the criteria are keep

satisfied. Use the largest KSPACING(the smallest k-point mesh) which satisfying the criteria assures fasted performance for other heavier calculations without compromising accuracy.

We performed test calculations of this scheme for more than 100 systems, and we found that 0.25 $Å^{-1}$ is good enough starting KSPACING. No insulating or semiconducting compound needs smaller KSPACING than 0.25, and we don't need to consider the calculation accuracy of metallic compounds because they don't have the band gap and sensible dielectric constant.

As discussed in section 2.1, hybrid functional gives better result for lattice parameters than DFT, but full ionic relaxation using hybrid functional is too expansive for high-throughput calculation. So the automation code only performs the ionic relaxation using GGA and LDA. GGA often gives a little better result than LDA, so we use GGA relaxed structure to calculate band gap. LDA relaxed structures are more proper for dielectric properties calculation. Using the already relaxed structure with GGA, the relaxation cost for LDA relaxation is negligible to whole automated step.

3.3 Calculating band gap

3.3.1 Searching band edge positions from symmetric k-points

To calculate the band gap energy of a system, we have to find E_{CBM} and E_{VBM}. The DFT calculation yields band energy levels(KS eigenvalue) at given k-points, but at which k-point the E_{CBM} and E_{VBM} are exist is hard to be determined explicitly. In many cases E_{CBM} and E_{VBM} of overall band structure are found at certain special k-points with high symmetry. It could be assumed that band edge positions are at high symmetric position of the Brillouin zone rather than randomly distributed, so people usually search the high symmetry line of the Brillouin zone to get the band structure and the band gap.

Taking this rough assumption, we make our code to search only the k-points on the high symmetric line of the Brillouin zone for each system. The spacing between nearest k-point for findng accurate enough band edge position is determined by below description.

The more dispersive the band is, the higher k-point density is needed to reduce error of predicted band gap. For oxides, the conduction band sometime has very dispersive curve while the valence band is mostly like flat line. The effective mass of conduction electron can be calculated from the conduction band

curve, and vice versa, the effective mass of the system describes the dispersiveness of the conduction band. The known smallest effective mass of oxides is $0.1-0.3\,m_e$, so we assumed that the oxide system which has $0.1\,m_e$ requires most dense k-point set along the symmetric line of the Brillouin zone.

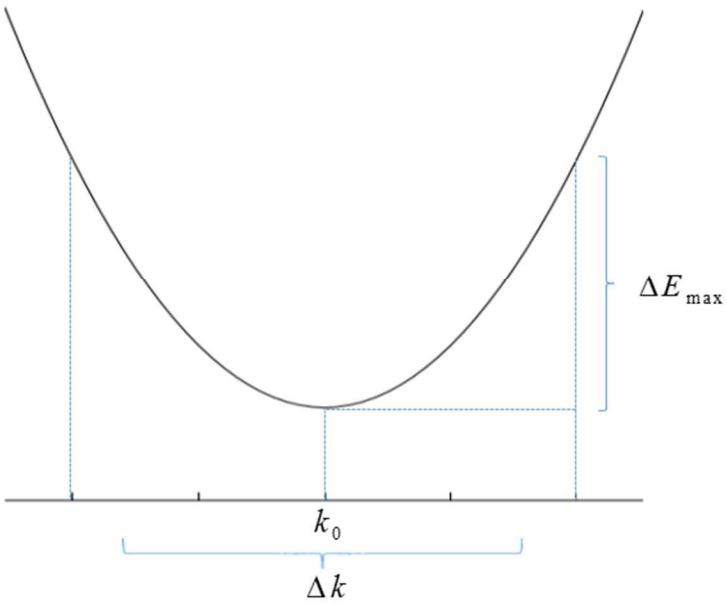

Figure 5. Schematic picture of a dispersive conduction band around the band edge position

Fig. 5 shows usual form of the conduction band near the band edge position of oxides. k_0 is minimum energy position of conduction band, and $\Delta E_{\max}$ is maximum error by spacing between nearest k-points Δk. The curve of band near the minimum position can be assumed quadratic. The maximum error for $\Delta E_{\max}$ from Δk can be calculated by eq.(3.1)

$$\Delta E_{\max} = \frac{\hbar^2}{2m^*}\left(\frac{\Delta k}{2}\right)^2 = \frac{\hbar^2 \Delta k^2}{8m^*} \tag{3.1}$$

where m^* is the effective mass.

If the effective mass m^* is 0.1 m_e and $\Delta E_{\max}$ is 0.1 eV, then $\Delta k \simeq 0.07 \text{Å}^{-1}$, which means that Δk should be smaller than 0.07 Å^{-1} to obtain E_{CBM} with in the error of 0.1 eV

Eq.(3.2) is valid for averaged error ΔE_{avg}

$$\Delta E_{avg} = \frac{1}{\Delta k}\int_{-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \frac{\hbar^2}{2m^*}\left(k' - \frac{\Delta k}{2}\right)^2 dk' = \frac{\hbar^2 \Delta k^2}{24m^*} \tag{3.2}$$

In this case $\Delta k \simeq 0.13 \text{Å}^{-1}$.

So we use 0.1 Å^{-1} for Δk to get the accuracy with error < ~0.1 eV.

We conducted test calculations for more than 30 structure set that randomly selected with various space group. Band gap prediction for every structures were performed by mesh search and line search method with varying k-point spacing. In many cases, the band edge positions are laid on Γ point or special symmetric k-point, so small number of k-point are needed to get converged result. But other cases, it is shown that line search method requires much less number of k-points to calculate and sometimes it gives more accurate position than mesh search method. Mesh search method sometimes fails to predict good band edge position with band gap error by 0.1~0.3 eV. Fig. 6 shows example of calculated band gap using two different k-points search method. Line search method find larger band gap when small number of k-points(large spacing between k-points) is searched but soon found minimum gap energy. In case of mesh search method, it also find minimum gap with very small number of k-points but the results are fluctuating as the number of k-points are increased. It requires very large number of k-points to get converged result and for this example it fails to find

minimum energy gap even with very dense k-point mesh used. Table 1 lists more examples for comparing mesh search and line search method.

