

# Pretraživanje svojstava molekularnih kristala pomoću strojno naučenih potencijala

---

**Ruža, Marko**

**Master's thesis / Diplomski rad**

**2020**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:189726>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-10-14**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

Marko Ruža

Pretraživanje svojstava molekularnih kristala  
pomoću strojno naučenih potencijala

Diplomski rad

Zagreb, 2020.

SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI STUDIJ  
FIZIKA; SMJER ISTRAŽIVAČKI

**Marko Ruža**

Diplomski rad

**Pretraživanje svojstava molekularnih  
kristala pomoću strojno naučenih  
potencijala**

Voditelj diplomskog rada: dr. sc. Ivor Lončarić

Ocjena diplomskog rada: \_\_\_\_\_

Povjerenstvo: 1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_

Datum polaganja: \_\_\_\_\_

Zagreb, 2020.

Zahvale mentoru dr. sc. Ivoru Lončariću za pomoć i savjete pri-  
likom izrade diplomskog rada te obitelji i prijateljima za dugo-  
godišnju podršku.

## Sažetak

Teorija funkcionala gustoće (DFT) najraširenija je metoda za računanje svojstava materijala, no za sustave s velikim brojem atoma ili molekula izračuni postaju računalno zahtjevni. Primjer takvog sustava predstavljaju molekularni kristali čije se ćelije često sastoje od desetaka ili stotina atoma. Uzmimo u obzir situaciju u kojoj je na raspolaganju velik broj poznatih materijala. Tražimo li jedno specifično svojstvo koje bi samo dio tih materijala mogao imati, provedba eksperimenata za sve materijale bila bi dugotrajna i skupa, te na kraju i neisplativa. Praktično rješenje u takvim slučajevima su računalne simulacije, no već spomenuti DFT izračuni također su neisplativi za kompleksnije strukture. Stoga se za simulacije traže alternativne metode. U zadnje vrijeme strojno naučeni potencijali se pokazuju kao dobra alternativna metoda. Takvi računi neusporedivo su brži u odnosu na DFT, a mogu pokazivati sličnu točnost. S obzirom da kompleksnost tih potencijala jako brzo rase s brojem različitih elemenata, oni su primijenjivi na klase materijala sastavljenih od nekoliko istih elemenata. Konkretno u ovom diplomskom radu proučavani su organski kristali sastavljeni od atoma C, H, N i O. Time je omogućena obrada velikog broja organskih molekularnih kristala sastavljenih od spomenutih atoma u puno kraćem vremenu. Za te kristale napravljeni su računi elastičnih tenzora kako bi se identificirali oni kristali koji pokazuju neuobičajeno svojstvo negativne linearne kompresibilnosti (NLC). Za izabrane materijale predviđanja su provjerena DFT računom.

Ključne riječi: molekularni kristali, neuronske mreže, tenzor elastičnosti, negativna linearna kompresibilnost (NLC), teorija funkcionala gustoće (DFT).

# Screening molecular crystals with machine learning potentials

## Abstract

Density functional theory (DFT) is the most widely used accurate method for predicting material properties, but its practicality is limited when used on systems with a large number of atoms or molecules. An example of such systems are molecular crystals as their unit cells usually have tens or hundreds of individual atoms. Let us consider a situation in which we have many known materials. If we want to identify certain materials which exhibit only one specific property, experimental research with such a goal often represents a technical problem. A large amount of experiments required to identify such materials can be time consuming and costly, in the end even unprofitable. Practical solution for such a problem are computer simulations, but DFT calculations of complex structures could also turn out to be impractical which is the main reason for developing alternative methods, like machine learning potentials. Lately, they have proven to be a good substitute. Calculations with machine learning potentials are much faster than DFT and their accuracy can be comparable. Certain potentials are constructed with a focus on a certain classes of materials, because complexity of such potentials grows with the number of different elements in a system. In this thesis, our focus is centered on organic crystals consisting only of C, H, N and O atoms. Such constraints make material property screening much faster and cost effective. Specifically, the main goal is to calculate elastic tensors for such crystals and identify those which could exhibit the property of negative linear compressibility (NLC). At the very end, several such results are verified by a DFT calculation.

Keywords: molecular crystals, neural networks, elastic tensor, negative linear compressibility (NLC), density functional theory (DFT).

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Strojno naučeni potencijali</b>	<b>3</b>
2.1	Neuronske mreže i simulacije materijala . . . . .	3
2.2	ANI1 strojno naučeni potencijali . . . . .	4
<b>3</b>	<b>Mehanička svojstva materijala</b>	<b>10</b>
3.1	Tenzor elastičnosti . . . . .	10
3.2	Linearna kompresibilnost . . . . .	12
<b>4</b>	<b>Teorija funkcionala gustoće (DFT)</b>	<b>16</b>
4.1	Uvod . . . . .	16
4.2	Hartreejeva i Hartree-Fockova aproksimacija . . . . .	17
4.3	Thomas-Fermi-Dirac aproksimacija . . . . .	19
4.4	Hohenberg-Kohn teoremi . . . . .	20
4.5	Kohn-Sham jednačbe . . . . .	21
4.6	LDA ( <i>Local density approximation</i> ) . . . . .	23
4.7	GGA ( <i>Generalized-gradient approximation</i> ) . . . . .	24
<b>5</b>	<b>Konstrukcija baza podataka za molekularne kristale i simulacija elastičnih svojstava</b>	<b>25</b>
5.1	ASE ( <i>Atomic Simulation Environment</i> ) i rad s datotekama . . . . .	25
5.2	Čitanje .cif datoteka . . . . .	25
5.3	Relaksacija strukture . . . . .	26
5.4	Račun elastičnog tenzora i kompresibilnosti . . . . .	27
5.5	Postavke DFT računa . . . . .	28
<b>6</b>	<b>Rezultati i diskusija</b>	<b>30</b>
6.1	Test ANI1 potencijala . . . . .	30
6.2	Relaksacije uz ograničenje na očuvanje simetrije . . . . .	33
6.3	Elastična svojstva ortorompskih kristala . . . . .	34
6.4	DFT račun . . . . .	40
<b>7</b>	<b>Zaključak</b>	<b>41</b>



# 1 Uvod

Svojestvo negativne linearne kompresibilnosti (NLC - *negative linear compressibility*) [1] zanimljivo je, ali rijetko svojstvo koje se pojavljuje samo kod specifičnih kristalnih struktura [2]. Zamislamo situaciju u kojoj se raspolaže s velikim brojem materijala, no potrebno je identificirati one kod kojih bi moglo biti prisutno neko rijetko ili anomalno svojstvo, poput NLC-a. Svojestva materijala mogu se eksperimentalno utvrditi, no sama provedba eksperimenata ima svoje tehnološke zahtjeve i troškove. Ako je unaprijed poznato da se želi identificirati svojstvo koje se relativno rijetko pojavljuje, provedba eksperimenata za spomenuti veliki broj materijala mogla bi biti neisplativa. U tom slučaju posebno su korisne računalne simulacije kojima bi se za određene materijale moglo predvidjeti traženo svojstvo s određenom točnošću samo korištenjem računala.

Na internetu se nalaze razne baze eksperimentalnih podataka za materijale. Jedna od takvih baza je *Crystallography Open Database* (COD) [3] na kojoj se nalaze difrakcijom dobiveni eksperimentalni podaci u .cif formatu koji se lagano može učitati u razne programe za potrebe računalnih simulacija. Time je moguće napraviti proizvoljan skup materijala u svrhu provedbe simulacija u potrazi za NLC svojstvom. Materijali koji će biti predmet razmatranja ovog diplomskog rada su molekularni kristali [4]. Njihove strukture mogu biti vrlo kompleksne, no samim time moguća su raznolika svojstva sa širokom primjenom u industriji, farmaciji, elektronici i medicini.

Precizna i najšire korištena metoda za simulacije materijala je teorija funkcionala gustoće (DFT - *density functional theory*), no u slučaju velikog broja materijala, velika je vjerojatnost da će određene strukture biti prekompleksne za DFT račun u razumnom vremenu. Zato su osmišljene alternativne metode, korištenjem strojno naučenih potencijala koji rade na principu neuronskih mreža. Kompleksnost tih metoda također raste s kompleksnošću strukture, prvenstveno s brojem različitih atoma u promatranom sustavu, zbog čega ćemo se ograničiti samo na molekularne kristale koji se sastoje od elemenata C, H, N i O.

Strojno naučeni potencijali pogodni za primjenu na definiran skup materijala su ANI1 [5] potencijali iz programskog paketa torchani, konstruirani isključivo za sustave s C, H, N i O atomima. Za što bolje konačne rezultate potrebno je prvo testirati točnost potencijala na eksperimentalnom skupu podataka. Nakon toga pokušat će se

pronaći materijali za koje ANI1 potencijali predviđaju NLC svojstvo.

## 2 Strojno naučeni potencijali

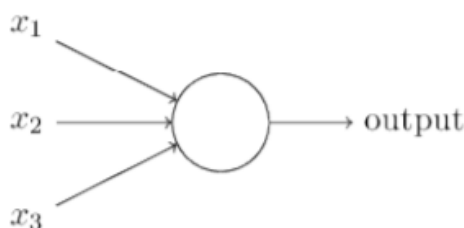
### 2.1 Neuronske mreže i simulacije materijala

Princip rada korištenog strojno naučenog potencijala ANAKIN-ME (*Accurate Neural network engINe for Molecular Energies*, kraće ANI) temelji se na neuronskim mrežama te služi za izračun energije molekula ili molekularnih kristala koji se sastoje od elemenata C, H, N i O. Općenito, umjetna neuronska mreža je model obrade podataka inspiriran ljudskim mozgom. Svaki neuron zasebno obrađuje ulazne podatke i vraća neku vrijednost. Takvi neuroni su, konkretno u *feed-forward* neuronskim mrežama, organizirani u više "skrivenih slojeva" i međusobno povezani "težinama" (*weights*, koje predstavljaju sinapse). U jednom sloju neuroni paralelno računaju na podacima iz prethodnog sloja te se prosljeđuju preko neke aktivacijske funkcije. Neuroni mogu imati definiranu tzv. pristranost ako obrađuju parametre od posebne važnosti.

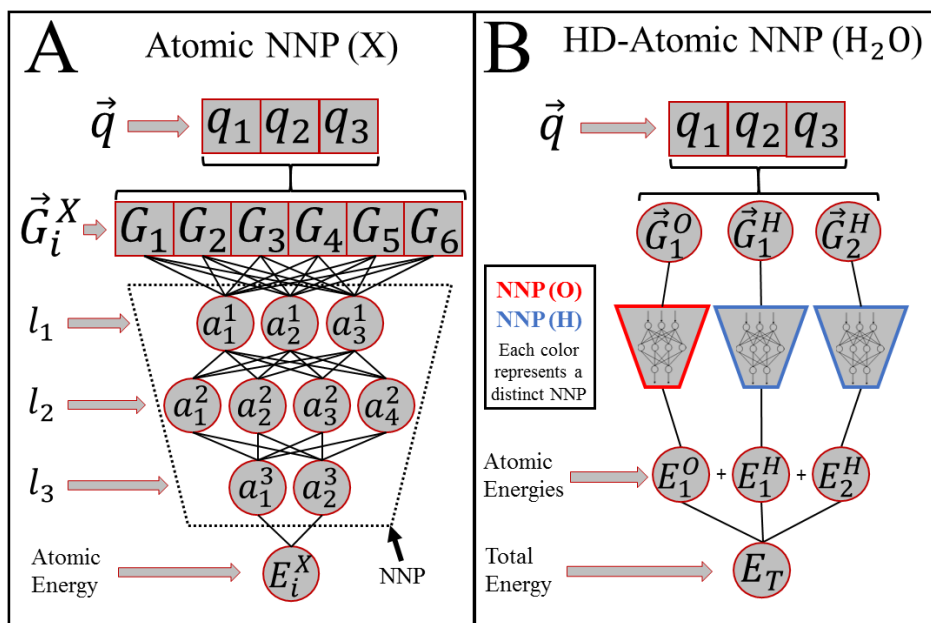
Jedan od jednostavnijih tipova neurona je tzv. perceptron, koji je 50-ih godina definirao Frank Rosenblatt [6]. Shema na slici 2.1 prikazuje perceptron s tri diskretna ulazna parametra  $x_i$  od kojih svaki ima pridruženu težinu (*weight*)  $w_i$ , realan broj koji određuje važnost svakog ulaznog parametra za izračun izlaza koji može poprimiti vrijednost 0 ili 1, ovisno o postavljenom pragu:

$$\text{izlaz} = \begin{cases} 0 & \text{if } \sum_j w_j x_j \leq \text{prag}, \\ 1 & \text{if } \sum_j w_j x_j > \text{prag}. \end{cases} \quad (2.1)$$

Tako definiran izlaz referira se na Heavisideovu funkciju koja se može koristiti kao aktivacijska funkcija za ovaj diskretni primjer. Za kontinuirane slučajeve mogu se koristiti gausijan ili sigmoidalna funkcija.



Slika 2.1: Shema perceptrona (preuzeto iz [6]).



Slika 2.2: Shematski prikaz rada neuronskih mreža kod konstrukcije potencijala (preuzeto iz [7]). Vektor atomskog okruženja  $\vec{q}$  s koordinatama  $q_i$  predstavlja ulazni parametar mreže koji se u procesu učenja provlači kroz "skriveno slojevo" mreže te kao izlazni parametar daje energiju. Na slici A prikazan je opći primjer ulaza i izlaza mreže, a na slici B dan je konkretni primjer za tri ulazna parametra kod molekule vode (jedan za svaki atom) koji se provlače kroz mrežu i kao izlazni parametar daje njihov zbroj energija.

## 2.2 ANI1 strojno naučeni potencijali

Referentni skup podataka koji ANI1 potencijal koristi izveden je iz GDB-11 [8] [9] baze koja sadrži DFT-om izračunate podatke o molekulama do maksimalno 11 atoma sastavljenih od elemenata C, N, O i F (sve njihove moguće kombinacije). Za potrebe ANI-ja eliminiraju se molekule koje sadrže fluor. Zbog velikog broja molekula ( $\approx 40$  milijuna), odabrana je 57,951 molekula s najviše 8 atoma. Podaci iz GDB-11 baze spremljeni su u SMILES [10] formatu koji se za potrebe ANI1 potencijala pretvaraju u 3D uz pomoć RDKit softvera te se popunjavanjem valentnih orbitala molekulama dodaju vodici. Nakon toga se optimizira struktura te generiraju normalni modovi i koordinate za provedbu *Normal Mode Sampling* (NMS) metode za generiranje različitih struktura i računanje pripadnih energija.

ANI1 podaci za različite konformacije molekula dobiveni su NMS-om. ANI1 koristi podatke u obliku para  $(\vec{q}, E_T)$ , odnosno neke prostorne raspodjele atoma u molekuli i energije za tu određenu konformaciju. Gledamo vibracije molekule optimizirane strukture preko normalnih koordinata pojedinih atoma u potencijalu harmonijskog oscilatora. Označimo ih s  $Q = \{q_1, q_2, \dots, q_{N_f}\}$ . Za linearne molekule s  $N_a$  atoma

postoji  $N_f = 3N_a - 5$  normalnih koordinata, a  $N_f = 3N_a - 6$  za sve ostale. Svaki  $i$ -ti atom koji je pomaknut od ravnotežnog položaja za neki  $R_i$  osjeća povratnu silu koja je određena pripadnom konstantom iz skupa  $K = \{K_1, K_2, \dots, K_{N_f}\}$ . Generiranjem slučajnih brojeva  $c_i$  uz uvjet da je vrijednost sume  $\sum_i^{N_f} c_i$  u intervalu  $[0, 1]$  definira se pomak  $i$ -tog atoma od ravnotežnog položaja:

$$R_i = \pm \sqrt{\frac{3N_a c_i kT}{K_i}} \quad (2.2)$$

za vrijednost ukupne prosječne energije molekule na određenoj temperaturi  $T$  skaliranu parametrom  $c_i$ . Odabir predznaka definiran je Bernoullijevom raspodjelom uz  $p = 0.5$  s namjerom da se vrijednostima popune i pozitivna i negativna strana potencijala harmonijskog oscilatora. Konačno, nove neravnotežne konformacije molekule dobivaju se zbrojem normalnih koordinata  $Q$  i pomaka  $Q^R$  čije su komponente skalirane pomakom na način  $q_i^R = q_i R_i$ . Obzirom da NMS služi za generiranje neravnotežnih struktura u kojima se mogu naći i strukture prevelike energije za upotrebu, postavljen je prag od 275 kcal/mol. Dodatni detalji opisani su na slici 2.3.

