



A new method for an old topic: Efficient and reliable estimation of material bulk modulus

Peng Wang^{a,b,*}, Yi Qin^b, Ming Cheng^b, Guanjie Wang^c, Dongbin Xiu^d, Zhimei Sun^{c,*}

^a Beijing Advanced Innovation Center for Big Data and Brain Computing, China

^b LMIB & School of Mathematics and System Sciences, Beihang University, Beijing, China

^c School of Material Science and Engineering, Beihang University, Beijing, China

^d Department of Mathematics, Ohio State University, Columbus, USA

ARTICLE INFO

Keywords:

Uncertainty quantification

Bulk modulus

Generalized polynomial chaos expansion

ab initio calculations

ABSTRACT

Bulk modulus is a key material property which provides fundamentals on the chemical bonding nature of a material as well as for deriving other related material properties, such as Young's modulus and Grüneisen constant. Traditional numerical methods of estimating bulk modulus often involve layers of approximations and cost considerably. In this work, we propose a novel and efficient numerical framework to estimate bulk modulus using the *ab initio* calculated data. Based on the generalized polynomial chaos expansion (gPC) method, our approach does not impose any physical priori and is mathematically rigorous and versatile. We have demonstrated the reliability and efficiency of the proposed method by estimating the bulk modulus of Ti_3SiC_2 and Sb_2Te_3 with comparison to conventional methods.

1. Introduction

Bulk modulus, a key material parameter, is defined as the ratio of infinitesimal pressure change to the fractional volume compression. It provides fundamental insights to the chemical bounding nature of a material and is instrumental in deriving other key material properties. Consequently, it has been widely used as benchmarks in materials discovery and design [1,2]. Traditionally, numerical methods have been used to estimate the bulk modulus using the calculated data of energy versus volume from *ab initio* calculations [3]. Normally, there are two distinct approaches and both start with an initial estimation of the equilibrium energy and volume ($E_{0\text{in}}$, $V_{0\text{in}}$) from *ab initio* calculations. The Voigt-Reuss-Hill [4] approach obtains elastic constants of the crystal system by applying a given homogeneous deformation (strain) to calculate the overall stress matrix [5,6]. The Voigt [7] and Reuss bound [8] can be then determined from those elastic constants and treated as upper and lower limits of the bulk modulus, respectively. Their arithmetic average, the Hill average [9], is then taken as the bulk modulus. Alternatively, one can employ a usual energy-volume approach and fit a set of *ab initio* data points by applying a small deformation of the initial equilibrium volume to an Equation of State (EoS), for example, the Birch-Murnaghan model [10]. The bulk modulus is then determined from the model coefficients.

Although the aforementioned numerical approaches have proved

effective in calculating the bulk modulus for an individual material, their associated costs and uncertainty from numerous assumptions may become problematic for high-throughput analysis in the new age of material genome. For example, many methods requires *ab initio* data as inputs whose Perdew-Burke-Ernzerhof (PBE) exchange–correlation functional is often determined with generalized gradient approximations (GGA) or the local density approximation (LDA). Without any update, those approximation errors would propagate as parametric uncertainty and might even be amplified in the process, and therefore reduce the reliability of our estimated results. Meanwhile, current EoS are derived under specific physical conditions and may lead to errors for new materials with unknown properties. For instance, the popular Birch-Murnaghan model [10] performs well for moderate values of the Grüneisen constant [11], but deteriorates under high pressure or large change of volume [12].

To address such challenge, we propose a novel numerical framework to estimate the bulk modulus. Similar to the energy-volume method that requires an initial guess of the equilibrium volume from *ab initio* calculation, our numerical approach employ the generalized polynomial chaos method from uncertainty quantification and aims to: (1) reduce the overall computational costs for a desired accuracy; (2) relax any physical assumptions except those in the *ab initio* calculations; and lastly, (3) provide a reliability estimation of the results.

The remainder of this paper is organized below: in Section 2, we

* Corresponding authors.

E-mail addresses: wang.peng@buaa.edu.cn (P. Wang), zmsun@buaa.edu.cn (Z. Sun).

<https://doi.org/10.1016/j.commsci.2019.04.022>

Received 12 January 2019; Received in revised form 29 March 2019; Accepted 13 April 2019

Available online 19 April 2019

0927-0256/© 2019 Elsevier B.V. All rights reserved.

provide the theoretical background of our approach. Its effectiveness would be investigated in Section 3 via two benchmark material systems: one on Ti_3SiC_2 with comparison to results from the popular the Birch-Murnaghan EoS, and one on Sb_2Te_3 using the Voigt-Reuss-Hill method. A thorough numerical study on the effectiveness of the proposed scheme regarding the number, locations and range of input data is also included. The final conclusions are drawn in Section 4.

2. Methods

In this section, we present our numerical framework that utilized the initial guess of equilibrium volume from *ab initio* calculation. We first explain the energy-volume (V_{oin}) approximation with generalized polynomial chaos method and then show the estimation procedure of bulk modulus. Reliability analysis on the estimated results are also included.

