



Cite this: DOI: 10.1039/d5tc02957k

Exploration of nonlinear optical materials by introducing information science

Kosuke Shirai,^a Tomoyuki Tamura,^{id} *^a Ming-Hsien Lee,^{id} *^b Bingbing Zhang^{id} ^c
and Masayuki Karasuyama^{id} ^d

A material search scheme is proposed herein to discover novel deep ultraviolet nonlinear optical (NLO) crystals. The scheme combines first-principles calculations of the optical properties, including birefringence and second-harmonic generation (SHG) susceptibility based on the density functional perturbation theory (DFPT), and prediction of the NLO properties using graph neural networks (GNNs). Systematic comparisons using a common computational code revealed that the DFPT method reproduced the experimental values better than the conventional sum-over-states (SOS) method. The accuracy of the scheme for predicting NLO properties was improved by including angle information in the crystal graph representation. By screening 6660 oxides with bandgaps greater than 4.0 eV in the Materials Project (MP) database, nine promising NLO crystals with dynamic stability were discovered. Moreover, new crystals are expected to be identified by extending the range of applications of the developed scheme from oxides to nitrides and sulfides.

Received 5th August 2025,
Accepted 10th October 2025

DOI: 10.1039/d5tc02957k

rsc.li/materials-c

1. Introduction

Nonlinear optical (NLO) second-harmonic generation (SHG) crystals of various wavelengths are important for frequency conversion in all-solid-state lasers, and are widely used in scientific and industrial applications.^{1,2} These crystals are classified according to their operating wavelengths, ranging from the ultraviolet (UV, 200–400 nm)/deep ultraviolet (DUV, <200 nm) to middle- and far-infrared (vis-NIR, 0.4–3 μm) regions. NLO crystals that operate efficiently in the DUV region are core components of frequency conversion to generate a DUV laser, which plays an important role in cutting-edge laser technology and fundamental science.³ Therefore, the exploration of new materials with suitable NLO properties is essential to meet the growing demands in this field.^{3–7}

Multiple properties of NLO crystals should be simultaneously optimised. These include a wide bandgap (E_g), large enough SHG response ($d_{ij} = \chi_{ij}^{(2)}/2$), and large birefringence (Δn). The recent introduction of computational science has promoted efforts to reduce the number of experiments required. Databases of properties calculated using density

functional theory (DFT), such as the materials project (MP)⁸ and the open quantum materials database (OQMD),⁹ have been developed for the predictive identification of novel crystals with desired properties. Although physical properties such as E_g , d_{ij} , and Δn can be calculated using DFT, there is no database with sufficient DFT calculations for d_{ij} and Δn . The high-throughput calculation (HTC) screening of NLO crystals using their optical properties has been reported for borates,¹⁰ germanates,¹¹ and silicates.¹² The application of machine learning (ML) frameworks in materials science has attracted considerable attention in recent years, and the screening of NLO crystals by combining HTC and ML has been reported.¹³

The NLO susceptibility can be calculated using the sum-over-states (SOS) method as a straightforward extension of the band structure and density-of-states calculations within DFT. In the SOS method, the momentum matrix elements of the valence and conduction bands must be calculated and used in the band summation formalism.^{14,15} Recently, optical property calculations based on the density functional perturbation theory (DFPT) have become available within the same CASTEP code.¹⁶ Therefore, the two methods and experiments can be directly compared.¹⁷ The HTC of the NLO susceptibility and dielectric tensor of 579 inorganic semiconductors were performed¹⁸ based on the DFPT using the ABINIT code.¹⁹ In this study, the data were systematically compared with previous high-throughput results. However, the computational cost of the DFPT method is significantly higher than that of the SOS method, making the application of the former to HTC screening resources demanding.

^a Division of Applied Physics, Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan. E-mail: tamura.tomoyuki@nitech.ac.jp

^b Department of Physics, Tamkang University, Tamsui, New Taipei 25137, Taiwan. E-mail: mhslee@mail.tku.edu.tw

^c College of Chemistry and Environmental Science, Hebei University, Baoding 071002, China

^d Division of Computational Intelligence, Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan

The combination of HTC and ML can reveal fundamental structure–property relationships between materials and provide an extensive chemical space for exploring new materials at a low computational cost. Traditionally, prediction methods have been based on general regression methods, such as linear regression, support vector machines (SVM), and random forest regression (RFR), using structural and compositional descriptors. New models such as deep learning using universal atomic graph representations have been developed. The application of the former regression to NLO crystals for ML screening has been reported;¹³ however, the NLO properties were calculated using the standard SOS method. Furthermore, the DFPT method is several times more computationally expensive than the SOS method, and there are no reported examples of ML screening.

In this study, new NLO crystals that satisfy the required E_g , d_{ij} , and Δn were searched by combining the HTC of DFPT optical property calculations with ML predictions based on atomic graph representations. All the oxides registered in the crystal database were comprehensively screened for NLO crystals.

