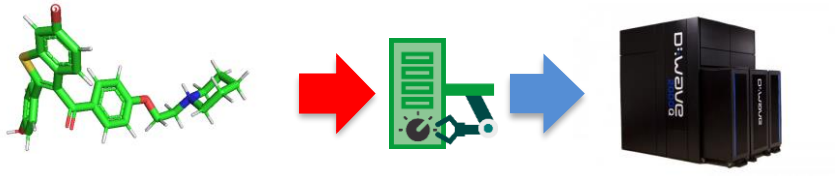
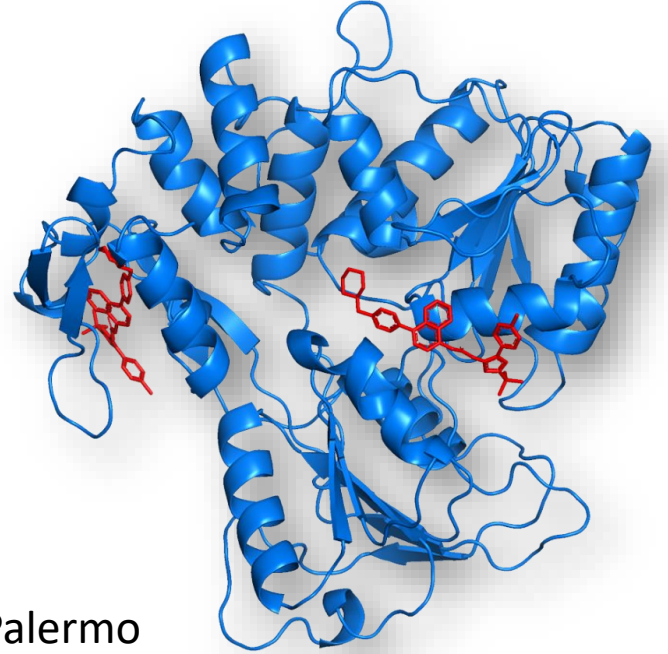


POLITECNICO
MILANO 1863

CINECA



Riccardo Mengoni, Kevin Mato, Daniele Ottaviani, Gianluca Palermo

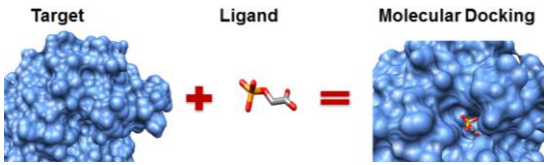
Molecule Unfolding with Quantum Annealing

HPC and Quantum Computing 2020

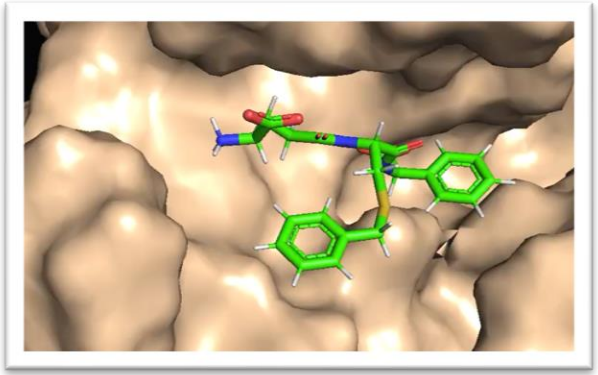
15 Dec 2020

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.

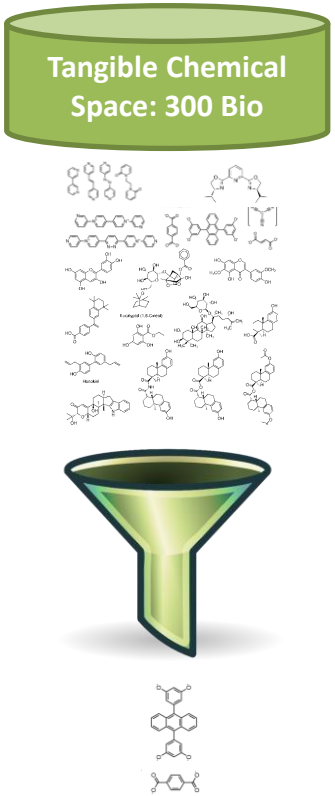


Geometrical approaches require a shape-bias removal.



