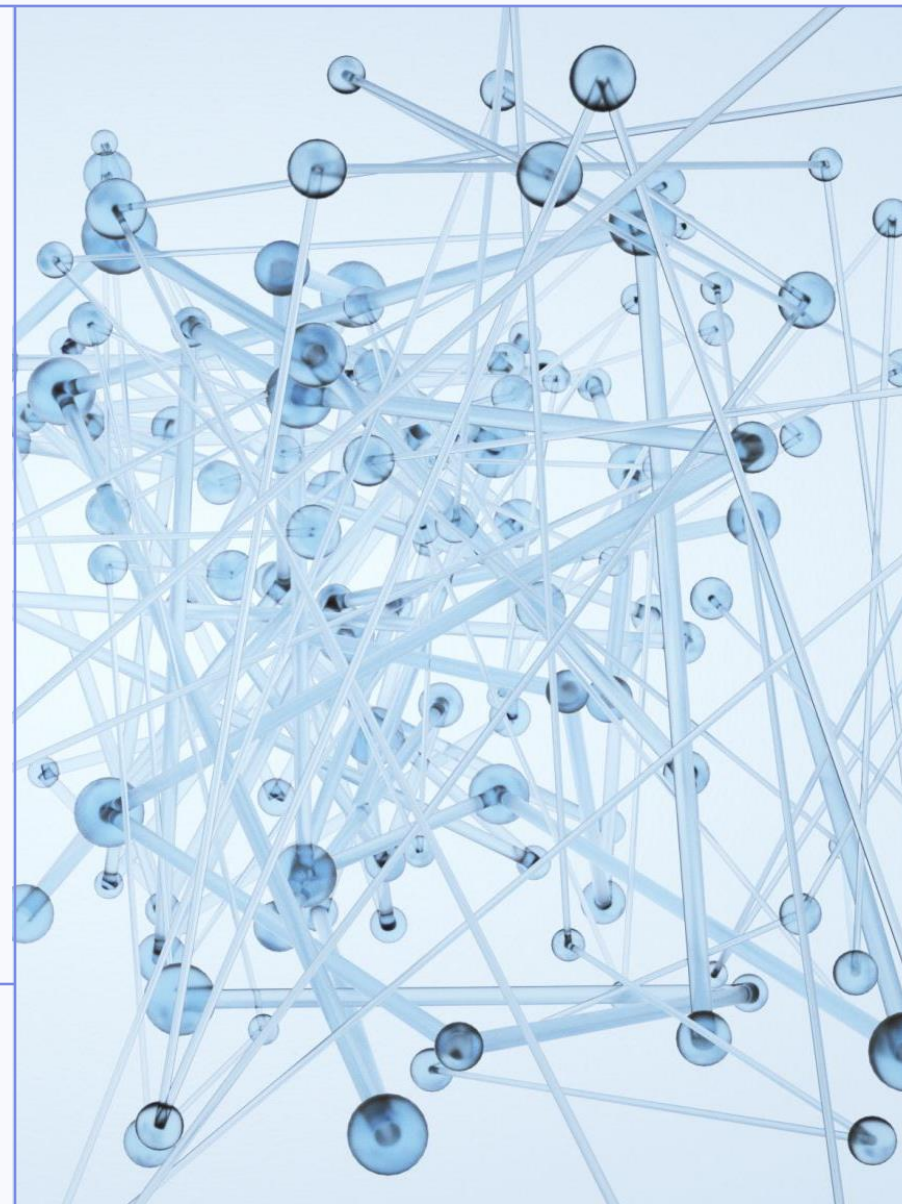


# Strojno naučeni potencijali za organske materijale

Ivan Žugec

Mentor: dr.sc Ivor Lončarić

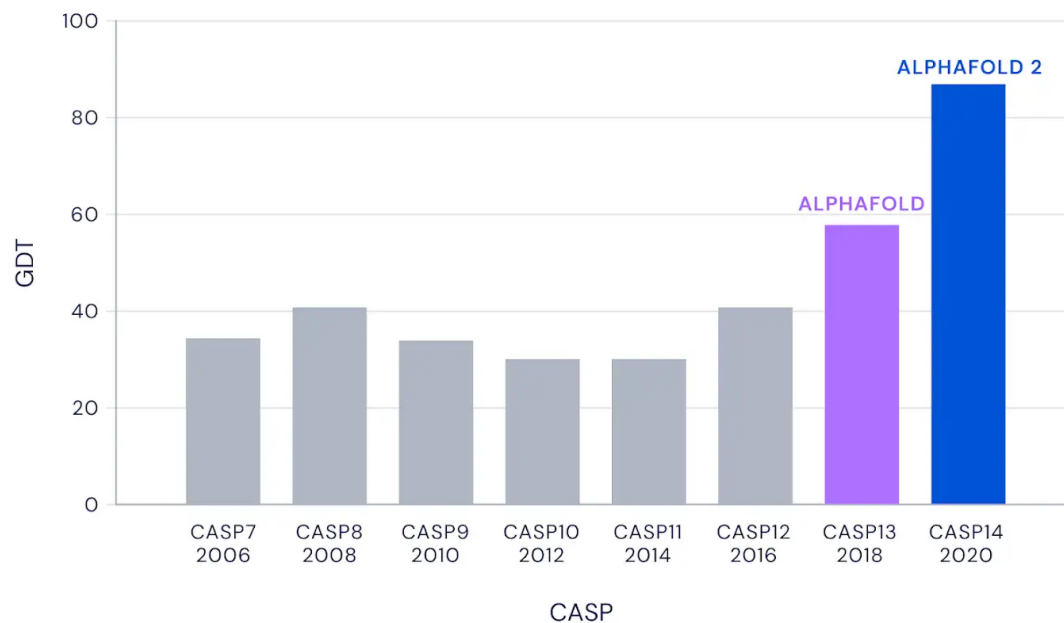


Grupa za fiziku kondenzirane tvari i statističku fiziku na Institutu Ruđer Bošković

# UVOD

- Metode strojnog učenja do sada najuspješnije u industrijskoj primjeni (npr. autonomna vozila ili prepoznavanje glasa)