2. Computational methods

A process flow for exploring NLO materials was proposed, in which optical property calculations based on DFPT and ML were combined (Fig. 1). d_{ij} and Δn were determined *via* predictive modelling and applied to the search for NLO crystals to reduce the search time. The predictive modelling of the NLO properties consists of three steps: (1) acquisition of crystal structure data, (2) HTC of d_{ij} and Δn , and (3) ML prediction of d_{ij} and Δn . The process for screening the NLO materials consists of five steps: (4) retrieving candidate structures for exploration, (5) screening based on E_g , (6) determination of the non-inversion centre symmetry (NICS), (7) screening based on the predicted d_{ij} and Δn , and (8) high-precision calculation of

the optical properties of the screened crystal structures using the DFPT method.

CASTEP code¹⁶ was used to calculate the self-consistent field (SCF), optical parameters, and phonons with norm-conserving pseudopotentials. The Perdew–Burke–Ernzerhof (PBE) functional²⁰ within the generalised gradient approximation (GGA) was used for material screening and the Heyd–Scuseria–Ernzerhof (HSE06) hybrid functional²¹ was used for high-precision calculations after material screening. The scissor operator, which was set as the difference in the bandgap between the GGA and HSE, was used to correct the SHG coefficients. The plane-wave energy cutoff was set to 800 eV. SCF calculations were performed with a convergence criterion of 1×10^{-6} eV per atom on the total energy. The Monkhorst–Pack k -point separation was set as $0.07 \text{ \AA}^{-1}$ in the Brillouin zone for the SCF and $0.04 \text{ \AA}^{-1}$ for the calculation of the optical parameters and phonons.

Based on the $2n + 1$ theorem,²² the derivative of the $2n + 1$ -th order of energy with respect to the perturbation (which can be an atomic displacement or electric field) depends only on the n -th order perturbed wavefunctions. Within the DFPT framework, the conjugate-gradient minimisation algorithm^{23,24} can be used to find solutions iteratively with high efficiency. In the birefringence (Δn) calculation, the refractive indices are determined by taking the square root of the three (diagonal) elements of the dielectric tensor on the principal axis. (As long as the crystal is properly oriented, as is usually the case with symmetry-annotated data from popular crystal structure databases, the calculated dielectric tensor is already diagonal). The dielectric functions and d_{ij} are related to the second- and third-order energy perturbation with respect to the external electric field; therefore, a search of the first-order ($n = 1$) wavefunction only is sufficient for both, which is why the scheme is also called the “Linear Response” method. CASTEP implements a linear-response DFPT that supports the GGA exchange–correlation functional.¹⁷ The formalism suggested by Miwa²⁵ was adapted for the NLO susceptibility coefficient, d_{ij} .

Recently, methods based on crystal graph convolutional neural networks (CGCNNs)²⁶ have been widely utilised to explicitly describe the local environment of crystals through graphical representations and predict the properties of solids and molecules. However, conventional graph neural networks (GNNs) used for crystalline systems, such as CGCNNs, do not encode interatomic angle data. The linear and nonlinear optical properties considered in this study are expected to be sensitive to three-body-term information, such as the bond angles between atoms and local polyhedral anisotropy. In contrast, the atomic line graph neural network (ALIGNN) approach,²⁷ which explicitly represents the bond angles as edges of line graphs, provides a general formulation that can be applied to periodic crystal graphs. It has been reported that explicitly including angle-based information in ALIGNNs and expressing the importance of such many-body interactions improve the accuracy of property predictions. Therefore, in this study, ALIGNNs that take the angle term into account are used to predict d_{ij} and Δn . The angle dependence of these property predictions is also discussed in comparison with those of the CGCNN.

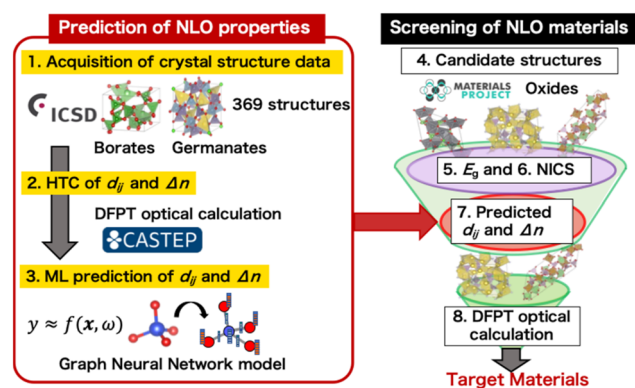


Fig. 1 Schematic of exploration of nonlinear optical crystals based on machine learning. (1) Acquisition of crystal structure data, (2) high-throughput calculations of d_{ij} and Δn , and (3) optimisation of parameters of the ML prediction model for d_{ij} and Δn . (4) Screening and retrieval of candidate structures for exploration (5) based on E_g , (6) non-inversion center symmetry, and (7) predicted d_{ij} and Δn . (8) Calculation of optical properties with high precision using the DFPT method for screened crystal structures.