2.1. Energy-volume relationship

The generalized polynomial chaos (gPC) method [13] is an extension of the seminal work on Hermite polynomial chaos [14] and has become one of the most popular uncertainty quantification methods. To be specific, we construct a N -th order gPC expansion $E_N(V)$ to approximate the relation between uncertain system input (the volume V) and system output (the energy E):

$$E(V) \approx E_N(V) = \sum_{i=0}^N c_i \phi_i(V), \quad (1)$$

where i refers to the degree of the Wiener-Askey polynomials $\phi_i(V)$ and c_i is the corresponding coefficient.

In contrast to conventional polynomials, such as, $\{x^0, x^1, \dots, x^N\}$, the generalized polynomials chaos are orthogonal basis in the Euclidean L_2 space.

$$\mathbb{E}[\phi_i(V)\phi_j(V)] = \int_{\Omega} \phi_i(V')\phi_j(V')\omega dV' = \delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}, \quad 0 \leq i, j \leq N, \quad (2)$$

where $\mathbb{E}[\cdot]$ is the expectation function, Ω is the outcome space (e.g. all possible values) of volume, ω is the weight function and δ_{ij} is the Kronecker delta function. In order to construct an expansion that closely approximates the target function $E(V)$, both conventional and orthogonal polynomials require one to compute their coefficients c_i , $i = 1, 1, \dots, N$. This amounts to the computation of the inner product (e.g. integral) between the target function and the polynomials:

$$c_i = \frac{1}{\|\phi_i\|_{L_2}} (E, \phi_i)_{L_2}, \quad i = 0, 1, \dots, N, \quad (3)$$

where $\|\cdot\|_{L_2}$ is the (Euclidean) L_2 -type of norm, $(\cdot, \cdot)_{L_2}$ refers to the inner product. Since many quadrature rules are based on the orthogonal polynomials, for instance, the Gauss-Legendre rule, the coefficients can be computed with sufficient accuracy with less data by the orthogonal polynomials.

Exact forms of the Wiener-Askey polynomials $\phi_i(V)$ are dependent on the weight function of the input variable [15], i.e., the material volume V . With no prior knowledge on which volume samples to take around the initial guess of the equilibrium volume V_{oin} , (in other words, every sample in the domain has the same chance to be selected), we prescribe identical weight ($\omega = 1$) to each sample V_i , $i = 1, \dots, M$. Consequently, the orthogonal polynomials $\phi_i(V)$ become Legendre polynomials [15] and can be defined by the following three-term recurrence formula:

$$(i+1)\phi_{i+1} = (2i+1)V\phi_i - i\phi_{i-1}, \quad (4)$$

In Table 1, we tabulate its first few terms:

One may also prescribe other types of weight function to the input

Table 1

The first few orders of Legendre polynomials $\phi_i(V)$.

index i	Legendre polynomial $\phi_i(V)$
− 1	0
0	1
1	V
2	$(3V^2 - 1)/2$
3	$(5V^3 - 3V)/2$
4	$(35V^4 - 30V^2 + 3)/8$
5	$(63V^5 - 70V^3 + 15V)/8$

variable. For example, if V is assumed as a Gaussian random variables, its weight function would be the Gaussian probabilistic density function and its corresponding Wiener-Askey polynomial is the Hermite polynomials. The choice of distribution allows one to accurately compute the statistical moments (e.g., mean, variance).

2.2. Estimation of bulk modulus

Having constructed the gPC model E_N (1), we can now determine its equilibrium volume V_0 as the point for global minima energy:

$$\left. \frac{dE}{dV} \right|_{V=V_0} \approx \left. \frac{dE_N}{dV} \right|_{V=V_0} = \sum_{i=0}^N c_i \left. \frac{d\phi_i}{dV} \right|_{V=V_0} = 0, \quad (5)$$

where the derivative of the Legendre polynomials can be found as [15]:

$$\frac{d\phi_i}{dV} = \frac{iV\phi_i - i\phi_{i-1}}{V^2 - 1}. \quad (6)$$

Following its definition as the second-order derivative of energy with respect to volume, the bulk modulus is determined as:

$$B = V \left. \frac{d^2 E_N}{dV^2} \right|_{V=V_0} = V \sum_{i=0}^N c_i \left. \frac{d^2 \phi_i}{dV^2} \right|_{V=V_0}, \quad (7)$$

whereas higher-order derivatives of the Legendre polynomials are found by taking the recursive derivative of its first-order derivative (6).