- Međutim sve više nalaze svoje mjesto kao snažan alat u znanosti
- Deepmind rješenje za namatanje proteina AlphaFold 2

Median Free-Modelling Accuracy



# UVOD

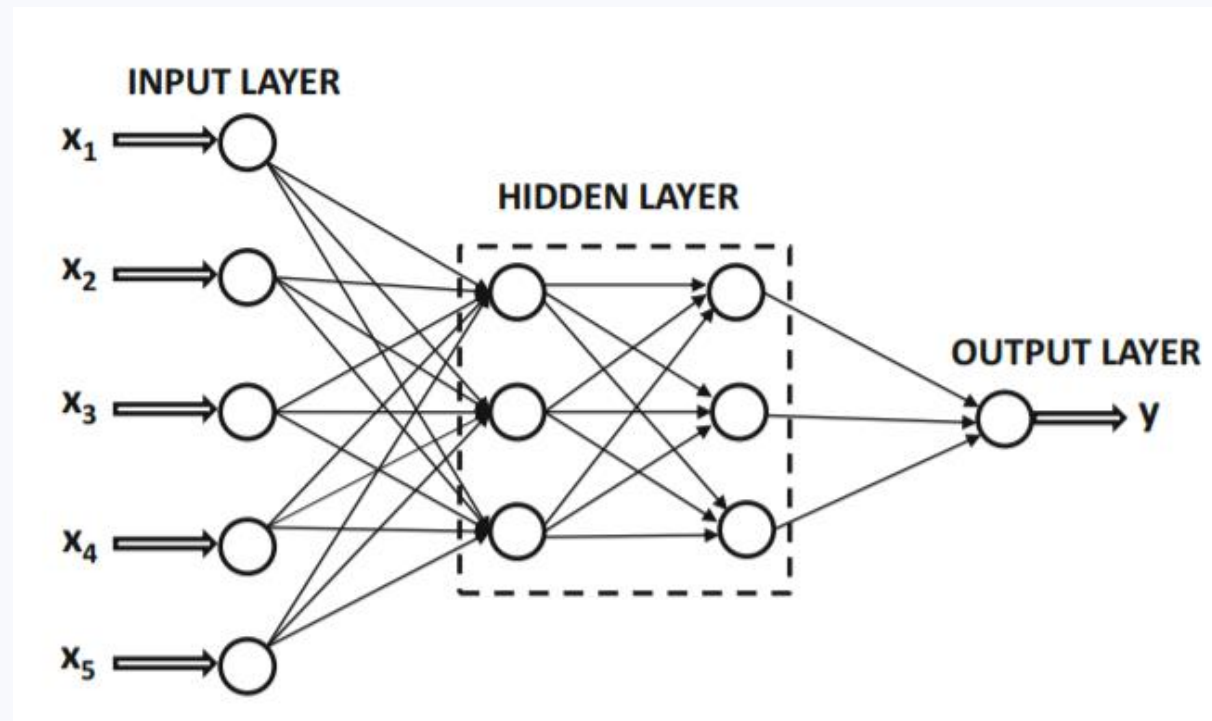
<b>DFT</b>	<b>Precizan</b>	<b>Spor</b>	<b>Primjenjiv</b>
<b>Polja sile</b>	<b>Često neprecizan</b>	<b>Brz</b>	<b>Često neprimjenjiv</b>
<b>DNN</b>	<b>Precizan</b>	<b>Brz</b>	<b>Primjenjiv</b>