3. Results and discussion

3.1. Prediction of NLO properties

To clarify the differences in the methods of calculating the optical properties, the DFPT method was used to calculate d_{ij} and Δn for ten representative NLO crystal structures for which the experimental values were obtained. The d_{ij} coefficient represents the maximum of $|\chi_{ij}^{(2)}|/2$ and was corrected using a scissor operator. The results are shown in Fig. 2. Here, the values calculated using the SOS method with scissor correction and experimental values were obtained from a published paper on NLO crystals (Table 1 in Ref. 10). The detailed values are presented in Table S1 (SI). The values obtained by LDA-based DFPT calculations¹⁸ are also listed. The mean absolute percentage errors (MAPEs) for the optical calculations using the SOS and DFPT methods with scissor correction compared to the experimental values were 32.8 and 16.7% for d_{ij} and 32.0 and 26.3% for Δn , respectively. In the SOS formalism for calculating d_{ij} ,^{14,15} the band energy differences between the valence and conduction bands appear in the denominator. Therefore, an underestimation of the band gap inevitably leads to an overestimation of d_{ij} . Compared with the experimental values, the mean percentage errors (MPEs) of the band gaps obtained using the GGA and HSE functionals were -28.8% and $+3.1\%$, respectively. Within the SOS framework, it has been demonstrated that the application of scissor correction brings d_{ij} into good agreement with experimental values.¹⁰ For GaAgS_2 , GaAgSe_2 , and ZnGeP_2 , which exhibit large d_{ij} values, LDA-based DFPT calculations overestimate the experimental results by more than 50%. In contrast, the DFPT calculations with scissor correction performed in this study showed deviations of no more than 16% from the experimental results for the three materials. Thus, scissor correction was demonstrated to be effective within the DFPT framework, yielding d_{ij} values that are closer to the experimental values than those obtained using the SOS method with scissor correction. As confirmed for borates, the d_{ij} value decreased markedly as the band gap increased.¹⁰ This implies that the impact of the bandgap underestimation diminished as the band gap increased. Consequently, in the DUV region, the d_{ij} value is expected to be small, and the absolute error relative to the experimental value remains limited.

A total of 369 borate and germanate structures from previous papers^{10,11} were selected for HTCS. Structural data for

these crystals were obtained from the Inorganic crystal structure database (ICSD)²⁸ and preliminarily screened for centre-of-inversion symmetry. Disordered and incomplete structures such as those of solid solutions with disordered positions and structures containing positionless H atoms were excluded. For these 369 structures, high-throughput optical property calculations were performed using the DFPT method to obtain the d_{ij} and Δn values. During the preparation of this manuscript, an in-depth comparison of the birefringence determined using the SOS and DFPT methods was published by one author.²⁹

The crystal-system classification of d_{ij} and Δn is shown by the violin diagram in Fig. 3(a) and (b). The distribution of these 369 crystal structures covers the trigonal, triclinic, tetragonal, orthorhombic, monoclinic, and hexagonal systems; hence, the dataset can be regarded as reasonable training data. Most of the birefringence (Δn) values in the dataset are less than 0.1. This bias in the data has a significant potential to affect the accuracy of predictions. Therefore, the birefringence on a natural logarithmic scale ($\ln(\Delta n)$),¹³ for which the dataset is closer to a normal distribution, was used as a label for the ML predictions to improve the results of the ML model. The data distribution of the d_{ij} is also similar to that of Δn ; therefore, a logarithmic transformation was performed. Data with d_{ij} and Δn values close to zero were excluded. Ultimately, 287 structures for d_{ij} and 246 structures for Δn were employed in the regression prediction. The respective data distributions are shown in Fig. 3(c) and (d). The corresponding conversions from the logarithmic to linear scale are shown in Fig. S1.

A regression model based on a GNN was constructed using a crystal graph as the input and d_{ij} and Δn as objective variables. The training and test data were randomly divided in an 8 : 2 ratio. CGCNN, which contains information on elements and distances, was applied to the crystal graphs along with ALIGNN, which contains information on the angles, elements, and distances. The regression conditions are listed in Table S2. The MAE transitions for d_{ij} and Δn during the learning process of CGCNN and ALIGNN are shown in Fig. S2. When the number of optimisation epochs was set to 200, the MAE of the test data for $\ln(d_{ij})$ was 0.53 for CGCNN and 0.47 for ALIGNN. Meanwhile, the MAE for $\ln(\Delta n)$ was 0.41 and 0.35 for CGCNN and ALIGNN, respectively. Fig. 4 presents a comparison between the predicted and calculated log-scaled values for the training and test data obtained using CGCNN and ALIGNN. The corresponding conversions from the logarithmic to linear scale are shown in Fig. S3. Taken together, these results demonstrate that the regression performance of ALIGNN is superior to that of CGCNN. Most of the test data were accurately predicted, although outliers were present. The outliers can be attributed to extrapolation due to a lack of training data. Further validation is required to improve the accuracy of the prediction by increasing the amount of training data.