As a conclusion we adopted line search method to find band edge positions with high efficiency and even better accuracy.

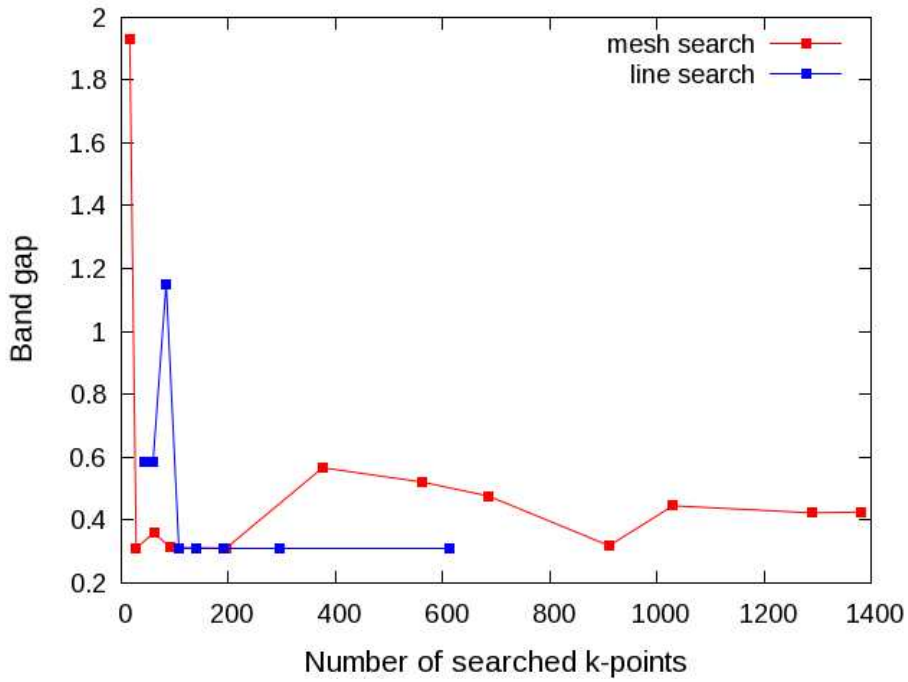

Figure 6. Calculated band gap of Al_2O_3 (ICSD#: 173014) by by mesh search and line search

structure	number of k-points		predicted gap (eV)	
	mesh	line	mesh	line
Al_2O_3_173014	1030	107	0.45	0.31
$Na_2[SO_4]$_27955	530	12	4.96	4.88
HfO_2_53033	104	20	3.81	3.77
Li_2O_108886	69	42	5.00	4.76
ZrO_2_647689	56	20	3.16	3.08
BeO_163824	72	27	6.84	6.83
$KClO_3$_9483	288	177	5.13	5.13
BaO_173922	18	46	2.66	2.66
MgO_2_41732	11	7	3.75	3.75
SeO_2_99464	18	20	3.34	3.32
...				

Table 1. Comparison for number of k-points and calculated band gap by mesh search method and line search method

3.3.2 Band gap by HSE06 using GGA structure

As explained in section 2.1.2, the hybrid functional is best choice for our automated code for band gap calculation. It gives accurate predictions for band gap as GW approximation while computing processes are cheaper and simpler. But full relaxation using hybrid functional is still too much expensive for applying to high-

throughput calculation. It is known that band gap energy is not much sensitive to lattice parameters, so it is expected hybrid calculation using relatively accurate structure can yield similar result for band gap. Table 2 and Fig. 7 shows our test calculation result of band gap using 3 different method and experimental data.

	Auto calc.			Reference	
	[1)]GGA	[2)]HSE@ GGA	[3)]HSE	GGA (ref)	Expt.
Li_2O	4.96	6.57	6.64	4.96	6.7
Na_2O	1.99	3.41	3.52	1.83	3.84
K_2O	1.63	2.83	2.92	1.71	3.33
Rb_2O	1.24	2.4	2.51	1.31	2.3
BeO	7.35	9.26	9.54	7.31	10.6
MgO	4.54	6.24	6.57	4.8	6.4
CaO	3.63	5.31	5.31	3.7	5.7
SrO	3.27	4.88	4.87	3.4	5.9
BaO	2.08	3.38	3.38	2.0	3.88

1) GGA relax + GGA band gap
2) GGA relax + HSE(one-shot)
3) HSE relax + HSE band gap

Table 2. Calculated Band gap by automation code reference calculatation data and experimet data

