الجمـــهوريـــة الجـــزائريـة الديمقراطيـــة الشـــعبيـــة

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

وزارة التعـلـيـــم العـــالـــي والبـحث الـعلـــمـي

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جـامعـــة عـمـــار ثـلـيـجـي بالأغواط

UNIVERSITY OF AMAR TELIDJI LAGHOUAT

كلـيـــــة العـلـــــوم

FACULTY OF SCIENCES

قسم علوم المادة

DEPARTEMENT OF MATERIAL SCIENCES

Master's dissertation

Sector: Material Sciences

Option: Materials Physics

Presented by: Ms. MECHRAOUI Kaima Bouchra

THEME

Data mining of physical properties and non-linear optical parameters for materials enhancement

Board of jury:

LEFKAIER Iben Khaldoun	University of Laghouat-ALGERIA	Professor	President
HEBBOUL Zoulikha	University of Laghouat-ALGERIA	MCA	Examiner
BENGHIA Ali	University of Laghouat-ALGERIA	MCA	Supervisor
FERCHANE Selmane	University of Freiburg-GERMANY	PhD.Student	Co-supervisor

2021/2022

Acknowledgments

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

I would like to express my deep and sincere gratitude to my research supervisor, Dr.BENGHIA Ali and Co-supervisor PhD.FERCHANE Selmane, for giving me the opportunity to do research and providing invaluable guidance throughout this research. Their dynamism, vision, sincerity and motivation have deeply inspired me. They has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work under their guidance.

I would like to thank the members of my dissertation committee Mr.LEFKAIER Ibn Khaldoun, professor at university of Laghouat and Mrs.HEBBOUL Zoulikha master conference at university of Laghouat, for accepted to be members of my dissertation committee and for their time and extreme patience.

My sincere thank goes to all my lab-mates in the laboratory of Materials physics, especially Dr.FADLA Mohamed abdelilah for his help and support.

Last but not the least, I would to thank all the professors of department of material sciences, especially Pr.HELIFA Bachir. And all those who have helped me and supported me in this work.

To my family, without their help, love, and devotion, this work would not exist. Thank you

For Amina, Ahmed, and Assil: Remember that the quickest path to easy is through hard.

Figures list

Tables list

Abbreviations list

NLO: Non-Linear optics

SHG: Second harmonic generation

Mid-IR: Middle infrared

SEM: Scanning electronic Microscopy

MRI: Magnetic resonance imaging

LIDT: Laser damage induced threshold

KDP: Monopotassium phosphate KH_2PO_4

KTP: Potassium titanyl phosphate $KTiOPO_4$

BBO: Barium borate β-Ba_2BO_4

LBO: Lithium triborate LiB_3O_5

DL-Structure: Diamond-Like structure

Eg: gap energy

PCA: Principal component analysis

PLS: Partial least square

Summary

General Introduction

The discovery of materials with a second harmonic generation (SHG) by Franken and his colleagues in 1961[1], had opened a new era of non-linear optical materials in research and technological applications. Second order non-linear materials are of a great importance for many applications in civil and military technological sectors. Particularly, Mid-Infrared NLO crystals have attracted much of interest due to their indispensable role in obtaining laser pulses with Mid-Infrared wavelengthens. As this is still active research area to deal with the challenges to improve and develop new NLO materials.

The experimental synthesis and characterization combined with the first principal calculations of NLO crystals have been one the main research interests in our university since two decades. The first successful synthesis of $Hg(IO_3)_2$ was made by D. Benbertal and B. Bentria in 2001, whom they had been engineering quadratic nonlinear optical iodate [2, 3]. A.Gacemi and his co-workers later investigated $NH_4[HgCl_2(SCN)_2]$, $NH_4[HgCl(SCN)_2]$, and $K_2[HgBr_2(SCN)_2]$ [4, 5]

The preparation and the XRD characterization of the following materials, namely; $M(IO_3)_2$ with (M: Ni, Mg, Mn, Co, Zn) [6], Cadmium Iodate Monohydrate $Cd(IO_3)_2 \cdot H_2O$ [7], and polymorphism of Anhydrous Cadmium Iodate crystals, ε-$Cd(IO_3)_2$ [8] by B. Bentria *et al*, helped to extend the family of the examined NLO materials to cover more materials like the iodates seeking materials with advantageous higher band gap but still to enhance their SHG effect. Moreover, among many materials that was examined experimentally, where M. B. Taouti and his collaborators worked on many materials to mention the relevant once, like the synthesis of $La(IO_3)_3(HIO_3)$ which characterized by single crystal XRD [9]**,** and the characterization of the solid solution $Zn_{1-x}Cu_x(IO_3)_2$ using the Kurtz-Perry method [10] to test the SHG effect and compared to with KH_2PO_4 crystal (KDP), where the SHG effect of each material is in the following order: KDP ≈ α-$Cu(IO_3)_2$ < $Zn_{1-x}Cu_x(IO_3)_2$ < α-$LiIO_3$ < $Zn(IO_3)_2$ [11].

The rise of powerful computers in the recent years helped scientists to implement theoretical methods using sophisticated algorithms, to be what is known today as *ab initio* methods, very like an experimental set up inside a computer. This third pole called simulation, is to bridge theory and experiment for faster predictions and low cost effort with good agreement as in experiment, reducing the high cost to synthesis and characterize NLO materials with targeted properties, including time, study effort using semi-empirical models, plus the price tag of the experiment …etc.

The Materials Physics Laboratory in the University of Laghouat invested wisely to offer such research tools to help to combine *ab initio* methods with experimental work for low cost and high research performance. Materials studio, is one of the main packages used to study new crystal models and analyzing their properties. Including many modules with multiple applications. Plus other open source packages used to help to foster the discovery of new materials with NLO properties.

The nice idea of mixed ligand that has been brought to the iodates materials by Z. Hebboul and co-workers [12], plus the synthesis where two methods are used to obtain and characterize the new phase ζ - $Cd(IO_3)_2$ [13], plus the great work that was dedicated to study the iodates family of materials, both experimentally and DFT calculations, namely $LiZn(IO_3)_3$ with a wide transparency and chemical stability which make it a good candidate for NLO applications, and $La(IO_3)_3$ with a new simple low cost method of characterization [14].

On the other hand and for the effort to search for better NLO materials with enhanced properties other research groups worked on the investigation of new NLO materials using only *ab initio* methods without experimental work for low cost search. To name few of these efforts. B. Lagoun *et al*, conducted a first principal investigation of the structural, electronic, and optical properties of Hg(IO3)2 [15], where it was in a good agreement with the experimental work on the same crystal [3]. Thus, theoretical modeling becomes of great importance that can help to back up experimental work and beyond to nice predictions. Using such advantage to go beyond verifying experimental findings and try to search for new class of NLO family such as chalcogenides. A. Benghia *et al*, investigated the barium-based chalcogenides family, and aluminum-thiophosphate [16, 17]. As the calculation of SHG coefficients is computationally expensive to calculate, while using the DFPT functional takes around 6-12 weeks roughly.

From the era of *ab initio* methods that runs in powerful computers to the new era of machine learning and data mining. Where machine learning models is used for the prediction of new materials with NLO properties, opening a new axis of materials discovery using faster and efficient methods based on advanced algorithms that uses machine learning models. A promising research field to discover new materials with enhanced NLO properties and computationally cheaper than traditional *ab initio* methods.

The aim of this master project is to apply data mining methods such as Principal Component Analysis PCA as one of dimensionality reduction methods on a large amount of data harvested from previous experimental and simulation studies, to determine useful patterns between the atomic position of the compounds and materials properties like bandgap, and NLO coefficient. **Also to investigate a candidate NLO materials that can achieve balance between Eg and the SHG effect. Moreover, based on machine learning models, we employ partial least square PLS to predict the band gap of NLO crystals.**

After this general introduction, this manuscript organized as follow: first chapter I presents briefly a general introduction to the theory of Non-linear optics, and giving examples of materials with NLO properties. Then, chapter II, gives a general idea on data mining methods such as principal component analysis PCA and partial least square PLS. After that, chapter III assembling the essential work, the analyses developed by data mining methods. Finally, a conclusion summarizes the principal points discussed and lists the most significant results, while proposing avenues for future explorations.

Chapter 1

Nonlinear optical theory, & their materials

The first part of this chapter is a brief introduction to the theory of non-linear optical phenomena, the quadratic nonlinear optics, particularly the second harmonic generation, its coefficient, and the five conditions for the mid-infrared (IR) zone. In the second part, we shade light on some well-known materials with NLO properties and their applications. Families of materials such as the family oxides including borates, niobates, phosphates, and iodates. Moreover, the chalcogenides family including ternary, and quaternary chalcogenides will be discussed as well.

1.1. Linear and Nonlinear optics phenomena

1.1.1. Linear optics

An electromagnetic wave propagating through a condense medium, e.g. crystal, in which it excites the electrons in the ground state to the excited energetic levels creating an electrical dipole moment (e.g. excitons). Whose macroscopic manifestation will be the polarization ***P***. The relaxation of the excited electrons back to the ground states will result in a light emission (known as fluorescence) equally of amplitude to the initial absorbed energy [18], and the polarization **P** remains proportional to the electric field E as given in the following form:

$$\vec{P} = \vec{P}_0 + \varepsilon_0 \chi^{(1)} \vec{E} \tag{1.1}$$

Such as:

P_0 : Spontaneous polarization of the material. (≈ zero for the majority of mediums) [19]

$\chi^{(1)}$: First order electrical susceptibility tensor.

ε_0 : Permittivity of vacuum.

1.1.2. Nonlinear optics

From linear to non-linear optics, the electrons will absorb more than one photon with same or different energies, after emitting one or more photons after relaxation back to the initial state. The order of the non-linear optical process is determined by the number of waves that take part in that process. A process involving (n+1) waves will be an n-th order process. And by the waves involved, both the absorbed and emitted waves have to be considered [20]. The polarization **P** is no longer linear and must be developed according to the increasing power of the electric field:

$$\vec{P} = \vec{P}_0 + \varepsilon_0 \chi^{(1)} \vec{E} + \varepsilon_0 \chi^{(2)} \vec{E}^2 + \varepsilon_0 \chi^{(3)} \vec{E}^3 + \cdots \quad (1.2)$$

$\chi^{(2)}, \chi^{(3)}$, and so on are the nonlinear dielectric susceptibility coefficients

In general, $\chi^{(1)}E \gg \chi^{(2)}EE \gg \chi^{(3)}EEE$ which indicate that the nonlinear behavior are trivial until the electric field is large enough.

As an example of the third non-linear order, we mention two photon absorption phenomena where it can occur in centrosymmetric materials, therefore instead of second order as shown in **Fig.1(a).** 2PA leaves the material in an excited state from which it often relaxes internally before decaying radiatively, meaning it can deposit some energy in the absorbers. Also, the emission resulting from 2PA is incoherent and is emitted sometime after the absorption. In addition, the absorption occurs according to:

$$I(z) = I_0 e^{-\alpha z} \quad \text{where } \alpha = \alpha_0 + \beta I \quad (1.3)$$

And with the condition for 2PA : $\frac{Eg}{2} < \hbar\omega < Eg$

1.1.3. Second order nonlinear optics

The second order non-linear optics is a process where the valance electrons absorbs double the incident photons and emits a photon with the frequency summing the initial absorbed two photon energies, as it involves three waves in the process and therefore is a 2nd order process. The coherent waves excited in the medium will have frequencies produced by a mixture of frequencies; second order polarization is given in the frequency domain by:

$$P^{(2)}(\omega_3) = \varepsilon_0 \chi^{(2)} E_1(\omega_1) \otimes E_2(\omega_2) \quad (1.4)$$

Where $\omega_3 = \omega_1 \pm \omega_2$ which is the origin of second order nonlinear phenomenon

1.1.3.1. Second harmonic generation

Second harmonic generation is a non-linear phenomenon permitting the doubling frequency ω of an incident wave. In other words, it is the absorption of two photons $h\upsilon$ than the emission of one single photon $2h\upsilon$. The absorption of the second photon is on the assumption that the first energy level is populated as shown in both **Fig.1(b)**

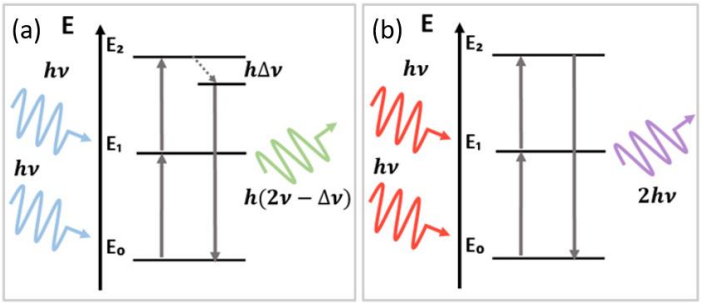

Figure 1: *Tow different Nonlinear optical phenomena's, (a) two photon absorption (2PA), (b) second harmonic generation (SHG).*

For simplification, in the second harmonic generation we write the susceptibility tensor χ in **eq.1.3.** In a different form. As the last indices *j* and *k* can be exchanged, there will be 18 different elements left from the 27 set in the original tensor as in **eq.1.3.** These 18 indices combination can be written in a 2-dimensional 3×6 matrix, called d_{ij} tensor, ranging from d_{11} to d_{36} (it used to be called the d_{il} tensor in the literature). These ingredient are related to the susceptibility $\chi^{(2)}$ by the formula: $d_{ij}=\frac{\chi^{(2)}_{ijk}}{2}$ as the last two indices j and k will be contracted to one index called *j* ($jk\rightarrow j$) such correspondence to $\left(xx=1, yy=2, zz=3, yz=zy=4, xz=zx=5, \text{ and } yx=xy=6\right)$ which gives the second harmonic polarization as follows:

$$\begin{bmatrix} P_x(2\omega) \\ P_y(2\omega) \\ P_z(2\omega) \end{bmatrix} = \varepsilon_0 \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix} \cdot \begin{bmatrix} E_x^2(\omega) \\ E_y^2(\omega) \\ E_z^2(\omega) \\ 2E_y(\omega)E_z(\omega) \\ 2E_x(\omega)E_z(\omega) \\ 2E_x(\omega)E_y(\omega) \end{bmatrix} \tag{1.5}$$