2.3. Reliability analysis

We choose the gPC expansion as the energy-volume model because of its excellent accuracy. It provides the best approximation of $E(V)$ in the weighted Euclidean (L_2) norm as the *orthogonal projection* of $E(V)$ onto the Wiener-Askey polynomials space of V via the inner product $(\cdot, \cdot)_{L_2}$ [15]:

$$\lim_{N \rightarrow \infty} \|E - E_N\|_{L_2} = \inf \|E - \sum_{i=0}^N c_i \phi_i(V)\|_{L_2}, \quad c_i = \frac{1}{\|\phi_i\|_{L_2}} (E, \phi_i)_{L_2},$$

where $\{c_i\}_{i=0}^N$ are the (generalized) Fourier coefficients. The truncation error of E_N with respect to the expansion order N is:

$$\lim_{N \rightarrow \infty} \|E - E_N\|_{L_2} = \lim_{N \rightarrow \infty} \int_{\Omega} |E(V) - E_N(V)| \omega dV \leq \eta N^{-\alpha} \rightarrow 0. \quad (8)$$

Here η is a constant and $\alpha > 0$ is a measure on the actual smoothness of the energy as a function of volume. In other words, for a relatively smooth function $E(V)$, sufficient accuracy and low computational cost may be achieved using a low-degree gPC expansion $E_N(V)$.

In practice, with a given set of *ab initio* data $\{(E_k, V_k)\}_{k=1}^M$, one can write:

$$bfAc = \mathbf{d}, \quad \mathbf{A} = \begin{bmatrix} \phi_0(V_1) & \phi_1(V_1) & \cdots & \phi_N(V_1) \\ \phi_0(V_2) & \phi_1(V_2) & \cdots & \phi_N(V_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(V_M) & \phi_1(V_M) & \cdots & \phi_N(V_M) \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} E_0 \\ E_1 \\ \vdots \\ E_M \end{bmatrix}.$$

and the gPC coefficients $\mathbf{c} = \{c_i\}_{i=0}^N$ can be obtained as:

- For $M = N + 1$: $\mathbf{c} = \mathbf{A}^{-1}\mathbf{d}$.
- For $M > N + 1$: \mathbf{c} , $\min\|\mathbf{A}\mathbf{c} - \mathbf{d}\|_{L_2}$, e.g. a minimization problem.
- For $M < N + 1$: a unique and sparsest solution via compressive sampling [16–18].

A more detailed description on how the gPC coefficients are computed in these cases can be found at [19]. Unless specified otherwise, we focus on the simple case: $M \geq N + 1$, and solve the coefficients $\{c_i\}_{i=0}^N$ with least orthogonal interpolation [20,21].

If the *ab initio* data contains errors can be represented by a row vector $\epsilon_{ab} = [\epsilon_1, \dots, \epsilon_M]$, one can write the error upper bound of the gPC model in the Euclidean (L_2) norm $\|\cdot\|_{L_2}$:

$$\epsilon_E = \|E - E_N\|_{L_2} \leq \left[\sum_{i=N+1}^{\infty} \frac{(\int_{\Omega} E \phi_i \omega dV)^2}{\|\phi_i\|_{L_2}^2} \right]^{\frac{1}{2}} + \|\mathcal{F}(\mathbf{A}, \epsilon_{ab})\Phi\|_{L_2} + \|\epsilon_Q\|_{\infty}, \quad (9)$$

$$\mathcal{F}(\mathbf{A}, \epsilon_{ab}) = \begin{cases} \mathbf{A}^{-1} \epsilon_{ab}, & \text{if } M = N + 1 \\ \mathbf{c}', \frac{\min}{c} \|\mathbf{A}\mathbf{c}' - \epsilon_{ab}\|_2, & \text{if } M > N + 1 \end{cases} \quad (10)$$

where $\Phi = [\phi_0, \dots, \phi_N]^T$ is a column vector of the Legendre polynomials. In all, its first term represents the truncation error, the second is due to *ab initio* data error and the last term ($\|\epsilon_Q\|_{\infty}$) denotes aliasing error introduced by interpolation on a discrete mesh. Since Gauss points often yield an aliasing error the same order as that of the projection, the continuous and discrete expansions thus exhibit similar qualitative behavior for locally smooth interval [15]. Hence we will not distinguish between the two but group them together as the truncation error to facilitate our presentations.

Consequently, for locally smooth interval, the error upper bound of the first-order derivative of our gPC model can be obtained by taking the first-order derivative of ϵ_E with respect to volume:

$$\epsilon_{E'} = \left\| \frac{dE}{dV} - \frac{dE_N}{dV} \right\|_{L_2} \leq \left\| \sum_{i=N+1}^{\infty} c_i \frac{d\phi_i}{dV} \right\|_{L_2} + \|\mathcal{F}(\mathbf{A}, \epsilon_{ab})\|_{L_2} \frac{d\Phi}{dV} \quad (11)$$

and the error upper bound of the second-order derivative can be found as:

$$\epsilon_{E''} = \left\| \frac{d^2E}{dV^2} - \frac{d^2E_N}{dV^2} \right\|_{L_2} \leq \left\| \sum_{i=N+1}^{\infty} c_i \frac{d^2\phi_i}{dV^2} \right\|_{L_2} + \|\mathcal{F}(\mathbf{A}, \epsilon_{ab})\|_{L_2} \frac{d^2\Phi}{dV^2} \quad (12)$$

The error magnitude of our equilibrium volume from the gPC model would largely depend on its initial value from *ab initio* calculations. This is because all energy-volume approaches are based on the fundamental assumption that *ab initio* calculation provides relatively accurate estimation of the equilibrium volume at global minima energy. Although such initial guess may contain error, it will be updated as we take a number of its adjacent points: $|V_i - V_{0in}| \leq |V_{0in} P\%|$ and construct a model of $E(V)$ to compute its minima with a numerical scheme. In practice, since *ab initio* calculation and the numerical scheme to compute zeros of a function are both sufficiently accurate, one may argue the maximum error of the equilibrium volume between truth and that of the gPC model, or of any energy-volume approach, would be less than the maximum volume change $\epsilon_{V_0} \leq |V_{0in} P\%|$, as long as the adjacent points do not fall into other local energy minima.