ANI1 potencijal kao ulazni parametar uzima vektor atomskog okruženja ("atomic environment vector")  $\vec{G}_i^X = \{G_1, G_2, G_3, \dots, G_M\}$  čije su komponente produkti tzv. *cutoff* i simetrijskih funkcija [11] koje sadrže prostorne informacije o okolini nekog atoma rednog broja  $X$  u odnosu na druge u molekuli. Definirajmo komponente vektora atomskog okruženja. *Cutoff* funkcije definirane su na sljedeći način:

$$f_C(R_{ij}) = \begin{cases} 0.5 \cos\left(\pi \frac{R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C, \\ 0 & \text{for } R_{ij} > R_C. \end{cases} \quad (2.3)$$

Parametrom  $R_S$  određena je radijalna udaljenost do koje se okolina oko atoma uzima u obzir, a  $R_{ij}$  predstavlja relativnu udaljenost atoma.

Radijalna simetrijska funkcija je produkt Gaussijana i *cutoff* funkcije oblika:

$$G_m^r = \sum_{\substack{\text{svi atomi} \\ j \neq i}} e^{-\eta(R_{ij}-R_S)^2} \cdot f_C(R_{ij}) \quad (2.4)$$

Parametar  $\eta$  služi za određivanje širine, a  $R_S$  za pomicanje centra Gaussijana. Model ANI1 je konstruiran tako da kombinira jedan parametar  $\eta$  s više različitih  $R_S$ . Time se izbjegavaju manje vrijednosti  $\eta$  koje bi dovele do nepotrebno velikih vrijednosti

Number of heavy atoms	Total Molecules	Max Temperature	S value	Energies < 275 kcal × mol <sup>-1</sup>	Energies > 275 kcal × mol <sup>-1</sup>	Total data
1	3	2,000.00	500	10,800	0	10,800
2	13	1,500.00	450	50,962	398	51,360
3	20	1,000.00	425	151,200	0	151,200
4	61	600	400	651,936	6,144	658,080
5	267	600	200	1,813,151	9,889	1,823,040
6	1,406	600	30	1,682,245	29,963	1,712,208
7	7,760	600	20	6,460,162	869,222	7,329,384
8	47,932	450	5	11,236,918	1,714,819	12,951,737
Total	57,462	—	—	22,057,374	2,630,435	24,687,809

Slika 2.3: U tablici su prikazani parametri korišteni za generiranje struktura za ANI1 potencijal korištenjem GDB-11 baze. U prvom stupcu nalaze se brojevi atoma za pojedinu molekulu, u drugom njihov broj mogućih kombinacija, a u trećem maksimalne korištene temperature za perturbaciju molekule.  $S$  vrijednost je empirijski parametar koji određuje broj potrebnih struktura za molekulu po stupnju slobode prema jednadžbi  $N_{struktura} = S \times N_{dof}$ , gdje  $N_{dof}$  označava broj stupnjeva slobode molekule [5]. Ukupan broj na taj način generiranih podataka dan je u donjoj lijevoj ćeliji tablice od kojih su prema dogovoru upotrebljive samo one strukture čija je ukupna energija manja od 275 kcal/mol, što znači da je od 24,687,809 struktura upotrebljivo 22,057,374. (Preuzeto iz [5])

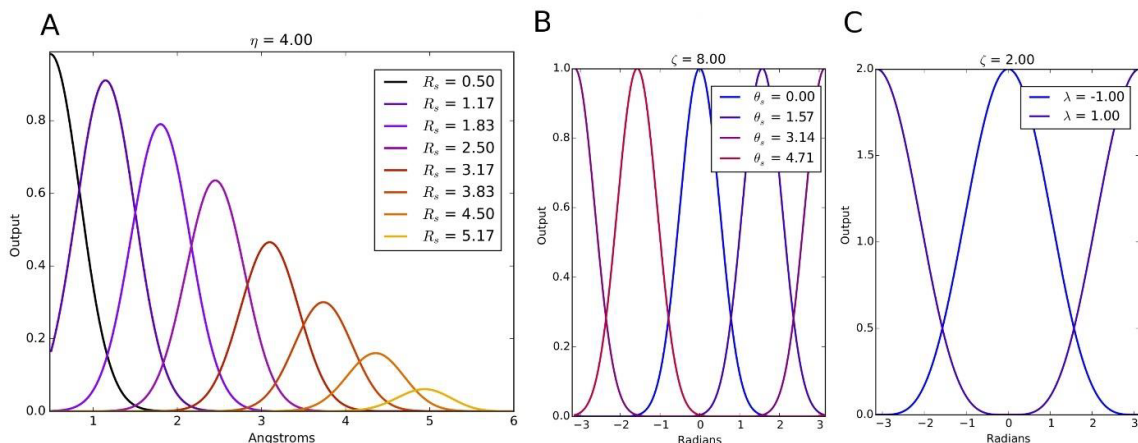
komponentata vektora  $\vec{G}$ , a različiti  $R_S$  omogućavaju ispitivanje cijelog prostora oko atoma.

Ukupna funkcija, ujedno i komponenta vektora  $\vec{G}$  za atom rednog broja  $X$  je oblika:

$$G_m^X = 2^{1-\zeta} \sum_{\substack{\text{svi atomi} \\ j,k \neq i}} [1 + \lambda \cos(\theta_{ijk} - \theta_s)]^\zeta \times \exp \left[ -\eta \left( \frac{R_{ij} + R_{ik}}{2} - R_S \right)^2 \right] f_C(R_{ij}) f_C(R_{ik}). \quad (2.5)$$

Gledaju se interakcije i-tog atoma s j-tim i k-tim kao problem interakcije tri tijela. Kutni parametar  $\theta_S$  analogan je radijalnom  $R_S$ , a  $\zeta$  određuje širinu vrhova u kutnom dijelu. Ponašanje funkcije za različite vrijednosti  $\zeta$ ,  $\theta_S$  i  $R_S$  prikazano je na slici 2.4.

Pokazalo se da je za ANI1 neuronsku mrežu najbolja piramidalna struktura oblika 768:128:128:64:1 (768 ulaza i 1 izlaz, a u sredini 3 skrivena sloja od 128:128:64 neurona). Skriveni slojevi koriste gausijan kao aktivacijsku funkciju, a izlaz linearnu funkciju. Težine su nasumično određene normalnom distribucijom u intervalu  $(-1/\sqrt{d}, 1/\sqrt{d})$ , gdje je  $d$  broj ulaza u neuron, a pristranosti su inicijalno jednake



Slika 2.4: Grafovi ovisnosti simetrijskih funkcija o parametrima  $R_S$ ,  $\lambda$  i  $\zeta$  (preuzeto iz [7]).

nuli. Težine se optimiziraju pomoću funkcije gubitka:

$$C(\vec{E}^{ANI}) = \tau \exp \left[ \frac{1}{\tau} \sum_j (E_j^{ANI} - E_j^{DFT})^2 \right]. \quad (2.6)$$

U programu se koristi poseban testni skup od 1024 molekula čije su energije izračunate DFT-om. Veličina  $\tau = 0.5$  empirijski je određena.  $\vec{E}^{ANI}$  je vektor energija izračunatih ANI-jem čije se vrijednosti uspoređuju s testnima uz podešavanje težina i pristranosti sve dok se vrijednost funkcije gubitka ne minimizira.

Konačno, strojno naučeni potencijal kao izlaz vraća energiju. Ukupna energija određena je kao zbroj izlaza iz neuronske mreže za svaki obrađeni vektor, odnosno atom.

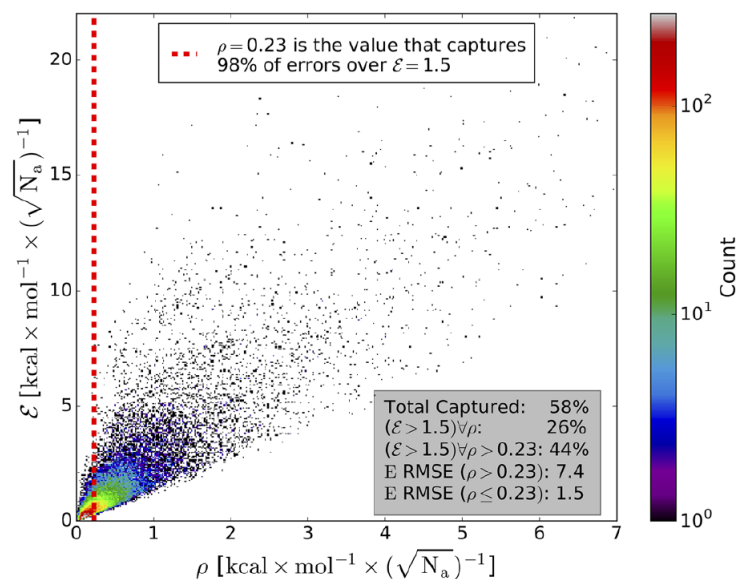
$$E_{tot} = \sum_i^{\text{svi atomi}} E_i \quad (2.7)$$

U okviru ovog rada korišteni su potencijali ANI1-x i ANI1-ccx koji su izvedeni iz skupa podataka za ANI1 potencijal.

### ***ANI1-x potencijal***

ANI1-x potencijal izveden je iz podataka korištenih u ANI1 potencijalu, no njegov je opseg podataka otprilike četiri puta manji i istovremeno daje bolje rezultate što je postignuto aktivnim učenjem. [12]

Veliki opseg podataka koji se koristi za ANI1 mora se optimizirati uklanjanjem skupova podataka koji uzrokuju velike greške. To se postiže metodom *Query By Com-*



Slika 2.5: Empirijsko određivanje  $\varepsilon_i$  i  $\hat{\rho}$ . Odabirom  $\varepsilon_i < 1.5$  kcal/mol i  $\hat{\rho} = 0.23$  kcal/mol eliminira se 98% grešaka iznad zadane vrijednosti  $\varepsilon$ . (Preuzeto iz [12].)

*mittee* (QBC). Provjeravaju se podaci u bazi koji imaju najveću standardnu devijaciju u ANI1 bazi. Kriterij po kojima je neki podatak prihvatljiv određen je kao:

$$\hat{\rho} = \frac{\sigma_i}{\sqrt{N_i}}, \quad (2.8)$$

gdje je  $N_i$  broj atoma u određenoj molekuli u bazi podataka. Definirajmo još jednu veličinu, pogrešku po atomu za  $i$ -tu molekulu na sljedeći način:

$$\varepsilon_i = |MAX(\{E_T^{ANI}\}_i^{ens} - E_{T,i}^{REF})|/\sqrt{N_i}. \quad (2.9)$$

Referentna energija uzima se za  $i$ -tu molekulu iz GDB07to09 *benchmarka* koji sadrži bazu sličnih molekula. Pokazalo se da ako se odredi  $\varepsilon_i < 1.5$  kcal/mol, za vrijednost  $\hat{\rho} = 0.23$  kcal/mol eliminira se 98% grešaka iznad dozvoljene vrijednosti što je vidljivo na slici 2.5.

Aktivno učenje počinje s filtriranom bazom, odnosno 2% ukupne početne ANI1 baze koja se dalje uspoređuje s DFT rezultatima te je ovaj put kriterij za eliminaciju molekule  $|E_{ANI} - E_{DFT}|/\sqrt{N} > 0.04$  kcal/mol. Program se prekida kada 5% ukupnih podataka ne zadovoljava kriterij te se te molekule svejedno dodaju u konačni skup podataka.

U sljedećoj fazi određuje se prostorna konfiguracija molekula. Uzimaju se podaci iz neke vanjske baze, primjerice GDB-11, iz koje se nasumične molekule prevode



u 3D prostor i optimiziraju uz pomoć UFF-a (*Universal Force Field*) [13]. Zatim se uzima ansambl od pet različitih ANI potencijala koji koriste skup podataka definiran u prošlom odlomku i testiraju se na nasumičnom skupu molekula. Kriterij daljnjeg filtriranja skupa podataka je prema već prije definiranoj veličini, odnosno ako vrijedi  $\hat{\rho} > 0.23$  za danu molekulu. Pri tome se srednja vrijednost energije koja ulazi u standardnu devijaciju jednostavno računa preko

$$E = \frac{1}{5} \sum_{i=1}^5 E_i, \quad (2.10)$$

gdje je  $E_i$  energija molekule izračunata i-tim ANI modelom (korišteno ih je pet). Sve prošlim kriterijem izbačene molekule testiraju se još i na torzijske konformacije čime se dobiva konačan skup podataka za ANI1-x potencijal.

Apsolutna greška (*Mean Absolute Error*) ovakvog modela u odnosu na DFT podatke na kojem je potencijal treniran za tzv. GDB-10to13 bazu molekula iznosi MAE=1.98 kcal/mol, a srednje kvadratno odstupanje (*Root Mean Squared Error*) RMSE=2.80 kcal/mol.

### ***ANI1-ccx potencijal***

ANI1-ccx potencijal dobiven je usporedbom skupa podataka ANI1-x s podacima dobivenima preciznim CCSD(T) (*Coupled Cluster*) i CBS (*Complete Basis Set*) metodom. [14] Glavni ciljevi su sažeti ANI1-x bazu i poboljšati prepoznavanje torzijskih konformacija molekula.

Počinja se s manjim uzorkom molekula iz ANI1-x baze te se QBC metodom određuju molekule s najvećom greškom koje se računaju CCSD(T)/CBS metodom i spremaju u bazu (konačno njih oko 480 tisuća). Posebno se provodi aktivno učenje nasumičnim uzorkovanjem torzijskih konformacija (konačno oko 200 tisuća). Transferno učenje postignuto je implementiranjem objašnjenog postupka unutar skrivenih slojeva ANI1-x neuronske mreže. Navedenim postupkom modificirana su dva skrivena sloja.

Apsolutna greška ovakvog modela u odnosu na CCSD(T)/CBS iznosi MAE=1.46 kcal/mol, a srednje kvadratno odstupanje RMSE=2.07 kcal/mol.

## 3 Mehanička svojstva materijala

### 3.1 Tenzor elastičnosti

Mehanikom čvrstih tijela opisuje se odnos između naprezanja i deformacija tijela pod utjecajem vanjskih sila koje na njih djeluju. Ako u jednoj dimenziji gledamo silu malog iznosa koja djeluje na komad izotropnog materijala duž neke osi, odnos naprezanja  $\sigma$  i male deformacije  $\epsilon = \Delta l/l$  duž te osi dan je jednostavnim linearnim izrazom:

$$\sigma = c\epsilon, \quad (3.1)$$

gdje  $c$  predstavlja faktor proporcionalnosti, odnosno Youngov modul elastičnosti koji određuje krutost tijela (engl. *stiffness*). Njemu recipročan izraz  $s = 1/c$  naziva se podatljivost (engl. *compliance*).