Commonly there is a principal coefficient associated with a direction of the propagation according to which the harmonic generation is maximum. If we introduce the Kleinman [21] symmetry condition in **Table.1**., not all of the 18 elements of d_{il} are independent ($d_{12}=d_{26}$; $d_{14}=d_{25}$). Then *dil* has only 10 independent elements blue colored in the following matrix:

$$d_{il} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{16} & d_{22} & d_{23} & d_{24} & d_{14} & d_{12} \\ d_{15} & d_{24} & d_{33} & d_{23} & d_{13} & d_{14} \end{bmatrix} \quad (1.6)$$

Table 1: *Nonlinear optical tensors for the crystallographic point groups (with Kleinman symmetry conditions)[22]*

Point group	NLO tensor	Point group	NLO tensor
-1, 2/m, mmm, 4/m, 422, 4/mmm, -3, -3m, -6/m, 622, 6/mmm, m-3, 432, m-3m	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	1	$\begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{16} & d_{22} & d_{23} & d_{24} & d_{14} & d_{12} \\ d_{15} & d_{24} & d_{33} & d_{23} & d_{13} & d_{14} \end{pmatrix}$
2	$\begin{pmatrix} 0 & 0 & d_{13} & d_{14} & 0 & d_{16} \\ d_{16} & d_{22} & d_{23} & 0 & d_{14} & 0 \\ 0 & 0 & 0 & d_{23} & 0 & d_{14} \end{pmatrix}$	m	$\begin{pmatrix} d_{11} & d_{12} & d_{13} & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & d_{12} \\ d_{15} & d_{24} & d_{33} & 0 & d_{13} & 0 \end{pmatrix}$
mm2	$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{15} & d_{24} & d_{33} & 0 & 0 & 0 \end{pmatrix}$	-4	$\begin{pmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & -d_{15} & d_{14} & 0 \\ d_{15} & -d_{15} & 0 & 0 & 0 & d_{14} \end{pmatrix}$
6, 422, 4mm, 6mm	$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{15} & d_{15} & d_{33} & 0 & 0 & 0 \end{pmatrix}$	3	$\begin{pmatrix} d_{11} & -d_{11} & 0 & 0 & d_{15} & -d_{16} \\ -d_{16} & d_{16} & 0 & d_{15} & 0 & -d_{11} \\ d_{15} & d_{15} & d_{33} & 0 & 0 & 0 \end{pmatrix}$
32	$\begin{pmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	-6	$\begin{pmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & -d_{16} \\ -d_{16} & d_{16} & 0 & 0 & 0 & -d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
222, 4-3m, 23, 4-2m	$\begin{pmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{14} \end{pmatrix}$	6m2	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -d_{16} \\ -d_{16} & d_{16} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
3m	$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & -d_{16} \\ -d_{16} & d_{16} & 0 & d_{15} & 0 & 0 \\ d_{15} & d_{15} & d_{33} & 0 & 0 & 0 \end{pmatrix}$		

Among several applications of the non-linear optics is in the biological and medical sectors, where the SHG commonly applies in biological tissue visualization, **Fig.2** shows that we can class existing approaches of the imaging, the depth and the resolution are in negative correlation. We remark that the most resolute methods, such as the electronic microscopy (i.e. SEM), authorize to identify the nanometric details uniquely in the massive and well prepared samples, conversely in imaging methods that is used in medical sector such as MRI and ultrasonography enable to imaging the organs with order of millimeters resolution, between this two, the multiphoton methods in the scale of micro-meter which is the scale of the optical wavelength, the NLO microscopy used in active biological tissue for example the SHG visualize the fibrillary structure like collagen.[23]

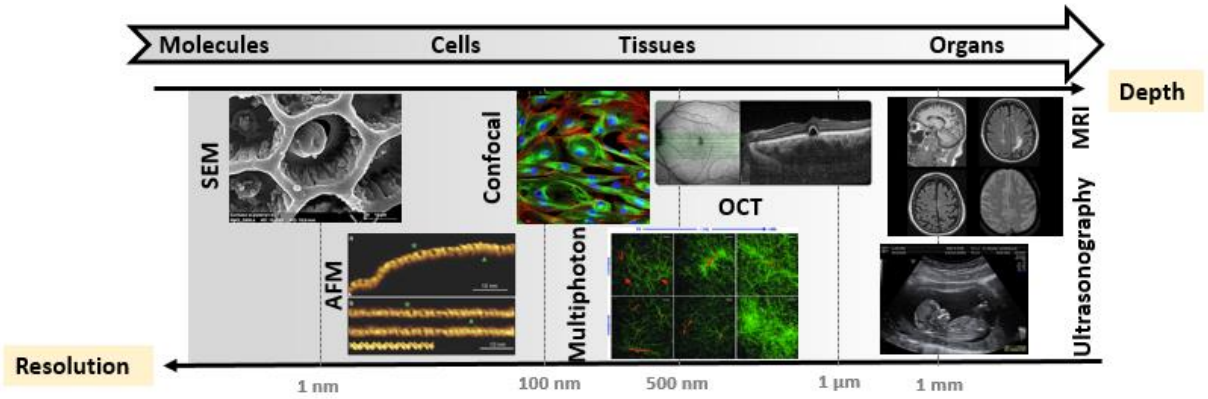

Figure 2: *Different medical imaging classified based in consideration of the depth and resolution*

1.2. Five conditions for good functionality in the mid-IR zone

We should take into consideration that the SHG can occur in non-centrosymmetric crystals, e.g, crystals that do not possess inversion symmetry. Centrosymmetric materials cannot generate second-order non-linear optical interactions since the susceptibility tensor are identically equal to zero. Therefore, amorphous solids (such as glass, ceramics…etc.) and crystals with symmetry class 432 422, 622, liquids, and gases as well they do not display SHG phenomenon. Since 11 out of the 32 crystal classes possess inversion symmetry as presented in **Fig.3.**, , this rule is very powerful, as it eliminates all crystals belonging to these classes from consideration for second-order nonlinear

effects. While third-order nonlinear optical interactions can occur in both centrosymmetric and non-centrosymmetric media.

For a good functionality and performance of NLO materials in the mid-IR region, is agreed

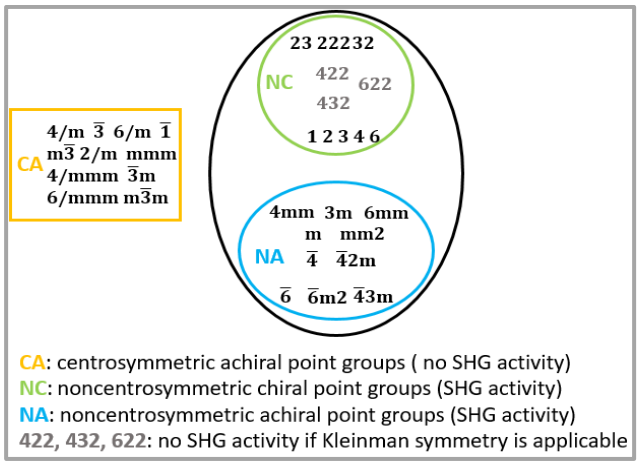

Figure 3: *SHG activity among the 32 crystallographic classes*

in the NLO scientific community, that the materials have to meet five standard conditions [24] as discussed as following:

1.2.1. Transparency in the mid-IR

The activity in the mid-IR band, where the material is able to run across two important atmospheric transparent windows, 3−5 µm and 8−12 µm as pointed in **Fig.4.** Giving an example of materials that operates in the mid-IR band, like the family of oxides, although the transparency range of oxides is very narrow in the infrared region, where chalcogenides are transparent in 6-10 µm for sulfur-based crystals, 11 µm for selenium-based crystal, and 18-25 µm for tellurium-based crystals [24-26].

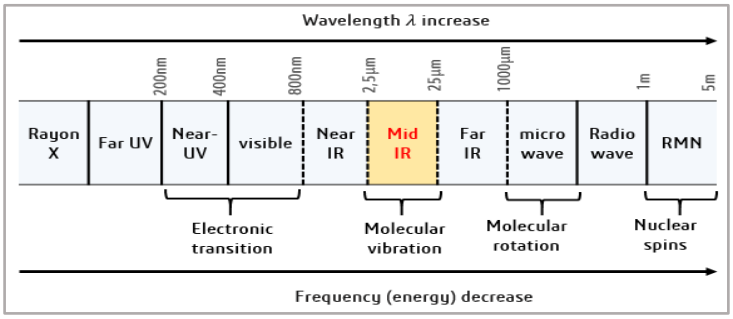

Figure 4: *Infrared region in the electromagnetic spectrum*

1.2.2. Large SHG effect

In sequence of obtain high SHG conversion rate, the SHG effect for mid-IR NLO crystals should be at least 10 times (preferably 20 times) larger than that of the KDP crystal of value d_{36} = 0.39 pm/V [27].

1.2.3. High laser damage threshold

The third condition of a good NLO functionality in the mid-infrared range, is the high laser damage threshold LIDT (laser induced damage threshold), instead of the LIDT we prefer to choose as a condition the value of the bandgap Eg > 3.5 eV because there is a direct (linear behavior)

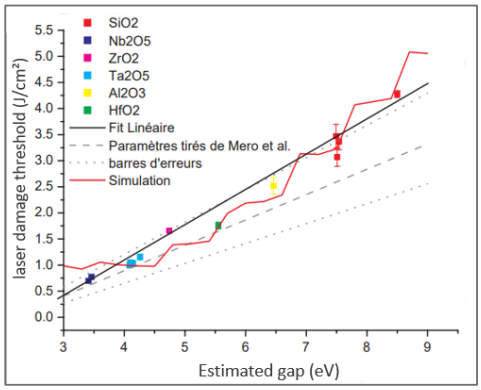

Figure 5: *damage threshold of oxides layers measured at 1030 nm, 500fs, in single pulsed depending on the gap. A linear adjustment of this data is also represented, as well as data from the literature (800 nm) with their error bar.*

correlation between the two, as in **Fig.5.**[28]

Laser damage is the result of the light-matter interaction effect, which results in the degradation of the material degradation as shown in **Fig.6.** [29]. Laser damage is generally defined as any irreversible modification of the component induced in an unintentional manner.

For a high laser damage threshold, the bandgap should be at least greater than 3.0 eV (preferably beyond 3.5 eV). Moreover, there is a relation between the atomic radius and bandgap, thus we can predict new materials that exhibit an important value of bandgap, by replacing the heavy atoms with lighter once, the replacing should be either from the same column in periodic table or the same external layer of the electronic configuration.

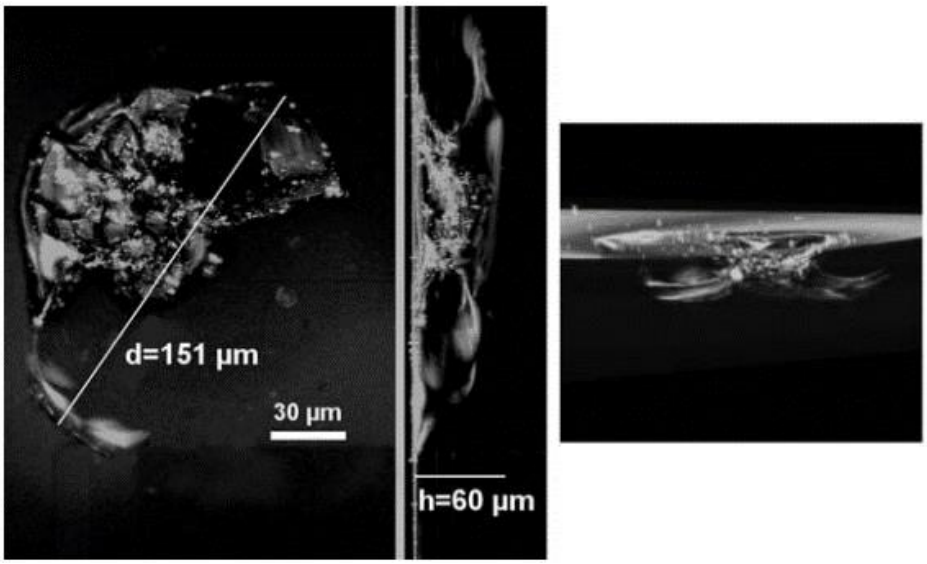

Figure 6: *Observation by confocal microscopy of damage created by a nanosecond laser on the back side of silica glass*

1.2.4. Birefringence

For the mid-IR NLO materials require a moderate birefringence (difference between the ordinary and extraordinary refractive index $\Delta n=n_o-n_e$). The small birefringence (e.g., $\triangle n < 0.04$) would not be suitable to achieve the phase-matching condition in the mid-IR region, while the too large birefringence (e.g., $\triangle n > 0.10$) would result in the destructive optical behaviors such as self-focus and walk-off effects in the conversion process. Preferably, the birefringence for a good mid-IR material should be in the range of 0.04 – 0.10.

1.2.5. Chemical and thermal stability and good mechanical properties

The good chemical stability, thermal stability, and mechanical property such as young modulus and compressibility modulus are very beneficial to the practical applications of a NLO crystal.

1.3. Non-linear optical materials

Since the discovery of second harmonic generation behavior in quartz crystal by Franken et al [1], then the nonlinear optics witnesses a rapid development with characterization of different non-centrosymmetric materials such as KH_2PO_4 (KDP), $KTiOPO_4$ (KTP), $LiNbO_3$, …etc. These NLO crystals are useful in near-IR, visible, and ultraviolet regions (wavelength in the range of 0.2

- 2 μm), which are widely used in scientific and technological fields, including visible laser generation [30] artificial nuclear fusion [31] precision scientific instruments [32] .However, NLO materials which can efficiently generate high-power mid-IR lasers in the spectral range of 2−25 μm are scarce. Inorganic materials are candidates for NLO applications including chalcogenides and oxides.

1.3.1. Iodates

Iodates (with the formula XO_3^{-n}) are provedto be a candidates of good NLO properties, since they often have a large bandgaps (high LIDT). For instance, the $RbIO_3$, $Rb_2BiI_5O_{15}$, and $LiCd(IO_3)_3$ crystals have a large bandgap of 4.0, 4.6, and 4.18 eV respectively. It also shows excellent transparency in the range of 2.5 to 13.6 μm, 2.5−10.8 μm, up to 5.5 μm respectively. However, the SHG coefficients of iodates are generally much smaller compared with commercially available mid-IR NLO crystals AgGaS2 (AGS) (d_{ij}=13.7 pm/V). As reported in the literature, $RbIO_3$, $Rb_2BiI_5O_{15}$, and $LiCd(IO_3)_3$ exhibit a phase-matchable SHG efficiency of only about 1.75, 1.17, and 4.68pm/V, respectively[33-35].