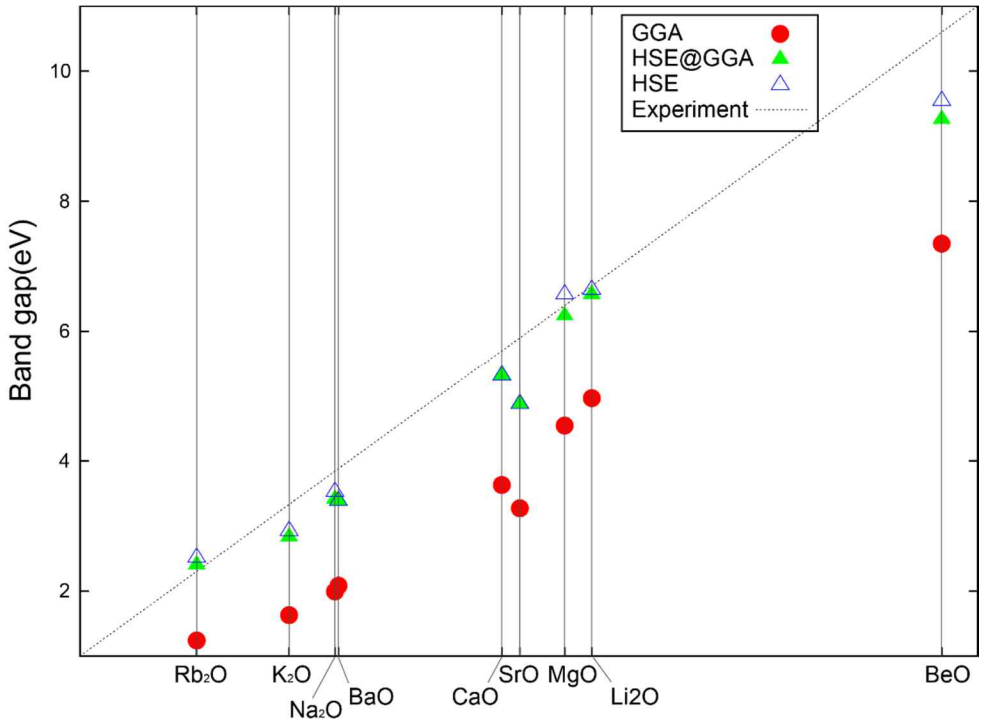

Figure 7. Calculated band gap by automation using three different method versus experiment data

In table 2, it can be found that our automation code result almost same date compared to reference calculation by hand. Compared to GGA result, remarkable improvements of band gap in HSE and HSE@GGA are observed and both result are very similar, so it is assumed that HSE method can be substituted with HSE@GGA method with no harmful error. But it should be noticed that

HSE@GGA method (and also HSE method) still has a little underestimating tendency by ~10%.

To achieve high efficient automation code we choose HSE@GGA calculation for predicting band gap.

3.3 Calculating dielectric constant

Static dielectric constant of bulk crystal using density functional perturbation theory can be easily calculated by VASP. One thing that should be considered carefully is the calculation is much more sensitive to k-point mesh selection that required much more k-points than other calculation. In addition, the calculation of dielectric function intrinsically requires much more computing times for considering every degree of freedom. It need to be carefully tested how large the k-point mesh is required but such test calculations consume too much computing cost since massive number of k-points is calculated each test step. But dielectric constant is unit-less values so converging criteria doesn't need to be set with specific small value.

To avoid terrible computation cost of k-point dependence test, we simply multiplied size of the k-point mesh that used for relaxation

for each direction in our scheme. That means 8 times more k-points are calculated when calculating dielectric constant and we quite assure by our experience that resulting data will doesn't change more than a few percent when more dense k-point mesh is used.

Table 3 and figure 8 shows the test result of the automation code with using both LDA and GGA functional and experimental references.

	Auto calc.		Ref.
	GGA	LDA	Expt.
BeO	7.23	6.83	7.01
Li_2O	7.97	6.74	7.72
Al_2O_3	10.80	9.72	9
MgO	11.10	9.33	9.8
CaO	17.26	15.16	11.8
SrO	22.24	18.23	13.3
BaO	68.36	58.20	34
TiO_2	562.8	119.9	114

Table 3. Calculated static dielectric constant with two XC functional and experimental data

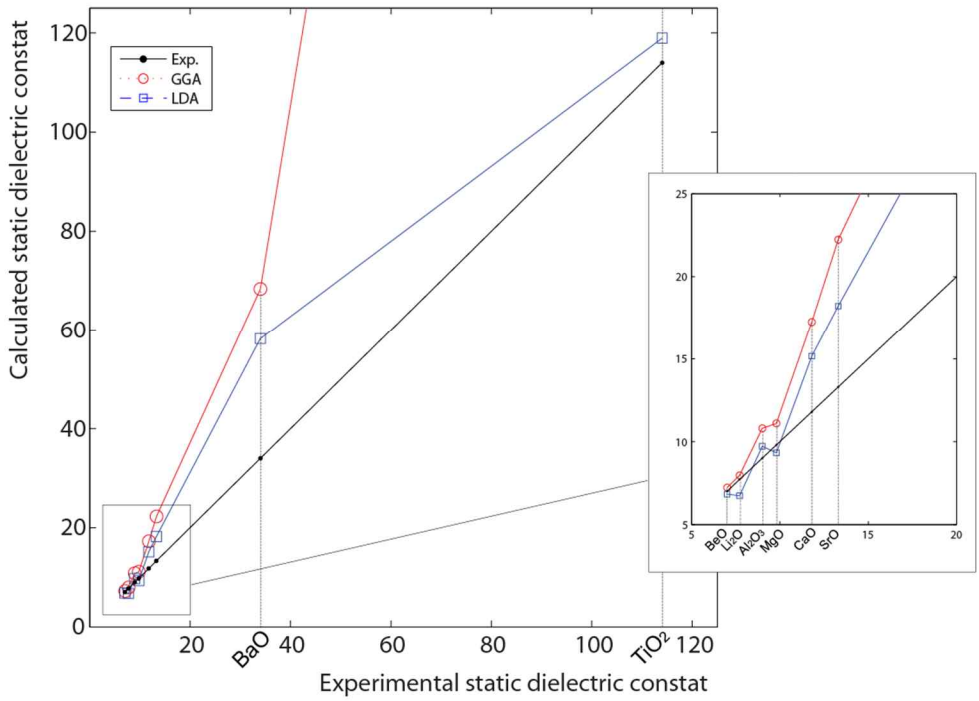

Figure 8. Calculated static dielectric constant with two XC functional versus experiment

It is confirmed that DFT result using both XC functional have tendency of overestimation for dielectric constants, which is consistent with discussion in section 2.4. Comparing LDA and GGA result, LDA functional gives better result especially for high dielectric constant material. It seems that overestimating tendency of lattice parameter in GGA functional give more positively deviated dielectric constant since the it makes the phonon modes softer. The

largest deviation is observed in BaO. This large overestimation is also reported by ref.[21].

Even though the LDA result often have overestimation tendency, it has still qualitative meaning for comparing each other. So, instead considering much expensive quasiparticle calculations, we choose the DFPT scheme with LDA functional to get comparative better result with fast computing performance.