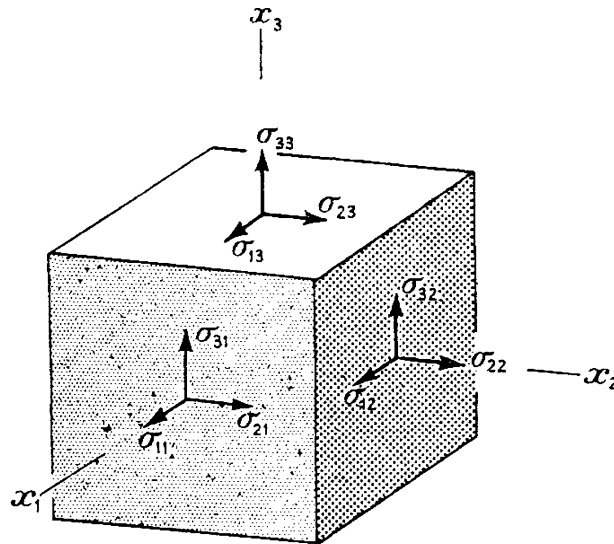
U trodimenzionalnom slučaju potrebno je definirati tenzor naprezanja. Sile koje djeluju na materijal mogu uzrokovati dvije vrste naprezanja: normalno, koje uzrokuje duljinske deformacije (i promjenu volumena) i posmično, koje uzrokuje kutne deformacije. Ako u prostoru gledamo kocku materijala infinitezimalnih dimenzija kao na slici 3.1, možemo definirati 3 različita naprezanja u smjeru normala na svaku površinu duž svake osi koordinatnog sustava. Pri tome uz normale na svaku površinu možemo definirati dva različita posmična naprezanja paralelna s tom površinom. Uvedemo li notaciju

$$(x, y, z) = (1, 2, 3), \quad (3.2)$$

postoje 3 različita normalna naprezanja koje označavamo kao  $\sigma_{ii}$  i 3 različita posmična naprezanja  $\sigma_{ij}$  (vrijedi  $i \neq j$ ). Ukupno postoji 6 nezavisnih komponenata od kojih se može sastaviti  $3 \times 3$  simetričan tenzor naprezanja oblika:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}. \quad (3.3)$$

Primjerice, tenzor naprezanja kojim bi se mogao opisati hidrostatski tlak ima samo dijagonalne komponente različite od nule:  $\sigma_{11} = \sigma_{22} = \sigma_{33} = -p$ . Analogno se može definirati tenzor deformacije  $\epsilon_{ij}$ .



Slika 3.1: 3D naprezanja (preuzeto iz [15]).

Tablica 3.1: Voigtova notacija.

Matrično	11	22	33	23, 32	13, 31	12, 21
Voigt	1	2	3	4	5	6

Velik broj kristala pokazuje anizotropna svojstva. U tom slučaju, za opis odnosa naprezanja i deformacija potrebno je poopćiti izraz (3.1), na način da se pretpostavi da svaka od 6 različitih mogućih komponenta tenzora naprezanja može uzrokovati sve moguće deformacije. Uz tu pretpostavku i korištenjem Einsteinove konvencije o sumaciji, opći izraz Hookeovog zakona je oblika:

$$\varepsilon_{ij} = s_{ijkl}\sigma_{kl}. \quad (3.4)$$

Time je definiran 81 koeficijent podatljivosti  $s_{ijkl}$ , no vrijedi simetrija koeficijenata

$$s_{ijkl} = s_{ijlk} = s_{jikl} \quad (3.5)$$

te se broj koeficijenata smanjuje na 36. Obzirom da konačan tenzor mora biti simetričan, maksimalan broj nezavisnih koeficijenata je 21.

Koeficijenti  $s_{ijkl}$ , a time i  $c_{ijkl}$  čine tenzore četvrtog reda. U izrazu (3.4) pokazana je ovisnost između dva različita tenzora drugog reda preko tenzora četvrtog reda (detaljnije u [15]). Radi praktičnosti, često se koristi Voigtov način indeksiranja konstanata. Uz (3.2), notacija je prikazana u tablici 3.1. Relacija koja opisuje odnos

deformacija i naprezanja prema Voigtovoj konvenciji sljedećeg je oblika:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix}. \quad (3.6)$$

Materijali kristalne strukture mogu kristalizirati u 7 različitih kristalnih sustava: kubični, ortorompski, trigonski, tetragonski, heksagonski, monoklinski i triklinski. Svaki sustav ima različito definirane parametre jedinične ćelije. Obzirom da se radi o različitim geometrijama, logično je da će naprezanje koje djeluje na kristale za svaki sustav prouzročiti različite deformacije. Primjerice, kod triklinskog sustava unutar jedinične ćelije ne postoji simetrija, zbog čega je svaka od 21 konstante različita. Istovremeno, za kubični sustav  $s$  matrica podatljivosti iz (3.6) je oblika:

$$\begin{pmatrix} s_{11} & s_{12} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{12} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{44} \end{pmatrix}. \quad (3.7)$$

### 3.2 Linearna kompresibilnost

Linearna kompresibilnost  $\beta_L$  opisuje promjenu duljine u nekom smjeru između dvije točke unutar kristala koja je uzrokovana hidrostatskim tlakom. Takvo naprezanje definiramo kao  $\sigma_{ij} = -p\delta_{ij}$ . Iz toga je očito da kompresibilnost neće ovisiti o konstantama  $s_{44}, s_{55}, s_{66}, s_{56}, s_{46}, s_{45}$  iz donjeg desnog  $3 \times 3$  bloka  $s$  matrice. Uzevši to u obzir, pomak (deformacija) duž neke usmjerene linije unutar kristala u hidrostatskom tlaku može se zapisati kao [15]

$$\varepsilon_{ij}l_i l_j = -ps_{ijkk}l_i l_j, \quad (3.8)$$

gdje su  $l_i$  komponente jediničnog vektora. Iz te jednadžbe prepoznaje se opći izraz za linearnu kompresibilnost:

$$\beta_L = s_{ijkk} l_i l_j. \quad (3.9)$$

Općim izrazom definiran je slučaj najmanje simetrije, za triklinske kristale:

$$\begin{aligned} \beta_L = & (s_{11} + s_{12} + s_{13}) l_1^2 + (s_{16} + s_{26} + s_{36}) l_1 l_2 + (s_{15} + s_{25} + s_{35}) l_3 l_1 + \\ & + (s_{12} + s_{22} + s_{23}) l_2^2 + (s_{14} + s_{24} + s_{34}) l_2 l_3 + \\ & + (s_{13} + s_{23} + s_{33}) l_3^2. \end{aligned} \quad (3.10)$$

Linearnu kompresibilnost također možemo definirati pomoću tenzora 2. reda [1]:

$$\beta_L = -\frac{\partial}{\partial p} \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}. \quad (3.11)$$

Svojstveni vektori tog tenzora određuju glavne osi kompresibilnosti. Definiramo ih kao osi duž kojih hidrostatski tlak ne uzrokuje posmično naprezanje. Ako u jednadžbu (3.4) uvrstimo hidrostatski tlak i prijedemo u sustav glavnih osi, dobivamo izraz:

$$\varepsilon_i = -p \sum_{k=1}^3 s_{ik}. \quad (3.12)$$

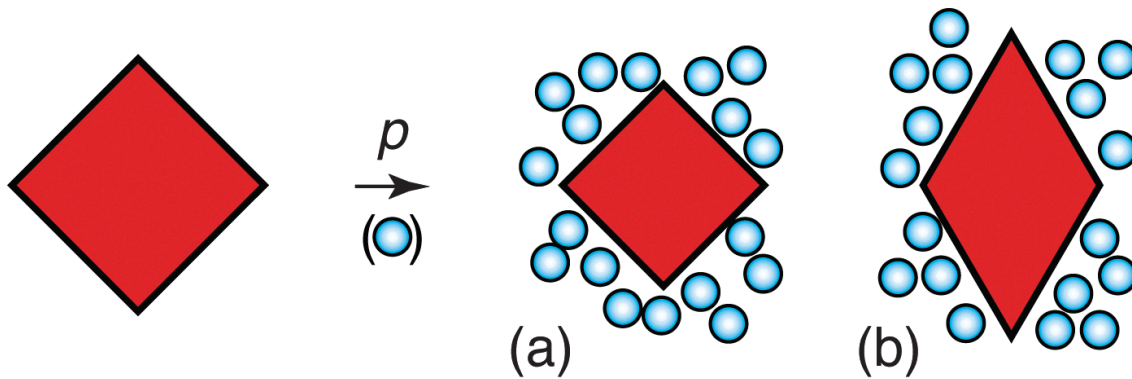
Time se dobiva konačan izraz za kompresibilnost u smjeru glavnih osi:

$$\beta_{L,i} = \sum_{k=1}^3 s_{ik}. \quad (3.13)$$

### **Negativna linearna kompresibilnost (NLC)**

Zamislimo materijal u hidrostatskom tlaku. Obzirom da na svaku njegovu stranicu djeluje jednaka sila prema unutra, očekuje se da će se i svaka stranica skratiti, a time i smanjiti ukupan volumen.

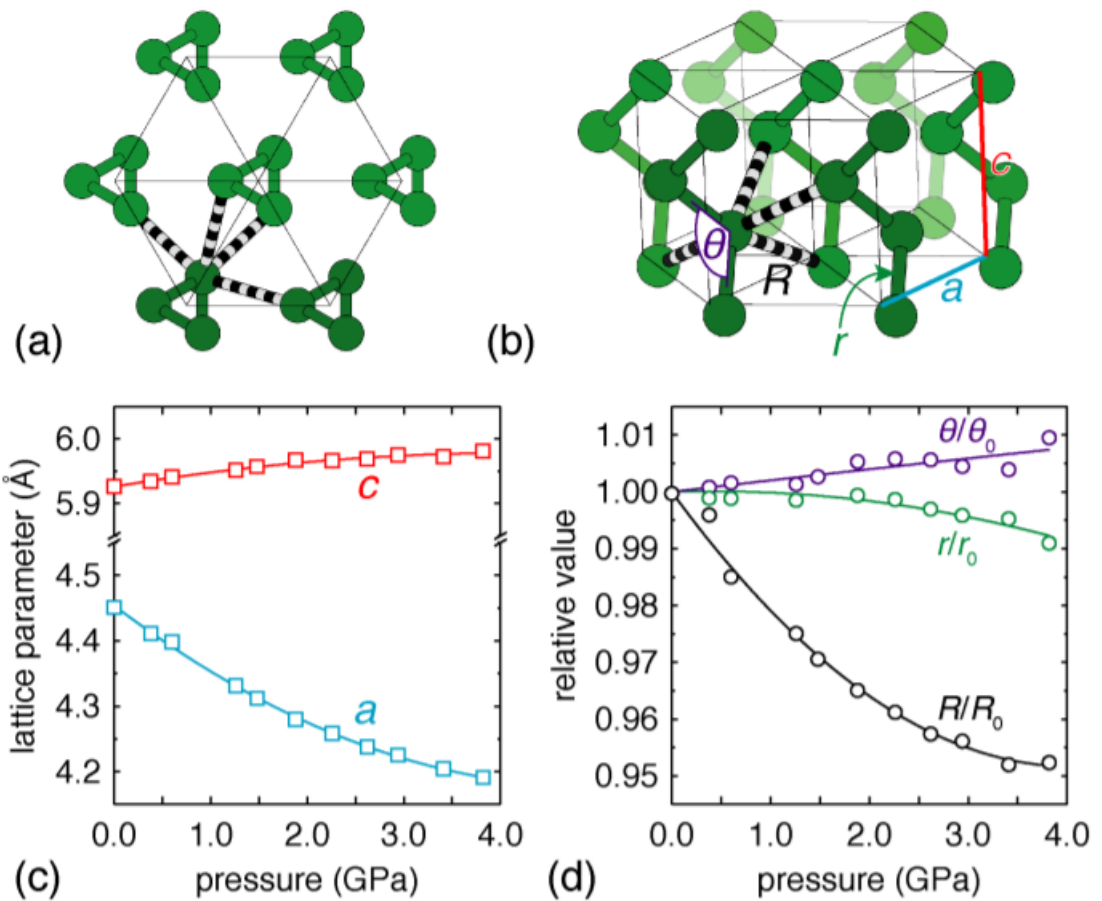
Zanimljivo svojstvo prisutno kod nekih materijala je pojava negativne linearne kompresibilnosti (NLC - Negative Linear Compressibility) u smjeru jedne od glavnih osi, uvijek popraćena velikim pozitivnim iznosom linearne kompresibilnosti u smjeru druge glavne osi. Kada se takav materijal stavi u hidrostatski tlak, njegove će se dimenzije mijenjati kao na slici 3.2.



Slika 3.2: Materijal u hidrostatskom tlaku: a) općenito b) NLC (preuzeto iz [1]).

Polimorfne strukture telurija i selenija [1] imaju izraženo NLC svojstvo. Mehanizam svojstva može se objasniti iz njihove strukture na slici 3.3. Za opis ćelije definirani su parametri  $r$ ,  $R$  i  $\theta$ . Uz primjenu hidrostatskog tlaka smanjuje se parametar  $R$  kojim je definirana udaljenost između atoma u spirali što uzrokuje produljenje tih spirala duž kristalne osi  $c$ , odnosno povećanje parametra  $\theta$  čime je savladan tlak u tom smjeru i umjesto kontrakcije, u tom smjeru potaknuto je produljenje.

Primjena NLC svojstava moguća je u industriji, konkretno u konstrukciji umjetnih mišića, kontrolnih sustava, aktuatora. Jedna od primjena je i u pojačavanju signala piezoelektričnih signala.



Slika 3.3: a) Polimorfna struktura telurija i selenija sastoji se od enantiomorfnih spirala atoma duž kristalne osi  $c$  sa slike b). b) Parametri  $r$ ,  $R$  i  $\theta$  i kristalne osi. c) Ovisnost kristalnih osi o tlaku. d) Ovisnost parametara  $r$ ,  $R$  i  $\theta$  o tlaku. (preuzeto iz [1])

## 4 Teorija funkcionala gustoće (DFT)

Teorija funkcionala gustoće je često korištena precizna metoda za predviđanje svojstava materijala u fizici kondenzirane materije (do 1 kcal/mol apsolutne greške).

### 4.1 Uvod

Ako gledamo općeniti slučaj, materijali se sastoje od mnoštva atoma, odnosno jezgara i elektrona pozicioniranih u prostoru. Potrebno je odrediti energiju takvog sustava pomoću njihovih međusobnih interakcija hamiltonijanom mnoštva čestica (4.1) [16].

$$\begin{aligned} \hat{H} = & -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ & - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \quad (4.1)$$

U jednadžbi su elektroni indeksirani malim, a jezgre velikim slovima. Od navedenih 5 članova, 2 su kinetičke energije (elektrona i jezgara), 2 su pozitivna doprinosa potencijalnoj energiji zbog odbijanja čestica istog naboja te jedan negativan zbog privlačenja čestica različitog naboja. Relativistički, magnetski (Zeemanov efekt) i QED efekti se u ovom koraku zanemaruju.

Pojednostavljenje hamiltonijana mnoštva čestica postiže se Born-Oppenheimerovom aproksimacijom. Pretpostavlja se da je gibanje jezgara zanemarivo sporo u odnosu na gibanje elektrona, odnosno da se kinetički član za jezgre u izrazu (4.1) može zanemariti. Hamiltonijan koji se sastoji od kinetičke energije elektrona i 3 doprinosa potencijalu možemo zapisati kao:

$$\hat{H} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{ne}, \quad (4.2)$$

gdje indeks  $e$  označava elektrone, a  $n$  jezgre.

Nerelativističke kvantne sustave opisujemo vremenski ovisnom Schrödingerovom jednadžbom:

$$i\hbar \frac{\partial \psi(\{\vec{r}_i\}, t)}{\partial t} = \hat{H} \psi(\{\vec{r}_i\}, t), \quad (4.3)$$

gdje  $\{\vec{r}_i\}$  označava skup radijvektora koji pokazuju  $i$ -ti elektron. Rješenja su anti-



metrične valne funkcije oblika:

$$\psi(\{\vec{r}_i\}, t) = \psi(\{\vec{r}_i\})e^{-i\frac{Et}{\hbar}}. \quad (4.4)$$

Poznavajući oblik rješenja, možemo izračunati očekivanu vrijednost opservable  $\hat{O}$ :

$$\langle \hat{O} \rangle = \frac{\langle \psi | \hat{O} | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (4.5)$$

Valne funkcije su po definiciji normalizirane pa je nazivnik jednak 1. Skalarni produkt i očekivana vrijednost u bra-ket notaciji podrazumijeva integraciju valnih funkcija po svim koordinatama.