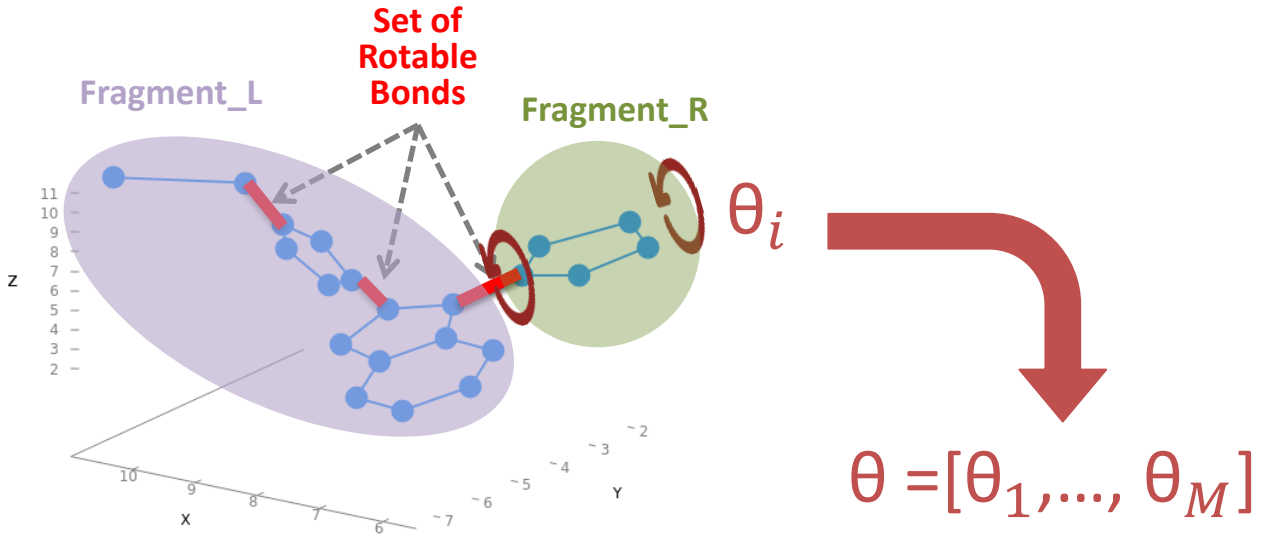
The initial shape impacts the final docking quality. This removal takes as much as the docking.

Induced by SMILE-to-3D tools.



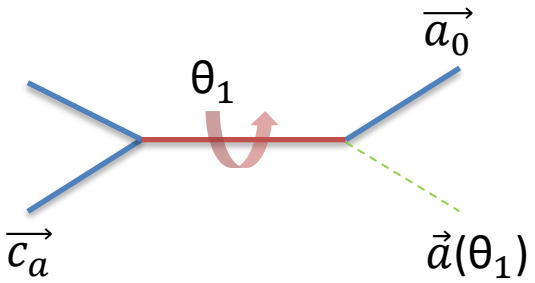
Problem Definition

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

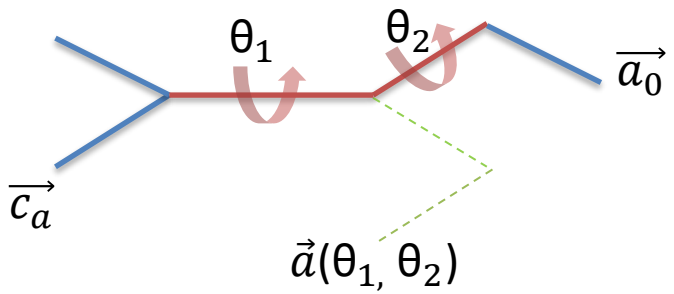


Problem Definition

To each torsion is associated is a rotation matrix R.



$$\vec{a}(\theta_1) = R(\theta_1)\overrightarrow{a_0}$$



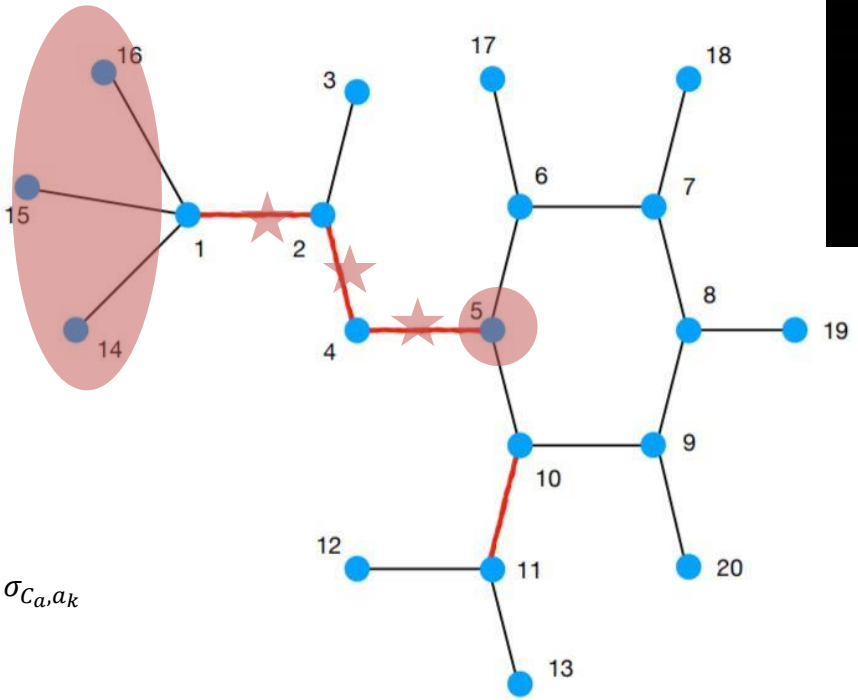
$$\vec{a}(\theta_1, \theta_2)=R(\theta_2)R(\theta_1)\overrightarrow{a_0}$$