4. Result and discussion

4.1 Performance of automation code

Our automation code has high performance with various optimizations with no compromising accuracy. The main optimization strategies are :

1) Screening starting data set : exclude overlapping or impractical structures
2) HSE@GGA method for band gap calculation
3) High symmetry line search is used for finding CBm and VBM

Until recently, we managed to calculate more than 1,000 binary and ternary oxide materials using our automation code. Fig. 9 and 10 shows the selection pool of cations for binary and ternary compounds. Most transition metal atoms are not included because it is hard to determine ground state magnetic structure of metal oxides containing transition metals and many of transition metal oxides(TMOs) are known as small band gap materials. We will also calculate these TMOs in the future.

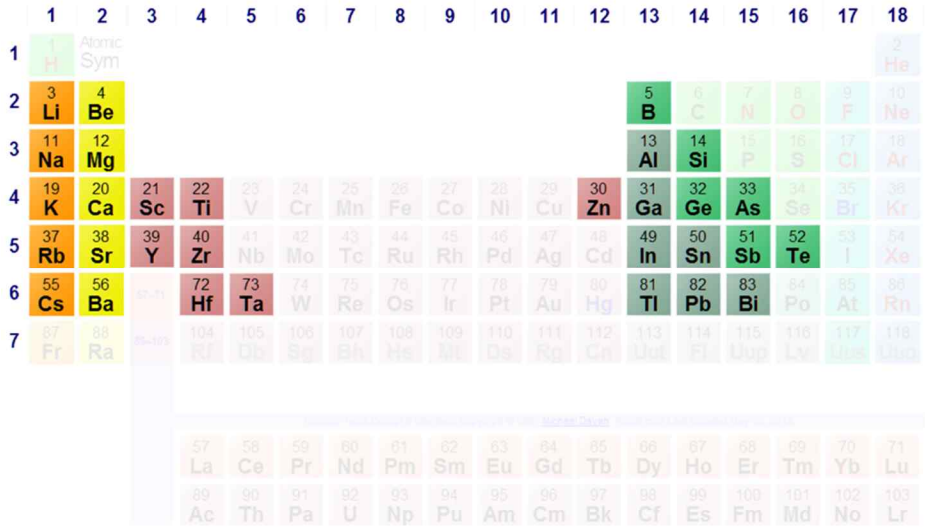

Figure 9. Selected atom A for AO binary compounds

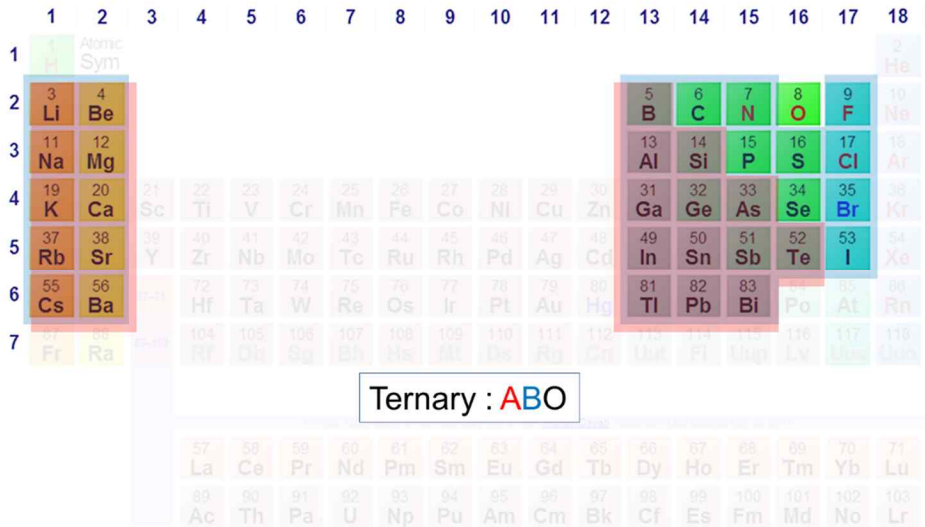

Figure 10. Selected atom A and B for ABO binary compounds

We used 64 core cluster for paralleled computing(8 nodes of cluster computer with 8 processors per each nodes). It takes For average, 130 CPU hours consumed per one structure which means 8 structures per one day were calculated. It is quite impressing performance for the relatively small computing resource used for this calculations. Table 4 shows the ratio of computing cost for each main process of automation.

Process	Structure relaxation	Band gap (GGA)	Band gap (HSE)	Dielectric constant
%	6%	21%	43%	30%

Table 4. Percentage of each automated process computing cost

4.2. Materials map

4.2.1 Overview of materials map

Fig. 11 is the " Materials map" recently generated from the obtained property database for about 1,000 binary and ternary oxides by our automation code. About 800 data plotted on the map except the materials predicted as metallic compounds.