3.2. Screening of NLO materials

NLO crystals were screened using the predictive model for d_{ij} and Δn , which was constructed as described above. The screening conditions are shown in Fig. 5. As mentioned above, it is

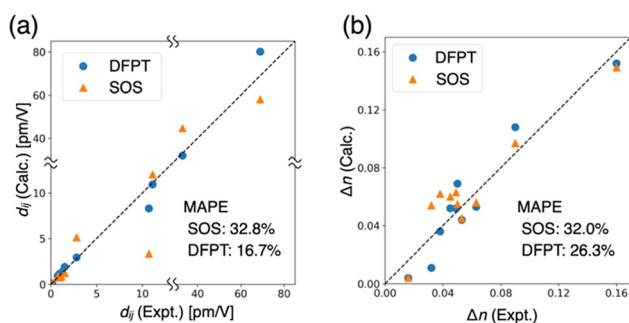


Fig. 2 Comparison of (a) d_{ij} and (b) Δn calculated using the SOS and DFPT methods with experimental values.

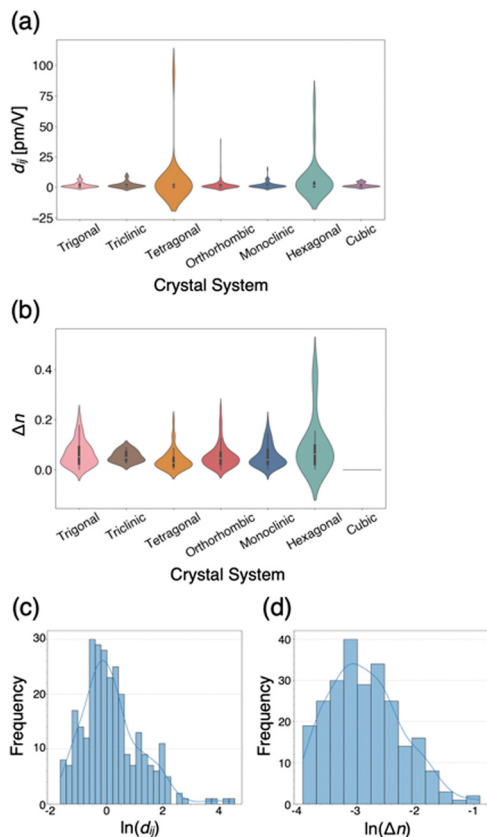


Fig. 3 Violin plot grouped by different crystal systems depicting distributions of (a) d_{ij} and (b) Δn calculated using the DFPT method. Histogram and density curves representing the distributions of calculated (c) log-scaled d_{ij} and (d) log-scaled Δn .

well-established that DFT-GGA underestimates the bandgap by approximately two-thirds. Because the wavelength of DUV light (200 nm) corresponds to 6.2 eV, two-thirds of this value is approximately 4.1 eV. Accordingly, the bandgap screening threshold for DFT-GGA was set to 4.0 eV. A total of 6660 oxide materials with bandgaps greater than 4.0 eV can be found in the MP database. A total of 2504 structures with non-inversion centre symmetry were retained based on the space group of the crystals. These structures were transformed into crystal graphs, and d_{ij} and Δn were predicted. For screening DUV NLO crystals, thresholds of $d_{ij} > 0.39 \text{ pm V}^{-1}$ and $\Delta n > 0.06$ have typically been employed.¹⁰ However, as demonstrated above, the predicted values for d_{ij} and Δn were systematically overestimated. Therefore, taking the experimental value of $d_{ij} = 1.23 \text{ pm V}^{-1}$ for $\beta\text{-BaB}_2\text{O}_4$ as a reference, we adopted revised thresholds of $d_{ij} > 2.0 \text{ pm V}^{-1}$ and $\Delta n > 0.05$. Based on these predictions, 155 structures were retained.

The band gaps of the screened structures were calculated. This step is usually unnecessary; however, incorrect data may exist in the MP database in rare cases, for example, owing to a lack of reference papers. To eliminate such data, crystal structures with a bandgap of 4.0 eV or greater were selected using DFT calculations employing GGA functionals. An error of $\pm 0.2 \text{ eV}$ was allowed, and 59 crystal structures were screened. Using the DFPT method, the