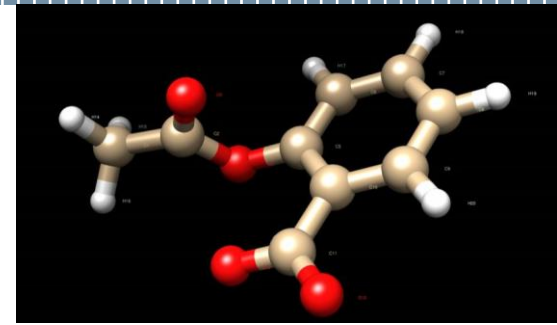
CINECA POLITECNICO MILANO 1863

Molecule Unfolding

Betweenness centrality:

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





Rotables Influence set:

$E_R = Rotable\ bonds;$
$E_{C_a,a_k} = Bonds\ on\ the\ shortest\ path\ \sigma_{C_a,a_k}$

$$\mathrm{I_S} = \mathrm{E}_{C_a,a_k} \cap E_R$$

CINECA POLITECNICO MILANO 1863

Combinatorial Optimization Problem Definition

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** θ_i **into** d **possible values**

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

And introduce a **binarv variable** x_{ik} with 1 ≤ k ≤ 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** θ_i becomes **a function of** all the **binary variables** x_{ik} needed to represent the angle θ_i

$$R(\theta_i) = R(x_{i1}, x_{i2}, ..., x_{id})$$

Constructing the High-order Unconstrained Binary Optimization (HUBO) problem

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 $\boldsymbol{D_{ab}(\theta)}$ **depends on *m* torsions**, $D_{ab}(\theta)$ contains terms up to the *m*-th order, hence the **highest order in the HUBO is *2m***

HUBO Problem Structure

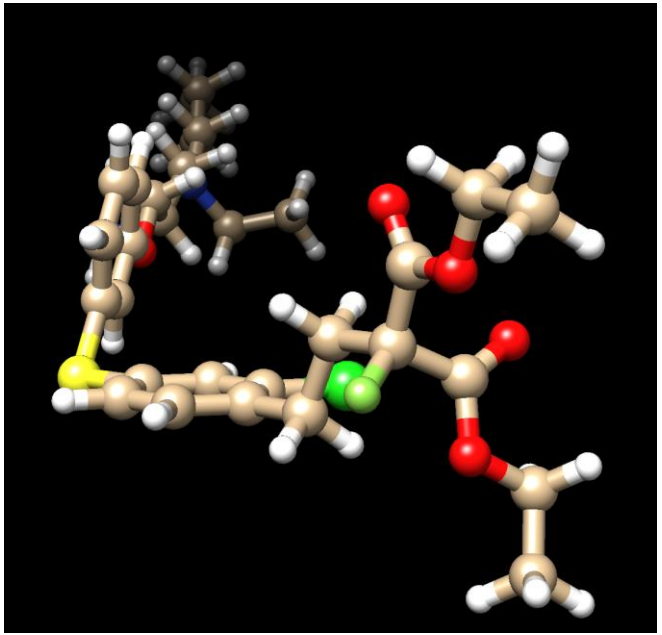
In order to obtain a **precision of $\Delta\theta_i$**, the **number of variables** needed **for each**

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

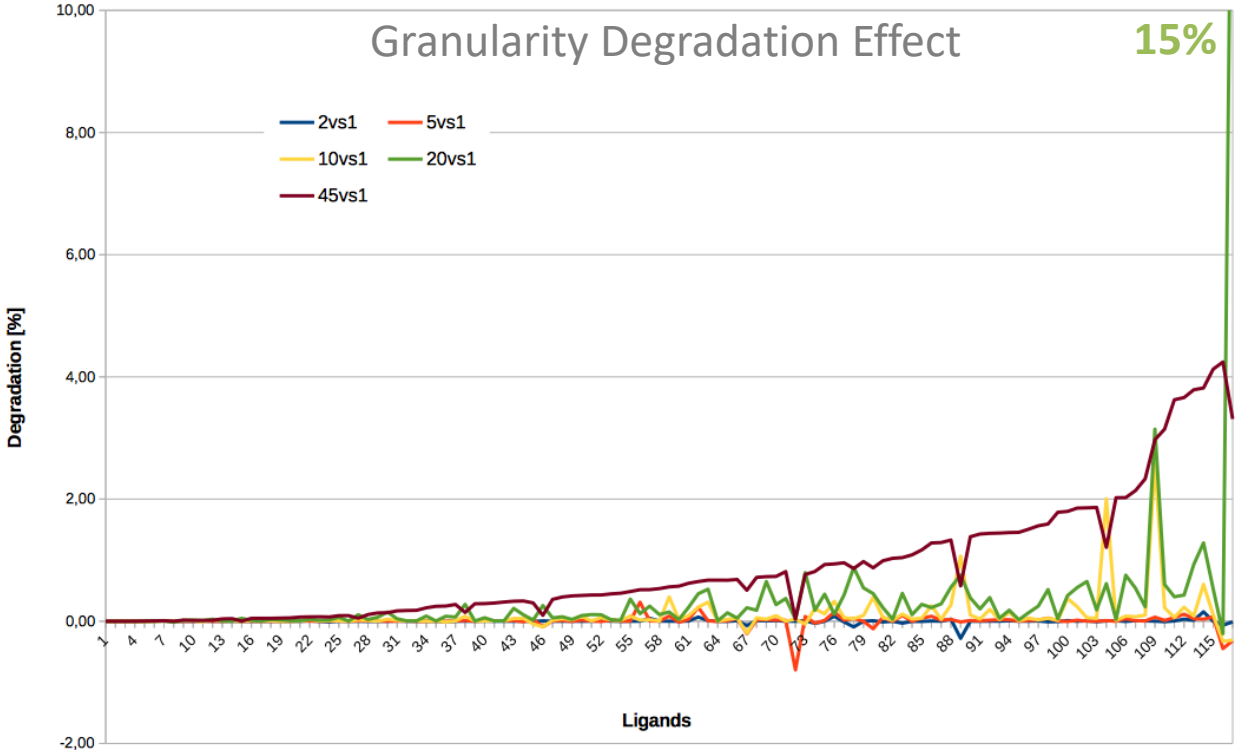
Given a molecule with ***M*** **torsions**, the total **number of binary variables x_{ik}** in the HUBO