Za opis interakcija s elektronima potrebno je definirati pozicije elektrona. Zato koristimo operator gustoće čestica definiran kao

$$\hat{n}(\vec{r}) = \sum_i^N \delta(\vec{r} - \vec{r}_i). \quad (4.6)$$

Ako uzmemo u obzir sve spomenuto i kao opservablu uzmemo energiju definiranu jednačbom hamiltonijana u Born-Oppenheimerovoj aproksimaciji (4.2), očekivanu vrijednost energije možemo zapisati kao funkcional na sljedeći način:

$$E = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \equiv \langle \hat{H} \rangle = \langle \hat{T}_e \rangle + \langle \hat{V}_{ee} \rangle + \langle \hat{V}_{nn} \rangle + \int V_{ne}(\vec{r})n(\vec{r})d^3r. \quad (4.7)$$

Bitno je uočiti da se u ukupnom izrazu za energiju interakcija jezgara i elektrona  $V_{ne}$  može zapisati kao integral umnoška potencijala i gustoće čestica. Doprinos  $V_{nn}$  izostavljamo jer se radi o aditivnoj konstanti.

Dva poznata pristupa teoriji mnoštva čestica su Hartreejeva i Hartree-Fockova aproksimacija.

## 4.2 Hartreejeva i Hartree-Fockova aproksimacija

Svojstvenim stanjima sustava pripadaju svojstvene energije izraza (4.7). One su određene minimumima te jednačbke i uz definiranu valnu funkciju mogu se računati varijacijskim principom:

$$\frac{\delta E}{\delta \phi_i^*} = 0 \quad (4.8)$$

U Hartreejevoj aproksimaciji varijacijska valna funkcija je produkt jednočestičnih valnih funkcija:

$$\psi(\{\vec{r}_i\}) = \prod_i \phi_i(\vec{r}_i). \quad (4.9)$$

Koristimo Hartreejeve atomske jedinice  $\hbar = m_e = e = 4\pi/\epsilon_0 = 1$ . Ubacivanjem u izraz za energiju i variranjem po  $\phi_i^*$  dobivamo:

$$\left\{ \sum_i \left[ -\frac{1}{2} \nabla_i + U_i(\vec{r}_i) \right] + \sum_{j \neq i} \int \frac{|\phi_j(\vec{r}_j)|^2}{|\vec{r}_j - \vec{r}_i|} d^3r \right\} \phi_i(\vec{r}_i) = \epsilon_i \phi_i(\vec{r}_i). \quad (4.10)$$

Zadnji se član s lijeve strane zove Hartreejev operator  $V_H(\vec{r}_i)$  i predstavlja pozitivan doprinos energiji zbog odbijanja i-tog elektrona sa svim ostalima j-tim elektronima. U toj aproksimaciji pojedini se elektron giba u kulonskom potencijalu jezgara i ostalih elektrona. Rezultat za ukupnu energiju u Hartreejevoj aproksimaciji je suma jednočestičnih energija umanjena za energiju međudjelovanja elektrona jer je ona dva puta uračunata u jednočestične energije:

$$E(\{\phi_i\}) = \sum_i \epsilon_i - \frac{1}{2} \sum_{i \neq j} \langle \phi_i \phi_j | V_{ij} | \phi_i \phi_j \rangle = \sum_i \epsilon_i - \frac{1}{2} \sum_i \langle \phi_i | V_H | \phi_i \rangle. \quad (4.11)$$

Nedostatak u Hartreejevoj aproksimaciji je to što varijacijska valna funkcija (4.9) ne zadovoljava Paulijev princip. Dodatni nedostaci su to što se njome ne mogu opisati zasjenjenja ni korelacije.

U Hartree-Fockovoj aproksimaciji (HFA) varijacijska valna funkcija opisana je Slaterovom determinantom:

$$\psi(\{\vec{r}_i\}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_1(\vec{r}_2) & \cdots & \phi_1(\vec{r}_N) \\ \phi_2(\vec{r}_1) & \phi_2(\vec{r}_2) & \cdots & \phi_2(\vec{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\vec{r}_1) & \cdots & \cdots & \phi_N(\vec{r}_N) \end{vmatrix}. \quad (4.12)$$

U tom slučaj može se pokazati da se dobiva izraz za međudjelovanje elektrona oblika:

$$\sum_{i \neq j} \langle \psi | V_{ij} | \psi \rangle = \int \frac{d^3r d^3r'}{|\vec{r} - \vec{r}'|} \sum_{i \neq j} \{ |\phi_i(\vec{r})|^2 |\phi_j(\vec{r}')|^2 - [\phi_i^*(\vec{r}) \phi_i(\vec{r}')] [\phi_j(\vec{r}) \phi_j^*(\vec{r}')] \}. \quad (4.13)$$

Prvi član u zagradi predstavlja Hartreejev operator, a dodatni član predstavlja energiju izmjene koja je posljedica antisimetričnosti valne funkcije. Izraz je primijenjiv na čestice istog spina.

Hamiltonijan interakcije jednostavnije je zapisati pomoću matrice gustoće. Matrica gustoće koja opisuje interakciju čestica različitog spina jednaka je nuli:  $\rho_{s,s'}(\vec{r}, \vec{r}') = \delta_{s,s'} \rho_s(\vec{r}, \vec{r}')$ . Ako to uzmemo u obzir, za hamiltonijan interakcije dobiva se izraz:

$$\langle \hat{H}_{int} \rangle = \frac{1}{2} \sum_{s,s'} \int \frac{d^3r d^3r'}{|\vec{r} - \vec{r}'|} \left[ \rho_{s,s'}(\vec{r}) \rho_{s,s'}(\vec{r}') - |\rho_{s,s'}(\vec{r}, \vec{r}')|^2 \right]. \quad (4.14)$$

Hartree-Fockova aproksimacija zadovoljava Paulijev princip i daje bolje rezultate, no za veliki broj čestica obje metode imaju prekomplikirane varijacijske valne funkcije koje je teško definirati. Jednočestične energije u HFA imaju logaritamski singularitet, što je nefizikalni rezultat. Popravlja se pretpostavkom da je međudjelovanje čestica zasjenjeno.

### 4.3 Thomas-Fermi-Dirac aproksimacija

Temeljna pretpostavka teorije funkcionala gustoće je da se svako svojstvo sustava interagirajućih čestica može zapisati kao funkcional osnovnog stanja gustoće čestica  $n_0(\vec{r})$  [16]. Thomas-Fermi-Dirac aproksimacija povijesno predstavlja prvi pristup problemu. Pretpostavlja se da je kinetička energija elektrona jednaka kinetičkoj energiji u homogenom neinteragirajućem plinu, gustoća jednaka lokalnoj gustoći neinteragirajućeg plina (prvi član u jednadžbi (4.15)). Energija izmjene elektrona je zamijenjena lokalnom izmjenom (treći član), a energija korelacije elektrona je zanemarena. Konačan izraz za energiju u obliku funkcionala je oblika:

$$E_{TF}[n] = C_1 \int d^3r n(\vec{r})^{5/3} + \int d^3r V_{ext}(\vec{r}) n(\vec{r}) + C_2 \int d^3r n(\vec{r})^{4/3} + \frac{1}{2} \int d^3r d^3r' \frac{n(\vec{r}) n(\vec{r}')}{|\vec{r} - \vec{r}'|}, \quad (4.15)$$

gdje su  $C_1$  i  $C_2$  poznate konstante [16]. Četvrti član je otprije poznata Hartreejeva energija.

## 4.4 Hohenberg-Kohn teoremi

Dva teorema koje su 1964. godine definirali Pierre Hohenberg i Walter Kohn predstavljaju temelj teorije funkcionala gustoće.

**Teorem 1.** Za svaki sustav čestica u vanjskom potencijalu postoji do na konstantu jedinstveno definiran potencijal  $V_{ext}(\vec{r})$  određen gustoćom čestica u osnovnom stanju  $n_0(\vec{r})$ .

*Dokaz.* Pretpostavimo da postoje dva različita vanjska potencijala  $V_1$  i  $V_2$  definirana istom gustoćom čestica. Obzirom da su potencijali različiti, u ta dva slučaja razlikuju se i hamiltonijani sustava i same valne funkcije za koje je pretpostavljeno da vode na istu gustoću čestica u osnovnom stanju  $n_0(\vec{r})$ . Označimo s  $\psi_1$  i  $\psi_2$  valne funkcije osnovnog stanja za svaki potencijal. Očito je da osnovno stanje jednog sustava nije istovremeno i osnovno stanje drugog. Stroga nejednakost vrijedi ako osnovna stanja nisu degenerirana:

$$E_1 = \langle \psi_1 | \hat{H}_1 | \psi_1 \rangle < \langle \psi_2 | \hat{H}_1 | \psi_2 \rangle. \quad (4.16)$$

Zadnji izraz može se zapisati kao:

$$\begin{aligned} \langle \psi_2 | \hat{H}_1 | \psi_2 \rangle &= \langle \psi_2 | \hat{H}_2 | \psi_2 \rangle + \langle \psi_2 | \hat{H}_1 - \hat{H}_2 | \psi_2 \rangle \\ &= E_2 + \int d^3r [V_1(\vec{r}) - V_2(\vec{r})] n_0(\vec{r}). \end{aligned} \quad (4.17)$$

Iz jednadžbe (4.16) vrijedi:

$$E_1 < E_2 + \int d^3r [V_1(\vec{r}) - V_2(\vec{r})] n_0(\vec{r}). \quad (4.18)$$

Nejednadžba za energiju  $E_2$  može se dobiti analogno. Izraz koji se dobiva sličan je gornjem, uz zamjenu indeksa 1 i 2. Izraz je oblika:

$$E_2 < E_1 + \int d^3r [V_2(\vec{r}) - V_1(\vec{r})] n_0(\vec{r}). \quad (4.19)$$

Sumiranjem zadnja dva izraza dobiva se kontradiktorna nejednakost:

$$E_1 + E_2 < E_1 + E_2. \quad (4.20)$$

Time je dokazano da je vanjski potencijal  $V_{ext}$  jednoznačno definiran. Slijedi da su ha-

miltonijan i višečestične valne funkcije jednoznačno definirane i sva svojstva sustava mogu se jedinstveno odrediti iz  $n_0(\vec{r})$ .  $\square$

**Teorem 2.** Funkcional energije sustava  $E[n]$  koji ovisi o gustoći čestica  $n(\vec{r})$  definiran je za svaki vanjski potencijal  $V_{ext}(\vec{r})$ . Energija osnovnog stanja sustava odgovara globalnom minimumu funkcionala koji je određen gustoćom čestica u osnovnom stanju  $n_0(\vec{r})$ .

*Dokaz.* Funkcional energije možemo zapisati kao:

$$E[n] = T[n] + E_{ee}[n] + \int d^3r V_{ext}(\vec{r})n(\vec{r}) + E_{nn}. \quad (4.21)$$

Prvi dio jednadžbe možemo zapisati kao zaseban Hohenberg-Kohnov funkcional

$$F_{HK}[n] = T[n] + E_{ee}[n] \quad (4.22)$$

kojim je univerzalno određen sustav elektrona, neovisno o samointerakciji jezgara ili vanjskom potencijalu u kojem se nalaze. Isti funkcional  $F_{HK}[n]$  vrijedi za svaki sustav međusobno interagirajućih elektrona.

Ponovno definirajmo sustav s vanjskim potencijalom  $V_1$ , pripadnom gustoćom čestica u osnovnom stanju  $n_1$  i valnom funkcijom  $\psi_1$ . Za takav sustav funkcional  $F_{HK}$  je jednak:

$$E_1 = E_{HK}[n_1] = \langle \psi_1 | \hat{H}_1 | \psi_1 \rangle. \quad (4.23)$$

Neko više stanje sustava određeno je s  $n_2$  i  $\psi_2$  iz čega slijedi nejednakost:

$$E_1 = \langle \psi_1 | \hat{H} | \psi_1 \rangle < \langle \psi_2 | \hat{H} | \psi_2 \rangle = E_2. \quad (4.24)$$

Time je dokazano da je energija sustava minimalna za pripadnu gustoću čestica u osnovnom stanju. Poznavanjem  $F_{HK}$  i variranjem ukupne energije (4.21) po gustoći čestica mogu se dobiti egzaktni vrijednosti gustoće i energije u osnovnom stanju.  $\square$

## 4.5 Kohn-Sham jednadžbe

Kohn i Sham su predložili model u kojem se umjesto problema mnoštva čestica gleda "pomoćni" sustav (engl. *auxiliary system*) nezavisnih čestica. Ako uzmemo u obzir da elektroni mogu imati dvije vrijednosti spina, novi hamiltonijan (u Hartreeje-

vim jedinicama) definiran je kao:

$$\hat{H}_{aux}^\sigma = -\frac{1}{2}\nabla^2 + V^\sigma(\vec{r}). \quad (4.25)$$

Ukupan broj elektrona je  $N = N^\uparrow + N^\downarrow$  te u osnovnom stanju elektroni popunjavaju najniža stanja energije  $\epsilon_i^\sigma$ .

Prvi član u hamiltonijanu odnosi se na kinetičku energiju:

$$T_s = -\frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N^\sigma} \langle \psi_i^\sigma | \nabla^2 | \psi_i^\sigma \rangle = \frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N^\sigma} \int d^3r |\nabla \psi_i^\sigma(\vec{r})|^2, \quad (4.26)$$

a ukupan funkcional energije s doprinosima zbog potencijala  $V^\sigma(\vec{r})$  je oblika:

$$E_{KS} = T_s[n] + \int dr V_{ext}(\vec{r})n(\vec{r}) + E_{Hartree}[n] + E_{nn} + E_{xc}[n]. \quad (4.27)$$

Integralni doprinos s  $V_{ext}$  posljedica je interakcije elektrona s jezgrama i vanjskog potencijala za koji se pretpostavlja da ne ovisi o spinu. Hartreejeva energija poznata je otprije i javlja se zbog međusobne interakcije elektrona:

$$E_{Hartree} = \frac{1}{2} \int d^3r d^3r' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|}, \quad (4.28)$$

a  $E_{nn}$  prestavlja konstantni doprinos zbog interakcije jezgara. Posljednji član  $E_{xc}$  predstavlja energiju izmjene i korelacije (engl. *exchange and correlation*). Energija izmjene pojavila se u HFA i posljedica je antisimetričnosti valne funkcije. Korelacijska energija je aproksimirani doprinos kojim se opisuje utjecaj elektrona na njihovo međusobno gibanje.