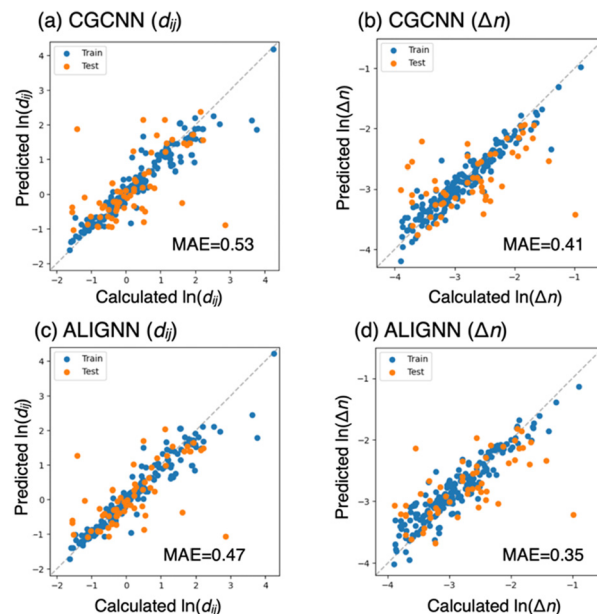


Fig. 4 Comparison between the log-scaled values predicted by CGCNN and ALIGNN and the corresponding calculated values. Panels (a) and (c) show d_{ij} , while panels (b) and (d) show Δn .

optical parameters were calculated for these crystal structures to evaluate the values of d_{ij} and Δn with good accuracy. The predicted and calculated values of log-scaled d_{ij} and Δn for the 59 screened structures are plotted in Fig. 6. For comparison, the predicted and calculated values of non-log-scaled d_{ij} and Δn are provided in Fig. S4. The predicted values of both d_{ij} and Δn were overestimated, indicating unsatisfactory prediction accuracy, which is likely attributable to the lack of training data. Nevertheless, given that the screening targets are crystal structures with inherently high d_{ij} and Δn , such overestimations can be regarded as acceptable.

The d_{ij} and Δn values obtained by DFPT calculations for the 59 screened crystal structures are plotted in Fig. 7. Defining the screening criteria $d_{ij} > 1 \times \text{KH}_2\text{PO}_4$ (KDP) = 0.39 pm V^{-1} ¹⁰ and $\Delta n > 0.04$, 13 structures were found to meet the criteria. Phonon calculations were also performed to assess the dynamic stability of the crystal

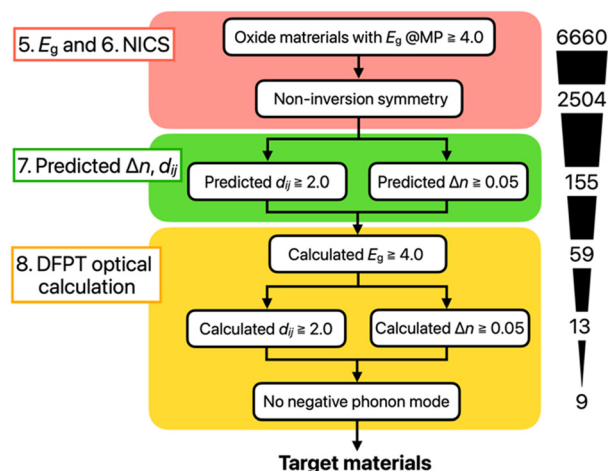


Fig. 5 Conditions for screening NLO materials. The labels are the same as in Fig. 1.

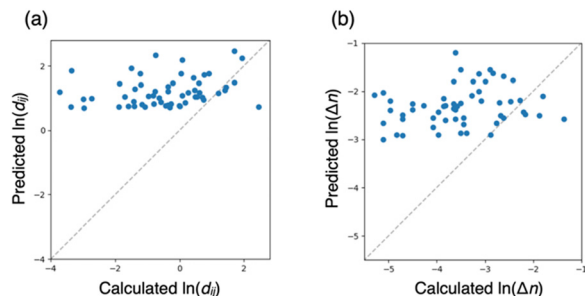


Fig. 6 Plots of predicted and calculated (a) log-scaled d_{ij} and (b) log-scaled Δn for the 59 screened structures.

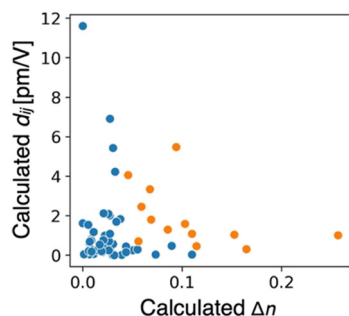


Fig. 7 Calculated d_{ij} and Δn for 59 screened structures. The orange dots indicate the 13 crystals satisfying the criterion values of $d_{ij} > 0.39 \text{ pm V}^{-1}$ and $\Delta n > 0.04$.

structures and verify the presence of negative modes. Consequently, nine structures with dynamic stabilities were identified. The phonon dispersion spectra of these structures are presented in Fig. S5–S7.

The physical properties of nine screened structures were determined using high-precision calculations. The bandgap was evaluated by band calculations using the HSE functional, and d_{ij} and Δn were determined by DFPT calculations with scissor correction. The results are summarised in Table 1. When the accuracy of the calculations increased, a small number of crystals did not satisfy the screening criteria. The structures identified in this study can be classified into three main types: (a) silicon oxynitride, (b) transition-metal oxides, and (c) arsenic sulfide crystals.