# UVOD

- Neuralne mreže imaju i svoje nedostatke
- Puno hiperparametara
- Ne počiva na fizikalnim načelima
- Kompleksnost prostora konfiguracija eksponencijalno raste sa brojem različitih kemijskih elemenata u sustavu
- ANAKIN-ME (Accurate Neural network engine for Molecular Energy)

# METODE

- Neuralne mreže su računalni sustavi dijelom inspirirani radom mozga
- Vrlo fleksibilna nelinearna funkcija
- Trening set  $(X,y)$
- Funkcija gubitka



# METODE

- Vektori okoline atoma
- Svaka komponenta pruža radijalnu i angularnu informaciju okoline atoma u sustavu
- Ulazni sloj neuralnoj mreži

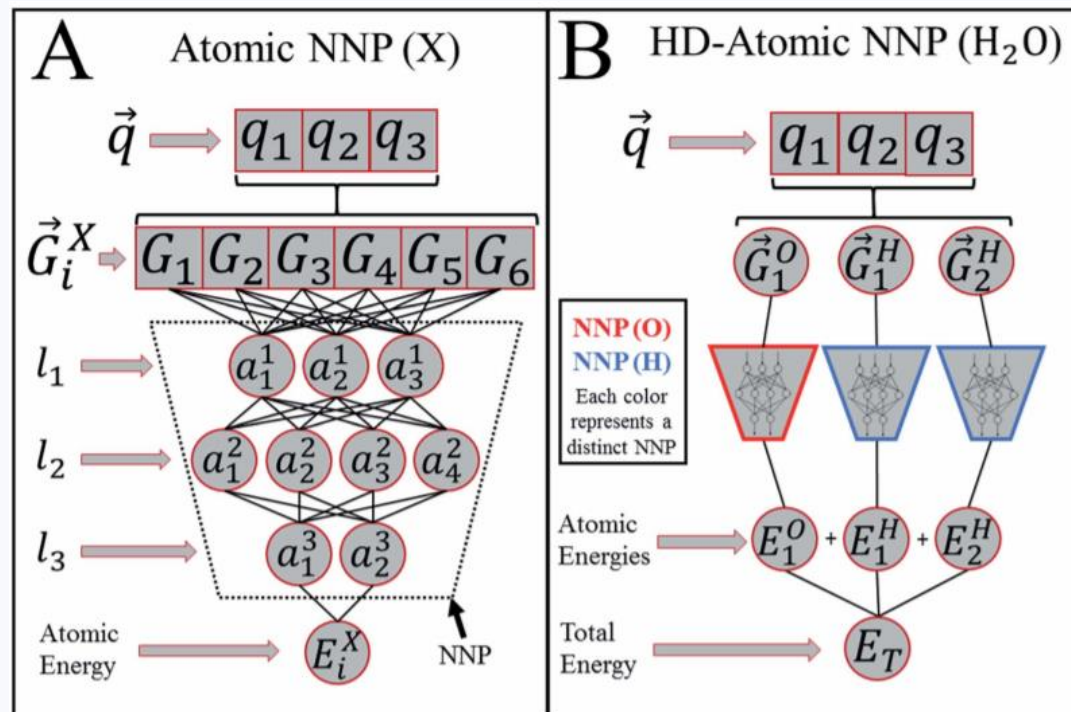
$$G_m = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{svi atomi}} (1 + \cos(\theta_{ijk} - \theta_s))^\zeta \exp \left[ -\eta \left( \frac{R_{ij} + R_{ik}}{2} - R_s \right)^2 \right] f_C(R_{ij}) f_C(R_{ik})$$