Following the bulk modulus definition from the energy-volume relationship, we can now find its error upper bound:

$$\begin{aligned} \epsilon_B &= \left| B - (V_0 + \epsilon_{V_0}) \frac{d^2E_N}{dV^2} \right| \approx \left| V_0 \frac{d^2E_N}{dV^2} - (V_0 + \epsilon_{V_0}) \frac{d^2E_N}{dV^2} - \epsilon_{V_0} \frac{d^3E_N}{dV^3} \right| \\ &= \left| V_0 \frac{d^2E_N}{dV^2} - V_0 \frac{d^2E_N}{dV^2} - \epsilon_{V_0} \frac{d^2E_N}{dV^2} - \epsilon_{V_0} \frac{d^3E_N}{dV^3} \right| \\ &\leq V_0 \epsilon_{E''} + |V_{0in} P\%| \left| \frac{d^2E_N}{dV^2} + \frac{d^3E_N}{dV^3} \right|, \end{aligned} \quad (13)$$

where $(V_0 + \epsilon_{V_0})$ refers to the equilibrium volume computed from the gPC model.

In summary, our general framework to compute the bulk modulus is:

1. Obtain an initial guess of the equilibrium volume V_{0in} and a number of data $\{(E_k, V_k)\}_{k=1}^M$ around it using *ab initio* calculations.
2. Construct a gPC model as an approximation of the energy-volume relationship: $E_N(V) \approx E(V)$.
3. Compute and update the equilibrium volume V_0 from the gPC model $E_N(V)$ (5).
4. Determine the bulk modulus from $E_N(V)$ using the expressions (7).

We note here that there are also other definitions of the bulk modulus from the energy-volume relation. Without loss of generality, our procedure could also incorporate those expressions as they do not affect the overall workflow, albeit their error estimation would be slightly different. In addition, other material properties based on the energy-volume relationship can also be computed from the gPC model provided with their definitions. Lastly, our proposed framework can be executed in both serial or parallel for high-throughput estimations.

3. Results and discussion

In this section, we provide two benchmark cases to elucidate the general procedure of our proposed framework. To be specific, we have considered the material systems of Ti_3SiC_2 and compared our results with those from the Birch-Murnaghan EoS. A numerical analysis is also conducted to investigate the impacts of data points, e.g. the data number M , its locations and ranges, on the estimated bulk modulus. A comparison on the material systems of SB_2Te_3 with solutions from the Voigt-Reuss-Hill method is also included as further validation of our framework.

3.1. Bulk modulus for Ti_3SiC_2

Ti_3SiC_2 crystallizes in a layered hexagonal-close-packed structure with low-degree symmetry. And Ti_3SiC_2 has been characterized to have three types of chemical bonding of covalent, ionic and metallic, which hence provides an ideal benchmark case along with established experimental and computational results.

Following the proposed framework, we obtain an initial value of the equilibrium volume from the *ab initio* calculation: $V_{0in} = 14.522 \text{ nm}^3$. Here GGA approximation is employed for the exchange-correlation functional in the *ab initio* calculation. The lattice constants are $a = 0.3076 \text{ nm}$ and $c = 1.7718 \text{ nm}$, whose ratio $c/a = 5.7595$ is in good agreement with experimental data [22,23].

A total of 61 data points are generated from *ab initio* calculations for subsequent numerical analysis. Maximum volume change is set as 1% of the initial equilibrium volume V_{0in} . For superior convergence rate [24,25], we choose the zeroes of the Legendre polynomials as locations of 61 input data:

$$V_i = V_{0in} \left[1 - 1\% \cos\left(\frac{i-1}{M-1}\pi\right) \right], \quad i = 1, \dots, M, \quad (14)$$

Now we construct the gPC model $E_N(V)$ from the 61 data points using least orthogonal interpolation. Its order N can be determined from the Cauchy sequence of convergence, difference between two consecutive orders:

$$\delta E(N) = \|E_N - E_{N-1}\|_{\ell_2}, \quad N \geq 1 \quad (15)$$

where $\|\cdot\|_{\ell_2}$ refers to the ℓ_2 -norm computed by using 200 equidistance points in the domain $V_{0in}(1 \pm 1\%)$. We find in Fig. 1 that the truncation error, δE , rapidly drops as the gPC order N increases. Although it can be reduced to 10^{-6} at $N = 11$ and presumably to an even lower value at higher orders, we note the slope roughly converges at $N = 3$. Since the