1.3.2. Chalcogenides crystals

Among the non-linear materials used for frequency conversion in the mid-IR there are those whose structure derives from the structure of diamond which characterizes the element of group IV of the periodic table as illustrated in **Fig.7.** [25] This structure consists of two interpenetrating

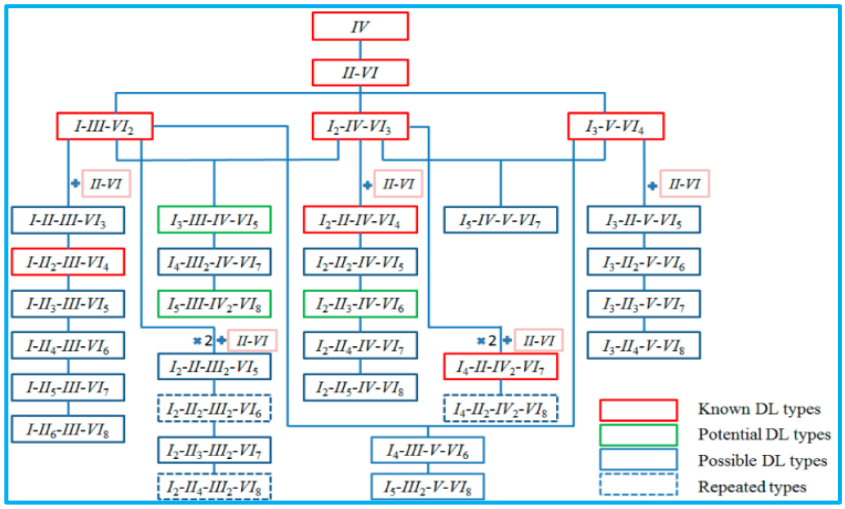

Figure 7: *Pyramidal graph of single element DL structure IV, and their extracted multi-element DL structure*

lattice of face centered cubic type (fcc) with space group Fd3m. all derives of these compounds of type III-V and II-VI which is obtained exchanging atoms with other similar atoms.

1.3.2.1. Ternary chalcogenides

1.3.2.1.1. Traditional chalcogenides

Traditional chalcogenides are the most used materials in nonlinear optics to name few, like $CdGeAs_2$, $CdSiP_2$, and $ZnGeP_2$ plus high NLO coefficients equals to 101, 91, and 75 pm/V respectively but with low value of bandgap 0.52, 2.2, and 2.1eV respectively, which is an indication of low value of LIDT [36-38].

1.3.2.1.2. Family of I-III-VI_2

The family of I_B-III-VI2 with (I: Ag, Cu ; III: Ga,In ; VI: S, Se, Te) is the best known and used for frequency shifting like AgGaS and AgGaSe, but their application are limited due to low resistance to laser flux (weak value of LIDT) most of them possess a low value of bandgap Eg>3 eV.[39, 40]

1.3.2.1.3. Lithium based chalcogenides

Li- III-VI2 compounds with (III: Ga,In ; VI: S, Se, Te) they have large bandgaps compared to I_B-III-VI_2 compounds since the Lithium(I_A: in periodic table) is very light compared to the cooper and silver (I_B in periodic table) as mentioned in condition III for good functionality in the mid-IR zone.

1.3.2.1.4. Chalcophosphates $M_xP_yX_z$

Metal chalcophosphate that contains phosphorous and chalcogen atoms with P-X, M-X bonds, as the importance of these family is the great structural variety (one dimensional, two and three dimensional). This variety shows an association of non-centrosymmetry in the crystral structure and reversible phase change behavior, which appear an important technological applications in different sectors, such as NLO devices, solar energy conversion and semiconductors batteries, memory storage (DVD), …etc [41]. for example, $APSe_6$ with (A: K, Rb) possess high NLO coefficient 75pm/V and large bandgap 2.16, 2.18 eV respectively[42].

1.3.2.1.5. Barium based chalcogenides

The recent barium based chalcogenides crystals like $BaGa4S_7$ and $BaGa4se_7$. We find that $BaGa4s_7$ possesses an attractive optoelectronic proprieties and achieve the balance between the bandgap and d_{ij} parameters (Eg=3.6 eV and d_{ij}=32×KDP)[43]. While $BaGa_4Se_7$ has a high laser damage threshold LIDT of a value equals to 557 MW/Cm^2 and large SHG effect (18 pm/V) [44]

1.3.2.2. Quaternary chalcogenides

1.3.2.2.1. Diamond-like metal chalcogenides

Diamond-like structure metal chalcogenides, where the structures of these compounds are derived from that of diamond, of either the cubic or hexagonal form, and consist of tetrahedral building blocks [45, 46] as demonstrated by the pyramidal graph of normal DL structures plotted in Fig.8. Compared with binary and ternary diamond-like phases, the compositional flexibility of quaternary materials increases and allows for the potential to tune the physical properties of the materials in nonlinear optics as well as other utilities.

The quaternary sulfide DL compound $AgGaGe_5Se_{12}$ reported to exhibit excellent performances with concurrently large NLO coefficients (d_{31}=29pm/V) and impressive LIDT, which satisfy the essential requirements of mid-IR NLO candidates [47].

1.3.2.2.2. Alkali-Metal-Based Quaternary Chalcogenides

The chalcogenides with the formula $\mathbf{A_xB_yM_zD_n}$, where **A** is selected from the group consisting of an alkali metal and a mixture of alkali metals, **B** is selected from the group consisting of mercury, zinc and manganese, **C** is a metal selected from the group consisting of germanium and tin and **D** is selected from the group consisting of sulfur and selenium.

To summary, the SHG is one of the most interesting applications of the laser which is the investigation of the nonlinear interactions of light with matter, SHG is observed in the non-centrosymmetric materials. From the beginning of this century, the rapid development of laser science and technology urgently demands new NLO crystals for wider application ranges, especially in the mid-IR region, thus several materials have been confirmed experimentally and theoretically, using first principal calculations. For instance, oxides and chalcogenides families. Herein, for our study,

we are interested in iodates and chalcogenides NLO crystals. In the following chapter, we explained the models which we employed in this study.

Chapter 2

Data mining ,Machine learning, and their models

"I never guess. It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

— Sir Arthur Conan Doyle,

author of Sherlock Holmes stories

Data mining is a composite discipline that overlaps other branches of science. In **Fig. 8**, we can see the contributions of many different fields in the development of the science of data mining. Because of the contributions of many disciplines, staying up to date on the progress being made in the field of data mining is a continuous educational challenge.

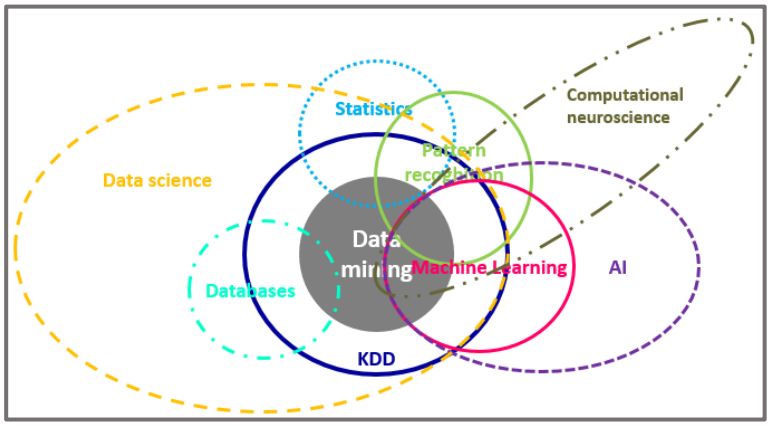

Figure 8: *Multidisciplinary nature of data mining; Source: SAS Enterprise Miner Training material from 1998*

2.1. Data mining

Data mining is the process of updating relations, correlations, dependencies, associations, models, structures, classes, and factors obtained by navigating to across long sets of data usually stored in databases, by using mathematical, statistics or algorithms[48].

Data mining, also popularly known as knowledge discovery in a database (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data.

2.2. Machine learning

On the other hand, machine learning is the process of discovering algorithms that have improved courtesy of experience derived from data. It's the design, study, and development of algorithms that permit machines to learn without human intervention. It's a tool to make machines smarter, eliminating the human element (but not eliminating humans themselves)[48, 49].

2.2.1. Machine-learning types

There are different types of machine-learning systems. We mention the supervised learning and unsupervised learning.

2.2.1.1. Unsupervised learning

In unsupervised learning, we lack supervisors or training data. In other words, all what we have is unlabeled data. The idea is to find a hidden structure in this data. There can be a number of reasons for the data not having a label. It can be due to unavailability of funds to pay for manual labeling or the inherent nature of the data itself. With numerous data collection devices, now data is collected at an unprecedented rate. The variety, velocity, and the volume are the dimensions in which Big Data is seen and judged. To get something from this data without the supervisor is important[50-52]. Very simple classic example of unsupervised learning are dimensionality reduction. Herein, we review one of the dimensionality reduction methods which represented in principal component analyses.

2.2.1.2. Supervised learning

In supervised learning, the target is to infer a function or mapping from training data that is labeled. The training data consist of input vector X and output vector Y of labels or tags as shown in Fig.9. A label or tag from vector Y is the explanation of its respective input example from input vector X. Together they forma training example. In other words, training data comprises training examples. If the labeling does not exist for input vector X, then X is unlabeled data [53].

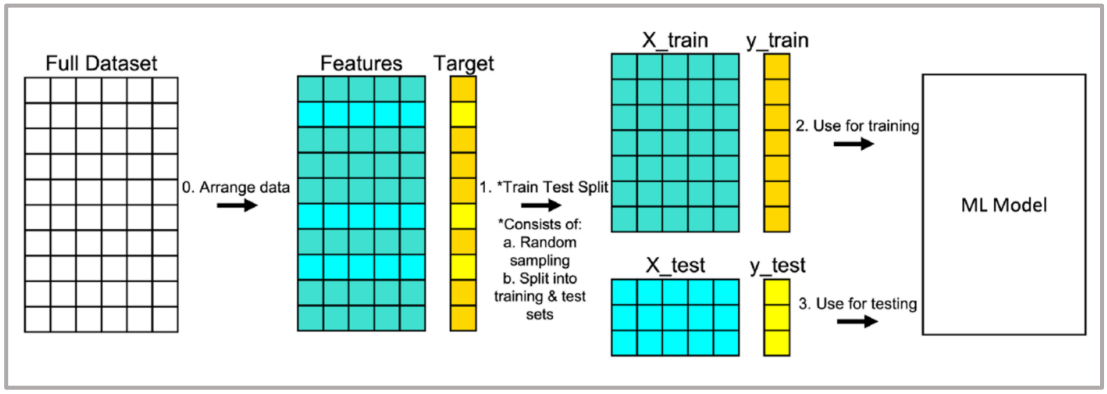

Figure 9: *The supervised learning steps*

2.2.2. Principal component analysis

The large dimensions of the information can pose problems of computational complexity and capacity storage. Thus, it's sometimes preferable to seek to reduce dimensionality. Hence, in material science, it becomes hard to visualize and describe the correlation among the selecting features (descriptors) in high dimensionality data.

Principal component analysis is statistical technique which uses sophisticated underlying mathematical principles to transforms a number of possibly correlated variables into a smaller number of variables called principal components. The origin of PCA goes back to Pearson (1901) [54]. But Hotelling (1933)[55] formalized the modern representation of PCA, also he coined the term "*principal component* ". PCA analyzes a large set of data representing observations described by inter-correlated variables. Targeted to extract the important information from the data set in a new orthogonal variables called *principal component* [56]

2.2.2.1. Theory of principal component analysis PCA

Assume that we start with a data set that is represented in terms of an $m \times n$ matrix, X where the n columns are the samples (in our case physical properties) and the m rows are the variables (e.g. NLO materials). We wish to linearly transform this matrix, X into another matrix, Y, also of dimension $m \times n$, so that for some $m \times n$ matrix, P

$$Y = PX \tag{2.1}$$

If we consider the rows of P to be the row vectors $p_1, p_2, \ldots, p_m$, and the columns of X to be the column vectors $x_1, x_2, \ldots, x_n$, then can be interpreted in the following way.

$$PX = \begin{pmatrix} p_1 \times x_1 & p_1 \times x_2 & \cdots & p_1 \times x_n \\ p_2 \times x_1 & p_2 \times x_2 & \cdots & p_2 \times x_n \\ \vdots & \vdots & \ddots & \vdots \\ p_m \times x_1 & p_m \times x_2 & \cdots & p_m \times x_n \end{pmatrix} = Y \quad ; \qquad p_i \cdot x_j \in \mathbb{R}^m \tag{2.2}$$

This tells us that the original data, X is being projected on to the columns of P

Principal component analysis defines independence by considering the variance of the data in the original basis. It seeks to de-correlate the original data by finding the directions in which

variance is maximized and then use these directions to define the new basis. Recall the definition for the variance of a random variable, Z with mean, μ

$$\sigma_Z^2 = E\left[\left(Z-\mu\right)^2\right] \tag{2.3}$$

Suppose we have a vector of n discrete measurements, $\tilde{r} = \left(\tilde{r}_1, \tilde{r}_2, \ldots\ldots, \tilde{r}_n\right)$, with mean μ_r. If we subtract the mean from each of the measurements, then we obtain a translated set of measurements $r = \left(r_1, r_2, \ldots\ldots, r_n\right)$,that has zero mean. Thus, the variance of these measurements is given by the relation:

$$\sigma_r^2 = \frac{1}{n}\mathbf{r}\mathbf{r}^T \tag{2.4}$$

Generalizing this idea to considering our $m \times n$ data matrix, X . Recall that m was the number of variables, and n the number of samples. We can therefore think of this matrix, X in terms of m row vectors, each of length n

$$C_X = \frac{1}{n\text{-}1}XX^T = \frac{1}{n\text{-}1}\begin{pmatrix} x_1x_1^T & x_1x_2^T & \cdots & x_1x_m^T \\ x_2x_1^T & x_2x_2^T & \cdots & x_2x_m^T \\ \vdots & \vdots & \ddots & \vdots \\ x_mx_1^T & x_mx_2^T & \cdots & x_mx_m^T \end{pmatrix} \in \mathbb{R}^{m\times m} \tag{2.5}$$

x_i is a vector of the n samples for the i^{th} variable

If we look closely at the entries of this matrix, we see that we have computed all the possible covariance pairs between the m variables. Indeed, on the diagonal entries, we have the variances and on the off-diagonal entries, we have the covariance. This matrix is therefore known as the Covariance Matrix

The covariance matrix C_Y:

$$\mathrm{C_Y} = \frac{1}{n-1}\mathrm{D} \tag{2.6}$$

Where D is a diagonal matrix which has the eigenvalues of XX^T as its (diagonal) entries.