3.3. Screened materials

3.3.1. Silicon oxynitride crystals

3.3.1.1. $\text{Si}_2\text{N}_2\text{O}$ and LiSiNO . As shown in Fig. 8, the crystal structures of the silicon oxynitrides consist of a Si tetrahedron

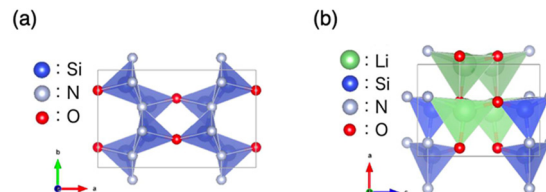


Fig. 8 Crystal structures of (a) $\text{Si}_2\text{N}_2\text{O}$ and (b) LiSiNO .

composed mainly of nitrogen and oxygen, which were identified experimentally. Both crystals exhibit no negative phonon modes and demonstrate favourable NLO properties, suggesting that silicon oxynitride-based crystals are promising NLO crystals with dynamic stability. However, $\text{Si}_2\text{N}_2\text{O}$ and LiSiNO have recently been reported as good candidates for NLO materials that are active in the UV range.¹² Therefore, the results indicate the usefulness of the material search scheme developed in this study, although it did not lead to the discovery of new materials.

3.3.1.2. Transition metal oxide crystals. Six transition-metal oxide crystal structures were also screened. These crystals are composed of transition metals such as Y, Sc, Hf, and Nb (Fig. 9), where most of the structures were obtained from first-principles calculations.

3.3.1.3. $\text{Y}_5\text{O}_4\text{F}_7$. This metal oxide crystallises in the monoclinic Pc space group and comprises five different yttrium polyhedra composed of oxygen and fluorine atoms. Although the birefringence is reduced by scissor correction, both the bandgap and SHG coefficient are sufficiently high to meet the criteria.

3.3.1.4. $\text{Y}_6\text{O}_5\text{F}_8$. This compound crystallises in the triclinic $P1$ space group and consists of six different yttrium polyhedra composed of oxygen and fluorine atoms. The maximum SHG coefficient, bandgap, and birefringence are -1.23 pm V^{-1} , 6.93 eV, and 0.078, respectively, making it a promising NLO material with a good bandgap, which satisfies all NLO criteria.

3.3.1.5. $\text{BaHf}(\text{BO}_3)_2$. This compound crystallises in the triclinic $R3c$ space group, with 6-coordinated Ba atoms and 8-coordinated Hf octahedra linked by BO_3 groups; here, parallel coordination of the BO_3 groups with the b axis is characteristic. This is the only one of the six transition-metal oxide structures discovered in this study that has been identified experimentally; although the SHG coefficient is reduced by scissor

Table 1 Calculated parameters for nine screened crystal structures

No.	Formula	MP-ID	SG	E_g (eV) GGA/HSE	$\Delta n/d_{ij}^{(2)}$ (pm V^{-1})
a1	$\text{Si}_2\text{N}_2\text{O}$	4497	$Cmc2_1$	5.10/6.74	$0.094/d_{13} = d_{35} = 0.89$
a2	LiSiNO	6015	$Pca2_1$	5.33/6.50	$0.108/d_{13} = d_{35} = 0.41$
b1	$\text{Y}_5\text{O}_4\text{F}_7$	675 109	$P1c1$	4.55/6.39	$0.023/d_{33} = -3.13, d_{31} = d_{15} = 2.89, d_{22} = -2.66$
b2	$\text{Y}_6\text{O}_5\text{F}_8$	676 605	$P1$	4.94/6.93	$0.078/d_{14} = d_{25} = d_{36} = -1.23, d_{11} = -0.63$
b3	$\text{BaHf}(\text{BO}_3)_2$	1 105 495	$R3c$	4.34/6.16	$0.121/d_{33} = 0.24$
b4	YScSi_2O_7	1 207 607	$C121$	4.51/6.53	$0.045/d_{32} = d_{24} = 0.49, d_{31} = d_{15} = -0.48$
b5	Hf_2GeO_8	1 212 158	$I42m$	4.34/6.14	$0.077/d_{22} = -3.92, d_{13} = d_{35} = -0.48, d_{11} = -2.64$
b6	NaHfScO_4	1 220 764	$Pmc2_1$	4.03/6.18	$0.123/d_{12} = -0.77, d_{23} = d_{34} = 0.48$
c1	As_2SO_6	27 230	$P2_12_12$	4.21/5.57	$0.213/d_{14} = d_{25} = d_{36} = 0.78$

