Molecule Unfolding with Quantum Annealing

HPC and Quantum Computing 2020

15 Dec 2020

Riccardo Mengoni, Kevin Mato, Daniele Ottaviani, Gianluca Palermo
Molecular Docking for Virtual Screening

Molecular docking is a method to calculate the preferred position and shape of one molecule to a second when bound to each other.

*Geometrical approaches require a shape-bias removal.*

The initial shape impacts the final docking quality. This removal takes as much as the docking.

Induced by SMILE-to-3D tools.
Problem Definition

**Objective:** find the unfolded torsion configuration that maximizes the molecular volume, or equivalently, that maximizes the distances between fragments.

\[ \theta = [\theta_1, \ldots, \theta_M] \]
Problem Definition

To each torsion is associated is a rotation matrix $R$.

\[ \hat{a}(\theta_1) = R(\theta_1)\vec{a}_0 \]

\[ \hat{a}(\theta_1, \theta_2) = R(\theta_2)R(\theta_1)\vec{a}_0 \]
Betweenness centrality:

\[ g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \]

Rotables Influence set:

\[ E_R = \text{Rotable bonds}; \]
\[ E_{C_a,a_k} = \text{Bonds on the shortest path } \sigma_{C_a,a_k} \]
\[ I_S = E_{C_a,a_k} \cap E_R \]
Objective: find the unfolded torsion configuration

\[ [\theta_1^{unfold}, \ldots, \theta_M^{unfold}] \]

that maximizes the sum of distances \( D_{ab}(\theta) \) between fragments \( a \) and \( b \)

\[
D(\theta) = \sum_{a,b} D_{ab}(\theta)^2
\]

where \( D_{ab}(\theta)^2 = \| \vec{a}_0 - R(\theta)\vec{b}_0 \|^2 \)
Constructing the High-order Unconstrained Binary Optimization (HUBO) problem

Consider a discretization of the torsion angle $\theta_i$ into $d$ possible values

$$\theta_i = [\theta_i^1, \theta_i^2, \theta_i^3, \ldots, \theta_i^d]$$

And introduce a binary variable $x_{ik}$ with $1 \leq k \leq d$, such that

$$x_{ik} = \begin{cases} 1 & \text{if } \theta_i = \theta_i^k; \\ 0 & \text{otherwise.} \end{cases}$$

with the constraint

$$\sum_{k=1}^{d} x_{ik} = 1$$

With such encoding, the rotation matrix $R(\theta_i)$ associated the torsion angle $\theta_i$ becomes a function of all the binary variables $x_{ik}$ needed to represent the angle $\theta_i$

$$R(\theta_i) = R(x_{i1}, x_{i2}, \ldots, x_{id})$$
The **general form of the HUBO** optimization function is

\[
O(x_{ik}) = A \sum_i \left( \sum_{k=1}^{d} x_{ik} - 1 \right)^2 - \sum_{a,b} D_{ab}(\theta)^2
\]

where the pairwise **distances** are expressed using the binary variables

\[
D_{ab}(\theta)^2 = \| \vec{a}_0 - R(\theta)\vec{b}_0 \|^2
\]

In general, if \( D_{ab}(\theta) \) depends on \( m \) torsions, \( D_{ab}(\theta) \) contains terms up to the \( m \)-th order, hence the **highest order in the HUBO is** \( 2m \)
In order to obtain a **precision of** $\Delta \theta_i$, the **number of variables** needed for each

$$d = \frac{2\pi}{\Delta \theta_i} = \frac{2\pi}{\theta_i^{k+1} - \theta_i^k}$$

Given a molecule with $M$ **torsions**, the total **number of binary variables** $x_{ik}$ in the HUBO

$$n = d \times M = \frac{2\pi}{\Delta \theta_i} \times M$$

82 atoms - 40 fragments - 20 torsions
Angle subsampling effect on the unfolding degradation

Granularity Degradation Effect 15%

Unfolded vs Initial

AVG Volume gain 17%
HUBO Problem Structure at $\triangle \theta_i = \pi/4$

**HUBO linear terms**

- Number of torsionals: 2, 4, 5, 6, 8, 10
- Values: 16, 32, 40, 48, 64, 80

**HUBO high order terms (Log-scale)**

- Number of torsionals: 2, 4, 5, 6, 8, 10
- Values: 296, 6745, 65936, 196282, 2028178, 12187955
HUBO Problem Approximation

Delete HUBO terms below a certain threshold $T$:

- **Speed up** the construction of the HUBOs;
- **Speed up** the transformation of HUBOs into QUBOs (done via dimod.make_quadratic);
- Make problem solvable with **DW_2000Q and Advantage**

---

**Full HUBO VS Approximated HUBOs**

<table>
<thead>
<tr>
<th>Number of torsionals</th>
<th>Full HUBO</th>
<th>ApproxHUBO T=100</th>
<th>ApproxHUBO T=1000</th>
<th>ApproxHUBO T=1500</th>
<th>ApproxHUBO T=500</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>296</td>
<td>353</td>
<td>447</td>
<td>352</td>
<td>281</td>
</tr>
<tr>
<td>4</td>
<td>6745</td>
<td>65936</td>
<td>196282</td>
<td>4599</td>
<td>12187955</td>
</tr>
<tr>
<td>5</td>
<td>65936</td>
<td>196282</td>
<td>2028178</td>
<td>265142</td>
<td>8566438</td>
</tr>
<tr>
<td>6</td>
<td>196282</td>
<td>2028178</td>
<td>265142</td>
<td>8566438</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2028178</td>
<td>265142</td>
<td>8566438</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8566438</td>
<td>265142</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From ApproxHUBOs to QUBOs

**QUBO linear terms**

- Full Hubo
- ApproxHUBO $T=100$
- ApproxHUBO $T=500$
- ApproxHUBO $T=1000$
- ApproxHUBO $T=1500$

**Number of binary variables**
- $1000$
- $100$
- $10$

**Number of torsionals**
- $2$
- $4$
- $5$
- $6$
- $8$
- $10$

**QUBO quadratic terms**

- **Number of quadratic terms**
  - $1000$
  - $10000$

- **Number of torsionals**
  - $100$
  - $1000$
  - $10000$

- $38$
- $54$
- $80$
- $58$
- $86$
- $371$
- $774$
- $143$
- $5324$
- $6566$
- $302$
- $364$
- $504$
- $323$
- $337$
- $618$
Embeddings DW_2000Q_6 & Advantage

Number of physical qubits variables

AVG maximum chain length

Number of torsionals

Number of torsionals

![Bar chart showing the comparison between Advantage and DW_2000Q_6 for the number of physical qubits variables and AVG maximum chain length.](chart.png)
Scaling DW_2000Q_6, Advantage & Classical

Optimal Configuration Volume

- **DW_2000Q_6**
- **Advantage**
- **Classical**

Volume vs. Number of torsionals:

- DW_2000Q_6:
  - 2 torsionals: 361
  - 4 torsionals: 853
  - 5 torsionals: 853
  - 6 torsionals: 1471
  - 8 torsionals: 2381

- Advantage:
  - 2 torsionals: 361
  - 4 torsionals: 853
  - 5 torsionals: 853
  - 6 torsionals: 1496
  - 8 torsionals: 2139

- Classical:
  - 2 torsionals: 361
  - 4 torsionals: 853
  - 5 torsionals: 1471
  - 6 torsionals: 1793
  - 8 torsionals: 2856

Runtime scaling:

- N-times:
  - 2 torsionals: 1.00
  - 4 torsionals: 1.00
  - 5 torsionals: 1.00
  - 6 torsionals: 0.95
  - 8 torsionals: 1.50

Number of torsionals:

- Volume:
  - 2 torsionals: 361
  - 4 torsionals: 853
  - 5 torsionals: 853
  - 6 torsionals: 1496
  - 8 torsionals: 2139

- N-times:
  - 2 torsionals: 1.00
  - 4 torsionals: 1.00
  - 5 torsionals: 1.00
  - 6 torsionals: 0.95
  - 8 torsionals: 1.50

CINECA

POLITECNICO MILANO 1863
Conclusions

• We tackled the problem of molecular unfolding.

• New HUBO and then QUBO formulations that can be solved on DWAVE quantum annealers.

• We have observed that by increasing the terms cut-off threshold as the problem size becomes bigger, it becomes possible to embed formulations that couldn’t be otherwise.

• Embedding our problems on Advantage, compared to the DW2000Q, cost 52% less in terms of physical qubits and with chains 49.7% shorter.

• Advantage performed better in terms of quality, since the percentage difference of Advantage and DW2000Q is 5.91%.

• DW2000Q has runtimes in average 2.46 times those of Advantage, and the runtimes growth in average of DW2000Q is 2.68 times that of Advantage, when we increase the size of the problem.
Thanks.