To conclude, using this method we get information about the relative importance of each principal component from the variances. The largest variance correspond to the first principal component, the second largest to the second principal component and so on..

2.2.3. Partial least square

Several problems in materials science especially in the context of new materials, properties prediction can be described in the form of regression model, where we have variables X on which we can more or less intervene and Y variables that can only be observed.

The objective is then to describe the relation between Y and X in the absence of theorical model. Partial least square is method of analysis data particularly constructed for the study of these type of problem.

2.2.3.1. Theory of Partial least square PLS

The origin of PLS regression can be traced back to social science precisely in the economics , Herman Wold 1966 [57], become popular in chemo-metrics (i.e., computational chemistry) due in a part of Herman's son Svante Wold [58].

PLS regression was born from the association of the NIPALS algorithm (Nonlinear Iterative Partial Least Squares) by Herman Wold [57].

PLS regression is the simplest form of the linear regression model, a linear model specifies the (linear) relationship between one (or more) dependent variables (responses) Y and a set of predictors' variables X such as:

$$Y = \left(b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p\right) + E \tag{2.7}$$

Which b_i the regression coefficients, E the error term

To construct the model, PLS regression produces a weight matrix for X such that $T = XW$, i.e. the columns of W are weight vectors for the columns of X producing the corresponding score factor matrix.

These weights are calculated in such a way that they maximize the covariance between the response and the corresponding scoring factors.

The OLS (Ordinary Least Squares) procedure for the regression of Y on T is then used to produce Q, such that $Y = TQ + E$. Once Q is calculated, we have $Y = XB + E$ where $B = WQ$, and the prediction model is complete.

An additional matrix is needed for a complete description of the procedure for PLS regression is the factor matrix P which gives the model $X = TP + F$ where F is the part unexplained score of X [59-61].

In this work we use python programming language for generating the algorithms and extracting patterns.

2.3. Why python?

The simplicity of python language make it the ideal language for beginners to learn. Several python programs can be written in few lines of code, also python can be used for a large variety of tasks such as for desktop applications, database applications, network programing, etc. Finally the written code on windows will work on Mac OS or Linux without making any changing to the python code which means python is a cross platform language [62].

Chapter 3

Results and discussion

In order to evaluate the prospect of both chalcogenides and iodates crystals for the mid-IR NLO applications. We select 140 quaternary and 91 ternary compounds from experimental data and DFT calculations.

All the data in the references [6, 11, 23, 24, 27, 32-36, 43, 44, 46, 63-142]

First of all, the distribution of Eg with the respect to SHG effect d_{ij} for ternary and quaternary compounds are represented **Fig.10** which attributed to the investigation of L. Kang *et al* **(Metal Thiophosphates with Good Mid-Infrared Nonlinear Optical Performances: A First-Principles Prediction and Analysis)** [24]**.** Which we can deduce that there is a negative correlation between the band gap and the SHG effect d_{ij}. Moreover, these study gives us two-dimensional visualization without taking into account other variables that can affect the band gap and the d_{ij} .

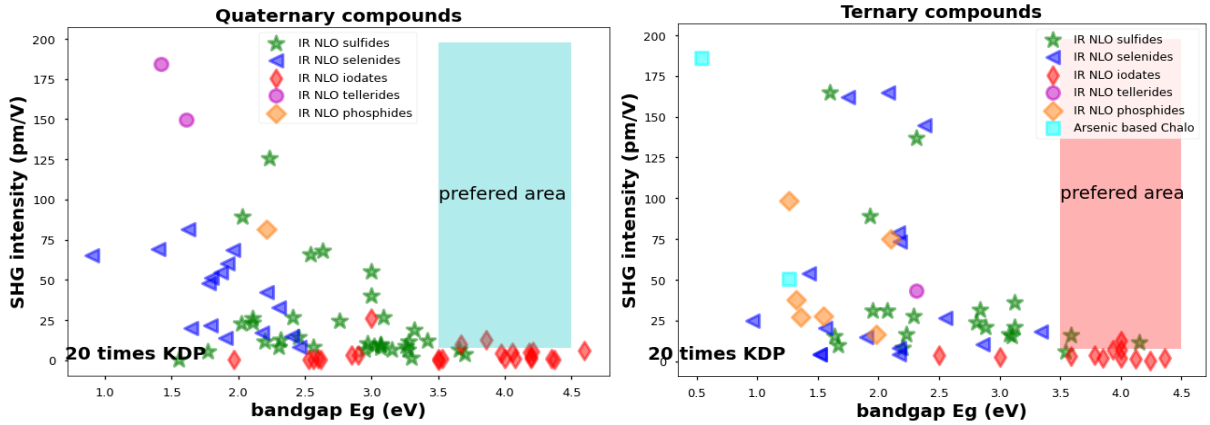

Figure 10: *The distribution of Eg with the respect to SHG effect dij for ternary and quaternary compounds*

3.1. Principal component analysis results

3.1.1. Selecting features

After the selection of the data set, we choose as descriptors the bandgap, SHG effect dij, and the number of period and group in periodic table **Fig.11.** The principal components do not necessarily have an obvious physical meaning, but rather are a combination of descriptors, which explain the largest variation in the data. The advantage of PCA is that, since each principal component uniquely captures the effect of a certain combination of relevant descriptors, typically a few principal components are sufficient for describing a system.

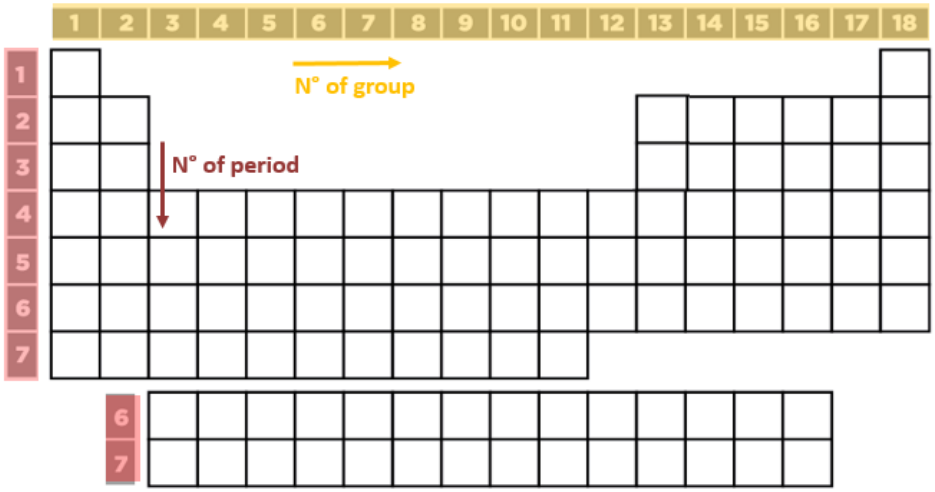

Figure 11: *Periodic table showing the number of group and period which we chose in this study*

Note: **PX** with **(X: I, II, III, IV)** number of period of the elements (I: First element in the compound, II: second elements in the compound, and so on) same abbreviation for the number of group for example (AgGaS2) the silver belongs to the group N°11 and to the period N°5 and the gallium belongs to the the group N°13 and to the period N°4 and for the sulfur the group N°16 and to the period N°3

3.1.2. Axes choice

The resulting scores plot of this analysis is shown in **Fig.12** and **Fig.13** for ternary and quaternary data. We notice that for ternary compounds dataset PC1 captures 27 % of the variance, whereas PC2 captures 20 %. The two PCs together capture ~47% of the variance of the data, furthermore the PC1 captures25 % of the variance, whereas PC2 captures 15 %. The two PCs together capture ~40 % of the variance for the quaternary compounds dataset. Therefore, a dataset of n-dimensions (8 and 10 initial descriptors for ternary and quaternary compounds respectively), can be reduced to a few dimensions (2 PCs) while capturing ~40 % of the original information. The reduction in dimensionality makes trends and correlations, which are “hidden” in the data, become easily visualized and described in PC space.

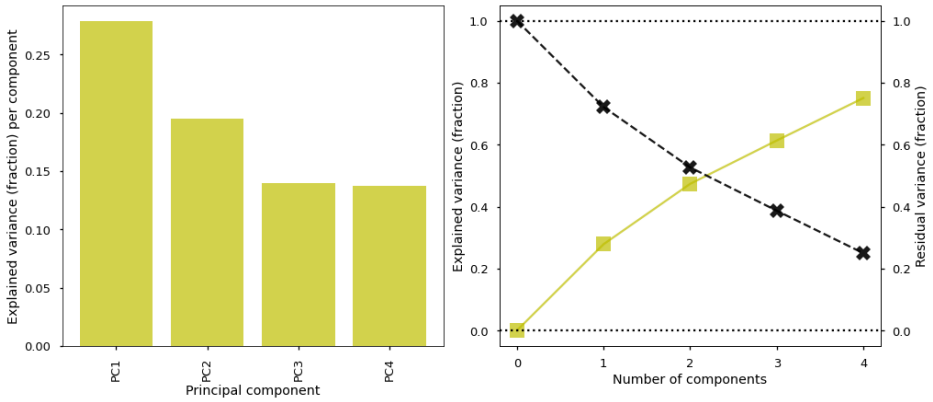

Figure 12: *The score plot for ternary data set*

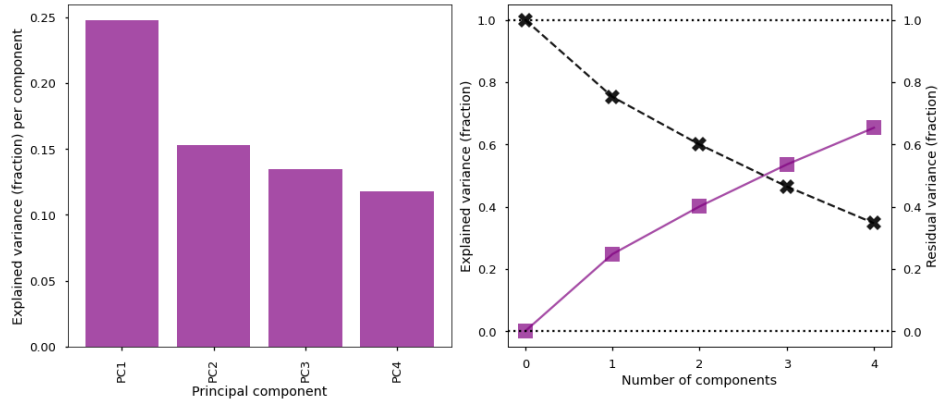

Figure 13: *The score plot for quaternary data set*

The contribution of the principal components from the original variables are represented in a bubble heat map **Fig.14** and **Fig.15**. Herein, both ternary and quaternary results are similar which we notice that the PC1 and PC8 in ternary data and PC1 and PC10 in the case of quaternary data contributes greatly from the band gap, also the PC1, PC5, PC6, and PC7 contributes from the d_{ij} for the ternary data while PC1, PC6, and PC7 dispenses from the d_{ij} that is for the quaternary data. While the PC2 contribute in the number of period and group in both ternary and quaternary data set, which make the first two PCs a good choice for visualizing the correlation between the variables

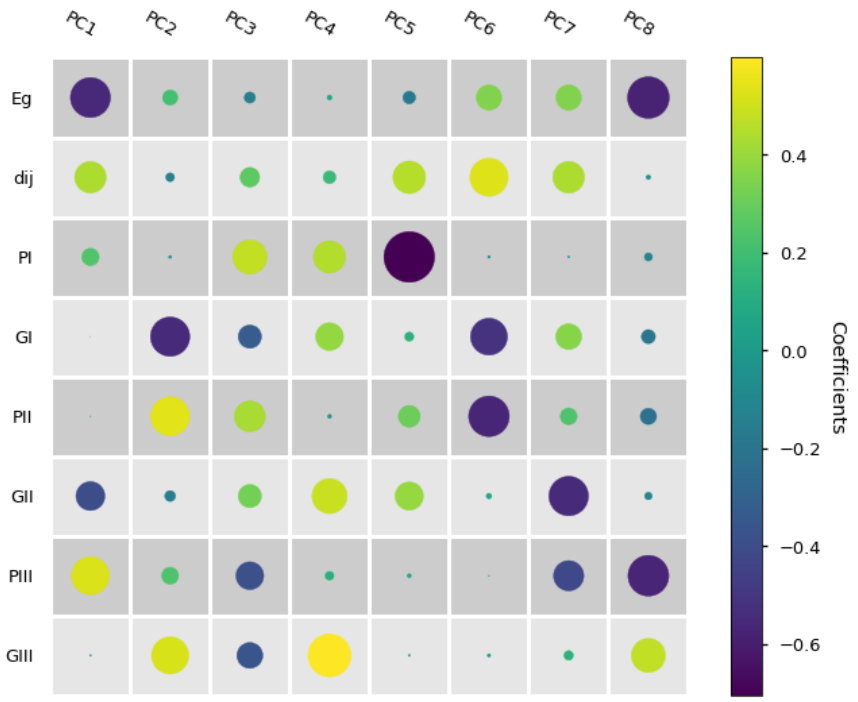

Figure 14: *The contribution of the principal components from the variables for ternary data*