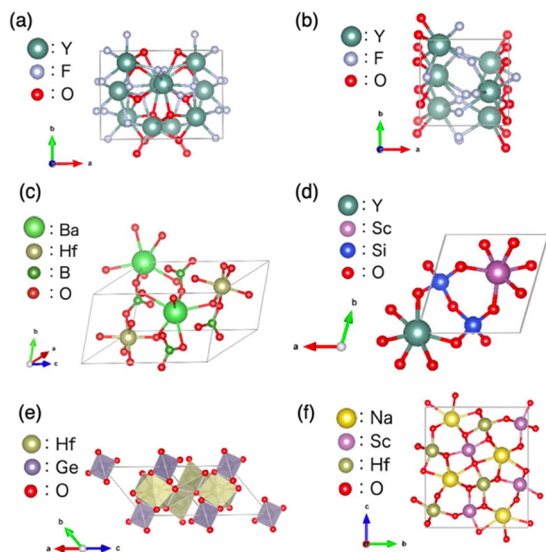


Fig. 9 Crystal structures of (a) $\text{Y}_5\text{O}_4\text{F}_7$, (b) $\text{Y}_6\text{O}_5\text{F}_8$, (c) $\text{BaHf}(\text{BO}_3)_2$, (d) YScSi_2O_7 , (e) Hf_3GeO_8 , and (f) NaHfScO_4 .

correction, $\text{BaHf}(\text{BO}_3)_2$ has a bandgap of 6.16 eV and large birefringence of 0.121.

3.3.1.6. YScSi_2O_7 . This compound crystallises in the monoclinic $C121$ space group, with 6-coordinated Y atoms sharing corners with six equivalent SiO_4 tetrahedra to form a distorted YO_6 pentagonal pyramid with three equivalent ScO_6 octahedra. It is a promising material that satisfies all NLO properties.

3.3.1.7. Hf_3GeO_8 . The compound crystallises in the tetragonal $I\bar{4}2m$ space group and consists of eight O atoms, Hf atoms bonded in a distorted body-centred cubic structure, and Hf atoms bonded in a geometric structure connected to a tetrahedral GeO_4 group. The maximum SHG coefficient, bandgap and birefringence of -3.92 pm V^{-1} , 6.14 eV, and 0.053, respectively, make it a promising material that satisfies all NLO properties. This crystal is not among the ones screened by the high-throughput computational screening of germanates.¹¹

3.3.1.8. NaHfScO_4 . This compound crystallises in the orthorhombic $Pmc2_1$ space group and consists of a Hf octahedron and a Sc octahedron linked to an O atom, a Na atom with seven coordination positions, and six O atoms that share edges and corners. The maximum SHG coefficient, bandgap, and birefringence are -0.77 pm V^{-1} , 6.18 eV, and 0.123, respectively, making it a promising NLO material with a large birefringence.

3.3.2. Arsenic sulfide crystals

3.3.2.1. As_2SO_6 . The structure crystallises in the orthorhombic $P2_12_12$ space group and consists of a S tetrahedron with four O atoms; the As atoms are connected to form a triangular pyramid composed of three O atoms, as shown in Fig. 10(a). The crystal structures were determined experimentally. The values of E_g , d_{ij} , and Δn are 5.57 eV, 0.78 pm V^{-1} , and 0.213, respectively, making As_2SO_6 a promising NLO material with a very high Δn value, although E_g is slightly smaller than 6.2 eV (=

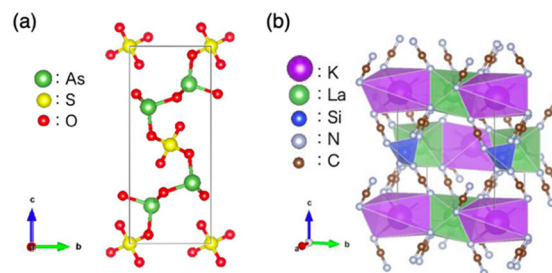


Fig. 10 Crystal structures of (a) As_2SO_6 and (b) $\text{KLaSi}(\text{CN})_4$.

200 nm) in the DUV region. However, as the structure contains As atoms, the safety of the material must be considered.

To search for alternative materials with similar structures, inorganic crystals in the same space group ($P2_12_12$) as As_2SO_6 were investigated using the MP database, where the carbonitride crystal system $\text{KLaSi}(\text{CN})_4$ (mp-567129) showed high NLO properties. The crystal structure is shown in Fig. 10. This crystal, called tetracyanamidosilicate, is an experimentally identified inorganic crystal in which nitrogen-coordinated cations are linked *via* carbon. The calculated E_g -GGA, d_{ij} , and Δn are 4.02 eV, 1.02 pm V^{-1} , and 0.34, respectively. $\text{KLaSi}(\text{CN})_4$ has a high Δn , similar to As_2SO_6 . Although $\text{KLaSi}(\text{CN})_4$ was not included as a screening target in this study because it is not an oxide, promising materials are expected to be discovered in the future by expanding the compositional range to include nitrides and carbon-based materials.

4. Conclusions

A material search scheme was proposed for the discovery of novel DUV NLO crystals by combining first-principles calculations of the optical parameters (birefringence and SHG susceptibility) based on DFPT and the prediction of NLO properties based on GNNs. Systematic comparisons using a common computational code revealed that the DFPT method reproduced the experimental values better than the SOS method. The accuracy of the scheme for predicting the NLO properties was improved by including angle information in the crystal graph representation. A total of 6660 oxides with E_g greater than 4.0 eV were selected from the MP database, among which nine were promising NLO crystals exhibiting dynamic stability. The present study focuses on oxides, and new crystals are expected to be unearthed by extending the range of applications of the developed scheme to nitrides and sulfides. The proposed material design scheme is expected to accelerate the rational design of NLO materials as well as various functional materials, thereby contributing to the realisation of desirable composite properties.