Kohn-Sham jednadžbe dobivaju se minimizacijom funkcionala energije po gustoći čestica ili efektivnom potencijalu. Obzirom da kinetička energija  $T_s$  kao funkcional eksplicitno ovisi o valnoj funkciji, a ostali su članovi funkcionali gustoće čestica, variranjem izraza i korištenjem lančanog deriviranja dobiva se izraz:

$$\frac{\delta E_{KS}}{\delta \psi_i^{\sigma*}(\vec{r})} = \frac{\delta T_s}{\delta \psi_i^{\sigma*}(\vec{r})} + \left[ \frac{\delta E_{ext}}{\delta n(\vec{r}, \sigma)} + \frac{\delta E_{Hartree}}{\delta n(\vec{r}, \sigma)} + \frac{\delta E_{xc}}{\delta n(\vec{r}, \sigma)} \right] \frac{\delta n(\vec{r}, \sigma)}{\delta \psi_i^{\sigma*}(\vec{r})} = 0. \quad (4.29)$$

Zbog različito usmjerenih spinova i ortogonalnosti valnih funkcija vrijedi uvjet:

$$\langle \psi_i^\sigma | \psi_j^{\sigma'} \rangle = \delta_{ij} \delta_{\sigma\sigma'}. \quad (4.30)$$

Koristeći sljedeće izraze:

$$\frac{\delta T_s}{\delta \psi_i^{\sigma*}(\vec{r})} = -\frac{1}{2} \nabla^2 \psi_i^\sigma, \quad (4.31)$$

$$\frac{\delta n(\vec{r}, \sigma)}{\delta \psi_i^{\sigma*}(\vec{r})} = \psi_i^\sigma(\vec{r}), \quad (4.32)$$

jednadžba (4.29) može se svesti na problem svojstvenih vrijednosti, kao kod Schrödingerove jednadžbe:

$$H_{KS}^\sigma = -\frac{1}{2} \nabla^2 + V_{KS}^\sigma(\vec{r}), \quad (4.33)$$

$$(H_{KS}^\sigma - \varepsilon_i^\sigma) \psi_i^\sigma(\vec{r}) = 0, \quad (4.34)$$

s potencijalom oblika:

$$\begin{aligned} V_{KS}^\sigma(\vec{r}) &= V_{ext}(\vec{r}) + \frac{\delta E_{Hartree}}{\delta n(\vec{r}, \sigma)} + \frac{\delta E_{xc}}{\delta n(\vec{r}, \sigma)} \\ &= V_{ext}(\vec{r}) + V_{Hartree}(\vec{r}) + V_{xc}^\sigma(\vec{r}). \end{aligned} \quad (4.35)$$

Jednadžbe (4.33), (4.34) i (4.35) su Kohn-Sham jednadžbe.

Ovim pristupom lakše je definirati energiju izmjene i korelacije kao posebni član:

$$E_{xc}[n] = \int n(\vec{r}) \epsilon_{xc}([n], \vec{r}) d^3r \quad (4.36)$$

koji se aproksimira lokalnom gustoćom. Veličina  $\epsilon_{xc}([n], \vec{r})$  ovisi o lokalnoj gustoći  $n(\vec{r}, \sigma)$  oko neke točke određene vektorom  $\vec{r}$ . Za poznati  $E_{xc}$  teorija daje egzaktno rezultate te su stoga razvijene razne aproksimacije za njegovo određivanje od kojih će jednostavnije biti objašnjene u nastavku.

## 4.6 LDA (Local density approximation)

Aproksimacija lokalne gustoće (LDA) [17] promatra elektrone u krutini blizu li-mesa homogenog elektronskog plina uz pretpostavku da su efekti izmjene i korelacije kratkodosežni. U tom slučaju se doprinosi energiji izmjene i korelacije mogu

aproksimirati integralom (4.36) po cijelom prostoru uz  $\epsilon_{xc}^{hom}([n], \vec{r})$  definiran kao kod homogenog elektronskog plina odgovarajuće gustoće u svakoj točki prostora. Ako se aproksimacija proširi uzimanjem spina u obzir, dobiva se LSDA (*Local Spin Density Approximation*). Izraz se može zapisati kao:

$$E_{xc}[n^\uparrow, n^\downarrow] = \int n(\vec{r}) \epsilon_{xc}^{hom}([n^\uparrow(\vec{r}), n^\downarrow(\vec{r})]) d^3r. \quad (4.37)$$

Funkcional  $E_{xc}$  je univerzalan jer se odnosi na opći slučaj homogenog elektronskog plina. Za dobivanje rezultata potrebno je izraziti energiju izmjene i korelacije kao funkciju elektronske gustoće, što se može do velike preciznosti dobiti Monte Carlo metodama [18].

#### 4.7 GGA (*Generalized-gradient approximation*)

LDSA aproksimacija daje dovoljno dobre rezultate, no moguća su daljnja poboljšanja. U GGA aproksimacijama [19]  $E_{xc}$  se definira kao funkcional gustoće i apsolutne vrijednosti gradijenta gustoće. Izraz je sljedeći:

$$E_{xc}^{GGA}[n^\uparrow, n^\downarrow] = \int n(\vec{r}) \epsilon_x^{hom}(n) F_{xc}(n^\uparrow(\vec{r}), n^\downarrow(\vec{r}), |\nabla n^\uparrow(\vec{r})|, |\nabla n^\downarrow(\vec{r})|, \dots) d^3r. \quad (4.38)$$

Veličina  $\epsilon_x^{hom}(n)$  odnosi se na energiju izmjene nepolariziranog plina, a  $F_{xc}$  je bezdimenzionalna funkcija koja ovisi o korištenom modelu GGA.

#### **Komentar**

U ovom poglavlju predstavljene su osnove teorije funkcionala gustoće čiju praktičnu primjenu omogućavaju programski paketi za računalnu fiziku. Najpoznatiji su *VASP*, *Siesta*, *Quantum Espresso*, *ABINIT*...

U okviru ovog diplomskog rada za provjeru rezultata izračunatih ANI1 potencijalima koristit će se GGA i programski paket *Quantum Espresso* [20].



## 5 Konstrukcija baza podataka za molekularne kristale i simulacija elastičnih svojstava

Za sve potrebe simulacija koristio se programski jezik *Python*, točnije *Jupyter Notebook*. Cijeli postupak je opisan u nastavku.

### 5.1 ASE (Atomic Simulation Environment) i rad s datotekama

COD baza podataka [3] sadrži strukturne eksperimentalne podatke dobivene iz difrakcije X zrakama (jedinična ćelija, koordinate atoma, simetrije...) za velik broj materijala u .cif formatu te su od njih odabrani oni koji se sastoje od elemenata C, H, N i O te je zadan uvjet da se moraju sastojati od minimalno dvije različite vrste atoma. U bazi postoji 42748 takvih datoteka za preuzimanje i daljnju obradu. ASE [21] je poznati programski paket za simulaciju atoma, molekula i materijala. U okviru ovog seminara korišteni su alati za učitavanje, optimizaciju i vizualizaciju materijala te paket za konstrukciju baze podataka u ASE sučelju (*ASE database*). Za optimizaciju strukture koristi se BFGS (Broyden–Fletcher–Goldfarb–Shanno) algoritam [22] iz istog paketa.

Naredbom `ase.io.read()` učitavaju se .cif datoteke. Učitavanje je u nekim slučajevima onemogućeno zbog grešaka ili višeznačnosti u .cif datotekama. Način filtriranja tih slučajeva bit će objašnjen u sljedećem pododjeljku. Time dobivamo `Atoms` objekt kojem se pridružuje kalkulator za izračun svojstava materijala i sprema u bazu kao `AtomsRow` objekt specifičan za ASE bazu podataka.

### 5.2 Čitanje .cif datoteka

Zbog velikog broja radnih datoteka i podataka velika je vjerojatnost da će se pojaviti i određene greške. Promotrimo neke od uočenih problema:

- Nemogućnost učitavanja pomoću naredbe `ase.io.read()`
- Nemogućnost jednoznačnog pozicioniranja pojedinih atoma
- Izostavljeni ili krivo pozicionirani vodici
- Besmislena struktura (nepotpuni podaci)

- Beskonačno vrijeme optimizacije BFGS algoritmom zbog nerealistične strukture

Prvi problem najlakše je uočljiv jer se pojavljuje kod učitavanja `Atoms` objekta te se rješava jednostavnom upotrebom iznimke (exception) kod upisivanja u bazu.

Drugi problem riješen je pomoću programskog paketa *Pymatgen* (*Python Material Genomics*) [23] koji sadrži algoritme za popravak mogućih grešaka u strukturi materijala i kompatibilan je s paketom ASE. Pokazalo se da se dana greška najlakše identificira ako se `.cif` datoteka učitava pomoću naredbe `CifParser()` te korištenjem metode `get_structures()`. *Pymatgen* u slučaju greške vraća `ValueError` te se u tom slučaju opet radi iznimka u iteraciji po materijalima. Potreba za time javila se zato što ASE samo upozorava na tu grešku, no dozvoljava učitavanje, što je nepoželjno.

Beskonačno vrijeme optimizacije može biti posljedica trećeg ili četvrtog navedenog problema te je stoga bilo potrebno napraviti još jedan filter baziran na uvjetu da relativna udaljenost dvaju različitih atoma mora biti veća od 0.8 Å. To je utvrđeno pregledavanjem pojedinih `.cif` datoteka i činjenicom da je kod difrakcije pozicije vodika relativno teško točno odrediti. Detalji vezani uz optimizaciju opisani su u sljedećem pododlomku.

### 5.3 Relaksacija strukture

Određeni potencijal iz `torchani` paketa za `Atoms` objekt pridružuje se metodom `set_calculator()`. Tim postupkom moguće je simulirati svojstva učitanih materijala poput volumena, energije, dimenzija jedinične ćelije, itd. Kako bi ti podaci bili što sličniji realnom slučaju, potrebno je provesti relaksaciju (optimizaciju) strukture, odnosno naći najnižu energiju sustava za dani potencijal. Time se postiže simulacija kojom bi se trebali dobivati podaci usporedivi s eksperimentalnima. Za optimizaciju strukture koristi se poznati BFGS algoritam. Uz sve navedeno, koristi se i `ExpCellFilter` klasa za svaki `Atoms` objekt iz modula `ase.constraints` kako bi se uz relaksaciju atoma, omogućila i relaksacija same jedinične ćelije.

Prethodno navedene greške odnose se na moguće smetnje kod samog izvođenja programskog koda potrebnog za izradu novih baza za svaki strojno naučeni potencijal te su eliminirane kako bi se program mogao nesmetano izvoditi. Preostale greške poput prevelikih numeričkih odskakanja od eksperimentalnih podataka predstavljaju korisne informacije za izradu statistike te utvrđivanja kvalitete primijenjenih strojno

naučenih potencijala.

### ***Kratka shema postupka***

1. Preuzimanje .cif datoteka s COD baze
2. Učitavanje i spremanje u prvobitnu bazu (ujedno i prvi filter za nemogućnost učitavanja pomoću `ase.io.read()`)
3. Dodatno filtriranje prvobitne baze, sortiranje i spremanje u novu "čistu" bazu za upotrebu
4. Optimizacija strukture kalkulatorom strojno naučenog potencijala i BFGS metodom (relaksacija) iteriranjem "čiste" baze i spremanje u novu bazu posebno za svaki kalkulator
5. Analiza podataka i izrada histograma

### ***5.4 Račun elastičnog tenzora i kompresibilnosti***

Relaksirane strukture uvjet su za daljnju simulaciju svojstava. Elastični tenzor računat će se pomoću dvije različite metode.

#### ***ASE elastic i Pymatgen***

Modul *elastic* [24] ekstenzija je ASE paketa koja sadrži potrebne alate za izračun elastičnog tenzora. Jednostavan primjer računa tenzora za materijal iz baze uz korištenje ANI1-x kalkulatora dan je u nastavku (varijabla `material` odnosi se na .cif datoteku za određeni materijal učitano u ASE kao `Atoms` objekt):

```
1 from elastic.elastic import get_elastic_tensor
2 from elastic import get_elementary_deformations
3 ...
4 material.set_calculator(torchani.models.ANI1x().ase())
5 systems = get_elementary_deformations(material, n=5, d=2)
6 Cij, Bij = get_elastic_tensor(material, systems=systems)
7 ...
```

Bilo kakav račun u ASE sučelju moguć je tek nakon pridruživanja kalkulatora. Nakon toga se funkcijom `elastic.elastic.get_elementary_deformations()` generiraju deformirane strukture ovisno o zadanim parametrima. Parametrom `d` definirana je apsolutna vrijednost maksimalne deformacije strukture, odnosno povećanje ili smanjenje dimenzija ćelije u postocima, dok parametar `n` određuje broj deformacija. Primjerice, za dane parametre u navedenom kodu, traži se 5 deformacija u intervalu od -2 do 2: `[-2, -1, 0, 1, 2]`. Primijenjene deformacije po parametrima ćelija ovise o njihovoj simetriji (detaljnije u [24]). Za svaku deformaciju računa se naprezanje te se linearnom prilagodbom dobivaju elastični koeficijenti  $C_{ij}$ .

Korištenjem modula `pymatgen.analysis.elasticity` u sklopu *Pymatgen* [23] paketa moguće je konstruirati kod za izračun elastičnog tenzora koji radi na sličnom principu. Razlika je u tome što je u ovom slučaju korištena nelinearna prilagodba pomoću metode `from_diff_fit()`.

## ELATE

ELATE [25] je besplatna internet aplikacija za analizu elastičnih tenzora i vizualizaciju mehaničkih svojstava u 3D. Iz koda su preuzete funkcije `maximize()` i `minimize()` za račun maksimalne i minimalne kompresibilnosti u svrhu dodatne provjere.

## 5.5 Postavke DFT računa

Programski paket *Quantum Espresso* (QE), koji se u ovom radu koristi za potrebe DFT računa, kao i mnogi drugi DFT paketi za simulacije koristi pseudopotencijale. Njihovo korištenje dodatno pojednostavljuje teoriju mnoštva čestica u svrhu ograničavanja simulacije na najbitnije parametre. Obzirom da su valentni elektroni atoma u materijalu oni koji bitno utječu na svojstva, elektroni u unutarnjim ljuskama zajedno s jezgrom čine zaseban sustav uklopljen u efektivni potencijal koji valentni elektroni u simulaciji osjećaju. U računima koristit će se GBRV [26] pseudopotencijali. Za aproksimaciju izmjene i korelacije koristit će se vdW-DF-cx [27] funkcional kojim će se u račun uključiti i van der Waalove sile za bolju preciznost GGA računa. QE koristi bazu ravnih valova. Što se više članova uzme u obzir točnost je bolja, no time račun postaje računalno skuplji. Stoga se definira maksimalna ki-

netička energija do koje će se ravni valovi s odgovarajućim  $\vec{G}$  uzimati u obzir (prema izrazu  $T = \hbar^2|\vec{G}|^2/2m$ ) te će u ovom slučaju ona iznositi 60 Ry ( $\approx 816$  eV). Integral recipročne rešetke unutar prve Brillouinove zone također se aproksimira sumom konačnog broja točaka. Koristili smo Monkhorst-Pack [28] mrežu s gustoćom točaka recipročnog prostora od 5 Å. Strukture se relaksiraju uz pomoć BFGS algoritma. sve dok sile na sve atome ne budu manje od `forc_conv_thr=1.0d-4` Ry/Bohr radius ( $\approx 2.571$  meV/Å), promjena energije ne bude manja od `etot_conv_thr=1.0d-5` Ry ( $\approx 136$   $\mu$ eV), te tlak ne bude manji od 0.1 kbar. Nakon toga provodi se račun elastičnih konstanta korištenjem `thermo_pw` [29] QE ekstenzije.