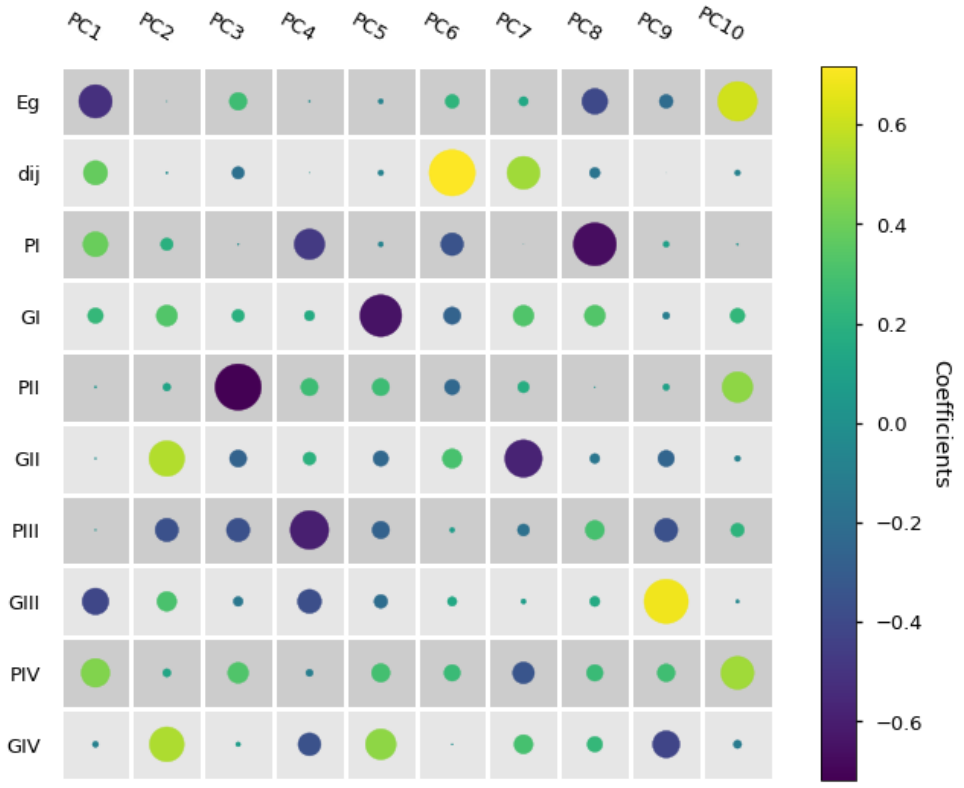

Figure 15: *The contribution of the principal components from the variables, for quaternary data*

3.1.3. The correlation matrix

On the other hand, the correlation matrix that shown in **Fig.16** and **Fig.17** explicate the correlation between our descriptors in all the principal components, from this results we notice that:

- The Eg are inversely correlated to d_{ij} with the value (-0.49, -0.36) for ternary and quaternary data respectively. That's means if the band gap increase, the d_{ij} decrease.
- Negative correlation between the band gap and the number of periods where the number of group has no correlation with Eg except the number of group of the second element GII in ternary compounds and the third element GIII in quaternary compounds, the second and third element that is in ternary and quaternary compounds arranged as follow: (Zn, Hg, In, Al, Ga, Ge, Sn, Si, P, Sb, and Iodine) which belong to the number of group (12, 12, 13, 13, 13, 14, 14, 14, 15, and 17 respectively) this correlation declare that more we use element with high number of group like iodine more the band gap increase
- Positive correlation between the d_{ij} and the number of period especially with last element (the third in ternary PIII and fourth in quaternary PIV compounds) namely the using elements: (O, P, S, Se, and Tellurium) which belong to the number of period (2, 3, 3, 4, and 5) that's means more we use element from the period 4 and 5 (Se and Te) more the value of the d_{ij} increase.

Remark: The data base that we have chosen for this study (91 ternary and 141 quaternary compounds) allows us to examine a large number of period and group in the periodic table such as (number of periods that we used ranging from 2 to 6, while the number of group ranging from 1to 17 except4, 5, 6, 7, 8, 9, and 10) of the chalcogenides and iodates compounds, scanning most of the period and group number which helped to make a global visualization of the correlation between the variables.

Figure 16: *Correlation matrix for ternary data*

Figure 17: *Correlation matrix for quaternary data*

3.1.4. The Biplot

The biplot results in **Fig.18** and **Fig.19** shows both the correlation between the descriptors and the samples (the NLO crystals) in the first and second principal components. Furthermore, for the interpretation of the biplot we should to obey the following rules:

- If two vectors are parallel, then there is a positive correlation between them.
- If they are in opposite directions, means there is a negative correlation between them.
- If they are perpendicular to each other which indicate is no linear relation between them.
- Last but not the least, if the length of the vector is too small then the variable doesn't have information in chosen principal axis[143].

Based on these rules we found that the biplot results are consistent with the correlation matrix results.

We notes from the biplot also the iodates crystals have the highest value of the bandgap and inversely the lowest value of d_{ij}, the tellurides and selenides based chalcogenides possess the largest value of the d_{ij} and the lowest value of bandgap, hence the materials which located in the center

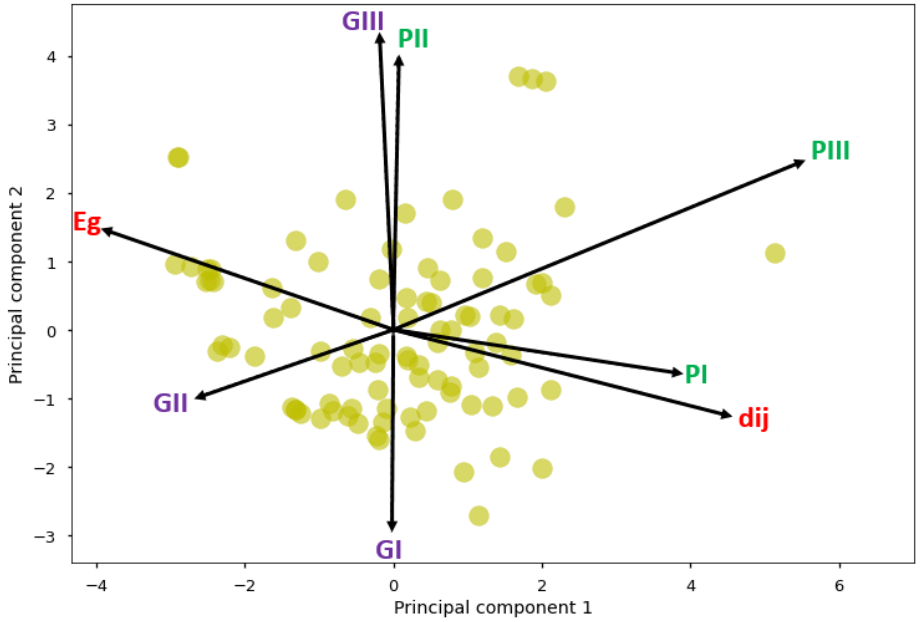

Figure 18: *The biplot of ternary data in the first and second principal components*

(most of them are sulfide based chalcogenides) of the d_{ij} and Eg vectors are candidates materials for achieving balance between Eg and d_{ij}.

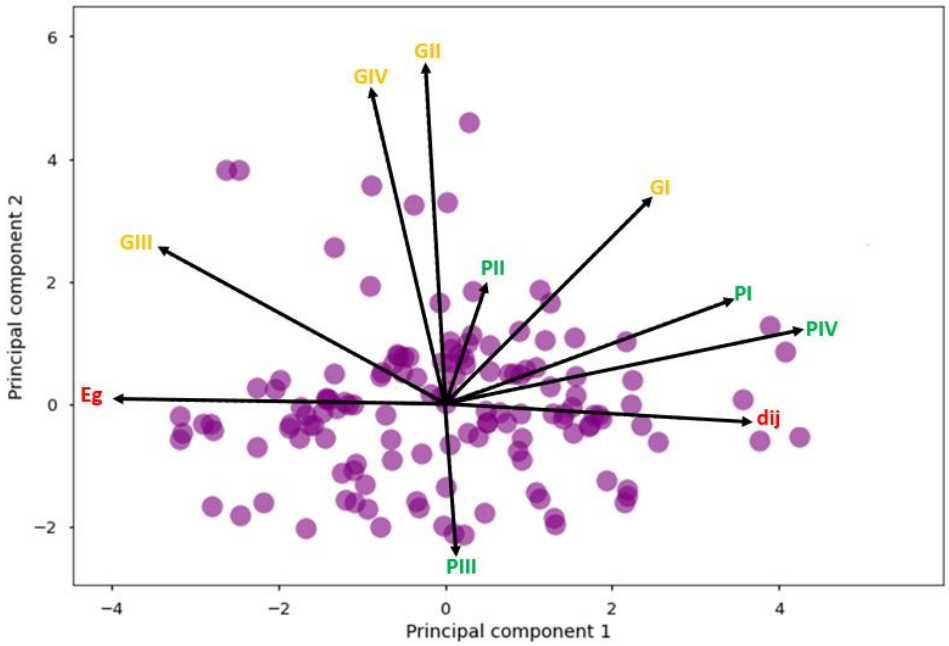

Figure 19: *The biplot of quaternary data in the first and second principal components*

Important note

As a next step, based on the PCA results (that the number of period inversely correlated with the band gap and vice versa for the d_{ij}), in order to achieve the balance between both Eg and d_{ij}, we study the normal distribution (or Gaussian distribution) of the number of the period of the ternary and quaternary elements. **Fig.20** represent the number of the period (N° period) with the respect to the count for ternary and quaternary data elements. We notice that for ternary data the period N°3 contribute with 40% from data and also the Gaussian shows a high density in the period N°3. Moreover, for the quaternary results shows that the number of period 3, 4, and 5 contribute with 30% from the data and the Gaussian shows a high density in both the period N°3 and 4.

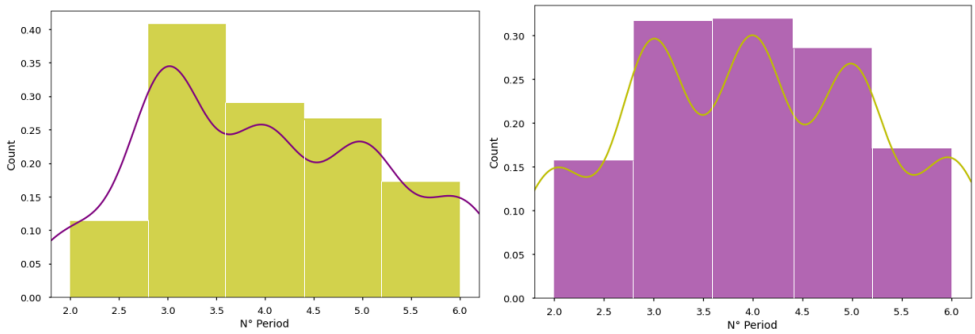

Figure 20: *The number of the period (N° period) with the respect to the count, the yellow one for ternary data, while the purple one for quaternary data.*

From the remarks above, we conclude that the fourth period possess the average of the contribution of the ternary and quaternary elements. The ternary compounds with the formula ABX, and quaternary compounds with the formula ABB'X, where A and X elements should be chosen from the fourth period. for a medium value of both the Eg and the dij, where if we select elements upper than the period four, then the value of the band gap becomes larger. In the other hand, the value of the d_{ij} becomes smaller, while the opposite if we select elements under the fourth period.

3.2. Partial least square results

3.2.1. The band gap prediction

In order to predict the value of the band gap, the selection of the descriptors or 'features' is an essential part for a good performance of the PLS model.

3.2.1.1. Taking in the consideration the PCA results

Based on the PCA results we chose four features for ternary data set (PI, GII, PIII, GIII) and five features for quaternary data set (PI, GI, PII, GIII, PIV)

The prediction was performed on the basis of both ternary and quaternary databases. Where 80% of the data used as training set and 20% as test set. **Fig.21** and **Fig.22** shows predicted against actual values of "Eg", for the training and test set for ternary and quaternary data respectively. However, without any information on electronic properties of the NLO crystals in the test set. It was possible to predict the band gap Eg using the prediction model for the training set.

The predicted models for ternary NLO crystals are:

$$Eg=-10.88583+(-0.19963\times PI)+(-0.03601\times GII)+(-0.75355\times PIII)+(1.08983\times GIII) \quad (3.1)$$

The predicted models for quaternary NLO crystals are:

$$Eg=3.08012+(-0.14785\times PI)+(-0.02879\times GI)+(-0,29877\times PII)+(0,18679\times GIII)+(-0,30165\times PIV) \quad (3.2)$$

The results of ternary and quaternary test data are presented in **Table 2** and **Table 4** respectively, we notice that the test set values are a good agreement with the recently reported data in literature where the average of the error is calculated to be the value 0.59 eV in ternary data while the average error for quaternary test set is around the value 0.45 eV

3.2.1.2. Including all the numbers of period and group

If we include all the variables such as the number of period and group in the models, **Fig.23** and **Fig.24** shows predicted against actual values of "Eg", for the training and test set for ternary and quaternary data respectively. We obtain the following equation for ternary compounds:

$$\begin{aligned} Eg=&-11{,}01069+(-0{,}19464\times PI)+(-0{,}0491\times GI)+(-0{,}25599\times PII)\\ &+(-0{,}00694\times GII)+(-0{,}73509\times PIII)+(1{,}15122\times GIII) \end{aligned} \tag{3.3}$$

For quaternary compounds we get the following equation:

$$\begin{aligned} Eg=&1{,}53375+(-0{,}11315\times PI)+(-0{,}03564\times GI)+(-0{,}29753\times PII)+(0{,}01108\times GII)\\ &(-0{,}20892\times PIII)+(0{,}17338\times GIII)+(-0{,}3611\times PIV)+(0{,}15645\times GIV) \end{aligned} \tag{3.4}$$

The results of ternary and quaternary test data are presented in **Table.2** and **Table.3** respectively, Since the mean the error is calculated to be the value 0.51eV in ternary data while the mean error for quaternary test set is around the value 0.49 eV

We notice that for both the ternary compounds the models with all the variables are more precise while of the quaternary compounds are more precise with taking in consideration the PCA results.