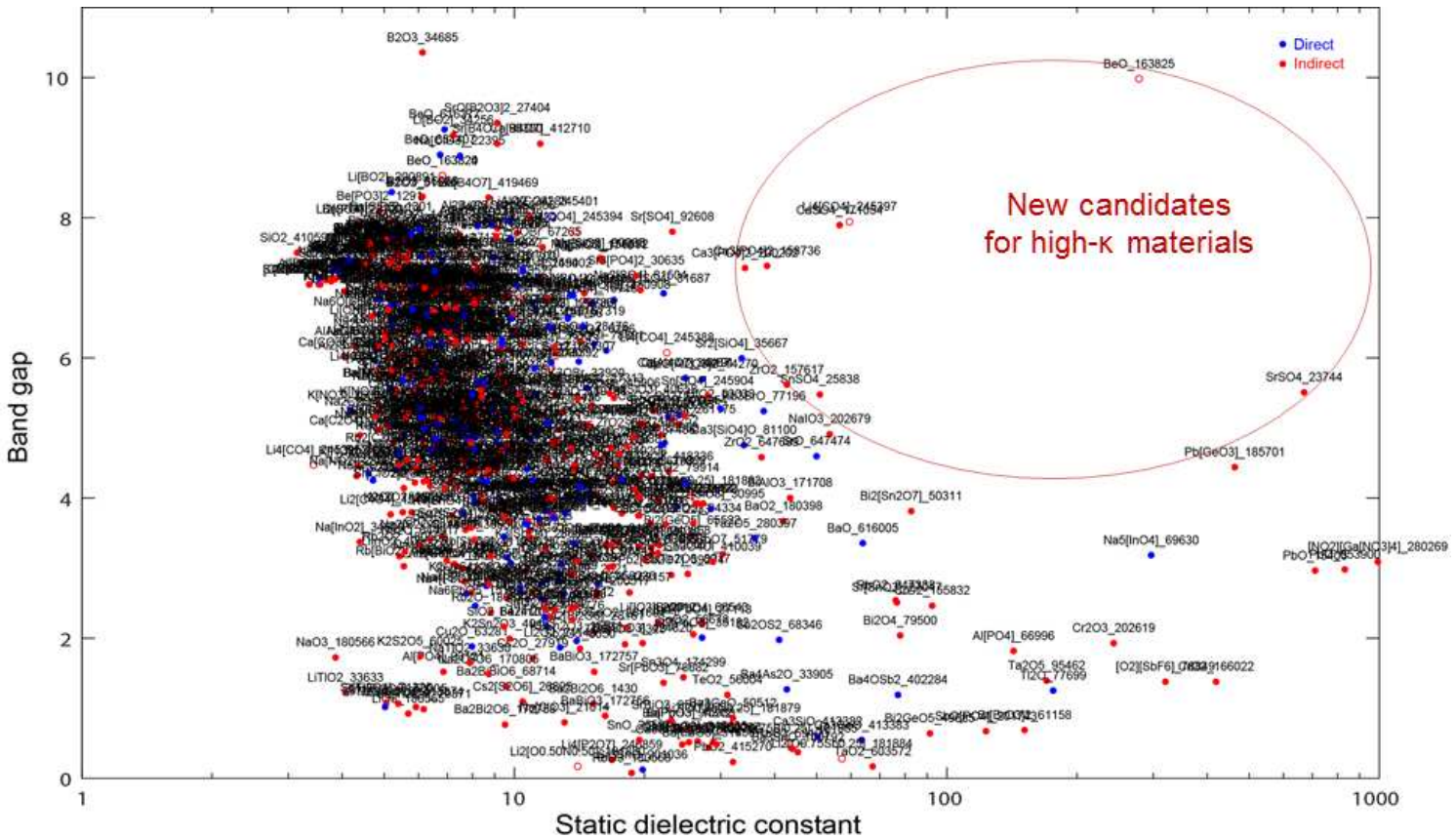

Figure 11. Materials map of band gap versus static dielectic constant (as of May 2013)

It can be found that trade-off relation of band gap and permittivity shown in Fig. 1 is also valid for our new mass database. Most structures which have large band gap tend to have small dielectric constant and vice versa. But there are also some interesting compounds which have both large band gap and permittivity from the materials map. These materials has new and unreported properties.

4.2.2 New candidate for high-k dielectrics

Table 5 shows list of the new candidate for high-k dielectrics found from the new materials map.

ICSD #	Material	Permittivity	Band gap	Info.
163825	BeO	276	10.01	Theoretical (High Pressure)
200202	Ca3[PO4]2	34.1	7.3	HT
158736	Ca3[PO4]2	38.15	7.33	2573K, 12Gpa
171054	CaSO4	56.13	7.91	1450K, 21GPa
245397	Li4[CO4]	59.39	7.95	Theoretical (High pressure)
202679	NaIO3	53.27	4.93	273K, atm
77196	Rb3BrO	37.61	5.25	RT, atm

Table 5. New candidate materials for high-k dielectrics

Unfortunately, many of these candidates are not stable materials at ambient pressure. Some of new candidates are only predicted by theoretical study. But there still are 3 interesting case, BeO, $NaIO_3$ and Rb_3BrO. We conducted calculations for these materials again more carefully by hand and the results were not different with automation data. Experimentally reported $NaIO_3$ and Rb_3BrO are stable materials in room temperature and atmosphere pressure. Those two materials have possibility to utilize for high-k dielectrics. But they are not superior to HfO_2, current most favored high-k dielectrics. (Predicted band gap of HfO_2 is 5.29 eV and permittivity is 29.9) BeO catch a more interest by its superior properties. In fact BeO is already studied as dielectric materials for its highest band gap among the oxides but dielectric constant of BeO reported as about 6~7[22]. But BeO has reported as a good oxygen diffusion barrier for Al_2O_3 and HfO_2 dielectrics Stable phase of BeO is wurtzite.

ICSD#	Structure	ΔE (eV/atom)	Band gap (eV)	Permittivity	Remarks
616377	wurtzite	‘	9.28	6.86	Expt (RT)
654407	tetragonal	0.02	8.92	6.73	Expt (high T)
163824	zincblende	0.01	8.62	6.80	Theory (high P)
163825	**rocksalt***	**0.48**	**10.01**	**276**	**Theory (high P)**

Table 6. Reported phase and calculated properties of BeO

The structure of BeO in table 5 is rocksalt structure which predicted by theoretical study as high pressure phase. But the phase transition from the wurtzite to rocksalt at 137 GPa is also reported in experiment by diamond anvil cell measurement[23]. Table 6 shows list of all experimental and theoretical phases of BeO and their calculated properties. By phonon calculation using Phonopy[24] we confirmed that the rocksalt structure is stable at high pressure(150 GPa) and it has imaginary phonon frequency without external pressure. We conducted some trials to stabilizing rocksalt BeO structure by doping but it was not succeed.

To find another phase transition from rocksalt BeO, we artificially moved the atoms of supercell, and found new tetrahedral structure by atomic relaxation using GGA scheme.(Fig.12) Newly found tetrahedral structure has almost same but a little bit higher total

energy. Disappointingly, the static dielectric constant of the new phase is calculated as

It is disappointing that there were no new superior materials for high-k dielectrics, but new candidate materials and phase found by our study still have good potential to utilize.