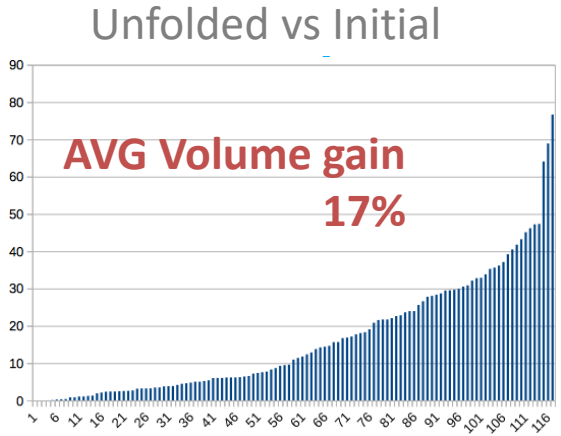
$$n = d \times M = \frac{2\pi}{\Delta\theta_i} \times M$$

82 atoms - 40 fragments - 20 torsions



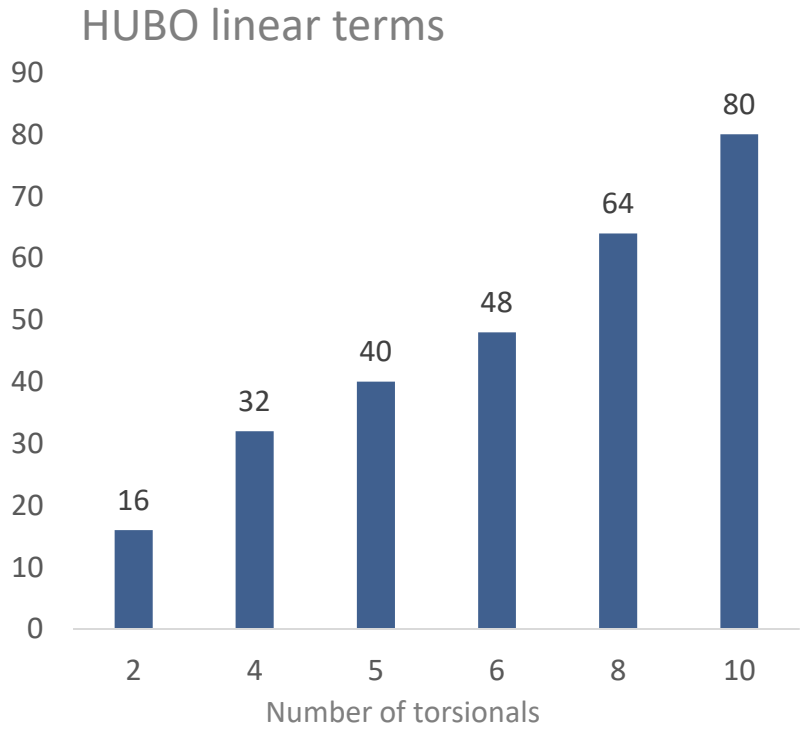
Angle subsampling effect on the unfolding degradation