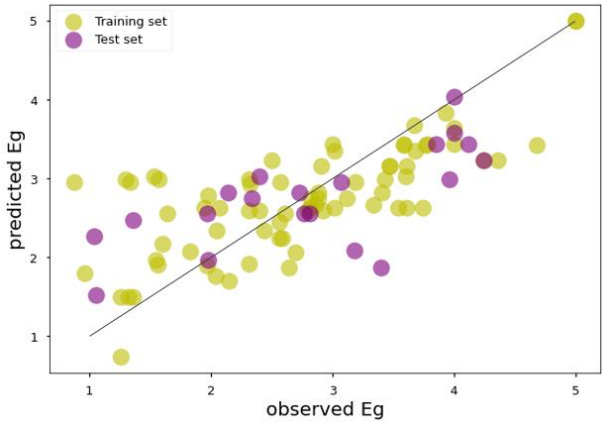

Figure 21: *The actual ternary data versus the predicted value via PLS model* ***(taking in consideration the PCA results)***

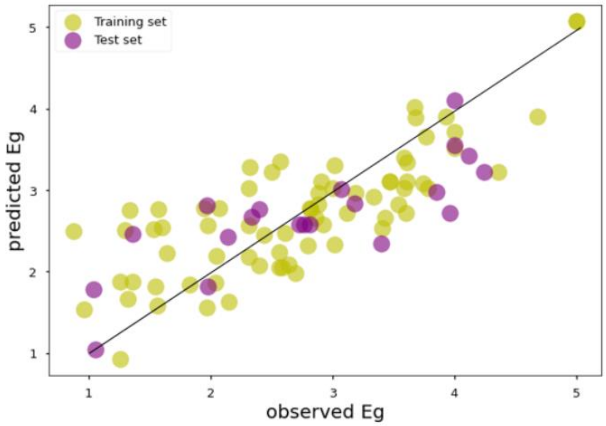

Figure 22: *The actual ternary data versus the predicted value via PLS model* ***(taking all the variables)***

Table2: *The 20% test data which chosen randomly in ternary data set the predicted and actual value are compared*

Compounds	Predic-tive Eg (eV)(PCA results)	Predictive Eg (eV) (all period & group)	Actual Eg (eV)[16, 40, 69-75]	Error (%) (PCA results)	Error (%) (all period & group)
$CuInSe_2$	2,27	1,78	1,04	1,23 (118,27)	0,74 (71,15)
$CuInTe_2$	1,52	1,04	1,06	0,46 (42,45)	0,02 (1,89)
$CuGaS_2$	3,02	2,77	2,4	0,62 (25,83)	0,37 (15,42)
$AgAlS_2$	2,82	2,83	3,18	0,36 (11,01)	0,35 (11,01)
$AgGaS_2$	2,82	2,58	2,73	0,09 (3,3)	0,15 (5,49)
Cu_3PS_4	2,95	3,01	3,07	0,12 (3,58)	0,06 (1,95)
Zn_2GeS_4	2,99	2,71	3,96	0,97 (24,49)	1,25 (31,57)
$Hg_2P_2S_7$	2,55	2,57	2,81	0,26 (8,9)	0,24 (8,54)
α-Ca_2CdP_2	1,97	1,81	1,98	0,01 (0,48)	0,17 (8,59)
$BaAl_4Se_7$	1,87	2,34	3,4	1,53 (44,71)	1,06 (31,18)
Hg_3P2S_8	2,55	2,57	2,77	0,22 (7,58)	0,2 (7,22)
$Sn_2Ga_2S_5$	2,82	2,43	2,14	0,68 (31,78)	0,29 (13,55)
$Sn_2P_2S_6$	2,75	2,67	2,34	0,41 (17,52)	0,33 (14,1)
$RbIO_3$	3,58	3,55	4	0,42 (10,5)	0,45 (11,25)
$NaIn_3Se_5$	2,47	2,47	1,363	1,107 (80,7)	1,107 (81,22)
$Ba_2As_2S_5$	2,55	2,81	1,97	0,58 (29,44)	0,84 (42,64)
α-$LiIO_3$	4,03	4,1	4	0,03 (0,75)	0,1 (2,5)
$Y(IO_3)_3(H_2O)$	3,43	3,42	4,12	0,69 (16,5)	0,7 (16,99)
$La_2(IO_3)_6(H_2O)$	3,23	3,23	4,24	1,01 (23,58)	1,01 (23,82)
$Cd(IO_3)_2$	3,43	2,98	3,85	0,42 (10,65)	0,87 (22,6)

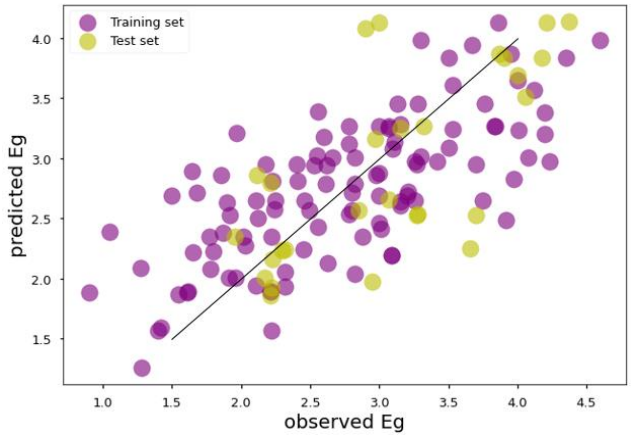

Figure 23: *The actual quaternary data versus the predicted value via PLS model* ***(taking in consideration the PCA results)***

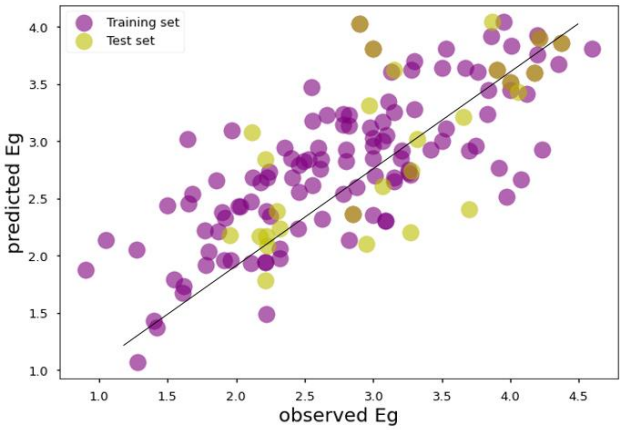

Figure 24: *The actual quaternary data versus the predicted value via PLS model* ***(taking all the variables)***

Table3: *The 20% test data which chosen randomly in quaternary data set the predicted and actual value are compared*

Compounds	Predicted Eg (eV)(PCA results)	Predicted Eg (eV)(all period & group)	Actual Eg (eV)[27, 39, 40, 44, 76-81]	Error (%) (PCA results)	Error (%) (all period & group)
Ag_2CdGeS_4	2,24	2,23	2,32	0,08 (3,45)	0,09 (3,88)
Li_2ZnSnS_4	3,27	3,01	3,32	0,05 (1,51)	0,31 (9,34)
Ag_2ZnSiS_4	2,54	2,73	3,28	0,74 (22,56)	0,55 (16,77)
$LiBSiS_4$	3,87	4,04	3,87	0 (0)	0,17 (4,39)
$ZnGa_2Ge_2S_8$	2,66	2,61	3,07	0,41 (13,36)	0,46 (14,98)
$BaGa_2SnSe_6$	2,35	2,17	1,95	0,4 (20,51)	0,22 (11,28)
$Mg_2In_3Si_2P_7$	2,8	2,84	2,21	0,59 (26,7)	0,63 (28,51)
$Ba_{10}In_6Zn_7S_{26}$	1,98	2,1	2,95	0,97 (32,88)	0,85 (28,81)
$TlBiP_2S_6$	1,92	2,16	2,22	0,3 (13,51)	0,06 (2,7)
$Ag_3Ga_3SiSe_8$	2,24	2,38	2,3	0,06 (2,61)	0,08 (3,48)

$PbGa_2SiSe_6$	2	2,16	2,17	0,17 (7,83)	0,01 (0,46)
Na_2BaSnS_4	2,53	2,19	3,27	0,74 (22,63)	1,08 (33,03)
Na_2BaGeS_4	2,53	2,4	3,7	1,17 (31,62)	1,3 (35,14)
$Ba_3RbGa_5Se_{10}$	1,86	1,78	2,21	0,35 (15,84)	0,43 (19,46)
$Ba_3KGa_5Se_{10}$	2,16	2,08	2,23	0,07 (3,14)	0,15 (6,73)
$LiCd_3PS_6$	3,16	3,31	2,97	0,19 (6,4)	0,34 (11,45)
$LiGa_2PS_6$	3,46	3,61	3,15	0,31 (3,17)	0,46 (14,6)
$BaLi_2GeS_4$	3,25	3,21	3,66	0,41 (38,52)	0,45 (12,3)
KAg_2PS_4	2,86	3,07	2,114	0,74 (35,48)	0,95 (45,22)
$K_2Mg(IO_3)_4.(H_2O)_2$	4,14	3,85	4,37	0,23 (5,26)	0,52 (11,9)
$BaGe(IO_3)_6$	3,51	3,43	4,06	0,55 (13,55)	0,63 (15,52)
$LiZn(IO_3)_3$	4,13	3,89	4,21	0,08 (1,9)	0,32 (7,6)
$LiCd(IO_3)_3$	3,83	3,6	4,18	0,35 (8,37)	0,58 (13,88)
$Li_2Ti(IO_3)_6$	4,13	3,8	3	1,13 (37,67)	0,8 (26,67)
$Li_2Sn(IO_3)$	3,83	3,62	3,9	0,07 (1,79)	0,28 (7,18)
$Na_2Sn(IO_3)$	3,68	3,5	4	0,32 (7,75)	0,5 (12,5)
$PbPt(IO_3)_6(H_2O)$	2,56	2,36	2,85	0,29 (9,82)	0,49 (17,19)
$CeF_2(IO_3)_2$	4,08	4,02	2,9	1,18 (40,69)	1,12 (38,62)

General conclusion

In summary, we have studied the dependencies between different variables of ternary and quaternary compounds on a large data set that contains 91 ternary compounds and 141 quaternary compounds. Unlike most of the previous studies in NLO materials development, herein, we made a study based on a mathematical models which is 'principal component analyses where it is normally used in a large data set for reducing dimensions. On the basis of our analysis and discussions, some general notes could be concluded:

- Negative correlation between the band gap Eg and the SHG effect d_{ij}
- Negative correlation between the band gap and the number of period of the elements and as well it is positive with the SHG effect d_{ij}
- Most the number of group of the elements has no correlation between both the band gap Eg and d_{ij}
- The distribution of the NLO compounds in the principal component space prove that the sulfides based chalcogenides and the elements with the third period in the periodic table are a good candidates for the non-linear optical applications.

In order to obtain materials that possess a moderate values of both the band gap and the d_{ij}, and based on the Gaussian distribution, the elements should be selected from the fourth period in the periodic table.

As a second part, we have demonstrated in this manuscript that based on Partial Least Square model and on the available knowledge whether experimental or theoretical on NLO materials, it is possible to predict new one that can achieve the balance between the band gap (or the laser damage threshold, the two ingredient are related) and SHG effect without any further experiments or calculations. Hence, the employ of the number of period and group in the periodic table for the prediction of the band gap Eg make problems, since the materials that possess the same element but different in the stoichiometric coefficients and also the different phases of compounds resulting the same predicting value. In the future work, adding a new parameters in the descriptors which represent the identity of each compound can solve this problem like the molar mass (M).

References:

1. Franken, P., et al., *Generation of optical harmonics.* Physical Review Letters, 1961. **7**(4): p. 118.
2. Bentria, B., et al., *Crystal structure of anhydrous bismuth iodate, Bi (IO3) 3.* Journal of chemical crystallography, 2003. **33**(11): p. 867-873.
3. Bentria, B., et al., *Crystal engineering strategy for quadratic nonlinear optics. Part II: Hg (IO3) 2.* Solid state sciences, 2003. **5**(2): p. 359-365.
4. Gacemi, A., et al., *Two mercury chloro thiocyanate complexes: NH4 [HgCl2 (SCN)] and NH4 [HgCl (SCN) 2].* Zeitschrift für anorganische und allgemeine Chemie, 2003. **629**(14): p. 2516-2520.
5. Gacemi, A., et al., *Crystal structure of dipotassium dibromo-bis (thiocyanato) mercurate (II), K2 [HgBr2 (SCN) 2].* Zeitschrift für Kristallographie-New Crystal Structures, 2005. **220**(1-4): p. 331-332.
6. Phanon, D., et al., *Crystal structure of M (IO3) 2 metal iodates, twinned bypseudo-merohedry, with MII: MgII, MnII, CoII, NiII and ZnII.* Zeitschrift für Kristallographie-Crystalline Materials, 2006. **221**(9): p. 635-642.
7. Bentria, B., et al., *Crystal Structure of a Cadmium Iodate Monohydrate: Cd (IO3) 2· H2O.* Zeitschrift für anorganische und allgemeine Chemie, 2004. **630**(6): p. 781-782.
8. Bentria, B., et al., *Polymorphism of Anhydrous Cadmium Iodate Structure of ε-Cd (IO3) 2.* Zeitschrift für anorganische und allgemeine Chemie, 2005. **631**(5): p. 894-901.
9. Taouti, M.B., et al., *Crystal structure of lanthanum triiodate iodic acid, La (IO3) 3 (HIO3).* Zeitschrift für Kristallographie-New Crystal Structures, 2008. **223**(3): p. 179-180.
10. Kurtz, S. and T. Perry, *A powder technique for the evaluation of nonlinear optical materials.* Journal of applied physics, 1968. **39**(8): p. 3798-3813.
11. Taouti, M., et al., *Synthèse et caractérisation de la solution solide ZN1-xCUx (IO3) 2.* Revue des sciences et sciences de l'ingénieur, 2010. **1**(1): p. 17-21.
12. Hebboul, Z., *Cadmium iodate syntheses and characterization.* Heavy Metals, 2018.
13. Hebboul, Z. and D. Benbertal, *Synthesis and characterization of new anhydrous cadmium iodate zeta polymorph ζ-Cd (IO3) 2.* J. Mater. Environ. Sci, 2018. **9**: p. 565.
14. Hebboul, Z., et al., *Simple new method for the preparation of La (IO3) 3 nanoparticles.* Nanomaterials, 2020. **10**(12): p. 2400.
15. Lagoun, B., B. Bentria, and I.K. Lefkaier, *Ab initio calculation of structural, electronic and optical properties of Hg (IO3) 2.* Physica B: Condensed Matter, 2014. **433**: p. 117-121.
16. Benghia, A., T. Dahame, and B. Bentria, *First principle calculation of physical properties of barium based chalcogenides BaM4S7 (M= Ga, Al); a DFT, DFT-D and hybrid functional HSE06 study.* Optical Materials, 2016. **54**: p. 269-275.
17. BENGHIA, A., T. DAHAME, and B. BENTRIA, *Ab initio investigation of the structural and optoelectronic properties of Aluminumthiophosphate AlPS4.*
18. Born, M. and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*2013: Elsevier.
19. Zheludev, I., *Theory of Spontaneous Polarization*, in *Physics of Crystalline Dielectrics*1971, Springer. p. 251-336.
20. Boyd, R.W., *Nonlinear optics*2020: Academic press.
21. Klienman, D.A., physical Review Letters, 1962. **20**.
22. Powell, R.C., *Symmetry, group theory, and the physical properties of crystals*. Vol. 824. 2010: Springer.
23. Dombeck, D.A., et al., *Uniform polarity microtubule assemblies imaged in native brain tissue by second-harmonic generation microscopy.* Proceedings of the National Academy of Sciences, 2003. **100**(12): p. 7081-7086.

