# Biasing in Latent space using two stage VGAE

## Model Architecture

The model consists of a hierarchical graph auto-encoder with two stages, this architecture combines both intra-molecular and inter-molecular interactions to provide a comprehensive understanding of molecular behavior

### Stage 1

In Stage 1, the input graph consists of water molecules, and each node in the graph is characterized by multiple features including the position (x, y, z) of the atom, atomic number, van der Waals (vdW) radius, and atomic mass. The stage 1 GAE focuses on capturing information about the confirmation/configuration of individual molecules ignoring its absolute positioning and interactions with neighborhood.

### Inter Molecular Graph Generation

The Inter molecular graph is constructed using the MD simulation data, each individual molecule is first encoded using Stage 1 GAE, some additional information like position of COM of respective molecule is added.

For every molecule a node is added in the inter molecular graph with features described above.

An edge is defined between two nodes if the distance between the COM of these molecules is within a certain CutOff distance (5 Angstrom).

### Stage 2

The Inter molecular graph is constructed in above stage is then utilized as input for Stage 2, the inter-molecular auto-encoder. This stage encodes the inter-molecular configuration graph, and primarily encodes the intermolecular interactions present between two molecules present in

neighborhood of each other.

Intra Molecular GAE

```
VGAE(
  (encoder): VariationalGCNEncoder(
    (conv1): GCNConv(5, 4)
    (bn1): BatchNorm(4)
    (conv2): GCNConv(4, 2)
    (bn2): BatchNorm(2)
    (linear1): Linear(in_features=2, out_features=2, bias=True)
    (conv_mu): GCNConv(2, 2)
    (conv_logstd): GCNConv(2, 2)
  )
  (decoder): VariationalGCNDecoder(
    (conv1): GCNConv(2, 10)
    (conv2): GCNConv(10, 10)
    (linear1): Linear(in_features=10, out_features=5, bias=True)
    (linear2): Linear(in_features=5, out_features=5, bias=True)
  )
)
```

Inter Molecular GAE

```
VGAE(
  (encoder): VariationalGCNEncoder(
    (conv1): GCNConv(3, 4)
    (conv2): GCNConv(4, 2)
    (pool): SAGPooling(GraphConv, 2, ratio=0.5, multiplier=1.0)
    (conv_mu): GCNConv(2, 2)
    (conv_logstd): GCNConv(2, 2)
  )
  (decoder): VariationalGCNDecoder(
    (conv1): GCNConv(2, 6)
    (conv2): GCNConv(6, 3)
  )
)
```

## Water

### Input files and simulation

- The input files contain the tip4p water molecules with 2 hydrogen 1 oxygen and 1 ghost atom
- Collected the tip4p.gro and single pdb of water Molecule from the gromacs share
- Using gromacs command we generate box.pdb and box.gro for the simulation and system setup and topology file is obtained by using the same
- The mdp files are obtained from previous methods and modified for the simulation wherever necessary
- The system consists of 100 water molecules with 4 atoms each and energy minimisation is done on system with periodic boundary conditions in all 3 directions for 100000 steps
- The nvt simulation is run at temperature 300 K and with periodic boundary conditions in all 3 directions for 100000 steps
- The npt simulation is carried out to generate 5,00,000 frames for data reference with dt = 0.002 and nsteps as 2500000

## Crystal

- 1h ice crystal is taken into consideration as for the crystal structure for this process and simulation

- Its made of tessellating hexagonal rings, with an oxygen atom on each vertex, and the edges of the rings formed by hydrogen bonds.
- Oxygen atoms form an ordered lattice, belonging to the hexagonal space group P63/mmc, with a four-site primitive unit cell.
- The gro file for the 101 molecules are obtained along with the topology file using gromacs commands and the energy minimisation simulation is run on the input files for nsteps as 100000 and with periodic boundary conditions in all 3 dimensions
- For the nvt simulation is carried out for nsteps as 10000 and dt = 0.002 with reference pressure as 1
- And the npt simulation is carried out for nsteps 10000 and dt = 0.002 for the number of output frames to be 100 and the output npt.xtc file is obtained

**Crystal Structure Ice 1H**