$$f_C(R_{ij}) = \begin{cases} 0.5 \cdot \cos \left( \frac{\pi R_{ij}}{R_C} \right) + 0.5 & R_{ij} \leq R_C \\ 0 & R_{ij} > R_C \end{cases}$$

# METODE

- Izlazni sloj sastoji se od jednog neurona koji računa energiju atoma
- Energija sustava tada je suma energija atoma od kojih se sastoji

$$E_{molekula} = \sum_i^{\text{svi atomi}} E_i$$



# METODE

- Osim izbora funkcije koju koristimo za ulazni sloj moramo odabrati broj skrivenih slojeva te broj neurona u svakom sloju
- ANI-1 potencijal ima piramidalnu strukturu 768:128:128:64:1
- 124 033 parametara za svaki atom

$$C(\vec{E}^{\text{ANI}}) = \tau \exp\left(\frac{1}{\tau} \sum_i (E_i^{\text{ANI}} - E_i^{\text{DFT}})^2\right)$$



# METODE

- Kreiranje pogodnog trening seta ključno je za optimalnu performansu neuralne mreže
- Trening set za ANI-1 baziran na GDB-11 bazi koja se sastoji od ~58k molekula
- Normal mode sampling
- Za svih ~17m struktura izračunati energije pomoću DFT-a
- Mnoge strukture redundantne te usporavaju generalizaciju

# METODE

- Iako se sama struktura ANI-1x potencijala ne razlikuje od ANI-1, postiže znatno bolje rezultate sa samo četvrtinom trening seta
- Takvu učinkovitost postizemo konceptom aktivnog učenja pomoću metode Query by Committee

$$\Delta E_i = |E_i^{ANI} - E_i^{DFT}| > \epsilon$$

$$\rho_i = \frac{\sigma_i}{\sqrt{N}}$$

# METODE

- Prirodna nadogradnja bila bi proširiti potencijal na primjenu više kemijskih elemenata
- ANI-2x treniran je na tri dodatna elementa u odnosu na ANI-1x
- (C, H, N, O, Cl, S, F) tvore ~90% molekula koje se nalaze u lijekovima
- Točnost usporediva sa DFT-om no ANI2-x brži za šest redova veličina