24. Kang, L., et al., *Metal thiophosphates with good mid-infrared nonlinear optical performances: A first-principles prediction and analysis.* Journal of the American Chemical Society, 2015. **137**(40): p. 13049-13059.
25. Liang, F., et al., *Mid-infrared nonlinear optical materials based on metal chalcogenides: structure–property relationship.* Crystal Growth & Design, 2017. **17**(4): p. 2254-2289.
26. Chung, I. and M.G. Kanatzidis, *Metal chalcogenides: a rich source of nonlinear optical materials.* Chemistry of Materials, 2014. **26**(1): p. 849-869.
27. Kang, L., et al., *First principles selection and design of mid-IR nonlinear optical halide crystals.* Journal of Materials Chemistry C, 2013. **1**(44): p. 7363-7370.
28. Gallais, L., *Endommagement laser-Présentation de la thématique.* 2012.
29. Bertussi, B., et al., *Initiation of laser-induced damage sites in fused silica optical components.* Optics Express, 2009. **17**(14): p. 11469-11479.
30. Wang, L., et al., *A widely tunable (5–12.5 μm) continuous-wave mid-infrared laser spectrometer based on difference frequency generation in AgGaS2.* Optics communications, 2011. **284**(1): p. 358-362.
31. Chernov, A., N. Zaitseva, and L. Rashkovich, *Secondary nucleation induced by the cracking of a growing crystal: KH2PO4 (KDP) and K (H, D) 2PO4 (DKDP).* Journal of crystal growth, 1990. **102**(4): p. 793-800.
32. Yong, Z., et al., *Sixth harmonic of a Nd: YVO4 laser generation in KBBF for ARPES.* Chinese Physics Letters, 2008. **25**(3): p. 963.
33. Wu, Q., et al., *RbIO3 and RbIO2F2: Two promising nonlinear optical materials in mid-IR region and influence of partially replacing oxygen with fluorine for improving laser damage threshold.* Chemistry of Materials, 2016. **28**(5): p. 1413-1418.
34. Huang, Y., et al., *A 2 BiI 5 O 15 (A= K+ or Rb+): two new promising nonlinear optical materials containing [I 3 O 9] 3− bridging anionic groups.* Journal of Materials Chemistry C, 2014. **2**(20): p. 4057-4062.
35. Jia, Y.J., et al., *LiMII (IO3) 3 (MII= Zn and Cd): Two Promising Nonlinear Optical Crystals Derived from a Tunable Structure Model of α-LiIO3.* Angewandte Chemie International Edition, 2019. **58**(48): p. 17194-17198.
36. Zawilski, K.T., et al., *Growth and characterization of large CdSiP2 single crystals.* Journal of crystal growth, 2010. **312**(8): p. 1127-1132.
37. Harrison, R., et al., *Efficient multikilowatt mid infrared difference frequency generation in CdGeAs 2.* IEEE Journal of Quantum Electronics, 1982. **18**(8): p. 1239-1242.
38. Nikogosyan, D.N., *Main Infrared Materials.* Nonlinear Optical Crystals: A Complete Survey, 2005: p. 75-114.
39. Zondy, J.-J., *Experimental investigation of single and twin AgGaSe2 crystals for CW 10.2 μm SHG.* Optics communications, 1995. **119**(3-4): p. 320-326.
40. Eckardt, R., et al., *Efficient second harmonic generation of 10-μm radiation in AgGaSe2.* Applied physics letters, 1985. **47**(8): p. 786-788.
41. Chung, I., *Exploratory synthesis in molten salts: Characterization, nonlinear optical and phase-change properties of new chalcophosphate compounds*2008.
42. Chung, I., et al., *APSe6 (A= K, Rb, and Cs): Polymeric selenophosphates with reversible phase-change properties.* Inorganic Chemistry, 2004. **43**(9): p. 2762-2764.
43. Lin, X., G. Zhang, and N. Ye, *Growth and characterization of BaGa4S7: a new crystal for mid-IR nonlinear optics.* Crystal Growth and Design, 2009. **9**(2): p. 1186-1189.
44. Yao, J., et al., *Growth and characterization of BaGa4Se7 crystal.* Journal of crystal growth, 2012. **346**(1): p. 1-4.

45. Lin, H., et al., *AX II4 X III5 Te 12 (A= Rb, Cs; X II= Mn, Zn, Cd; X III= Ga, In): quaternary semiconducting tellurides with very low thermal conductivities.* Dalton Transactions, 2016. **45**(44): p. 17606-17609.
46. Lin, H., et al., *Functionalization based on the substitutional flexibility: strong middle IR nonlinear optical selenides AXII4XIII5Se12.* Journal of the American Chemical Society, 2013. **135**(34): p. 12914-12921.
47. Schunemann, P.G., K.T. Zawilski, and T.M. Pollak, *Horizontal gradient freeze growth of AgGaGeS4 and AgGaGe5Se12.* Journal of crystal growth, 2006. **287**(2): p. 248-251.
48. Witten, I.H., et al. *Practical machine learning tools and techniques*. in *DATA MINING*. 2005.
49. Russell, R., *Machine Learning: Step-By-Step Guide to Implement Machine Learning Algorithms with Python*2020: (Knxb).
50. Barlow, H.B., *Unsupervised learning.* Neural computation, 1989. **1**(3): p. 295-311.
51. Ghahramani, Z. *Unsupervised learning*. in *Summer school on machine learning*. 2003. Springer.
52. Hastie, T., R. Tibshirani, and J. Friedman, *Unsupervised learning*, in *The elements of statistical learning*2009, Springer. p. 485-585.
53. Cunningham, P., M. Cord, and S.J. Delany, *Supervised learning*, in *Machine learning techniques for multimedia*2008, Springer. p. 21-49.
54. Pearson, K., *Principal components analysis.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901. **6**(2): p. 559.
55. Hotelling, H., *Analysis of a complex of statistical variables into principal components.* Journal of educational psychology, 1933. **24**(6): p. 417.
56. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.
57. Wold, H., *Path models with latent variables: The NIPALS approach*, in *Quantitative sociology*1975, Elsevier. p. 307-357.
58. Mateos-Aparicio, G., *Partial least squares (PLS) methods: Origins, evolution, and application to social sciences.* Communications in Statistics-Theory and Methods, 2011. **40**(13): p. 2305-2317.
59. Pirouz, D.M., *An overview of partial least squares.* Available at SSRN 1631359, 2006.
60. Cha, J., *Partial least squares.* Advanced Methods of Marketing Research, Richard P. Bagozzi, ed. Cambridge, MA: Blackwell, 1994: p. 52-78.
61. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics: A Journal of the Chemometrics Society, 2003. **17**(3): p. 166-173.
62. Chan, J., *Learn Python in one day and learn it well: Python for beginners with hands-on project: the only book you need to start coding in Python immediately*2015: CreateSpace Independent Publishing.
63. Wu, T., et al., *From Ce (IO 3) 4 to CeF 2 (IO 3) 2: fluorinated homovalent substitution simultaneously enhances SHG response and bandgap for mid-infrared nonlinear optics.* Journal of Materials Chemistry C, 2021. **9**(28): p. 8987-8993.
64. Zhou, M., et al., *Hg-based infrared nonlinear optical material KHg4Ga5Se12 exhibits good phase-matchability and exceptional second harmonic generation response.* Chemistry of Materials, 2017. **29**(18): p. 7993-8002.
65. Zhou, M., et al., *K 2 Sn 2 ZnSe 6, Na 2 Ge 2 ZnSe 6, and Na 2 In 2 GeSe 6: a new series of quaternary selenides with intriguing structural diversity and nonlinear optical properties.* Dalton Transactions, 2016. **45**(18): p. 7627-7633.
66. Zhou, M., et al., *Midinfrared nonlinear optical thiophosphates from LiZnPS4 to AgZnPS4: a combined experimental and theoretical study.* Inorganic Chemistry, 2016. **55**(8): p. 3724-3726.
67. Zhou, A.-Y., et al., *Ba10In6Zn7S26-n ZnS: An Inorganic Composite System with Interface Phase-Matching Tuned for High-Performance Infrared Nonlinear Optical Materials.* Inorganic Chemistry, 2019. **58**(6): p. 3990-3999.

68. Zhang, M.-J., et al., *Ln 3 GaS 6 (Ln= Dy, Y): new infrared nonlinear optical materials with high laser induced damage thresholds.* Dalton Transactions, 2013. **42**(39): p. 14223-14229.
69. Yin, W., et al., *Synthesis, structure, and properties of Li2In2MQ6 (M= Si, Ge; Q= S, Se): a new series of IR nonlinear optical materials.* Inorganic Chemistry, 2012. **51**(10): p. 5839-5843.
70. Yelisseyev, A., et al., *A New Nonlinear Optical Selenide Crystal AgLiGa2Se4 with Good Comprehensive Performance in Mid-Infrared Region.* Advanced Optical Materials, 2021. **9**(5): p. 2001856.
71. Yelisseyev, A., et al., *The optical properties of the nonlinear crystal BaGa4Se7.* Optical Materials, 2020. **99**: p. 109564.
72. Yang, H.D., et al., *The Rise of Infrared Nonlinear Optical Pnictides: Advances and Outlooks.* Chemistry–An Asian Journal, 2021. **16**(21): p. 3299-3310.
73. Yalcin, B.G., *Thermoelectric properties of stannite-phase CuZn2AS4 (CZAS; A= Al, Ga and In) nanocrystals for solar energy conversion applications.* Philosophical Magazine, 2016. **96**(21): p. 2280-2299.
74. Xue, D. and S. Zhang, *Linear and nonlinear optical properties of KNbO3.* Chemical physics letters, 1998. **291**(3-4): p. 401-406.
75. Xu, X., et al., *α-AgI3O8 and β-AgI3O8 with large SHG responses: polymerization of IO3 groups into the I3O8 polyiodate anion.* Chemistry of Materials, 2014. **26**(10): p. 3219-3230.
76. Xing, W., et al., *From AgGaS 2 to AgHgPS 4: vacancy defects and highly distorted HgS 4 tetrahedra double-induced remarkable second-harmonic generation response.* Journal of Materials Chemistry C, 2021. **9**(3): p. 1062-1068.
77. Xing, W., et al., *EuHgGeSe4 and EuHgSnS4: two quaternary Eu-based infrared nonlinear optical materials with strong second-harmonic-generation responses.* Inorganic Chemistry, 2020. **59**(24): p. 18452-18460.
78. Xing, W., et al., *Highly distorted [HgS4] motif-driven structural symmetry degradation and strengthened second-harmonic generation response in the defect diamond-like chalcogenide Hg3P2S8.* ACS Applied Materials & Interfaces, 2021. **13**(31): p. 37331-37338.
79. Xiao, L., et al., *A new cerium iodate infrared nonlinear optical material with a large second-harmonic generation response.* Journal of Materials Chemistry C, 2017. **5**(8): p. 2130-2134.
80. Wu, K., et al., *New compressed chalcopyrite-like Li2BaMIVQ4 (MIV= Ge, Sn; Q= S, Se): promising infrared nonlinear optical materials.* Journal of the American Chemical Society, 2017. **139**(42): p. 14885-14888.
81. Wu, K., Z. Yang, and S. Pan, *Na2BaMQ4 (M= Ge, Sn; Q= S, Se): infrared nonlinear optical materials with excellent performances and that undergo structural transformations.* Angewandte Chemie International Edition, 2016. **55**(23): p. 6713-6715.
82. Wu, K., Z. Yang, and S. Pan, *Na4MgM2Se6 (M= Si, Ge): the first noncentrosymmetric compounds with special ethane-like [M2Se6] 6–units exhibiting large laser-damage thresholds.* Inorganic Chemistry, 2015. **54**(21): p. 10108-10110.
83. Vasilyeva, L.I.I.a.I.G., *Nonlinear LiBIIICYI 2 crystals for mid-IR and far-IR:Novel aspects in crystal growth.* crystal growth, 2007.
84. Tsuji, I., et al., *Novel stannite-type complex sulfide photocatalysts AI2-Zn-AIV-S4 (AI= Cu and Ag; AIV= Sn and Ge) for hydrogen evolution under visible-light irradiation.* Chemistry of Materials, 2010. **22**(4): p. 1402-1409.
85. Toffoli, P., et al., *Structure du tétrathiophosphate (V) de zinc et d'argent, ZnAgPS4.* Acta Crystallographica Section C: Crystal Structure Communications, 1985. **41**(5): p. 645-647.
86. Taouti, M.B., et al., *Structures, thermal behaviors, and luminescent properties of anhydrous lanthanum iodate polymorphs.* Inorganic Chemistry, 2015. **54**(7): p. 3608-3618.

