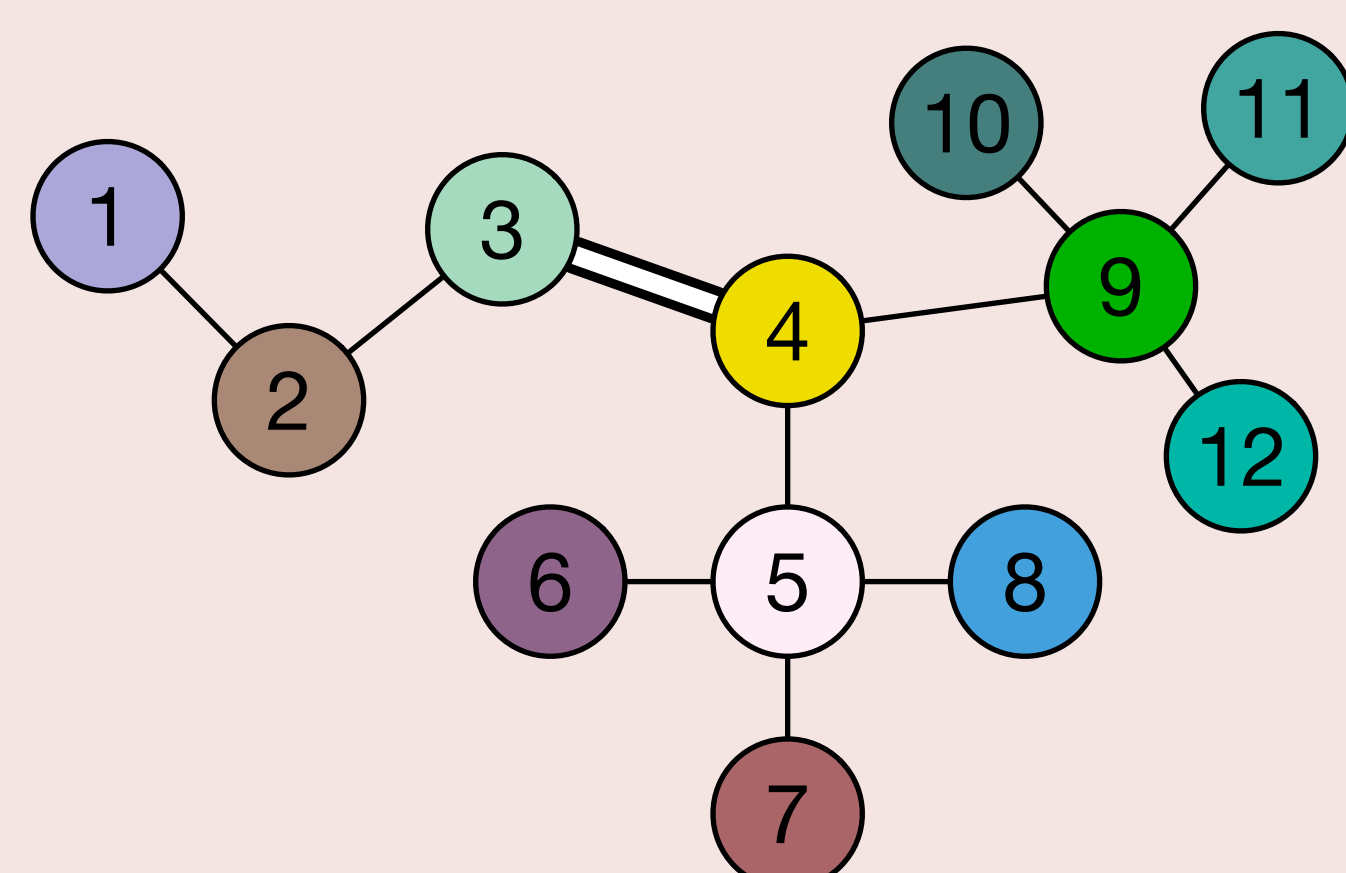