## 6 Rezultati i diskusija

### 6.1 Test ANI1 potencijala

U samom početku, broj preuzetih .cif datoteka s COD baze podataka iznosio je 42748. Nakon korištenih tehnika filtriranja spomenutih u prošlom poglavlju, konstruirana je opsežna baza molekularnih kristala sastavljenih od elemenata C, H, N i O uz uzlazno sortiranje. Za test točnosti ANI1-x i ANI1-ccx potencijala cilj je korištenjem početne baze s eksperimentalnim podacima struktura konstruirati dvije nove baze (jednu za svaki korišteni potencijal). Uzlazno sortiranje baze omogućuje izdvajanje prvih 3000 materijala iz baze koji imaju najjednostavniju strukturu (najmanju jediničnu ćeliju) za što kraće vrijeme relaksacije. Za prvih 3000 molekularnih kristala iz baze provedene su relaksacije te su izračunati histogrami za relativnu pogrešku  $\Delta X$  potencijalima simulirane veličine  $X_{x/ccx}$  u odnosu na eksperimentalnu  $X_{exp}$  za svaki potencijal:

$$\Delta X_{x/ccx} = \frac{X_{x/ccx} - X_{exp}}{X_{exp}}. \quad (6.1)$$

Prije je spomenuto da je jedan od problema koji se može javiti beskonačan broj koraka u optimizaciji, zbog čega ograničavamo maksimalni broj koraka BFGS algoritma za svaki korišteni potencijal na sljedeći način:

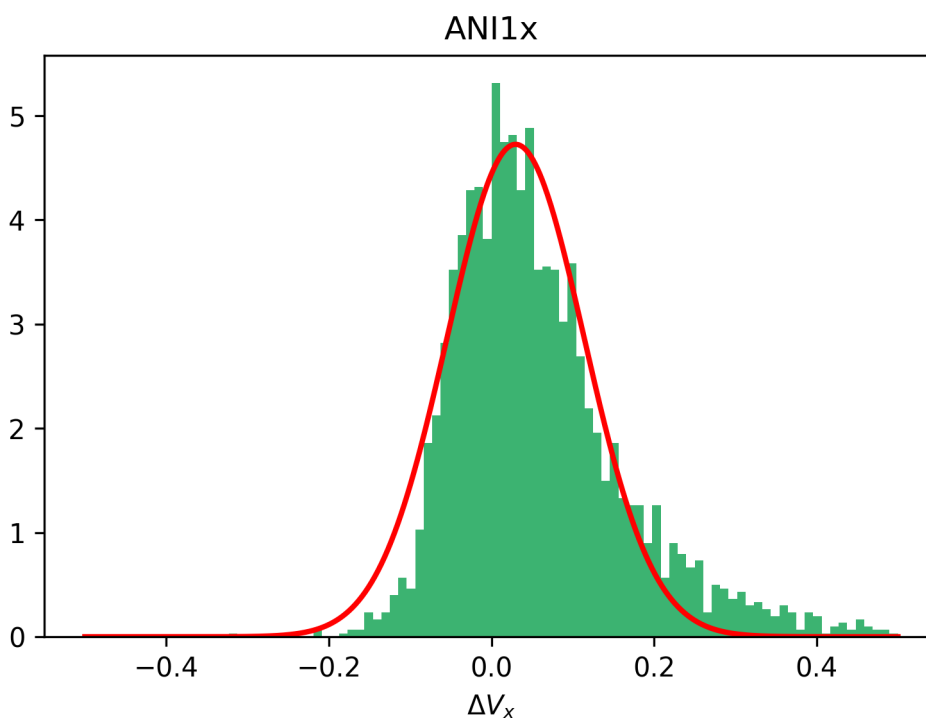
$$N_{steps} = 300 + x \cdot n_{atoms}, \quad (6.2)$$

gdje je  $N_{steps}$  broj koraka u relaksaciji,  $x = 6$  za ANI1-x kalkulator,  $x = 10$  za ANI1-ccx kalkulator, a  $n_{atoms}$  broj atoma u jediničnoj ćeliji kristala. Pretpostavlja se da trajanje relaksacije raste s kompleksnošću strukture materijala. Za ANI1-ccx potencijal dozvoljen je veći broj koraka jer je njegovim korištenjem relaksacija u prosjeku kraće trajala. Tako definirane relaksacije izvođene su paralelno te su se odvijale gotovo jednakim brzinama. Nakon relaksacije ukupno 3000 materijala, ANI1-x kalkulator je za 92 materijala (3.1%) dosegno maksimalan broj koraka, a ANI1-ccx za njih 49 (1.63%). Sve relaksirane strukture spremljene su u baze posebno za svaki potencijal. Maksimalna sila do koje se provodila relaksacija BFGS algoritmom iznosila je:

$$f_{max} = 0.0005 \text{ eV/Å}. \quad (6.3)$$

Tablica 6.1: Srednje vrijednosti i standardna devijacija odstupanja simuliranih ćelija od eksperimentalnih vrijednosti.

	$\mu$ [%]	$\sigma$ [%]
$\Delta V_x$	3.0	8.4
$\Delta V_{ccx}$	1.6	10.0
$\Delta  \vec{a}_1 _x$	1.2	6.1
$\Delta  \vec{a}_1 _{ccx}$	1.2	6.8

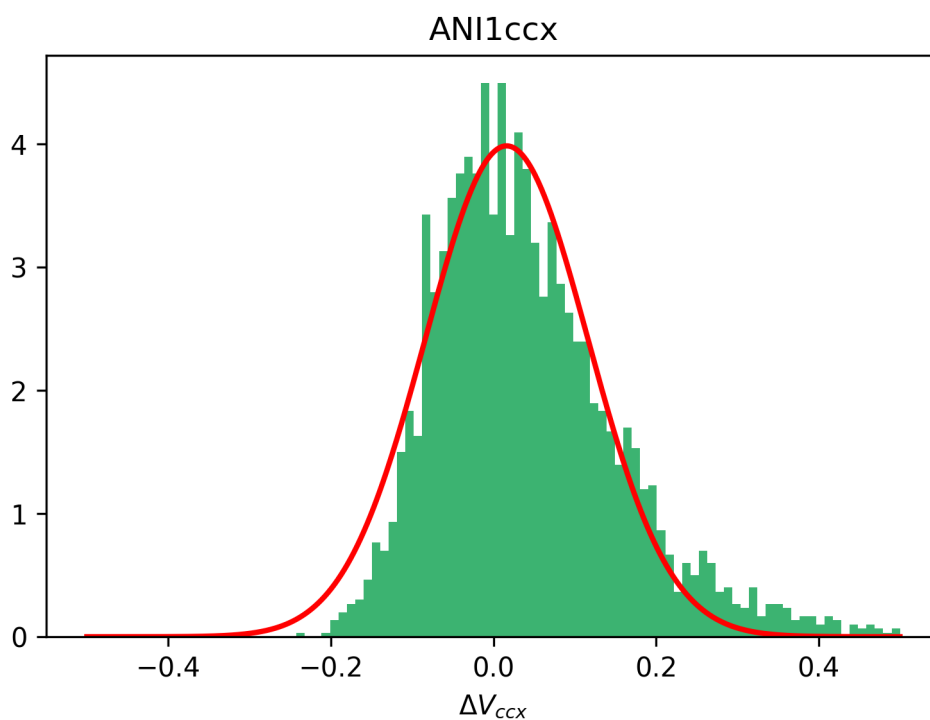


Slika 6.1: Histogram relativnih pogrešaka volumena izračunatih ANI1-x potencijalom u odnosu na eksperimentalne podatke.

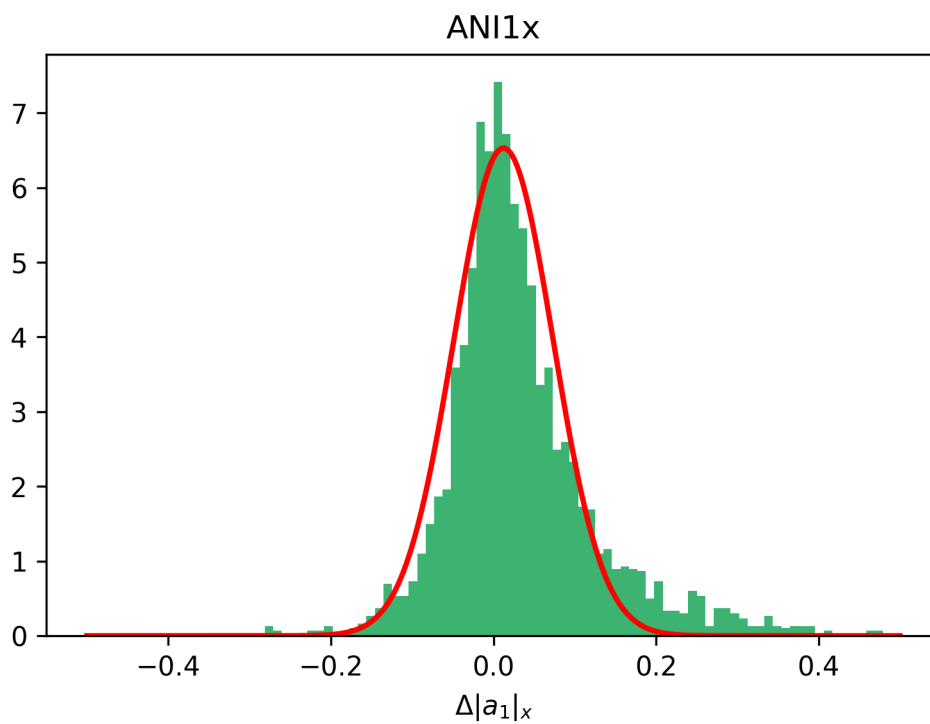
Prilagodбом Gaussove distribucije na histograme izračunate su srednje vrijednosti i standardne devijacije koje su za pripadni strojno naučeni potencijal prikazane u tablici 6.1. Histogrami na slikama 6.1 i 6.2 prikazuju relativnu pogrešku izračunatih volumena, a na slikama 6.3 i 6.4 relativnu pogrešku apsolutne vrijednosti vektora  $\vec{a}_1$  jedinične ćelije.

Iz rezultata je vidljivo da ANI1-ccx ima manju srednju vrijednost relativne pogreške, ali veću standardnu devijaciju. Zaključuje se da oba ANI1 potencijala daju dovoljno dobre i približne rezultate.

Bitno je uočiti asimetričnost histograma. Velike greške većinom su bile pozitivne, što prema jednadžbi (6.1) znači da je potencijal u većini slučajeva izračunao preveliku vrijednost. Razlog tome mogu biti nedostaci u filtriranju podataka ili višku

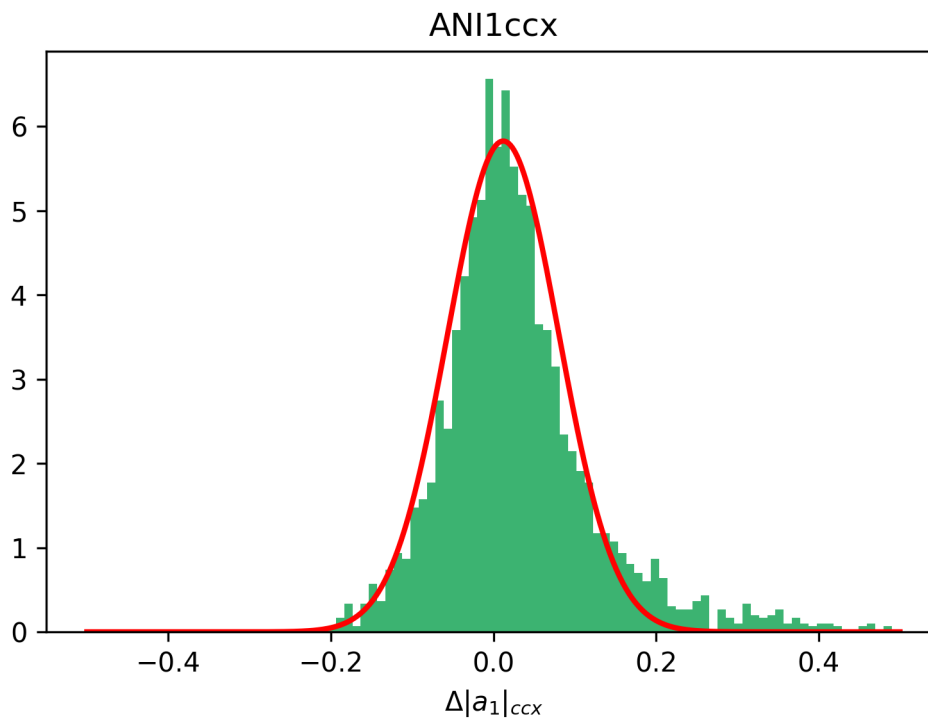


Slika 6.2: Histogram relativnih pogrešaka volumena izračunatih ANI1-ccx potencijalom u odnosu na eksperimentalne podatke



Slika 6.3: Histogram relativnih pogrešaka duljina jedne kristalne osi izračunatih ANI1-x potencijalom u odnosu na eksperimentalne podatke.





Slika 6.4: Histogram relativnih pogrešaka duljina jedne kristalne osi izračunatih ANI1-ccx potencijalom u odnosu na eksperimentalne podatke.

učitanih atoma (primjerice vodika) iz .cif datoteke. U ranijim pretraživanjima i vizualizaciji .cif-ova pronađeni su materijali kojima je učitani višak vodika čime je bio motiviran filter s uvjetom da relativna udaljenost između atoma mora biti manja od 0.8 Å. Dodatan razlog za nešto veće simulirane volumene je što korištenim potencijalima, prema samoj njihovoj konstrukciji preko *cutoff* funkcija, nedostaje dio dugodosežnih elektrostatskih interakcija poput dipol-dipol interakcije među molekulama, kao i dio disperzivnih van der Waalsovih interakcija između molekula. Ovi nedostaci će se pokušati otkloniti u daljnjem radu.

Dodatnom analizom utvrđeno je da postoje određene .cif datoteke kod kojih su u dokumentiranoj strukturi materijala izostavljeni atomi vodika. Neki od takvih materijala dostigli su maksimalan broj koraka u relaksaciji. Iz tog razloga je provedeno dodatno filtriranje baze čime su se uklonili takvi materijali.

## 6.2 Relaksacije uz ograničenje na očuvanje simetrije

U prošlom potpoglavlju opisan je postupak testiranja točnosti dvaju potencijala te su oba dala dovoljno dobre rezultate. Ideja je nastaviti koristiti te potencijale za simu-

Tablica 6.2: Srednje vrijednosti relativne pogreške i standardna devijacija simuliranih volumena ćelija u odnosu na eksperimentalne vrijednosti uz postavljen uvjet za čuvanje simetrije tijekom relaksacije.

	$\mu$ [%]	$\sigma$ [%]
$\Delta V_x$	2.5	8.5
$\Delta V_{ccx}$	0.9	9.6

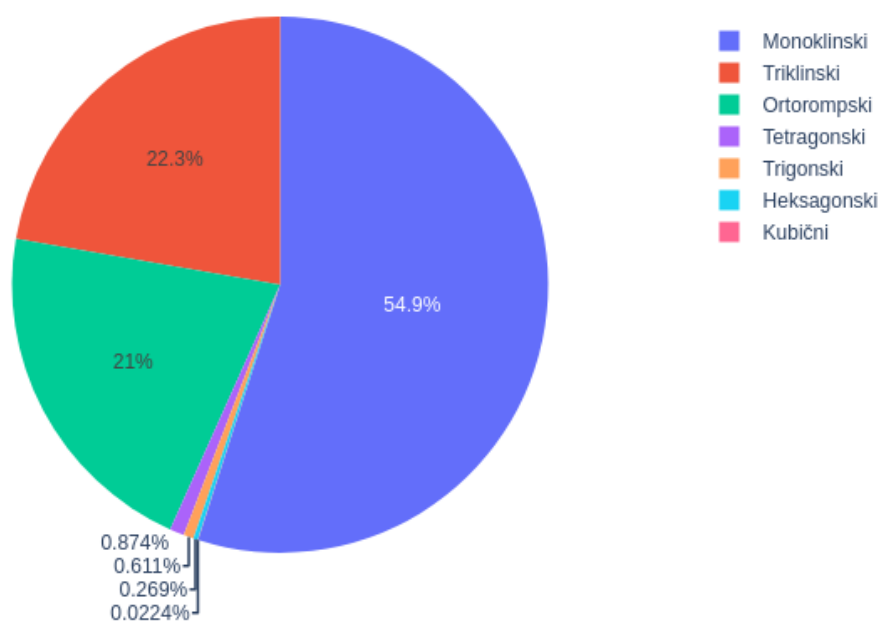
laciju elastičnih svojstava. Prema slici 6.5, u prvoj bazi s eksperimentalnim podacima o strukturama prevladavaju molekularni kristali monoklinske simetrije, (njih 54.9%), no nakon relaksacije gotovo svi prelaze u triklinski sustav. Kako bi se postigla što preciznija simulacija elastičnih svojstava, potrebno je definirati ograničenje u postavkama relaksacije da kristalna simetrija nakon relaksacije ostane ista, za što se koristi paket `ase.ase.spacegroup.symmetrize`. Takvo ograničenje za određeni materijal učitani u ASE sučelje zadaje se metodom `set_constraint(FixSymmetry(material))`, gdje `material` predstavlja učitani kristal kao `Atoms` objekt. Maksimalni broj koraka relaksacije definiran je opet prema jednadžbi (6.2), uz  $x = 10$  za oba potencijala u ovom slučaju.