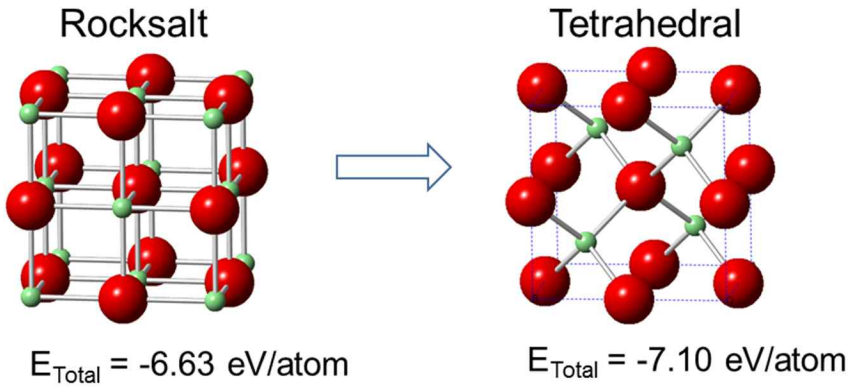

Figure 12. New predicted phase transition of rocksalt BeO

Structure	Type	ΔE (eV/atom)	Band gap	Permittivity
wurtzite	Expt. (RT)	,	9.28	6.86
rocksalt*	Theoretical (high pressure)	0.48	10.01	276
tetrahedral	New found	0.01	8.63	6.32

Table 7. Comparison for calculation result of tetrahedral BeO

4.3 Correlate relationship of calculated variables

From the materials map and property database, we can not only find some new promising materials but also find out correlations between calculated property variables. By searching the correlate relationships between variables, we can find descriptor for desired property.

From the beginning, we assumed that band gap and permittivity has trade-off relation. And same tendency with experimental data was found in our materials map. But the tendency is rough relation. It means permittivity is not a only function of band gap. Desired superior properties would be found from the materials which deviated from the tendency, so it is very important to find new descriptor that can determines or affects desired property.

Fig. 13 and Fig. 14 shows same materials map with Fig. 11, but each map considers only electronic and ionic contribution of dielectric constant respectively. It can be found that electronic contribution of dielectric constants are mostly inversely proportion to band gap which describe the trade-off relation very well. But for ionic contribution of permittivity, data are much randomly distributed. This lead to the conclusion that the ionic configuration does

important role to determine the ionic contribution of dielectric constant. This must be a key to finding abnormal behavior of new candidate material such as rocksalt BeO.

We also find interesting correlation of Born effective charge with dielectric constant. Born effective charge is the coefficient of proportionality between a change in macroscopic polarization in direction β caused by an atomic displacement in direction α under conditions of zero external field, shown in Eq.(4.1).

$$Z^*_{\kappa,\beta\alpha} = \Omega_0 \frac{\partial P_{mac,\beta}}{\partial \tau_{\kappa\alpha}(q=0)} \tag{3.3}$$

Fig. 15 shows correlation between Born effective charge and static dielectric constant of cubic binary oxides. There is apparent linear relationship with Born effective charge and electronic part of static dielectric constant, while ionic part show larger deviation to linear relationship. These correlate relation is obtained only by materials with similar symmetry phases.

We also calculated Bader charge for all binary materials but there is no correlation of Bader charge and dielectric constant, as shown in Fig.16.