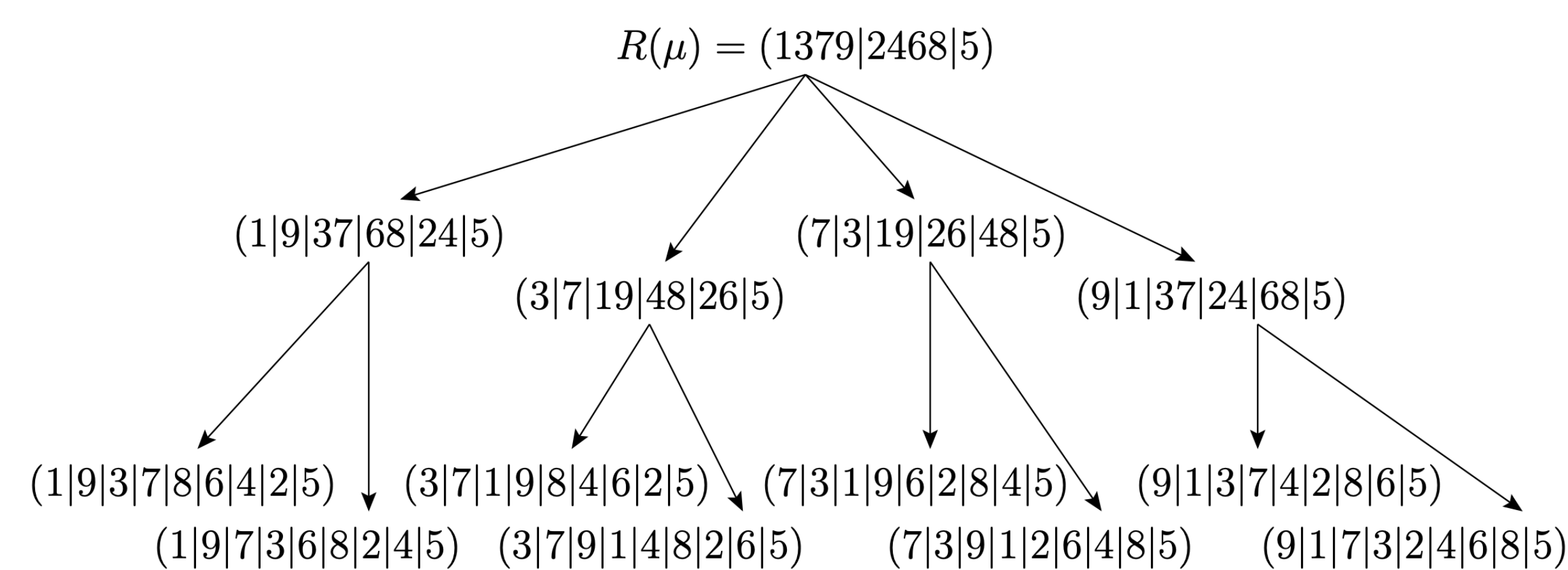