U tablici 6.2 navedene su gausijanom aproksimirane vrijednosti relativne pogreške i standardne devijacije, a na slikama 6.6 i 6.7 prikazani su histogrami relativnih pogrešaka za izračunate volumene jedinične ćelije. Nakon relaksacije materijali se ponovno spremaju u novu bazu za svaki potencijal, uz izostavljanje onih materijala kod kojih je postignut maksimalan broj koraka u relaksaciji.

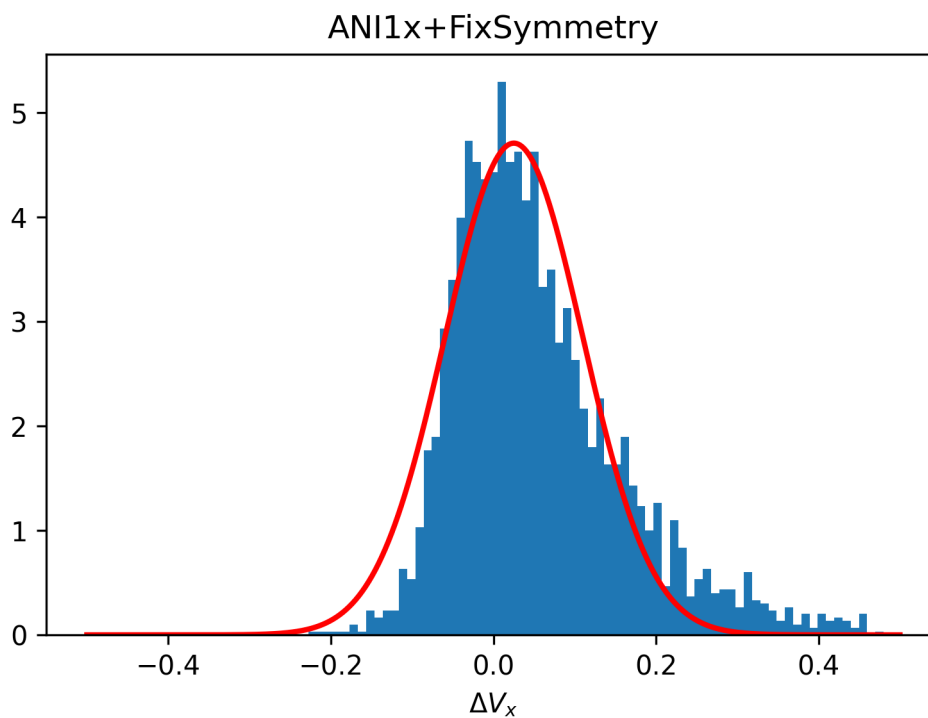
Usporedbom tablica 6.1 i 6.2 zaključuje se da relaksacije uz ograničenje za očuvanje simetrije daju preciznije rezultate kao što je i bilo očekivano. Na histogramima 6.6 i 6.7 ponovno se uočava asimetričnost kao i u prošlom potpoglavlju te je jedina razlika to što je vrh gausijana kod oba potencijala bliži nuli, što je posljedica manjih relativnih pogrešaka. Standardne devijacije nisu se bitno promijenile. Za potrebe simulacije elastičnih svojstava u obzir će se uzeti samo oni materijali čiji je iznos relativne pogreške simuliranog volumena manji od 15%.

### 6.3 Elastična svojstva ortorompskih kristala

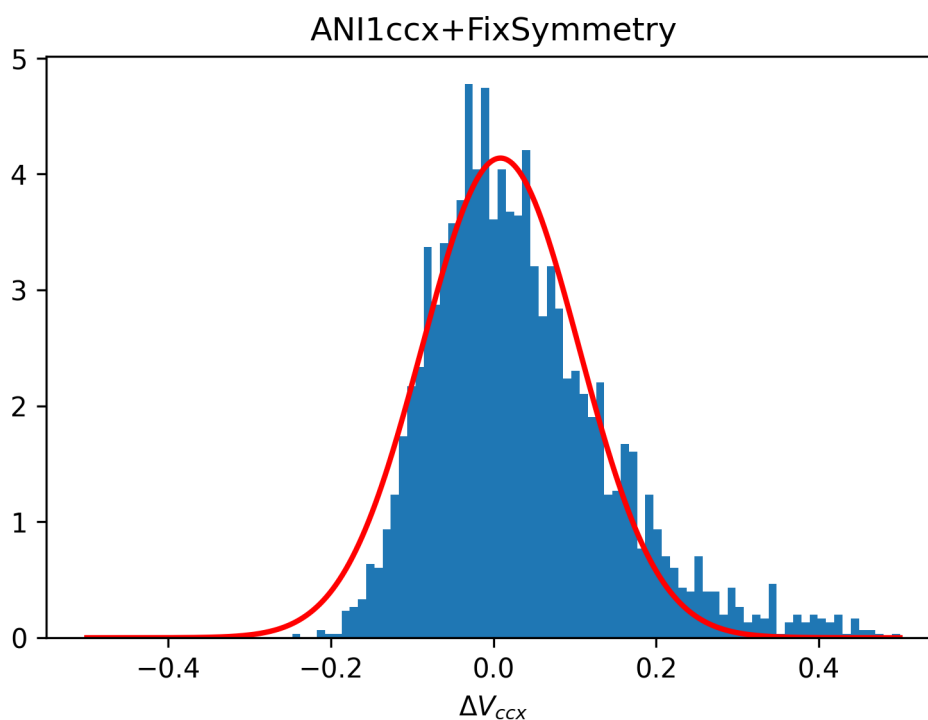
Račun elastičnog tenzora provodi se korištenjem paketa *Pymatgen* i *elastic*, a simulacija mehaničkih svojstava bit će provedena samo za ortorompske kristale. Razlog tome je to što u slučaju ortorompskih kristala račun jednostavniji zbog simetrije, a



Slika 6.5: Grafički prikaz zastupljenosti pojedinih kristalnih simetrija u bazi molekularnih kristala iz COD baze nakon filtriranja (konačan broj materijala nakon filtriranja je 35694, u odnosu na početnih 42748). Baza je sortirana uzlazno po broju atoma unutar jedinične ćelije materijala te svaka kristalna simetrija nije ravnomjerno zastupljena u svim dijelovima baze. Primjerice, kod prvih 8000 materijala prevladava triklinska simetrija (skoro 50%), dok u cijeloj bazi prevladava broj monoklinskih kristala (54.9% prema slici).



Slika 6.6: Histogram relativnih pogrešaka volumena izračunatih ANI1-x potencijalom u odnosu na eksperimentalne podatke uz ograničenje za očuvanje simetrije kristala tijekom relaksacije.



Slika 6.7: Histogram relativnih pogrešaka volumena izračunatih ANI1-ccx potencijalom u odnosu na eksperimentalne podatke uz ograničenje za očuvanje simetrije kristala tijekom relaksacije.

njihova zastupljenost u bazi (prema slici 6.5) dovoljno velika za izradu statistike. Krajnji cilj je iz elastičnih tenzora izračunati linearne kompresibilnosti te identificirati one materijale za koje ANI1 potencijali predviđaju svojstvo negativne linearne kompresibilnosti (NLC).

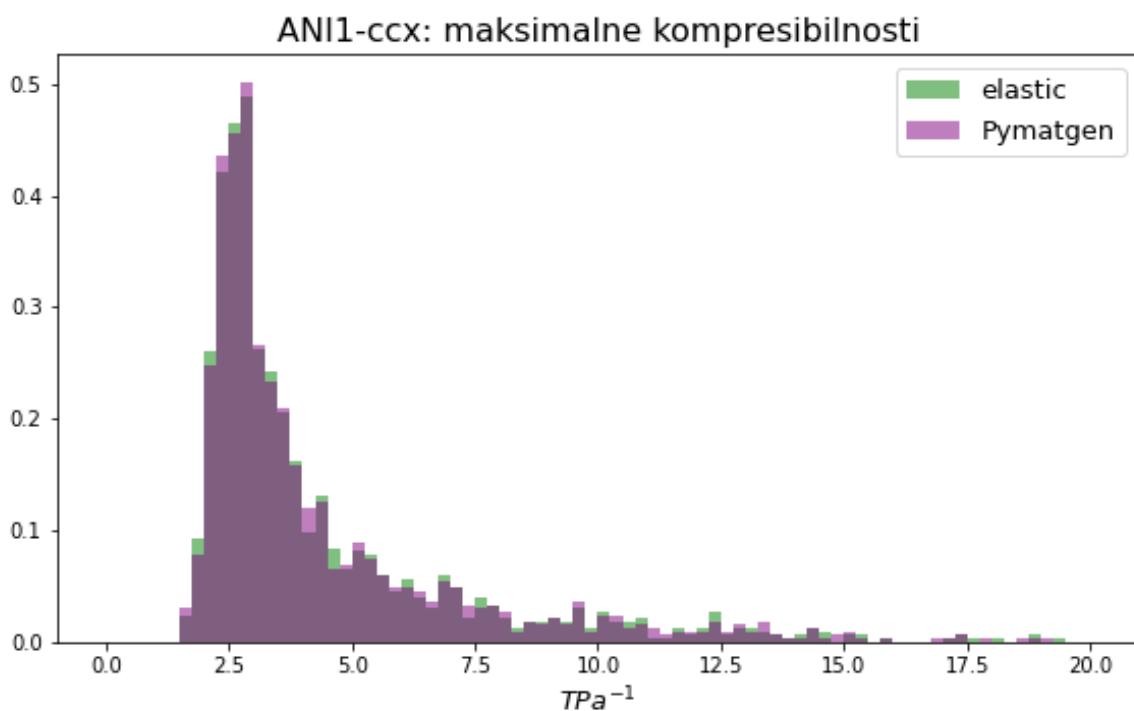
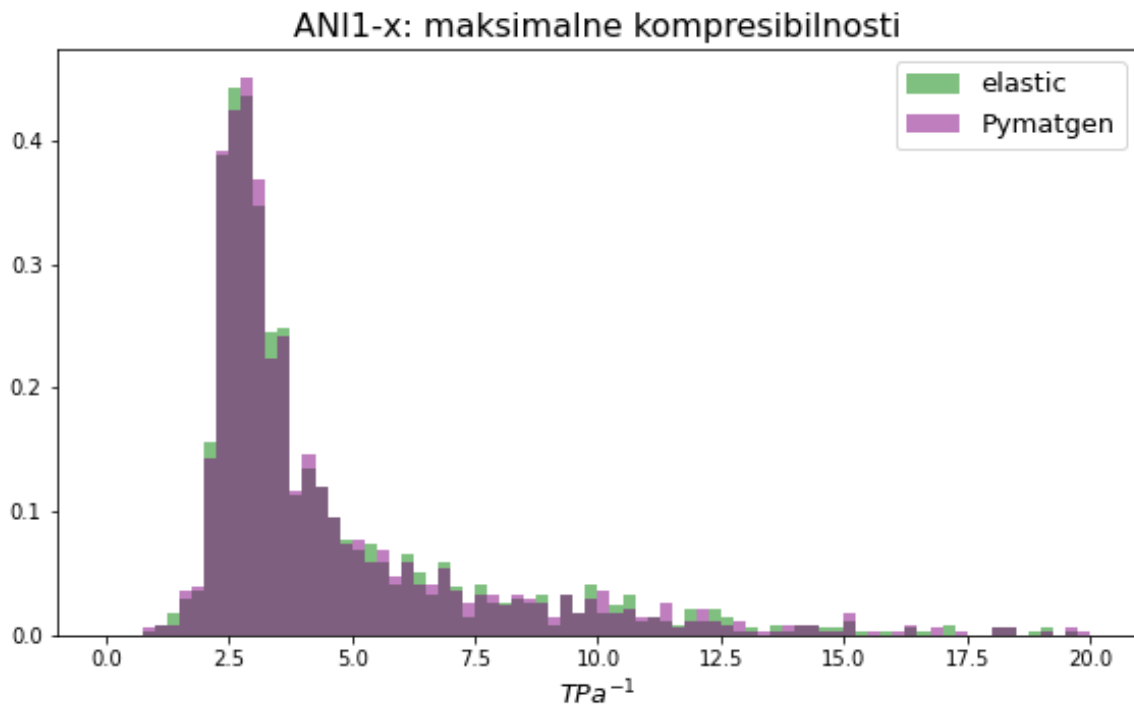
Obzirom da .cif datoteke ne sadrže eksperimentalne podatke za elastične konstante, dvije metode računanja konstanta omogućuju dva različita pristupa problemu čiji se rezultati moraju približno slagati kako bi se potvrdila točnost korištenih metoda (*Pymatgen* i *elastic*). Svakoj metodi pridružuju se posebno ANI1-x i ANI1-ccx potencijali, što znači da se za svaki materijal provode ukupno 4 računa.

Pomoću baze eksperimentalnih struktura konstruirane su još dvije nove baze relaksiranih struktura (za svaki ANI1 potencijal) uz čuvanje simetrije, isključivo za ortorompske kristale te je proveden filter za izdvajanje materijala kojima je volumen izračunat s relativnom pogreškom manjom od 15%. Konačan broj ortorompskih kristala u svakoj bazi je 1344. Za te strukture pripadnim potencijalom izračunati su elastični tenzori i glavne osi kompresibilnosti. U tablici 6.3 navedeni su udjeli materijala za koje određeni potencijal i metoda predviđaju pristustvo NLC svojstva. Za ANI1-ccx potencijal dobili su se jednaki rezultati - i *elastic* i *Pymatgen* jednako predviđaju NLC svojstvo te čine skup od 62 materijala (posljedica toga je i jednak udio NLC materijala u bazama). Za ANI1-x potencijal također se prepoznaje dobro slaganje metoda: materijali s NLC-om prema *elasticu* čine skup od 96 materijala, a prema *Pymatgenu* skup od njih 95. Presjek tih dvaju skupova sastoji se od 94 materijala.