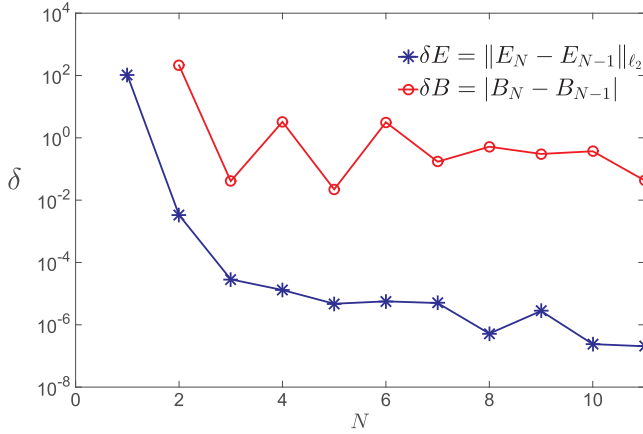


Fig. 1. Truncation errors of the gPC model between two consecutive orders N .

input data is accurate to the third decimal place, it is reasonable to choose a gPC model with a truncation error around 10^{-2} eV, which is $N = 2$ in our case. Consequently, difference of estimated bulk modulus between 2^{nd} - and 3^{rd} -order E_N is about 10^{-1} GPa as shown by δB in Fig. 1.

Using the 2^{nd} -order gPC model, we find the equilibrium volume and bulk modulus from their definitions (5) and (7) as 14.5224 nm^3 and 212.4 GPa , respectively. The former is computed numerically using Bisectional method with an error tolerance of 10^{-16} and showed a $<0.01\%$ difference than the initial guess from *ab initio* calculation. Since the *ab initio* calculation result is accurate to the third decimal place, i.e. $\epsilon_{ab} \sim 10^{-2}$ eV, the error of B due to input *ab initio* data is found less than 10^{-8} GPa (as shown in Fig. 2), which is significantly smaller than its truncation error. In Table 2, we present the estimated result of bulk modulus as $212.4 \pm 0.1 \text{ GPa}$ and find it agrees well with past experiments and theoretical works [26–28,23,29].

We now investigate the effect of input data number and location on the estimation result. This is conducted by comparing the bulk modulus from the gPC model using Legendre points, against those from the popular Birch-Murnaghan EoS using equidistance points. Fig. 3 presents B as a function of data size M using the 2^{nd} -order gPC model and Birch-Murnaghan model, respectively. It is clear that the estimated result from gPC model quickly converges around $M = 7$, whereas its counterpart from Birch-Murnaghan converges at about $M = 30$ points for an error of 0.1 GPa .

It is noted here that the Birch-Murnaghan model postulates a physical relationship between the material volume and its internal energy such that its expression incorporates various physical quantities, i.e. the equilibrium volume, equilibrium energy, bulk modulus and the Grüneisen constant, as its model parameters. In contrast, the gPC

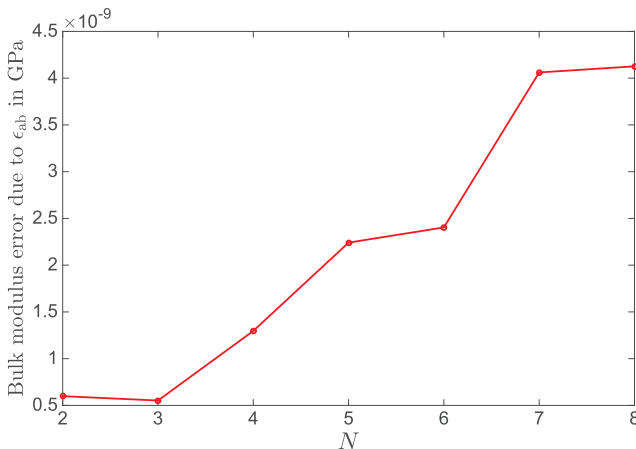


Fig. 2. The bulk modulus error due to *ab initio* data at various gPC order N .

Table 2

Estimated Ti_3SiC_2 bulk modulus from gPC model, Birch-Murnaghan EoS and past works.

Ti_3SiC_2	Bulk Modulus (GPa)
2^{nd} -order gPC model	212.4 ± 0.1
Birch-Murnaghan model	212.4
Reference [29]	204
Reference [23]	196
Reference [27]	206 ± 6
Reference [28]	225
Reference [26]	179