Input Molecule



## Serialization

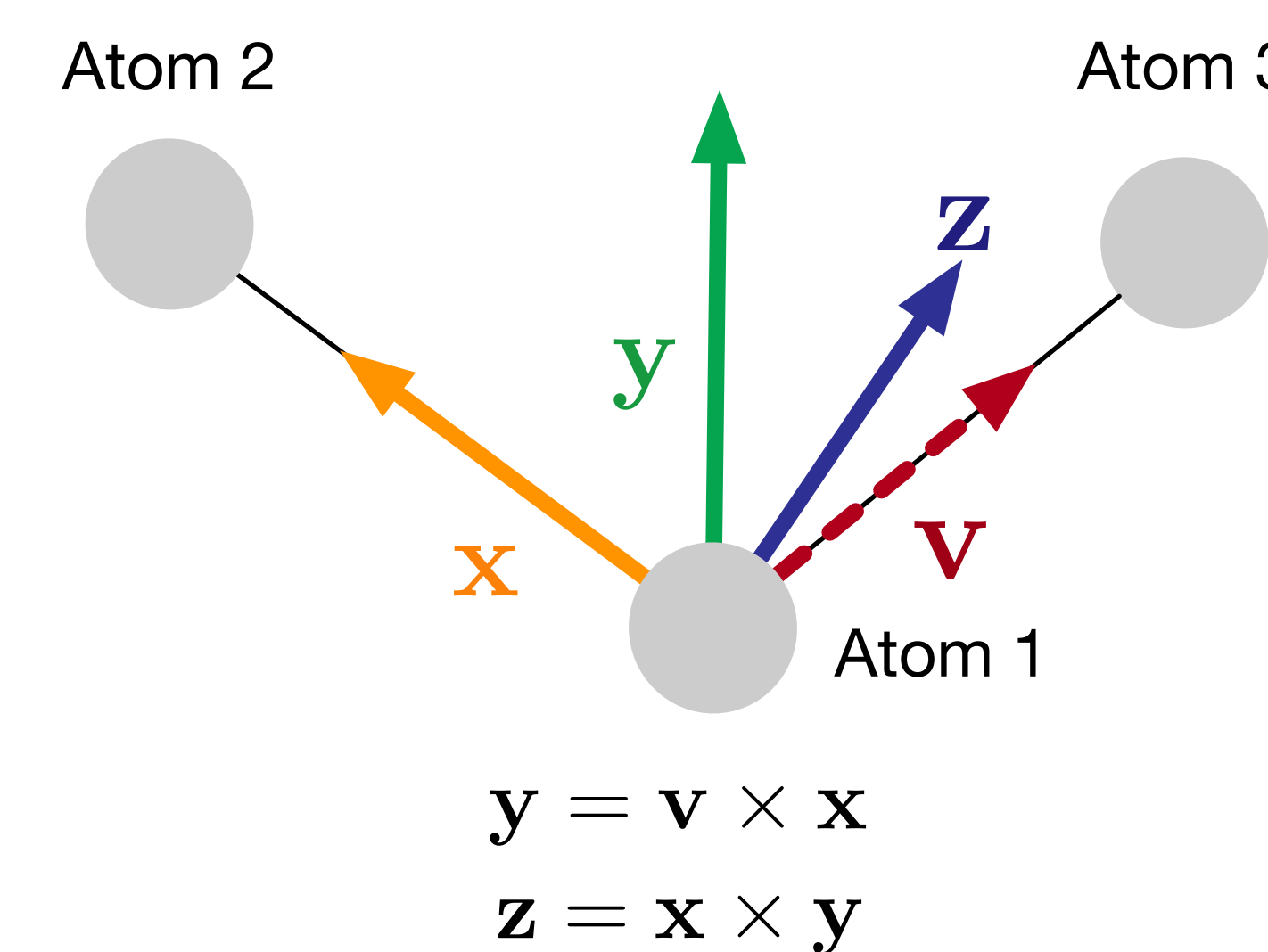
### Canonical Labeling



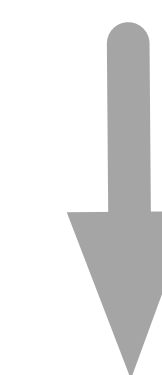
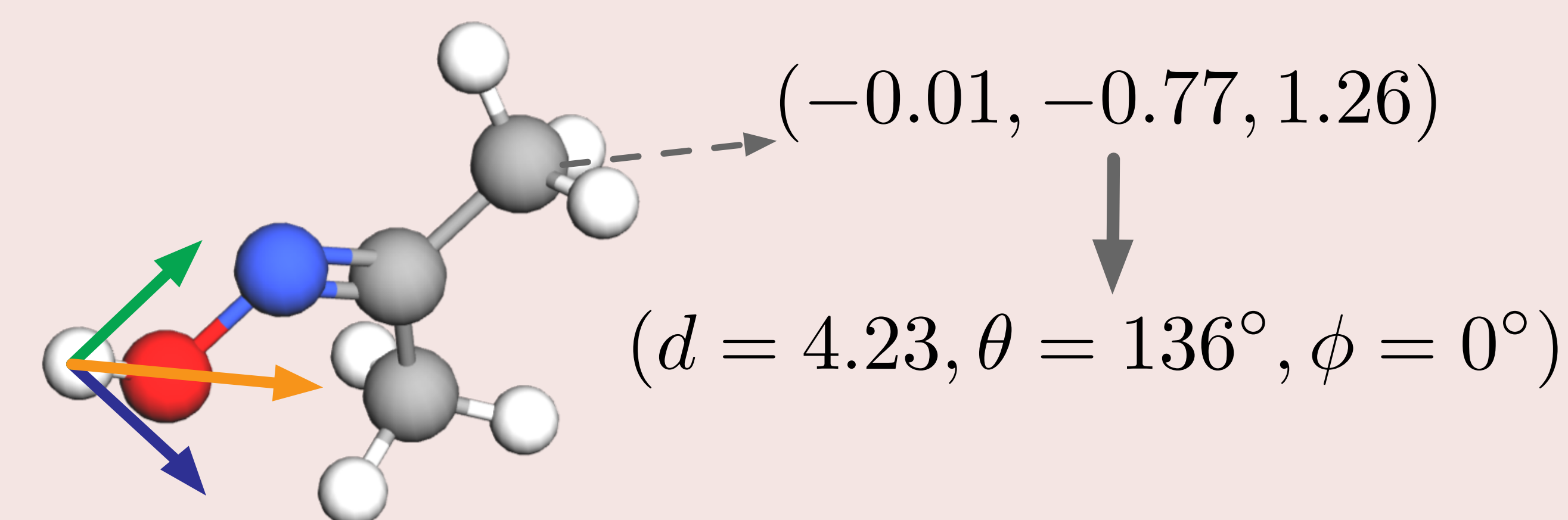
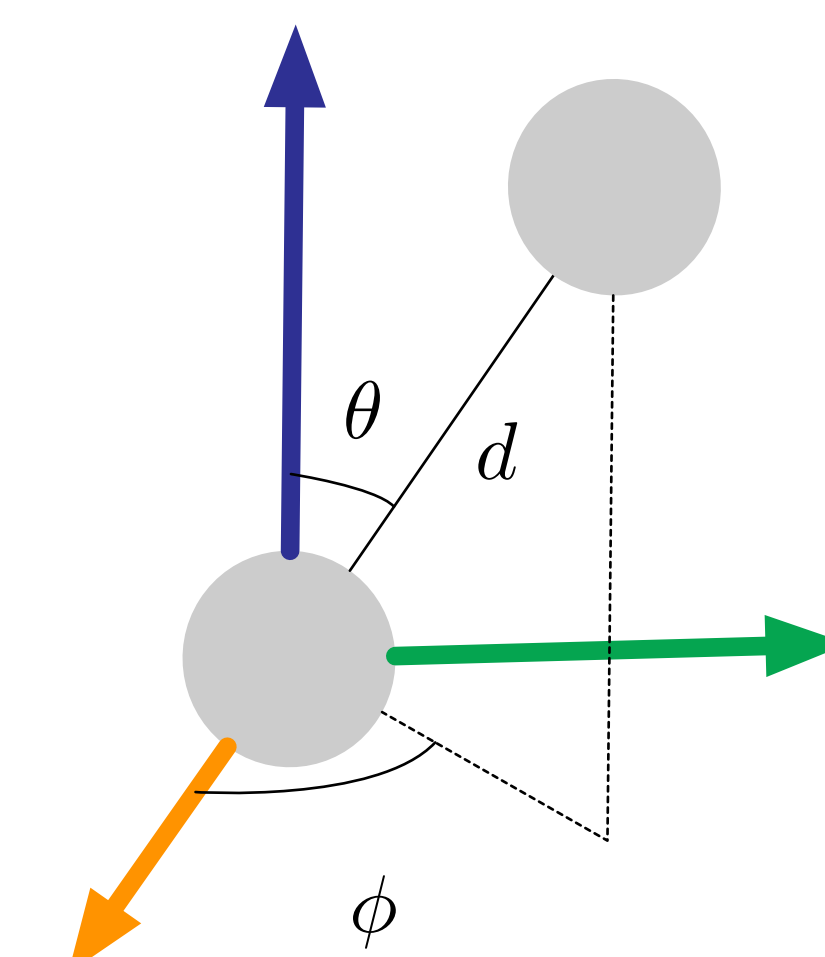
H O N C C H H H C H H H

## Spherical Representation

### Global frame



### Spherical coordinates



### 3D Sequence

H 0.00 0.00° 0.00°  
O 0.96 1.57° 0.00°  
N 1.86 2.40° 0.00°  
C 2.97 2.13° -0.01°  
C 3.58 1.71° -0.03°  
H 4.29 2.30° 0.25°  
H 5.25 2.20° -0.25°  
H 4.57 2.61° -0.01°  
C 4.23 2.37° -0.00°  
H 6.85 2.54° 0.18°  
H 6.59 2.81° -0.29°  
H 6.13 2.83° 0.24°