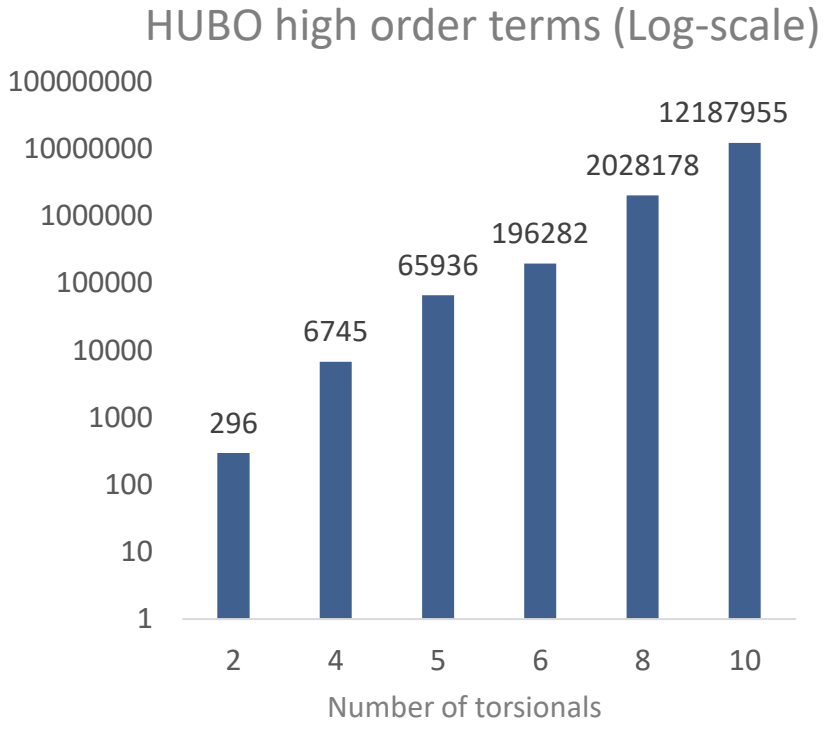




CINECA POLITECNICO MILANO 1863

HUBO Problem Structure at $\Delta\theta_i = \pi/4$



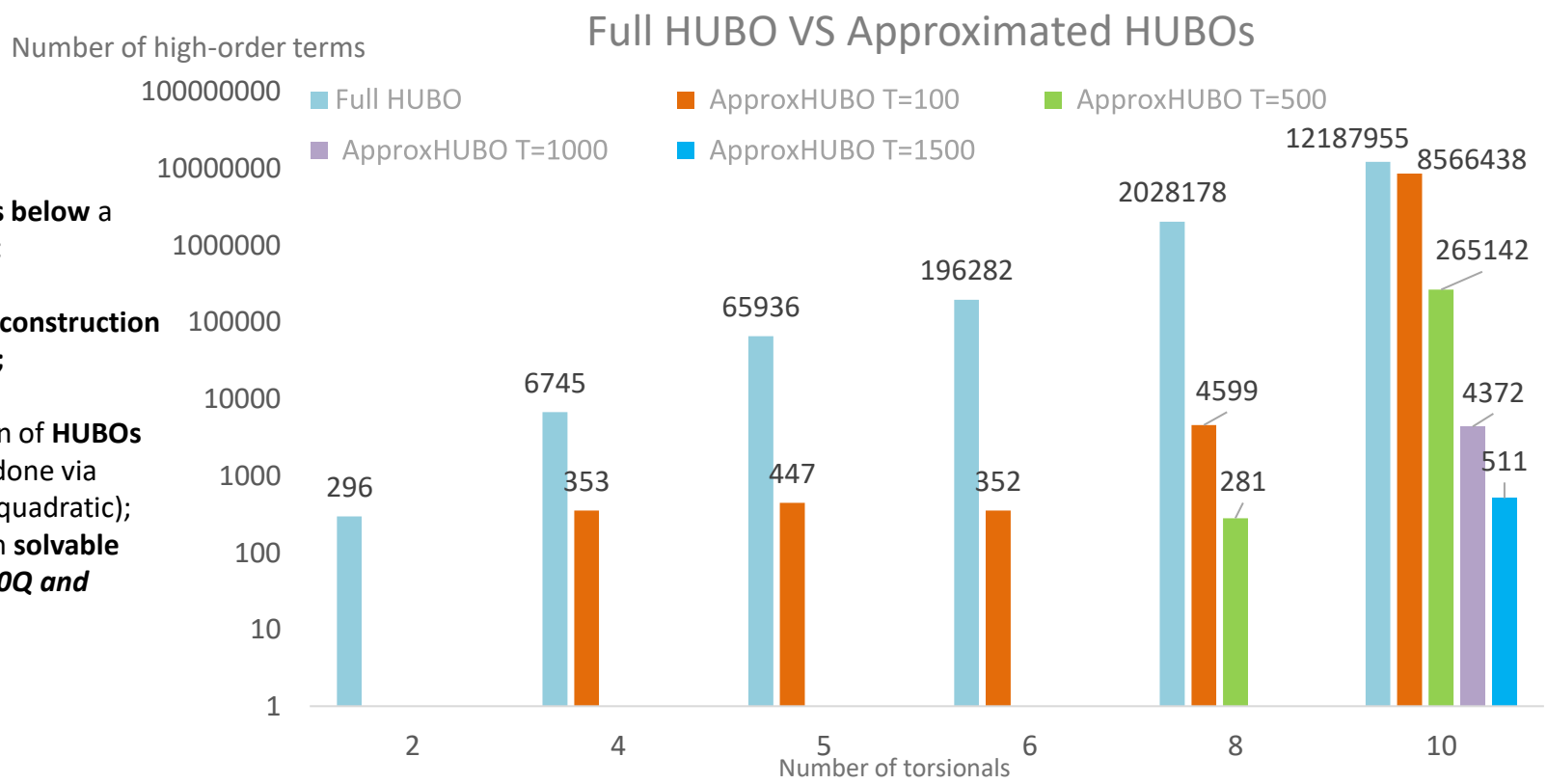


CINECA POLITECNICO MILANO 1863

HUBO Problem Approximation

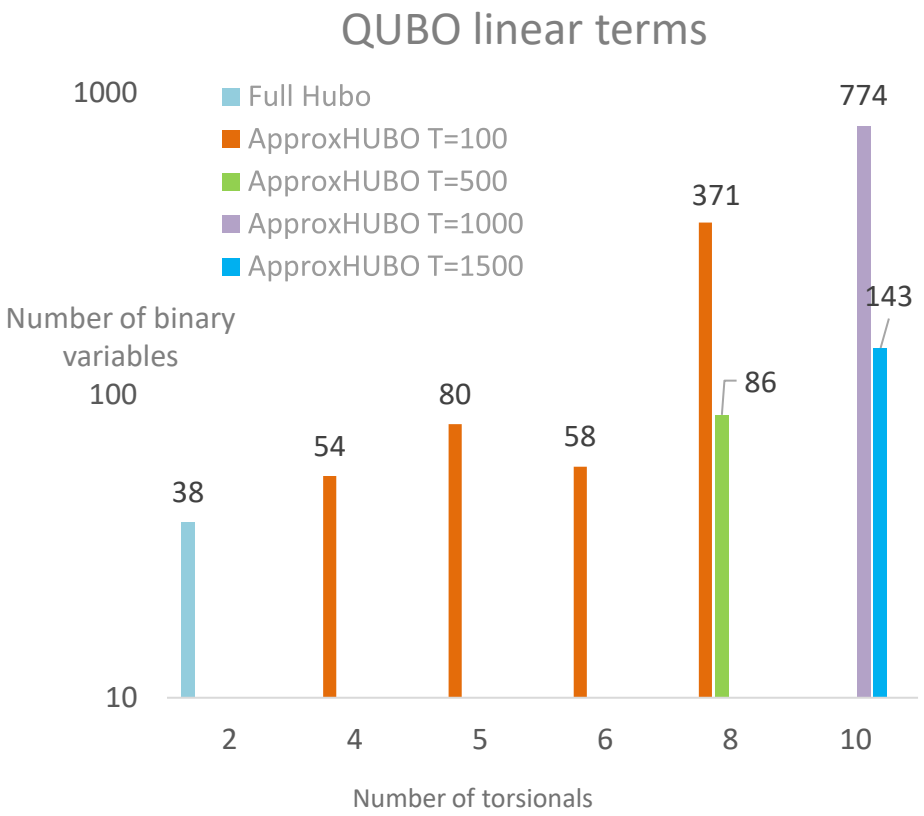