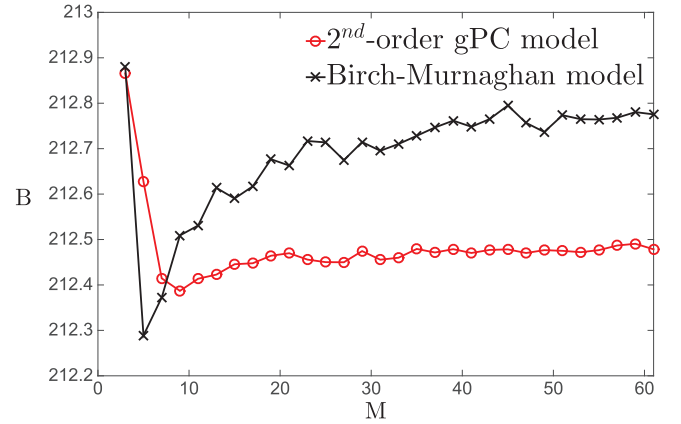


Fig. 3. Estimated bulk modulus B with respect to data number M , using a 2^{nd} -order gPC model with Legendre points and the Birch-Murnaghan model with equidistance points.

model is a numerical tool that provides the best approximation of the target function $E(V)$ in the weighted Euclidean norm. Its expansion coefficients are simply the products of numerical computation. Consequently, the gPC expansion does not incorporate any physical meaning alone in its representation of the energy-volume relationship. However, it becomes physically meaningful if one utilized it to estimate various physical quantities from $E(V)$ by following classical laws that all physical models agree upon. In such case, the gPC model is equivalent to the Birch-Murnaghan model in their pursuit of the equilibrium volume, the bulk modulus and Grüneisen constant from their physical definition derived from the energy-volume relationship.

In general, the choice of the gPC expansion order depends on two factors:

- Error tolerance for bulk modulus.

From earlier studies on Ti_3SiC_2 [27,26,28,23,29], the error tolerance for B in material science is about $<10^1 \text{ GPa}$. By plotting the Cauchy sequence of bulk modulus in Fig. 1, we found that such criteria can be satisfied from the 2^{nd} -order gPC expansion and onwards, since its difference with the 3^{rd} -order gPC is about $\sim 10^{-1} \text{ GPa}$.

- Computational cost of the input data.

Lower-order gPC expansion requires less data to construct than higher-order expansion. Such advantage would be significant when one conducts high-throughput material screening, often with thousands of tasks, but with limited computational resources. In our case, as shown in Table 3, the 2^{nd} -order gPC expansion needs 7 *ab initio* data points whilst 9 points are necessary for its 3^{rd} -order counterpart.

Although a minimum set of $M = N + 1$ data points are needed to construct a N^{th} -order gPC model using the least-square methods, there are other numerical methods available, such as compressive sensing, to estimate with even fewer data. However, they often require more

Table 3

Estimated bulk modulus (B) Ti_3SiC_2 , the number of data points for convergence (≤ 0.1 GPa) and the Grüneisen parameter at different orders of gPC model.

Ti_3SiC_2	gPC Order	Bulk Modulus (GPa)	No. of Data	Grüneisen Constant
	N = 2	212.4	7	-0.67
	N = 3	212.4	9	1.86
	N = 4	215.7	11	1.84
	N = 5	215.7	11	-0.18
	N = 6	218.9	21	-0.09
	N = 7	219.1	21	-5.95
	N = 8	219.6	21	-5.89

Table 4

Bulk modulus estimation using different range of data points, $V_{0\text{in}}(1 \pm P\%)$.

Maximum volume change	1%	10%	20%	30%
Bulk modulus (GPa)	212.44	211.22	210.59	210.3

Table 5

Estimated bulk modulus of Sb_2Te_3 using 2nd-order gPC model, recent numerical results using Voigt-Reuss-Hill approach and past experimental data.

Sb_2Te_3	Bulk Modulus (GPa)
2 nd -order gPC Model	39.9 \pm 0.1
Voigt bound [32]	39.4
Reuss bound [32]	37.3
Hill average [32]	38.4
Experiment study [33]	44.8

efforts to conduct and are beyond the scope of this study.

It is noted that the aliasing error can be amplified in its propagation to higher-order derivatives of the gPC expansion. This may prompt one to select higher-order expansions but it may also introduce overfitting. As an example, we compute the Grüneisen constant, defined as the third-order derivative of E_N [30,31]:

$$\begin{aligned}
 -\gamma &= \frac{1}{6} + \frac{V}{2B} \frac{dB}{dV} \bigg|_{V=V_0} = \frac{2}{3} + \frac{V^2}{2B} \frac{d^3 E_N}{dV^3} \bigg|_{V=V_0} \\
 &= \frac{2}{3} + \frac{V^2}{2B} \sum_{i=0}^N c_i \frac{d^3 \phi_i}{dV^3} \bigg|_{V=V_0},
 \end{aligned} \quad (16)$$

and list its values at different gPC order in Table 3. Using the two criteria above, we find the 3rd-order gPC expansion for E_N would yield a value of 1.86, which is in good agreement with past analysis [29]. But the result would deteriorate to faulty values at higher orders ($N \geq 5$).