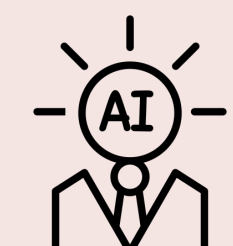
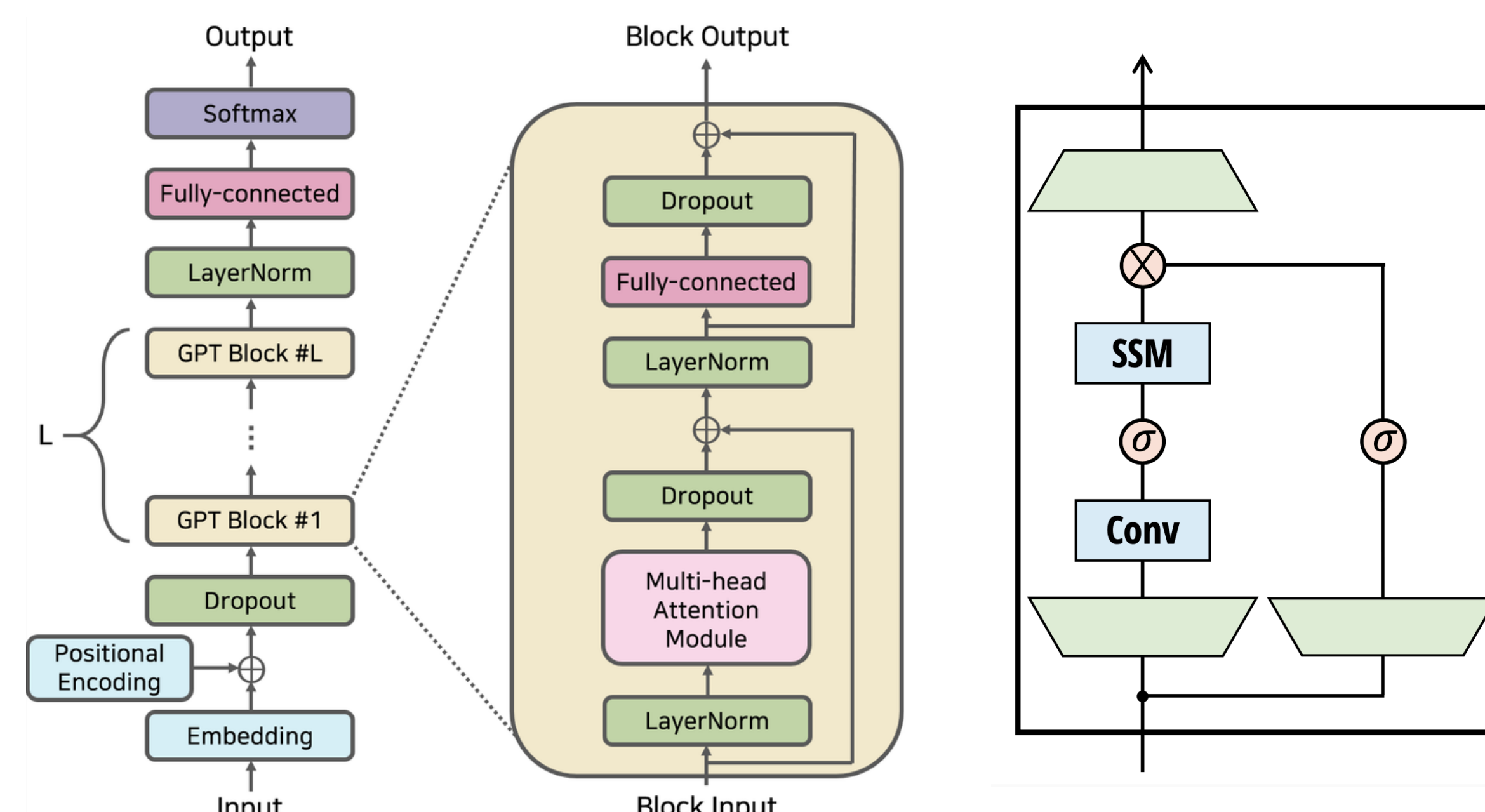
### Tokenization

Input\_IDs



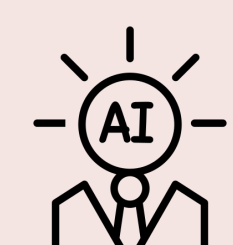
## Training

### Language Models



### Unconditional generation

Input: H 0.00 0.00° 0.00° O 0.96 1.57° ...



### Conditional generation

Condition:  $\varepsilon_{homo}$  Property \*-5.9484\*  
Input: H 0.00 0.00° 0.00° O 0.96 1.57° ...