Delete HUBO terms below a certain **threshold T**:

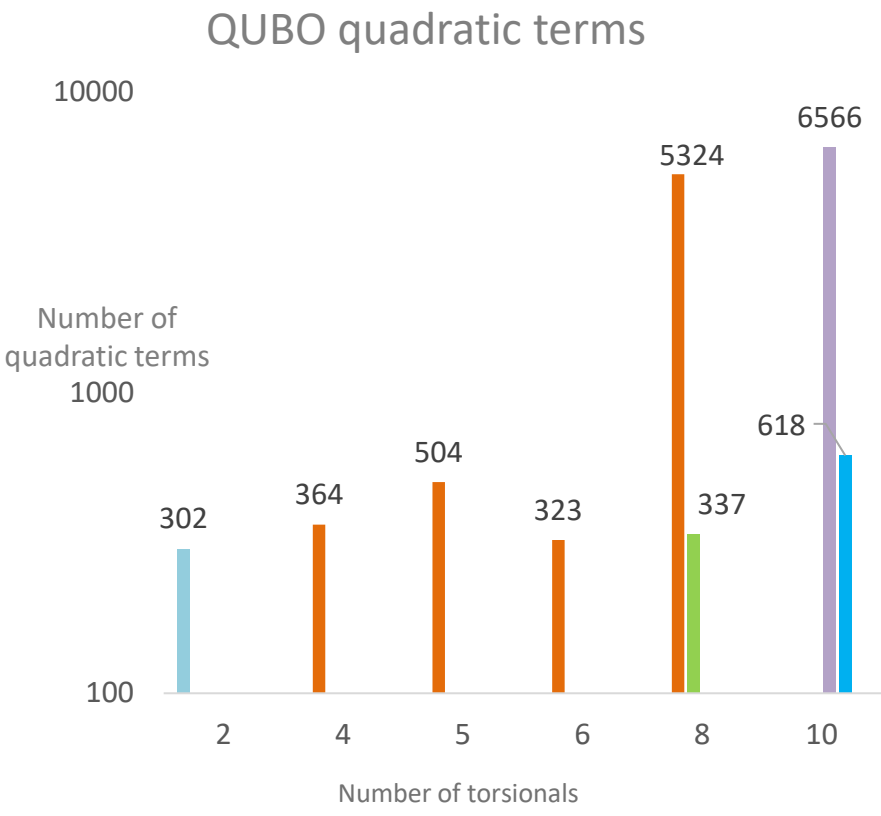
- **Speed up** the **construction** of the **HUBOs;**
- **Speed up** the transformation of **HUBOs into QUBOs** (done via dimod.make_quadratic);
- Make problem **solvable** with ***DW_2000Q and Advantage***