Author contributions

The manuscript was written with contributions from all authors. All the authors approved the final version of the manuscript. K. S.: data curation, formal analysis, visualisation, writing – original draft; T. T.: conceptualisation, methodology,

writing – review, and editing; M. L.: conceptualisation, methodology, writing – review, and editing; B. Z.: Data curation; T. K.: conceptualisation

Conflicts of interest

The authors declare no competing interests.

Data availability

The raw data required to reproduce these findings are available on request.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d5tc02957k>.

Acknowledgements

This work was supported by the MEXT KAKENHI grants awarded to T. T. (22H01808) and M. K. (21H03498). Figures were drawn using VESTA (<https://jp-minerals.org/vesta/en/>). The authors thank Editage (<https://www.editage.jp>) for the English proofreading.

References

- 1 T. T. Tran, H. Yu, J. M. Rondinelli, K. R. Poeppelmeier and P. S. Halasyamani, *Chem. Mater.*, 2016, **28**, 5238–5258.
- 2 S.-P. Guo, Y. Chi and G.-C. Guo, *Coord. Chem. Rev.*, 2017, **335**, 44–57.
- 3 L. Kang and Z. Lin, *Light: Sci. Appl.*, 2022, **11**, 206.
- 4 M. Mutailipu, M. Zhang, H. Wu, Z. Yang, Y. Shen, J. Sun and S. Pan, *Nat. Commun.*, 2018, **9**, 3089.
- 5 X. Liu, Z. Yang and S. Pan, *J. Appl. Phys.*, 2024, **135**, 120902.
- 6 J.-X. Zhang, Q.-G. Yue, S.-H. Zhou, X.-T. Wu, H. Lin and Q.-L. Zhu, *Angew. Chem., Int. Ed.*, 2024, **63**, e202413276.
- 7 P. Feng, J.-X. Zhang, M.-Y. Ran, X.-T. Wu, H. Lin and Q.-L. Zhu, *Chem. Sci.*, 2024, **15**, 5869.
- 8 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
- 9 J. E. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, *JOM*, 2013, **65**, 1501–1509.
- 10 B. Zhang, X. Zhang, J. Yu, Y. Wang, K. Wu and M.-H. Lee, *Chem. Mater.*, 2020, **32**, 6772–6779.
- 11 J. Yu, B. Zhang, X. Zhang, Y. Wang, K. Wu and M.-H. Lee, *ACS Appl. Mater. Interfaces*, 2020, **12**, 45023–45035.
- 12 X. Zhang, L. Guo, B. Zhang, J. Yu, Y. Wang, K. Wu, H.-J. Wang and M.-H. Lee, *Chem. Commun.*, 2021, **57**, 639–642.
- 13 M. Wu, E. Tikhonov, A. Tudi, I. Kruglov, X. Hou, C. Xie, S. Pan and Z. Yang, *Adv. Mater.*, 2023, **35**, 2300848.
- 14 J. Lin, M.-H. Lee, Z.-P. Liu, C. Chen and C. J. Pickard, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **60**, 13380.
- 15 B. Zhang, M.-H. Lee, Z. Yang, Q. Jing, S. Pan, M. Zhang, H. Wu, X. Su and C.-S. Li, *Appl. Phys. Lett.*, 2015, **106**, 031902.
- 16 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567–570.
- 17 K. Refson, P. R. Tulip and S. J. Clark, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 155114.
- 18 V. Trinquet, F. Naccarato, G. Brunin, G. Petretto, L. Wirtz, G. Hautier and G.-M. Rignanese, *Sci. Data*, 2024, **11**, 757.
- 19 A. H. Romero, *et al.*, *J. Chem. Phys.*, 2020, **152**, 124102.
- 20 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 21 A. V. Krukau, O. A. Vydrov, A. F. Izmaylov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 224106.
- 22 X. Gonze, *Phys. Rev. A*, 1995, **52**, 1086–1095.
- 23 X. Gonze, D. C. Allan and M. P. Teter, *Phys. Rev. Lett.*, 1992, **68**, 3603–3606.
- 24 X. Gonze, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1997, **55**, 10337–10354.
- 25 K. Miwa, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 094304.
- 26 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
- 27 K. Choudhary and B. DeCost, *npj Comput. Mater.*, 2021, **7**, 185.
- 28 National Institute of Standards and Technology (NIST), Inorganic Crystal Structure Database, DOI: [10.18434/M32147](https://doi.org/10.18434/M32147).
- 29 Y.-J. Hsieh and M.-H. Lee, *Comput. Condens. Matter*, 2025, **44**, e01066.