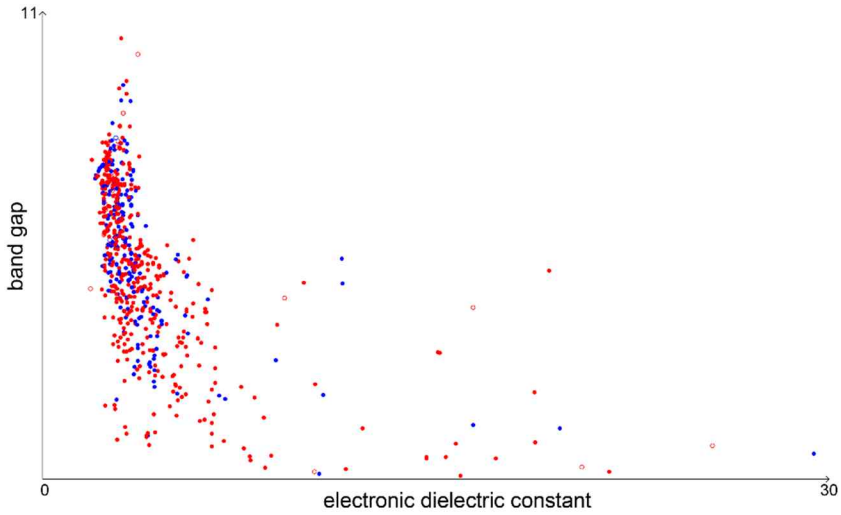

Figure 13. band gap versus electronic contribution of dielectric constant

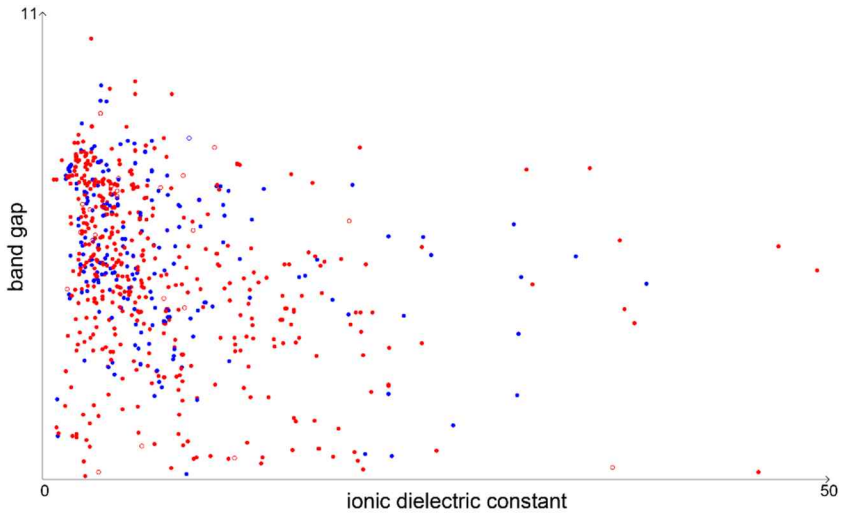

Figure 14. band gap versus ionic contribution of dielectric constant

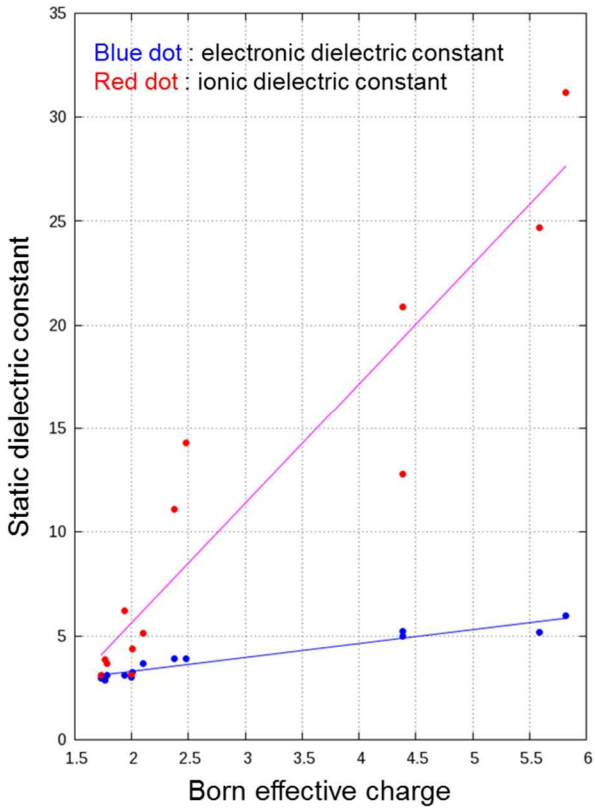

Figure 15. Correlation between Born effective charge and static dielectric constant (fitted line also plotted)

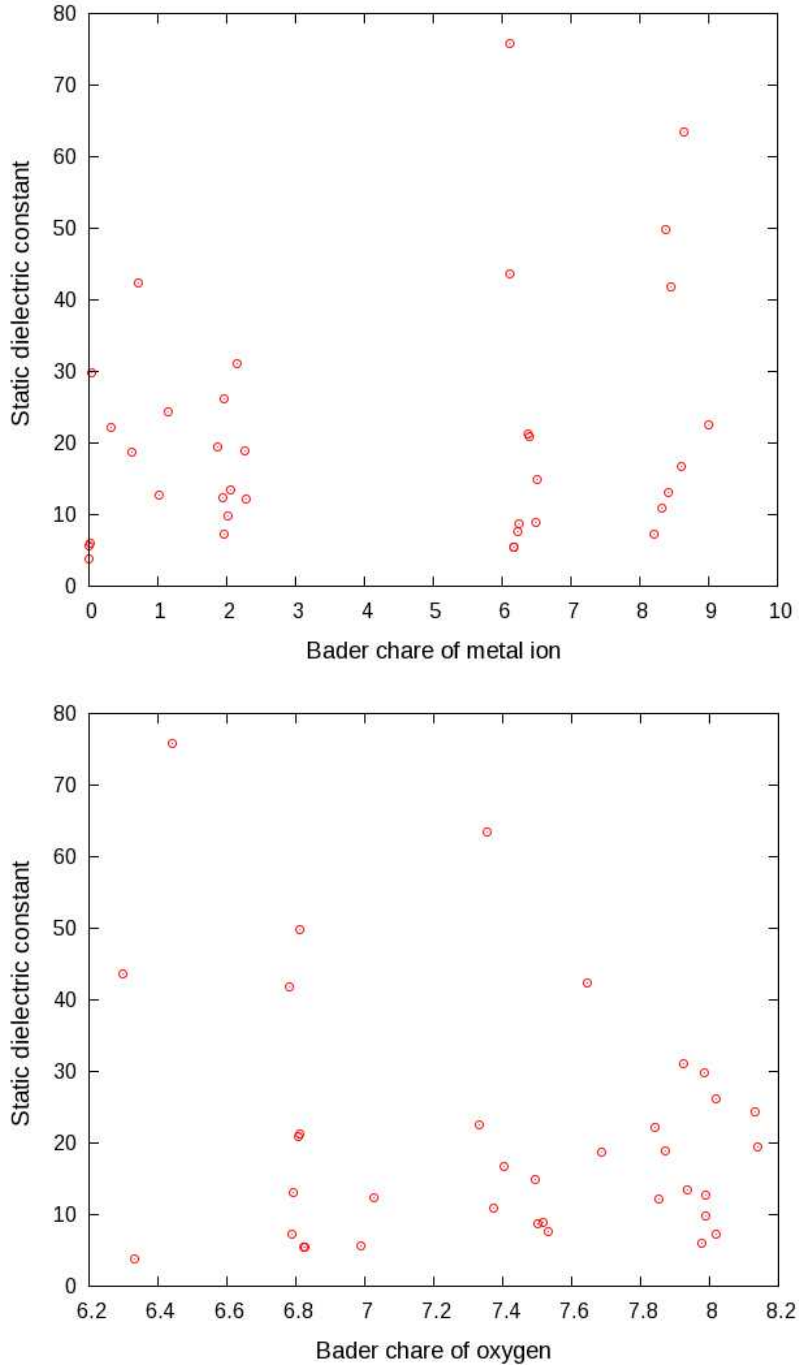

Figure 16. Corrlation of Bader charge and static dielectric constant (a) for metal ion (b) for oxygen

5. Conclusion

We conducted mass calculation of band gap and static dielectric constant for more than 1,000 binary and ternary oxides. This massive high-throughput calculation can be enabled by using newly developed automation code which has high efficiency with no compromising accuracy.

From the obtained property database, we generated the materials map of band gap versus static dielectric constant for high-k application. From the recently updated materials map, we found some new candidate materials for high-k dielectrics. The found candidates are mostly unstable structure at ambient pressure, but it seems worth to research more about these candidates to utilize their good properties. We also look into some correlate relationship among the calculated variables, and it is expected that structural variety is a important key to find superior property materials.