To investigate the impact of measurement range, e.g. the maximum volume change with respect to the initial guess $V_{0\text{in}}(1 \pm P\%)$, on the estimated bulk modulus, we conduct a set of experiments with various domain size $P\%$. Table 4 listed B for $P = 1\%$, 10%, 20% and 30% using a 2nd-order gPC expansion. It is found that there is a drop on the estimated value as the range increases, albeit very small ($\leq 0.5\%$). However, such result may be confined to the global minima range and a significant volume change may land the *ab initio* data to adjacent intervals with their own local minima. We therefore conclude that the measurement range has very limited influence on the estimation of bulk modulus as long as it is sufficiently small, i.e., not incorporating other intervals with local minima.

3.2. Bulk modulus for Sb_2Te_3

To demonstrate the efficiency of our method, a comparison on the material system Sb_2Te_3 with past experimental and computation results are also included in Table 5. Again, we have employed a 2nd-order gPC

expansion as an approximation to the energy-volume relation and compute the bulk modulus accordingly. We find our result agrees well with earlier numerical works using Voigt-Reuss-Hill approach and experimental study.

4. Conclusion

In this paper, we presented a numerical framework to estimate the bulk modulus. Our approach employs *ab initio* data to construct the generalized polynomial chaos (gPC) model as an approximation of the energy-volume relationship and is mathematically rigorous. To demonstrate the accuracy and efficiency of our approach, we compute the bulk modulus of the material systems Ti_3SiC_2 and Sb_2Te_3 and compare our results with those from conventional methods. Our analysis leads to the following conclusions:

- The gPC method provides excellent accuracy in estimating the energy-volume relationship and bulk modulus along with a quantification on the upper bound of their reliability;
- In contrast to standard practices, our approach imposes no physical assumptions and is purely based on the input *ab initio* calculated data;
- By choosing zeroes of the Legendre polynomials as locations of the input data points, our approach achieves fast convergence and requires less input data (and cost) than equidistance points.
- Accuracy of the gPC model could be enhanced if the accuracy of the input *ab initio* data improves.
- The proposed method can be incorporated to compute other material properties based on the energy-volume relationship and as a high-throughput screening framework in material design and discovery.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

CRediT authorship contribution statement

Peng Wang: Investigation, Methodology, Supervision, Funding acquisition, Visualization, Writing - original draft, Writing - review & editing. **Yi Qin:** Investigation, Formal analysis, Validation. **Ming Cheng:** Investigation, Formal analysis, Validation, Writing - review & editing. **Guanjie Wang:** Data curation, Software, Resources. **Dongbin Xiu:** Conceptualization, Methodology, Writing - original draft. **Zhimei Sun:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (Grant No. 2017YFB0701700), the National Natural Science Foundation of China (Grant No. 11571028) and the Recruitment Program of Global Experts.

References

- [1] S. Pugh, XCII. Relations between the elastic moduli and the plastic properties of polycrystalline pure metals, *Philos. Mag.* 45 (367) (2009) 823–843.
- [2] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. van der Zwaag, J.J. Plata, C. Toher, S. Curtarolo, G. Ceder, K.A. Persson, M. Asta, Charting the complete elastic properties of inorganic crystalline compounds, *Sci. Data* 2 (2015) 150009.
- [3] G. Kresse, J. Hafner, Ab initio molecular dynamics for open-shell transition metals, *Phys. Rev. B* 48 (17) (1993) 13115.
- [4] J.P. Watt, L. Peselnick, Clarification of the hashin-shtrikman bounds on the effective elastic moduli of polycrystals with hexagonal, trigonal, and tetragonal symmetries, *J. Appl. Phys.* 51 (3) (1980) 1525–1531, <https://doi.org/10.1063/1.327804>.