# METODE

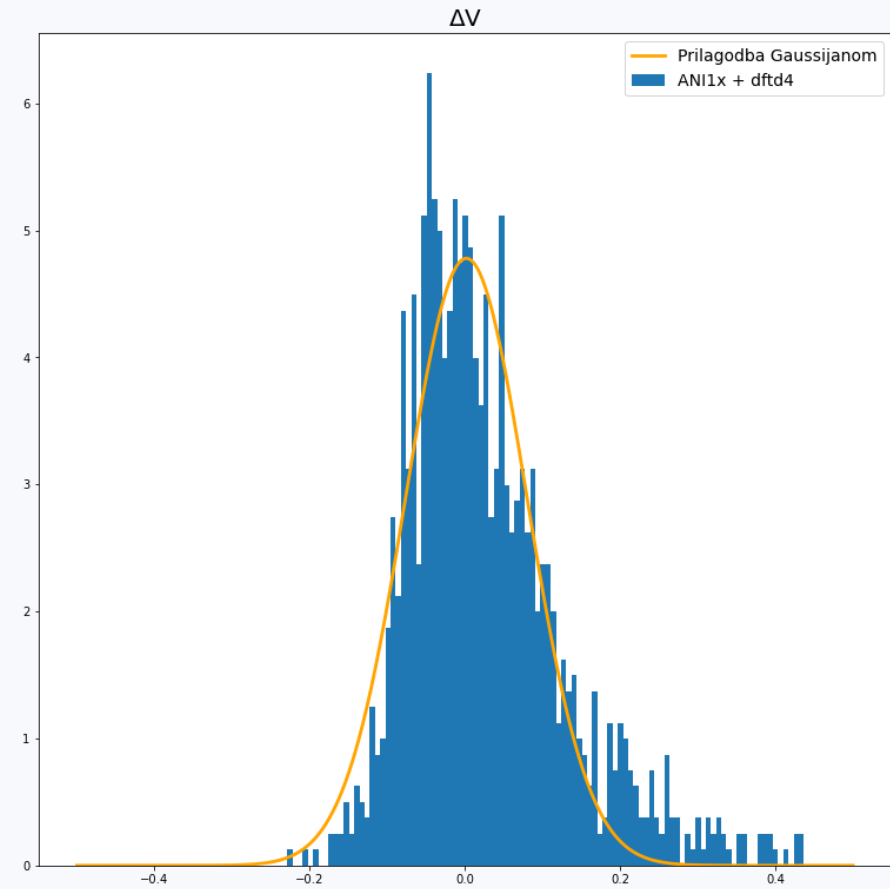
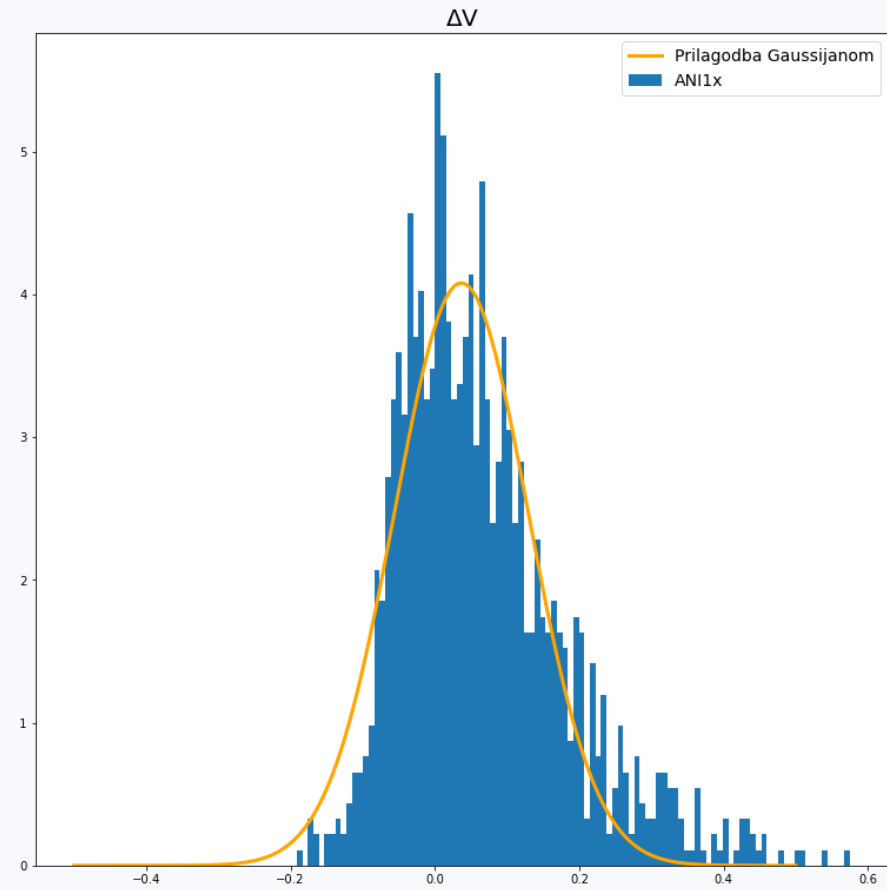
- Funkcionalni koje najčešće koristimo u DFT računima čine metodu lokalnom ili polulokalnom
- Ne uključuju dugodosežne interakcije kakva je primjerice Van der Waalsova
- Popravka u obliku aditivnog dodavanja energiji izračunatoj potencijalom
- Pokazano da model D4 ima uspjeha na različitim sustavima
- Implementaciju D4 pod imenom dftd4 koristimo kao popravku na našem potencijalu