This thesis also imply that this automation of *ab initio* calculation scheme of data mining can be applied to many other areas that need the new materials design. It must be emphasized that it is very important to carefully arranging the proper calculation method and optimization strategy for achieving every specific automation, but

once the automation is made well, it will have great potential to exploit the new materials design.

References

[1] ICSD : http://http://icsd.fiz-karlsruhe.de/

[2] VASP : http://vasp.at

[3] S.J. Wind, D.J. Frank, H.-S. Wong, Microelectronic Engineering 32 (1996) 271-282.

[4] Rovert W. Keyes, Proceedings of the IEEE, Vol. 63, No. 5, May (1975).

[5] John Robertson, Rep. Prog. Phys. 69 (2006) 327-396.

[6] John Robertson, J. Vac. Sci. Technol. B 18, 1785 (2000).

[7] P. Hohenberg, and W. Kohn, Phys. Rev. 136, B864 (1965).

[8] W. Kohn, and L. J. Sham, Phys. Rev. 140, A1133 (1965).

[9] J. P. Perdew, and W. Yue, Physical Review B 33, 8800 (1986).

[10] A. D. Becke, Phys. Rev. A 38, 3098 (1988).

[11] J. P. Perdew, M. Emzerhog, and K. Burke, Phys Rev Lett 77, 3865 (1996).

[12] J. P. Perdew *et al*, Phys. Rev. B 46, 6671 (1992).

[13] L. Hedin, Phys. Rev. 139, A796 (1965).

[14] A. V. Krukau *et al*, J Chem Phys 125 (2006).

[15] J. P. Perdew, M. Emzerhog, and K. Burke, J Chem Phys 105, 9982 (1996).

[16] J. P. Perdew, and M. R. Norman, Phys. Re. B 26, 5445 (1982).

[17] S. Park *et al*, Curr. Appl. Phys. 11, S337 (2011).

[18] M. S. Hybertsen and S. G. Louie, Phys. Rev. B. 35, 5585 (1987).

[19] M. Stadele, J. A. Majewski, P. Vogl, and A. Gorling, Phys. Rev. Lett. 79, 2089 (1997).

[20] L. J. Sham and T. M. Rice, Phys. Rev. 144, 708 (1966).

[21] I. Lukacˇ evic, Phys. Status Solidi B 248, No. 6, 1405– 1411 (2011)

[22] J. H. Yum *et al*. IEEE Transactions of electronic devices, Vol. 58. No. 12 (2011)

[23] B. Yu, Physica B. Condensed Matter. 404 (2009)

[24] A. Togo, *et al*. Phys. Rev. B 78 (2008)

국문초록

차세대 high-κ 물질 탐색을 위한 데이터 마이닝; 밴드갭 및 유전율 계산 자동화

필요로 하는 물성에 대해 존재하는 모든 물질에 대한 핸드북을 가지는 것은 모든 과학자들에게 꿈과 같은 과업니다. 최근 컴퓨터 성능의 비약적인 발전과 오랫동안 축적되어온 제일원리계산 기술의 발달은 방대한 이론적 물성 데이터를 만들 수 있는 새로운 길을 열어주고 있다. 각 물성에 적합한 계산 방법을 정교한 자동화 과정에 배치함으로써 빠르고 정확한 high-throughput 제일원리계산이 가능해질 수 있다. 사람의 직접적인 개입을 최소화함으로써 비교적 적은 전산자원으로도 대량의 물질들의 물성 데이터를 균일한 기준의 의해 얻어낼 수 있다. 물론 이때 주의할 점은 모든 계산 과정의 정확성과 최적화 과정이 세심한 테스트를 통해 검증되어야 한다는 점이다. 이러한 계산을 통해 얻어진 물성 데이터베이스를 이용하여 작성된 재료지도를 이용하면 기존에 기대하기 어려웠던 새로운 물성의 물질을 찾아내는 일이 가능해질 것이다. 그 첫번째 시도로써 우리는 high-k 유전물질에 적용하기 위한 새로운 물질을 찾기 위해 모든 산화물에 대한 밴드갭과 유전율을 계산하는 자동화 코드를 개발하였다. 이 코드는 ICSD 데이터베이스로부터 목표로 하는 모든 물질의 구조 데이터를 추출하여 제일원리계산을 위한 입력 데이터를 만들어내고, 그로부터 밴드갭과 유전율을 계산해낸다. 제일원리계산에는 널리 쓰이는 프로그램인 VASP 이 핵심 엔진으로 사용되었다. 계산 대상이 정해지면 모든 계산은 사람의 개입없이 완전히 자동적으로 이뤄지며, 각각의 물성 계산에 대한 신뢰성 테스트가 우선적으로 수행되었다. 그 결과 밴드갭 계산에는 우리가 새롭게 제안한 수정된 HSE06 방법이 사용되었고, 유전율 계산에는 LDA functional 계산을 사용하였다. 밴드갭 계산에 필요한 band edge position 을 찾는 데어도 새롭게 제안된 탐색 방법이 적용되었으며 각각의 계산이 실험과 비교하여 양질의 결과를 주는 것을 확인했다. 자동화 프로그램 전반에 걸쳐 중복된 데이터의 사전 필터링,

정확도를 유지하면서 최대한 계산 비용을 줄이는 과정 등, 주요 물성 계산을 위한 다양한 최적화 과정이 적용되어 있어 높은 효율과 정확도를 동시에 이뤄낼 수 있었다. 지금까지 약 1,000 개의 산화물에 대한 물성데이터가 축적되었으며 이러한 데이터로부터 새로운 재료지도를 작성하였다. 이 재료지도로부터 우리는 큰 밴드갭과 유전율을 동시에 가지는 새로운 물성의 후보 물질들을 찾을 수 있었으며, 계산된 물성들 사이의 상관 관계도 살펴볼 수 있었다.

주요어 : 데이터 마이닝, high-k, 밴드갭, 유전율, 제일원리계산

학번 : 2011-23320