1 Degree of freedom of a system with external forces

From MD simulations with different $N(CH_4)$, we compute average kinetic energy E_k of CH_4 as listed in **Table 1**.

Table 1: Average E_k of CH_4 for different $N(CH_4)$

N	$\frac{E_k}{N}$ (kJ/mol)
4	2.78
8	3.23
16	3.48
40	3.61
80	3.66
100	3.67
120	3.67
140	3.68
160	3.68

According to the equipartition theorem, energy carried from every degree of freedom (DOF) is equal to $\frac{1}{2}k_BT$, and the sum of the energies for all DOF ($N_{DOF} \times \frac{1}{2}k_BT$), is equal to total kinetic energy E_k. From this discussion, we obtain the relation between total kinetic energy E_k and temperature T as **Equation (1)**. To calculate total DOF (N_{DOF}) of CH_4 group, we assume it has the form of "$N_{DOF} = aN - b$" and perform a linear regression with data of E_k and N from **Table 1**, T=298 K, N_{DOF} is equal to $3N - 3$ (**Equation (2)**).

$$E_k = \sum_{N}^{i} \frac{1}{2} m v_i^2 = \frac{1}{2} k_B T \times N_{DOF} \tag{1}$$

$$N_{DOF} = 3.01 \times N - 3.00 \tag{2}$$

This result supports the idea that the calculation of degrees of freedom in LAMMPS is implemented in a way which keeps total linear momentum conserved and center of mass (COM) fixed, despite of the existence of an external force field. This correction of DOF has an effect on the velocities of CH_4, and, consequently, on the E_k distribution.

The conservation of total linear momentum determines total DOF of CH_4 is equal to $3N - 3$ instead of $3N$, which cause the anomalous relation. Indeed, to use the correct DOF estimation, we modify the source codes of LAMMPS related to the DOF calculations in the NVT simulation.

```
Lammps-version/src/:
    fix_nvt.cpp         \*call fix_nh.cpp*\
    fix_nh.cpp:         \*call "dof" variable from compute_temp.cpp*\
```

```
                    tdof=temperature->dof
                    kecurrent=tdof*boltz*t_current
compute.cpp:        \*set default value of "extra_dof"*\
                            \*and "fix_dof"*\
                    extra_dof=domain->dimension
                    fix_dof=0
compute_temp.cpp:   \*definition of "dof" variable*\
                    dof=domain->dimension*natoms_temp
                    dof-=extra_dof+fix_dof
```

For using Nose-Hoover thermostat, four `*.cpp` files are the most relevant. `fix_nvt.cpp` and `fix_nh.cpp` include all implementations of Nose-Hoover thermostat, and refer to variable `dof` to compute E_k. `compute.cpp` and `compute_*.cpp` include implementations of computation of certain property, e.g. `compute_temp.cpp` for temperature calculation.

In `compute_temp.cpp`, we find the definition of $N(\mathrm{CH_4})$ for a 3-dimensional system, `dof=3N-extra_dof-fix_dof`. In `compute.cpp`, default value of `extra_dof` is assigned to be equal to dimension, i.e. 3 for 3-dimensional system.

In conclusion, LAMMPS assigns the total DOF of our system to be $(3N-3)$ instead of $3N$, which causes the lower total kinetic energy E_k compared to the reference value, 3.70 kJ/mol (**Equation (3)**), and incorrect E_k distribution compared to the reference canonical distribution (**Equation (4)**).

Canonical distribution at 298 K

$$E_k \text{ per mol} = \frac{3}{2} k_{\mathrm{B}} T N_{\mathrm{A}} = 3.70 \text{ kJ/mol} \tag{3}$$

$$P = 2\left(\frac{E}{\pi}\right)^{\frac{1}{2}} \left(\frac{1}{k_B T}\right)^{\frac{3}{2}} \exp\left\{-\frac{E}{k_B T}\right\} \tag{4}$$

In other words, when we perform a NVT simulation at 298 K using LAMMPS, the target temperature is never achieved, more significant for small numbers of guest molecules, due to loss of energies of 3 DOF, and, in addition, the effective temperature of our simulation is $T_{\text{eff}} = \frac{N-1}{N} \times 298$ K. The effective temperature T_{eff} is lower than the imposed one, leading to slower diffusion of $\mathrm{CH_4}$ and smaller diffusion coefficients D_s, which explains the anomalous relation between $N(\mathrm{CH_4})$ and D_s.

To obtain the correct results of D_s at 298 K, we modify `compute_temp.cpp`, use `dof=3N-fix_dof` instead of `dof=3N-extra_dof-fix_dof`. By removing `extra_dof` variable, LAMMPS is able to calculate the total DOF of $\mathrm{CH_4}$ as $3N$.

The modifed Lammps is available in 'https://github.com/HengluXu/lammps.git'