# REZULTATI

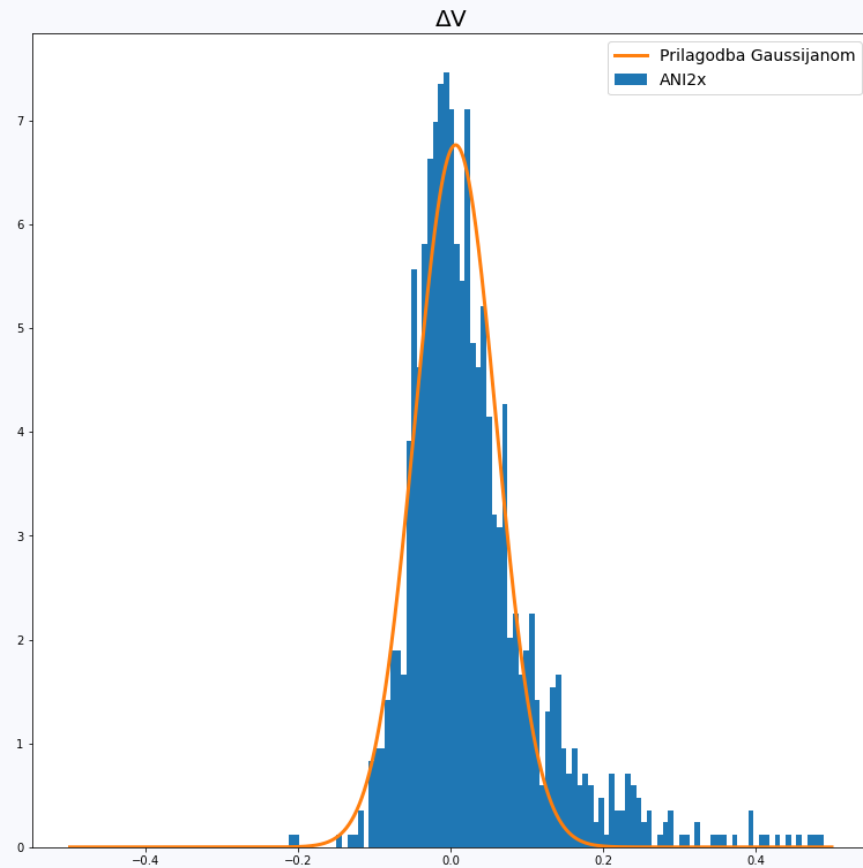
- Promatrat ćemo molekularne kristale
- Preuzeli smo ih iz COD baze
- Filtrirali smo samo one koji se sastoje od C, H, N i O atoma
- ASE (Atomic Simulation Environment)
- Relaksacija strukture za 1000 kristala
- Relativna pogreška  $\Delta V$

$$\Delta V = \frac{V_{model} - V_{exp}}{V_{exp}}$$

# REZULTATI



# REZULTATI



$\Delta V$	$\mu(10^{-3})$	$\sigma(10^{-3})$
ANI1-x	$37 \pm 4$	$93 \pm 4$
ANI2-x	$6 \pm 2$	$53 \pm 2$
ANI1-x+dftd4	$2 \pm 3$	$78 \pm 3$

# ZAKLJUČAK

- U ovome radu provjerili smo koliko se dobro potencijali ANI metodologije trenirani na molekulama generaliziraju na molekularne kristale
- Predikcija srednjih vrijednosti volumena ćelija ne razlikuju se za više od 4% od eksperimentalnih rezultata
- Treniranjem potencijala na molekulama sa težim atomima dobivamo bolju generalizaciju na molekularne kristale
- Pokazali smo značajno unaprijeđenje prediktivne moći dodavanjem dugodosežnih Van der Waalsovih interakcija



**HVALA NA PAŽNJI**