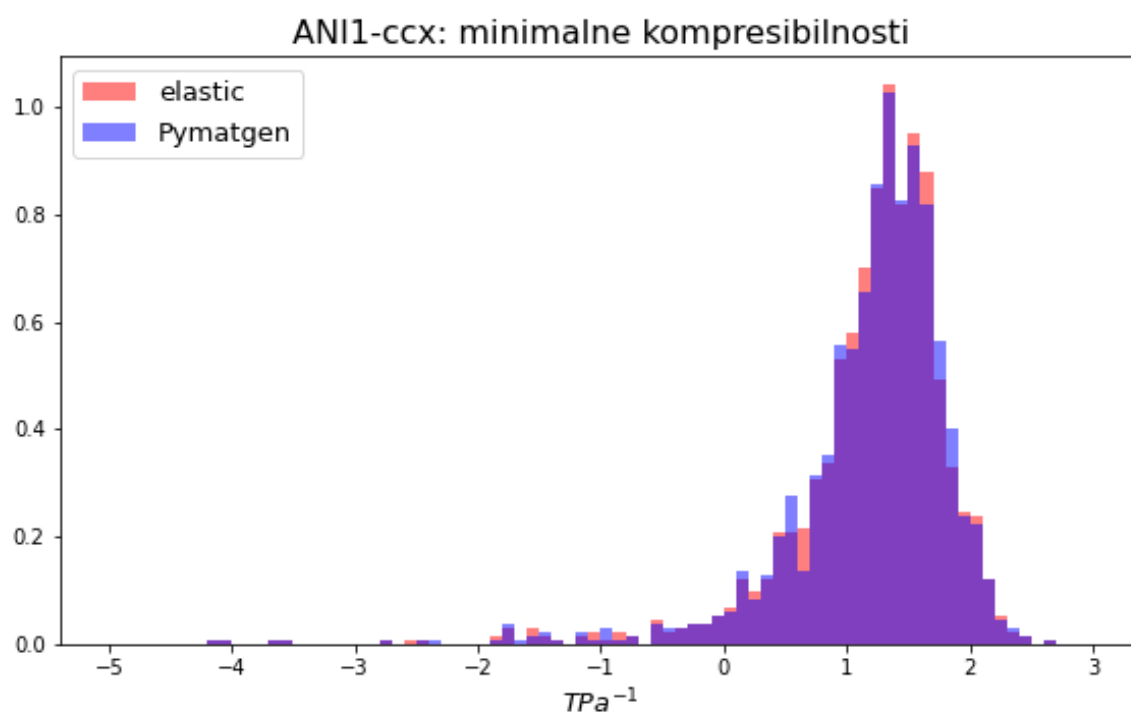
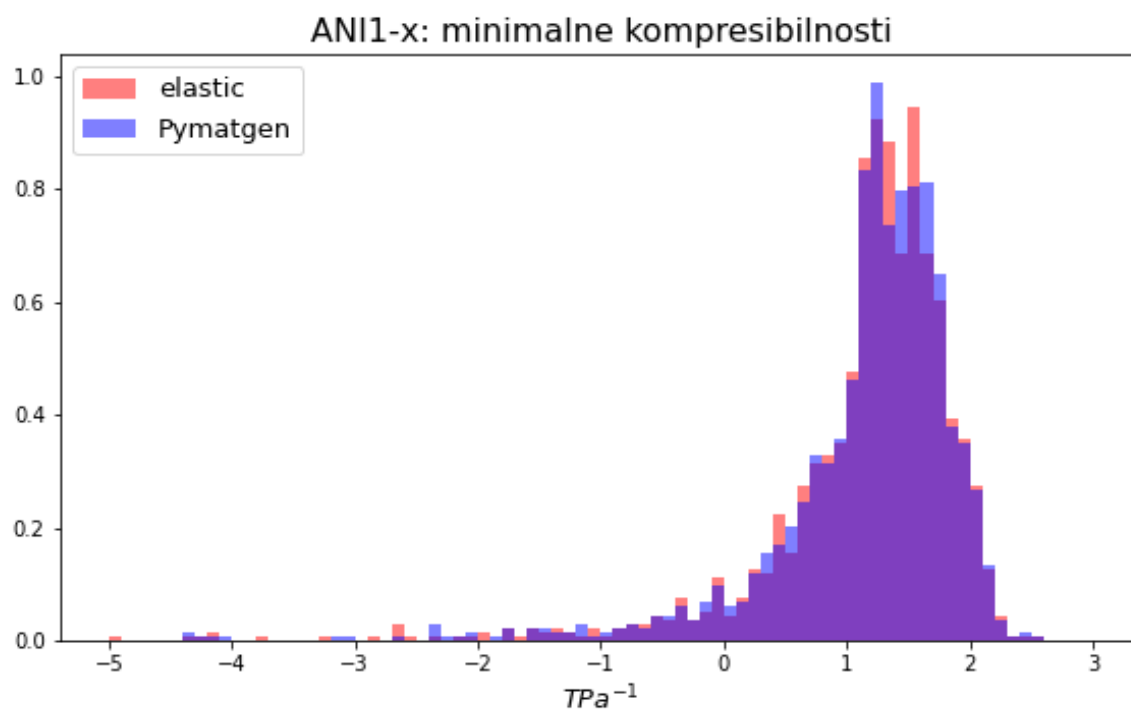
Tablica 6.3: Udio materijala za koje ANI1-x ili ANI1-ccx predviđaju NLC svojstvo pomoću metoda *elastic* ili *Pymatgen*. Rezultati za svaki potencijal dobiveni su korištenjem dvije baze (jedna za svaki potencijal) od 1344 relaksiranih struktura ortorompskih kristala. Za račun elastičnog tenzora koristio se pripadni potencijal.

	<i>ANI1x</i>	<i>ANI1ccx</i>
<i>elastic</i>	7.14%	4.61%
<i>Pymatgen</i>	7.07%	4.61%

Iskoristimo te skupove za usporedbu ANI1 potencijala. Presjek skupova materijala s NLC svojstvom prema ANI1-x i ANI1-ccx čini skup od 38 materijala. Time se od ukupno 1344 ortorompskih kristala dobio skup materijala za koji 4 različita računa predviđaju svojstvo negativne linearne kompresibilnosti. Time je udio ortorompskih kristala (koji prema tehnikama korištenima u okviru ovog diplomskog rada) pokazuju



Slika 6.8: Histogramski prikaz maksimalnih kompresibilnosti ortorompskih kristala izračunatih ANI1 potencijalima pomoću *elastic* i *Pymatgen* metoda.



Slika 6.9: Histogramski prikaz minimalnih kompresibilnosti ortorompskih kristala izračunatih ANI1 potencijalima pomoću *elastic* i *Pymatgen* metoda.

NLC svojstvo, jednak 2.83%.

Na slikama 6.8 i 6.9 prikazani su histogrami minimalnih i maksimalnih kompresibilnosti ortorompskih kristala iz baze izračunatih pomoću navedenih metoda. Iz oblika histograma vidljivo da sve 4 metode daju približno jednake rezultate za korišteni skup podataka.

## 6.4 DFT račun

Skup materijala za koji sve 4 metode predviđaju NLC svojstvo pogodan je za daljnje račune. Od početnih 42748 molekularnih kristala razmatranje se ograničilo na skup od 38 ortorompskih kristala. Za provjeru točnosti metoda koristit će se DFT račun pomoću programskog paketa *Quantum Espresso*. Kompleksnost i vrijeme izvođenja DFT računa za materijale brzo raste s povećanjem broja elektrona u sustavu. Iz tog razloga za račun su odabrana dva materijala iz spomenutog skupa s najjednostavnijom jediničnom ćelijom.

Prvi materijal nalazi se u COD bazi pod brojem 1505065 s empirijskom formulom  $C_6H_7N_3O_2$ . Molekula je prikazana na slici 6.10. Radi se spoju pod imenom 2-amino-4-nitroanilin te je i u samoj dokumentaciji navedeno da materijal pokazuje neočekivana svojstva. Eksperimentalno je utvrđeno da materijal pokazuje elektrooptička i nelinearna optička svojstva [30]. DFT računom potvrdilo se i svojstvo negativne linearne kompresibilnosti te su svi rezultati navedeni u tablici 6.4.

Drugi materijal je 4-aminofthalimid. Nalazi se u COD bazi pod brojem 2218877 s empirijskom formulom  $C_8H_6N_2O_2$  te pripada skupini n-aminofthalimida i pokazuje fotofizikalna svojstva [31]. Molekula je prikazana na slici 6.11. DFT računom također se potvrdilo svojstvo negativne linearne kompresibilnosti te su svi rezultati navedeni u tablici 6.5.

Vidljivo je da *elastic* i *Pymatgen* metode daju usporedive rezultate za isti korišteni potencijal, no same vrijednosti bitno odstupaju od DFT računa. Obzirom da su korištene metode razvijene samo kao filter za identifikaciju NLC svojstva koje se potvrdilo DFT-om u dva izolirana slučaja, može se zaključiti da su konačni rezultati zadovoljavajući.



Tablica 6.4: Tablica izračunatih vrijednosti kompresibilnosti u smjeru glavnih osi za 2-amino-4-nitroanilin.

Potencijal	Metoda	$K[TPa^{-1}]$		
		$K_1$	$K_2$	$K_3$
ANI1 – $x$	<i>elastic</i>	17.195	0.79	-2.176
	<i>Pymatgen</i>	17.381	0.775	-2.199
ANI1 – $ccx$	<i>elastic</i>	14.562	0.004	-0.038
	<i>Pymatgen</i>	14.779	-0.034	-0.038
<i>DFT</i>		36.941	1.063	-7.235

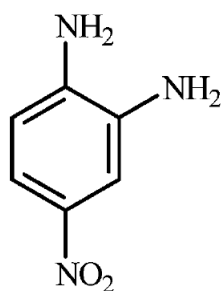
Tablica 6.5: Tablica izračunatih vrijednosti kompresibilnosti u smjeru glavnih osi za 4-aminoftalimid.

Potencijal	Metoda	$K[TPa^{-1}]$		
		$K_1$	$K_2$	$K_3$
ANI1 – $x$	<i>elastic</i>	7.532	0.852	-0.462
	<i>Pymatgen</i>	7.658	0.851	-0.517
ANI1 – $ccx$	<i>elastic</i>	11.509	1.088	-1.501
	<i>Pymatgen</i>	11.425	1.119	-1.495
<i>DFT</i>		57.024	1.362	-10.220

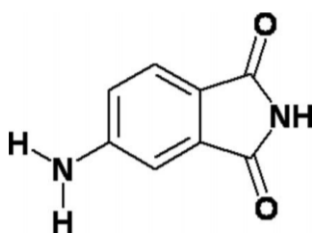
## 7 Zaključak

COD baza podataka iskoristila se za konstrukciju nove baze velikog broja molekularnih kristala sastavljenih od elemenata C, H, N i O. Filtriranjem te baze izdvojili su se materijali povoljni za simulaciju svojstava korištenjem ANI1 strojno naučenih potencijala. Korištenjem baze, testirala se točnost ANI1- $x$  i ANI1- $ccx$  potencijala uspoređivanjem simuliranih i eksperimentalno utvrđenih volumena te su postignuti zadovoljavajući rezultati. Dodatna poboljšanja postigla su se postavljanjem ograničenja za očuvanje simetrije u postupku relaksacije strukture te su se za konstrukciju baza s relaksiranim strukturama izdvojili oni materijali čiji je simulirani volumen unutar 15% vrijednosti relativne pogreške u odnosu na eksperimentalne vrijednosti. Za takve materijale provedena je simulacija elastičnih svojstava.

Elastična svojstva simulirana su računom elastičnog tenzora uz pomoć dvije različite metode, korištenjem dva različita *Python* paketa: *elastic* i *Pymatgen*. Time su, u kombinaciji s dva ANI1 potencijala, korištene 4 različite metode za račun elastičnog tenzora. Utvrdilo se da metode daju usporedive rezultate za ortorompske kristale. Izračunati tenzori omogućili su račun glavnih osi kompresibilnosti te se presjekom tih četiriju metoda iz baze od 1344 materijala identificiralo 38 materijala (ukupno 2.83%) s NLC svojstvom.



Slika 6.10: Molekula 2-amino-4-nitroanilina (preuzeto iz [30]).



Slika 6.11: Molekula 4-aminoftalimida (preuzeto iz [31]).

Obzirom da su se koristile metode za brze račune u svrhu dobivanja statistike za molekularne kristale, potrebno je dodatno provjeriti rezultate preciznijom metodom. DFT računom pomoću programskog paketa *Quantum Espresso* napravile su se provjere za dva materijala s najjednostavnijim jediničnim ćelijama za koje se predvidjelo NLC svojstvo. U oba slučaja potvrdilo se prisustvo NLC svojstva.

Iz svega navedenog zaključuje se da su ANI1 potencijali dobra zamjena za dugotrajne i komplicirane račune u svrhu simuliranja svojstava velikog broja materijala u kraćem vremenu. Ovaj rezultat otvara mogućnost za nalaženje kristala koji imaju zanimljiva svojstva.

## Literatura

- [1] Andrew Cairns and Andrew Goodwin. Negative linear compressibility. *Phys. Chem. Chem. Phys.*, 17, 02 2015.
- [2] Ray H. Baughman, Sven Stafström, Changxing Cui, and Socrates O. Dantas. Materials with negative compressibilities in one or more dimensions. *Science*, 279(5356):1522–1524, 1998.
- [3] Saulius Gražulis, Daniel Chateigner, Robert T. Downs, A. F. T. Yokochi, Miguel Quirós, Luca Lutterotti, Elena Manakova, Justas Butkus, Peter Moeck, and Armel Le Bail. Crystallography Open Database – an open-access collection of crystal structures. *J. Appl. Crystallogr.*, 42(4):726–729, Aug 2009.
- [4] Mérina K. Corpinot and Dejan-Krešimir Bučar. A practical guide to the design of molecular crystals. *Crystal Growth & Design*, 19(2):1426–1453, 2019.
- [5] Justin Smith, Olexandr Isayev, and Adrian Roitberg. Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific Data*, 4:170193, 12 2017.
- [6] Michael Nielsen. *Neural Networks and Deep Learning*.
- [7] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017.
- [8] Tobias Fink and Jean-Louis Reymond. Virtual exploration of the chemical universe up to 11 atoms of c, n, o, f: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *Journal of Chemical Information and Modeling*, 47(2):342–353, 2007. PMID: 17260980.
- [9] Tobias Fink, Heinz Bruggesser, and Jean-Louis Reymond. Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angewandte Chemie International Edition*, 44(10):1504–1508, 2005.

- [10] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28(1):31–36, February 1988.
- [11] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*, 134(7):074106, 2011.
- [12] Justin Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian Roitberg. Less is more: Sampling chemical space with active learning. *The Journal of Chemical Physics*, 148, 01 2018.
- [13] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.
- [14] Justin Smith, Benjamin Nebgen, Roman Zubatyuk, Nicholas Lubbers, Christian Devereux, Kipton Barros, Sergei Tretiak, Olexandr Isayev, and Adrian Roitberg. Outsmarting quantum chemistry through transfer learning. 01 2018.
- [15] J.F. Nye. *Physical Properties of Crystals*. Clarendon Press, 1957.
- [16] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.
- [17] Ivan Yakovkin and Peter Dowben. The problem of the band gap in lda calculations. *Surface Review and Letters (SRL)*, 14:481–487, 06 2007.
- [18] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, Aug 1980.
- [19] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Peder-son, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671–6687, Sep 1992.
- [20] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo,

A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio Jr, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29(46):465901, 2017.

- [21] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *J. Phys. Condens. Matter*, 29(27):273002, 2017.
- [22] R. Fletcher. *Practical Methods of Optimization; (2nd Ed.)*. Wiley-Interscience, USA, 1987.
- [23] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci*, 68:314–319, 2013.
- [24] Paweł T. Jochym and Codacy Badger. jochym/elastic: Maintenance release, May 2018.
- [25] Romain Gaillac, Pluton Pullumbi, and François-Xavier Coudert. ELATE: an open-source online application for analysis and visualization of elastic tensors. *Journal of Physics: Condensed Matter*, 28(27):275201, may 2016.
- [26] Kevin F. Garrity, Joseph W. Bennett, Karin M. Rabe, and David Vanderbilt. Pseudopotentials for high-throughput dft calculations. *Computational Materials Science*, 81:446 – 452, 2014.
- [27] Kristian Berland and Per Hyldgaard. Exchange functional that tests the robustness of the plasmon description of the van der waals density functional. *Physical Review B*, 89, 09 2013.

- [28] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, Jun 1976.
- [29] Andrea Corso. Elastic constants of beryllium: A first-principles investigation. *Journal of physics. Condensed matter : an Institute of Physics journal*, 28:075401, 01 2016.
- [30] Tsonko Kolev, Bojidarka Ivanova, Michael Spitteller, Heike Mayer-Figge, and William Sheldrick. 2-amino-4-nitroaniline, a known compound with unexpected properties. *The journal of physical chemistry. A*, 111:10084–9, 11 2007.
- [31] Moloy Sarkar. 4-aminophthalimide. *Acta crystallographica. Section E, Structure reports online*, 64:o1654, 08 2008.