87. TAOUTI, M., et al., *Synthèse et caractérisation de la solution solide ZN1-xCUx (IO3) 2.* Revue des sciences et sciences de l'ingénieur, 2010. **1**(1): p. 17-21.
88. Sun, Y., et al., *α-Ca2CdP2 and β-Ca2CdP2: Two Polymorphic Phosphide-Based Infrared Nonlinear Crystals with Distorted NLO-Active Tetrahedral Motifs Realizing Large Second Harmonic Generation Effects and Suitable Band Gaps.* Inorganic Chemistry, 2021. **60**(10): p. 7553-7560.
89. Sun, Y., et al., *LaSiP3 and LaSi2P6: Two Excellent Rare-Earth Pnictides with Strong SHG Responses as Mid-and Far-Infrared Nonlinear Optical Crystals.* Advanced Optical Materials, 2021. **9**(10): p. 2002176.
90. Sun, C.-F., et al., *Polar or Non-Polar? Syntheses, crystal structures, and optical properties of three new palladium (II) iodates.* Inorganic Chemistry, 2010. **49**(20): p. 9581-9589.
91. Sun, C.-F., C.-L. Hu, and J.-G. Mao, *PbPt (IO 3) 6 (H 2 O): a new polar material with two types of stereoactive lone-pairs and a very large SHG response.* Chemical Communications, 2012. **48**(35): p. 4220-4222.
92. Shi, T., et al., *Structural, electronic, and optical properties of Cu3-V-VI4 compound semiconductors.* Applied physics letters, 2013. **103**(15): p. 152105.
93. Schwarz, K., *DFT calculations of solids with LAPW and WIEN2k.* Journal of Solid State Chemistry, 2003. **176**(2): p. 319-328.
94. Rosmus, K.A., et al., *Optical nonlinearity in Cu2CdSnS4 and α/β-Cu2ZnSiS4: Diamond-like semiconductors with high laser-damage thresholds.* Inorganic Chemistry, 2014. **53**(15): p. 7809-7811.
95. Reshak, A.H. and S. Auluck, *Electronic structure, linear, nonlinear optical susceptibilities and birefringence of CuInX2 (X= S, Se, Te) chalcopyrite-structure compounds.* PMC Physics B, 2008. **1**(1): p. 1-17.
96. Phanon, D., et al., *Optical properties of Nd 3+ and Yb 3+-doped AgM (IO 3) 4 metal iodates: transparent host matrices for mid-IR lasers and nonlinear materials.* Journal of Materials Chemistry C, 2014. **2**(15): p. 2715-2723.
97. Phanon, D. and I. Gautier-Luneau, *Promising material for infrared nonlinear optics: NaI3O8 salt containing an octaoxotriiodate (V) anion formed from condensation of [IO3]− ions.* Angewandte Chemie International Edition, 2007. **46**(44): p. 8488-8491.
98. Phanon, D., et al., *New potential materials for infrared nonlinear optics. Preparation, characterisation and optical transparency of monometallic and bimetallic iodates.* Solid state sciences, 2006. **8**(12): p. 1466-1472.
99. Petrov, V., et al., *Second harmonic generation and optical parametric amplification in the mid-IR with orthorhombic biaxial crystals LiGaS2 and LiGaSe2.* Applied Physics B, 2004. **78**(5): p. 543-546.
100. Petrov, V., et al., *Growth and characterization of the chalcopyrite LiGaTe2: A highly non-linear birefringent optical crystal for the mid-infrared.* Journal of non-crystalline solids, 2006. **352**(23-25): p. 2434-2438.
101. Olekseyuk, I., et al., *Single-crystal growth and properties of AgCd2GaS4.* Journal of crystal growth, 2005. **279**(1-2): p. 140-145.
102. Morris, C.D., et al., *Molecular germanium selenophosphate salts: phase-change properties and strong second harmonic generation.* Journal of the American Chemical Society, 2012. **134**(51): p. 20733-20744.
103. Mei, D., et al., *Ag 3 Ga 3 SiSe 8: a new infrared nonlinear optical material with a chalcopyrite structure.* CrystEngComm, 2014. **16**(30): p. 6836-6840.
104. Mao, F.-F., et al., *A series of mixed-metal germanium iodates as second-order nonlinear optical materials.* Chemistry of Materials, 2018. **30**(7): p. 2443-2452.
105. Luo, M., et al., *Rational design of the nonlinear optical response in a tin iodate fluoride Sn (IO3) 2F2.* Chemistry of Materials, 2020. **32**(6): p. 2615-2620.

106. Lin, L., et al., *Incorporating rare-earth cations with moderate electropositivity into iodates for the optimized second-order nonlinear optical performance.* Inorganic Chemistry Frontiers, 2020. **7**(15): p. 2736-2746.
107. Lin, H., L.-J. Zhou, and L. Chen, *Sulfides with strong nonlinear optical activity and thermochromism: ACd4Ga5S12 (A= K, Rb, Cs).* Chemistry of Materials, 2012. **24**(17): p. 3406-3414.
108. Lin, H., et al., *Strong Infrared NLO Tellurides with Multifunction: CsXII4In5Te12 (XII= Mn, Zn, Cd).* Inorganic Chemistry, 2016. **55**(9): p. 4470-4475.
109. Li, S.-F., et al., *Syntheses, structures, and nonlinear optical properties of two sulfides Na2In2MS6 (M= Si, Ge).* Inorganic Chemistry, 2016. **55**(4): p. 1480-1485.
110. Li, S.-F., et al., *Strong infrared nonlinear optical efficiency and high laser damage threshold realized in quaternary alkali metal sulfides Na2Ga2MS6 (M= Ge, Sn) containing mixed nonlinear optically active motifs.* Inorganic Chemistry, 2018. **57**(12): p. 6783-6786.
111. Li, S.-F., et al., *Superpolyhedron-built second harmonic generation materials exhibit large mid-infrared conversion efficiencies and high laser-induced damage thresholds.* Chemistry of Materials, 2017. **29**(4): p. 1796-1804.
112. Li, P.-X., et al., *Explorations of new second-order nonlinear optical materials in the KI-MII-IV-O systems.* Inorganic Chemistry, 2010. **49**(10): p. 4599-4605.
113. Li, M.-Y., et al., *Sn2Ga2S5: a polar semiconductor with exceptional infrared nonlinear optical properties originating from the combined effect of mixed asymmetric building motifs.* Chemistry of Materials, 2019. **31**(16): p. 6268-6275.
114. Li, G.-m., et al., *Na 2 CdGe 2 Q 6 (Q= S, Se): two metal-mixed chalcogenides with phase-matching abilities and large second-harmonic generation responses.* Dalton Transactions, 2017. **46**(9): p. 2778-2784.
115. Li, G., et al., *Na2ZnGe2S6: a new infrared nonlinear optical material with good balance between large second-harmonic generation response and high laser damage threshold.* Journal of the American Chemical Society, 2016. **138**(23): p. 7422-7428.
116. Li, D., et al., *Theoretical studies of the structural, electronic, and optical properties of Cu {sub 2} HgGeS {sub 4}.* Physica Status Solidi B. Basic Solid State Physics, 2012. **249**.
117. Lekse, J.W., et al., *Second-harmonic generation and crystal structure of the diamond-like semiconductors Li2CdGeS4 and Li2CdSnS4.* Inorganic Chemistry, 2009. **48**(16): p. 7516-7518.
118. Lekse, J.W., et al., *Synthesis, physicochemical characterization and crystallographic twinning of Li2ZnSnS4.* Journal of Solid State Chemistry, 2008. **181**(12): p. 3217-3222.
119. Kumar, S.R., et al., *Crystal growth and characterization of monometallic NLO single crystals of Cd (IO3) 2.* Journal of crystal growth, 2009. **311**(2): p. 337-341.
120. Isaenko, L. and I. Vasilyeva, *Nonlinear LiBIIICYI2 crystals for mid-IR and far-IR: Novel aspects in crystal growth.* Journal of crystal growth, 2008. **310**(7-9): p. 1954-1960.
121. Huang, J.-b., et al., *Hierarchical active factors to band gap and nonlinear optical response in Ag-containing quaternary-chalcogenide compounds.* Journal of Solid State Chemistry, 2016. **239**: p. 30-35.
122. Huang, H., et al., *Hydrothermal synthesis, nonlinear optical property and photocatalytic activity of a non-centrosymmetric AgIO3 photocatalyst under UV and visible light irradiation.* Solid state sciences, 2015. **46**: p. 37-42.
123. Huang, C., et al., *Explorations of a series of second order nonlinear optical materials based on monovalent metal gold (III) iodates.* Inorganic Chemistry, 2013. **52**(19): p. 11551-11562.
124. Hu, Q., et al., *Synthesis and nonlinear optical properties of new gallium thiophosphate Rb 2 Ga 2 P 2 S 9.* New Journal of Chemistry, 2019. **43**(31): p. 12468-12474.
125. Ho, C.-H. and C.-C. Pan, *Optical behavior and structural property of CuAlS 2 and AgAlS 2 wide-bandgap chalcopyrites.* Applied Optics, 2014. **53**(22): p. E7-E13.

126. Hellstrom, E. and R. Huggins, *A study of the systems M2S · Al2S3, M= Li, Na, K; preparation, phase study and electrical conductivity.* Materials Research Bulletin, 1979. **14**(7): p. 881-889.
127. Haynes, A.S., et al., *Phase-Change Behavior and Nonlinear Optical Second and Third Harmonic Generation of the One-Dimensional K (1–x) Cs x PSe6 and Metastable β-CsPSe6.* Chemistry of Materials, 2015. **27**(5): p. 1837-1846.
128. Ghosh, A., R. Thangavel, and A. Gupta, *Chemical synthesis, characterization and theoretical investigations of stannite phase CuZn 2 AlS 4 nanocrystals.* New Journal of Chemistry, 2016. **40**(2): p. 1149-1154.
129. Du, X., et al., *Li2MTeO6 (M= Ti, Sn): Mid-Infrared Nonlinear Optical Crystal with Strong Second Harmonic Generation Response and Wide Transparency Range.* Angewandte Chemie International Edition, 2021. **60**(43): p. 23320-23326.
130. Devlin, K.P., et al., *Polymorphism and second harmonic generation in a novel diamond-like semiconductor: Li2MnSnS4.* Journal of Solid State Chemistry, 2015. **231**: p. 256-266.
131. Chen, M.-M., et al., *M 2 As 2 Q 5 (M= Ba, Pb; Q= S, Se): a source of infrared nonlinear optical materials with excellent overall performance activated by multiple discrete arsenate anions.* Journal of Materials Chemistry C, 2021. **9**(4): p. 1156-1163.
132. Chen, J., et al., *BaGe2Pn2 (Pn= P, As): Two congruent-melting non-chalcopyrite pnictides as mid- and far-infrared nonlinear optical materials exhibiting large second harmonic generation effects.* Chemistry of Materials, 2019. **31**(24): p. 10170-10177.
133. Chen, J., et al., *Mg2In3Si2P7: a quaternary diamond-like phosphide infrared nonlinear optical material derived from ZnGeP2.* Journal of the American Chemical Society, 2021. **143**(27): p. 10309-10316.
134. Chen, H.-Y., et al., *Ternary rare-earth sulfides RE3GaS6 (RE= Ho, Er): Crystal chemistry, second-order nonlinear optical properties and theoretical investigation.* Journal of alloys and compounds, 2021. **868**: p. 159112.
135. Chang, H.-Y., et al., *Polar or Nonpolar? A+ Cation Polarity Control in A2Ti (IO3) 6 (A= Li, Na, K, Rb, Cs, Tl).* Journal of the American Chemical Society, 2009. **131**(19): p. 6865-6873.
136. Chang, H.-Y., et al., *Alignment of lone pairs in a new polar material: synthesis, characterization, and functional properties of Li2Ti (IO3) 6.* Journal of the American Chemical Society, 2009. **131**(7): p. 2426-2427.
137. Brunetta, C.D., et al., *Cation ordering and physicochemical characterization of the quaternary diamond-like semiconductor Ag2CdGeS4.* Journal of Solid State Chemistry, 2012. **187**: p. 177-185.
138. Brunetta, C.D., et al., *The crystal and electronic band structure of the diamond-like semiconductor Ag2ZnSiS4.* Journal of alloys and compounds, 2012. **516**: p. 65-72.
139. Brant, J.A., et al., *Outstanding laser damage threshold in Li2MnGeS4 and tunable optical nonlinearity in diamond-like semiconductors.* Inorganic Chemistry, 2015. **54**(6): p. 2809-2819.
140. Boyd, G., H. Kasper, and J. McFee, *Linear and nonlinear optical properties of AgGaS 2, CuGaS 2, and CuInS 2, and theory of the wedge technique for the measurement of nonlinear coefficients.* IEEE Journal of Quantum Electronics, 1971. **7**(12): p. 563-573.
141. Bera, T.K., et al., *Soluble semiconductors AAsSe2 (A= Li, Na) with a direct-band-gap and strong second harmonic generation: a combined experimental and theoretical study.* Journal of the American Chemical Society, 2010. **132**(10): p. 3484-3495.
142. Abudouwufu, T., et al., *Ce (IO3) 2F2· H2O: The First Rare-Earth-Metal Iodate Fluoride with Large Second Harmonic Generation Response.* Chemistry–A European Journal, 2019. **25**(5): p. 1221-1226.
143. Yan, W. and N.A. Tinker, *An integrated biplot analysis system for displaying, interpreting, and exploring genotype× environment interaction.* Crop Science, 2005. **45**(3): p. 1004-1016.

International Congress of Energies and Engineering of Industrial ProcessesCEGPI'22

Algiers, Algeria 23 - 25 May 2022

CERTIFICATE OF PARTICIPATION

This certificate is proudly presented to

MECHRAOUI Kaima Bouchra

For her participation at International Congress of Energies and Engineering, of Industrial Processes, CEGPI'22, with ***Poster Presentation***

Title : **Data mining of physical properties and nonlinear optical parameters for mat enhancement.**

Co-Author: **A.BENGHIA, S.FERCHANE.**

23 - 25 May 2022

Date

Dr SAHRAOUI Naima

Chair of CEGPI'22

Dr SABBA Nassila

Honorary Chair of CEGPI'22

ملخص:

المواد الضوئية الغير خطية في الاشعة الحمراء المتوسطة ذات الأداء العالي لها أهمية كبيرة للتطبيقات التكنولوجية في العديد من المجالات المدنية والعسكرية، ومع ذلك لايزال من الصعب اكتشاف مواد ضوئية غير خطية جديدة تحقق لنا توازنا جيدا بين معامل الضوء الغير الخطي وطاقة الفجوة. في هذا العمل، درسنا العلاقة بين مواضع الذرات في الجدول الدوري (سطر وعمود) لمركبات الضوء الغير خطي وعلاقتها بطاقة الفجوة ومعاملات الضوء الغير خطي وذلك بواسطة تحليل المركبات الرئيسية. تم اختيار 91 مركب ثلاثي و141 مركب رباعي من بيانات الحسابات التجريبية وحسابات تعتمد على نظرية الكثافة الدالية. بناءً على هذه النتائج ونماذج التعلم الآلي، استخدمنا المربع الجزئي الاصغر للتنبؤ بطاقة الفجوة لبلورات الضوء الغير خطي، حيث توصلنا الى علاقة خطية تربط بين طاقة الفجوة ومواضع الذرات في الجدول الدوري.

الكلمات المفتاحية: مواد الضوء الغير خطي، تحليل المركبات الرئيسية، المربع الجزئي الاصغر، طاقة الفجوة.

Abstract:

Mid-infrared (IR) nonlinear optical (NLO) materials with high performance have a great importance for technological applications in many civil and military fields. However, it still a big challenge to discover new mid-IR NLO materials that achieve a good balance between the NLO effect and band gap. In this work, we study the correlation between the position of the atoms in the periodic table (period and group) and the band gap and nonlinear optical coefficients for the NLO compounds took advantage of the principal component analysis (PCA) .where 91 ternary and 141 quaternary compounds from experimental and DFT calculations data are selected. Based on these results and machine learning models, we employed partial least square (PLS) to predict the band gap of the NLO crystals. Since we get a linear relation that correlate the band gap and the position of the atoms in periodic table.

Keywords: NLO materials, principal component analysis (PCA), partial least square (PLS), band gap.