CINECA POLITECNICO MILANO 1863

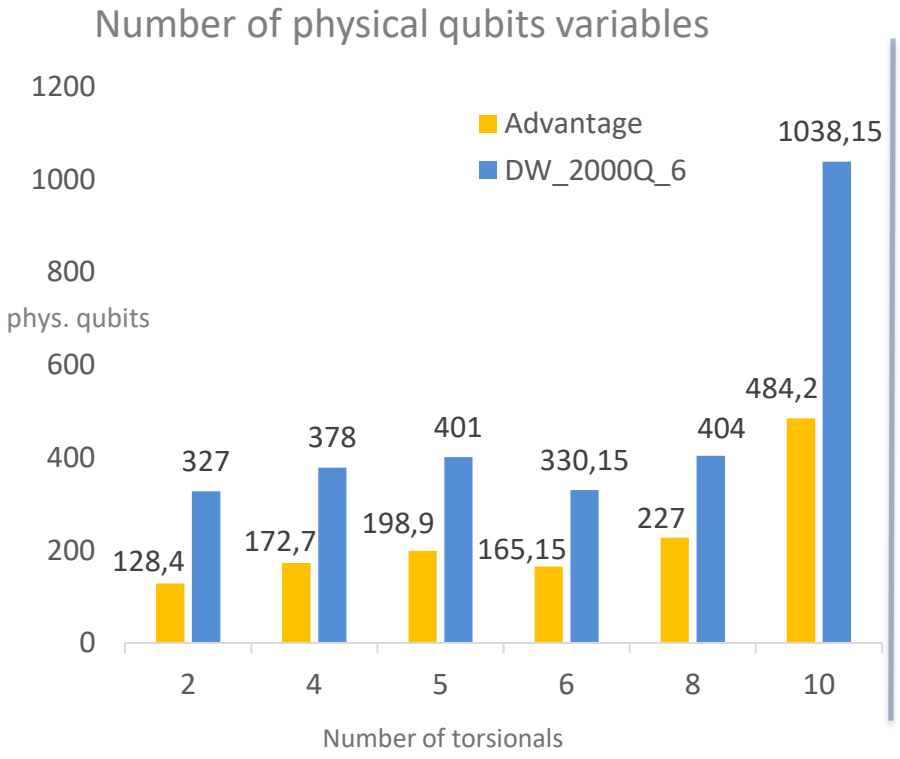
From ApproxHUBOs to QUBOs

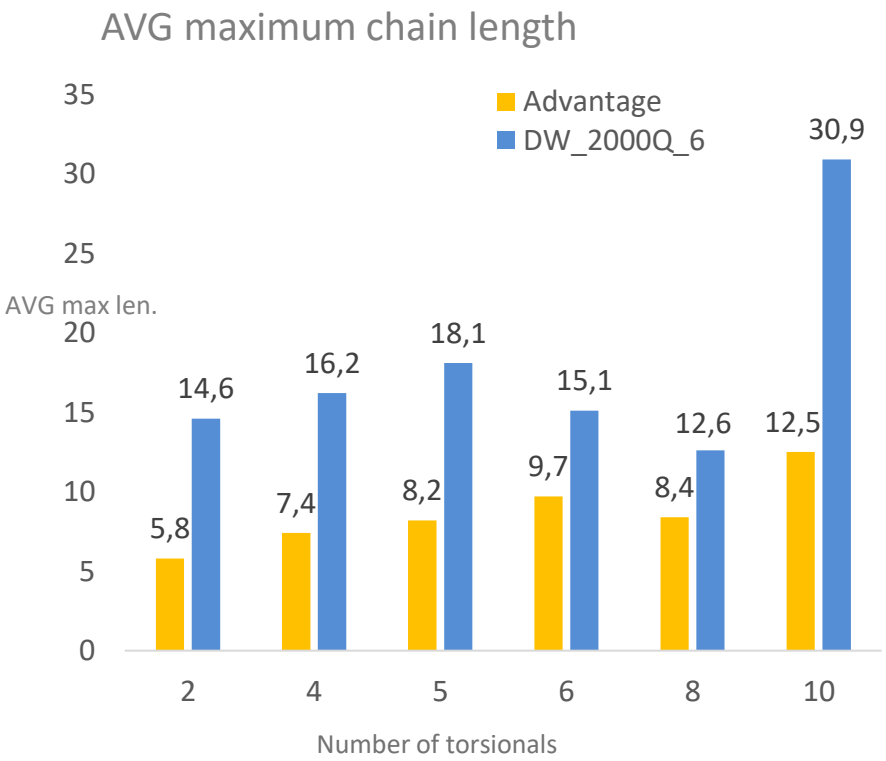




CINECA POLITECNICO MILANO 1863

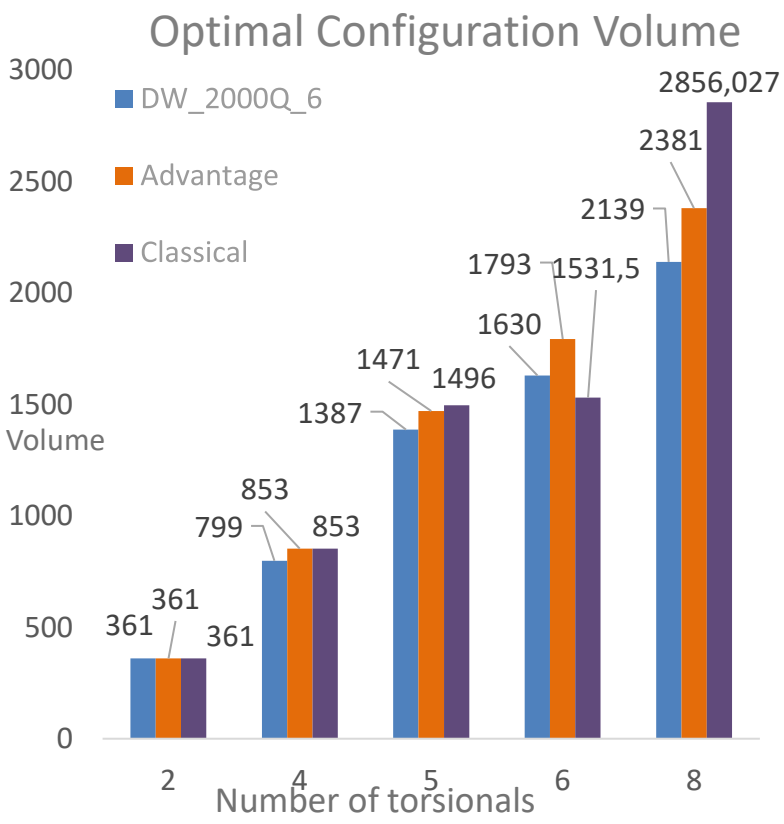