- [5] Y. Le Page, P. Saxe, Symmetry-general least-squares extraction of elastic coefficients from ab initio total energy calculations, *Phys. Rev. B* 63 (2001) 174103, <https://doi.org/10.1103/PhysRevB.63.174103>.
- [6] Z. Sun, R. Ahuja, J.E. Lowther, Mechanical properties of vanadium carbide and a ternary vanadium tungsten carbide, *Solid State Commun.* 150 (2010) 697–700.
- [7] W. Voigt, *Lehrbuch der kristallphysik (Mit Ausschluss der Kristalloptik)*, Springer-Verlag, 1928.
- [8] A. Reuss, Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle. *ZAMM-J. Appl. Math. Mech./ Zeitschrift für Angewandte Mathematik und Mechanik* 9(1) (1929,) 49–58, <https://doi.org/10.1002/zamm.19290090104>.
- [9] R. Hill, The elastic behaviour of a crystalline aggregate, *Proc. Phys. Soc. London* 65 (5) (1952) 350.
- [10] F. Birch, Finite elastic strain of cubic crystals, *Phys. Rev.* 71 (11) (1947).
- [11] A.M. Hofmeister, Interatomic potentials calculated from equations of state: Interatomic potentials calculated from equations of state: limitation of finite strain to moderate k' , *Geophys. Res. Lett.* 20 (7) (1993) 635–638.
- [12] F.D. Stacey, Theory of thermal and elastic properties of the lower mantle and core, *Phys. Earth Planet. Inter.* 89 (3–4) (1995) 219–245.
- [13] D. Xiu, G. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.* 24 (2) (2002) 619–644.
- [14] R. Ghanem, P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, 1991.
- [15] D. Xiu, *Numerical Methods for Stochastic Computations*, Princeton University Press, Princeton, New Jersey, 2010.
- [16] A. Doostan, H. Owahdi, A non-adapted sparse approximation of PDEs with stochastic inputs, *J. Comput. Phys.* 230 (8) (2011) 3015–3034, <https://doi.org/10.1016/j.jcp.2011.01.002> URL: <http://www.sciencedirect.com/science/article/pii/S0021999111000106>.
- [17] H. Rauhut, R. Ward, Sparse Legendre expansions via ℓ_1 -minimization, *J. Approximation Theory* 164 (5) (2012) 517–533, <https://doi.org/10.1016/j.jat.2012.01.008>.
- [18] L. Yan, L. Guo, D. Xiu, Stochastic collocation algorithms using ℓ_1 -minimization, *Int. J. Uncertainty Quantif.* 2 (3) (2012) 279–293, <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2012003925> URL: <http://www.dl.begellhouse.com/journals/52034eb04b657aea,3495dd1f6253a653,1447d224370655d6.html>.
- [19] A. Narayan, T. Zhou, Stochastic collocation on unstructured multivariate meshes, *Commun. Comput. Phys.* 18 (01) (2015) 1–36, <https://doi.org/10.4208/cicp.020215.070515a> arXiv:1501.05891 [math.NA]. URL: <http://journals.cambridge.org/article/S181524061500047X>.
- [20] A. Narayan, D. Xiu, Stochastic collocation methods on unstructured grids in high dimensions via interpolation, *SIAM J. Sci. Comput.* 34 (3) (2012) A1729–A1752, <https://doi.org/10.1137/110854059> URL: <http://epubs.siam.org/doi/abs/10.1137/110854059>.
- [21] A. Cohen, M.A. Davenport, D. Leviatan, On the stability and accuracy of least squares approximations, *Found. Comput. Math.* 13 (5) (2013) 819–834, <https://doi.org/10.1007/s10208-013-9142-3> URL: <http://link.springer.com/article/10.1007/s10208-013-9142-3>.
- [22] W.B. Pearson, *The Crystal Chemistry and Physics of Metals and Alloys*, Wiley-Interscience, New York, 1972.
- [23] Z. Sun, J. Zhou, D. Music, R. Ahuja, J. Schneider, Phase stability of Ti_3SiC_2 at elevated temperatures, *Scripta Mater.* 54 (2006) 105–107.
- [24] B. Alpert, V. Rokhlin, A fast algorithm for the evaluation of legendre expansions, *SIAM J. Sci. Stat. Comput.* 12 (1) (1991) 158–179, <https://doi.org/10.1137/0912009>.
- [25] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grid, *Adv. Comput. Math.* 12 (1999) 273–288.
- [26] P. Finkel, M.W. Barsoum, T. El-Raghy, Low temperature dependence of the elastic properties of Ti_3SiC_2 , *J. Appl. Phys.* 85 (7123) (1999).
- [27] A. Onodera, H. Hirano, T. Yuasa, N.F. Gao, Y. Miyamoto, Static compression of Ti_3SiC_2 to 61 gpa, *Appl. Phys. Lett.* 74 (3782) (1999).
- [28] R. Ahuja, O. Eriksson, J.M. Wills, B. Johansson, Electronic structure of Ti_3SiC_2 , *Appl. Phys. Lett.* (2000).
- [29] T.H. Scabarozzi, S. Amini, O. Leaffer, A. Ganguly, S. Gupta, W. Tambussi, S. Clipper, J.E. Spainer, M.W. Barsoum, J.D. Hettinger, S.E. Lofland, Thermal expansion of select $m_{n+1}ax_n$ (m = early transition metal, a = a group element, $x = c$ or n) phases measured by high temperature x-ray diffraction and dilatometry, *J. Appl. Phys.* 105 (2009) 013543 .
- [30] B.A. Weinstein, R. Zallen, *Light Scattering in Solids IV*, Springer, New York, 1982 Ch. Pressure Raman effects in covalent and molecular solids, p. 463.
- [31] W.A. Harrison, *Elementary Electronic Structure*, revised Edition, World Scientific, 2004, <https://doi.org/10.1142/5432>.
- [32] Z. Li, N. Miao, J. Zhou, H. Xu, Z. Sun, Reduction of thermal conductivity in $\text{Y}_2\text{Sb}_{2-x}\text{Te}_3$ for phase change memory, *J. Appl. Phys.* 122 (19) (2017).
- [33] X. Chen, H.D. Zhou, A. Kiswandhi, I. Miotkowski, Y.P. Chen, P.A. Sharma, A.L. Lima Sharma, M.A. Hekmaty, D. Smirnov, Z. Jiang, Thermal expansion coefficients of Bi_2Se_3 and Sb_2Te_3 crystals from 10 k to 270 k, *Appl. Phys. Lett.* 99 (26) (2011) 261912, <https://doi.org/10.1063/1.3672198>.