Embeddings DW_2000Q_6 & Advantage

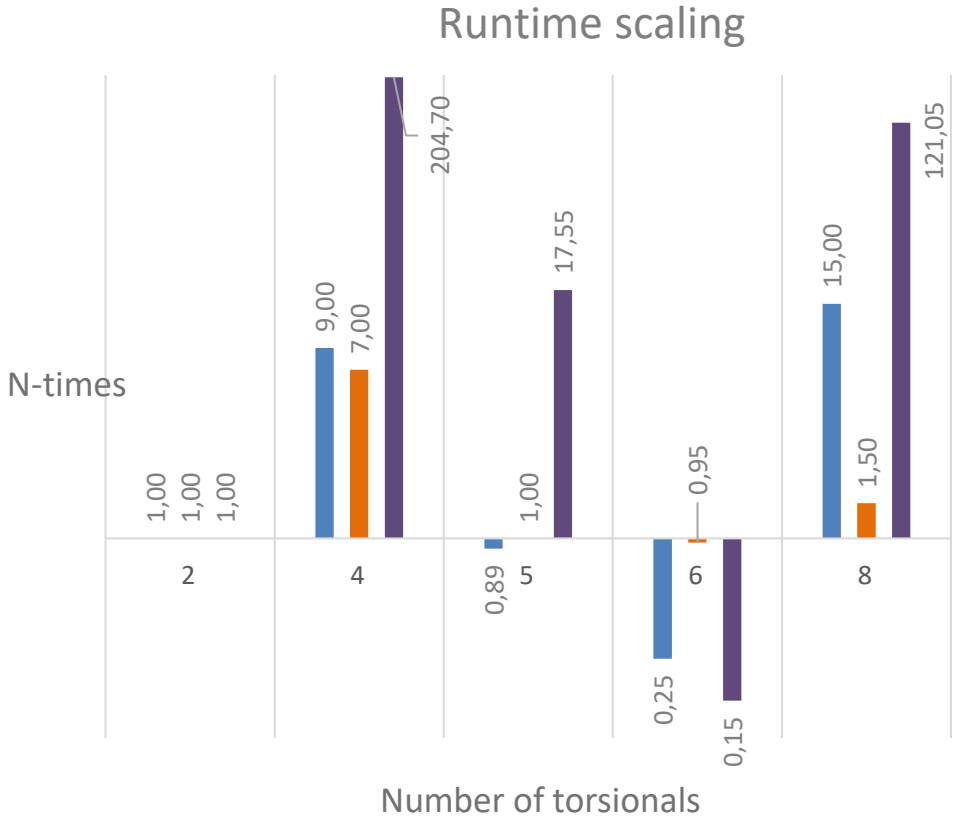




CINECA POLITECNICO MILANO 1863

Scaling DW_2000Q_6, Advantage & Classical





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.