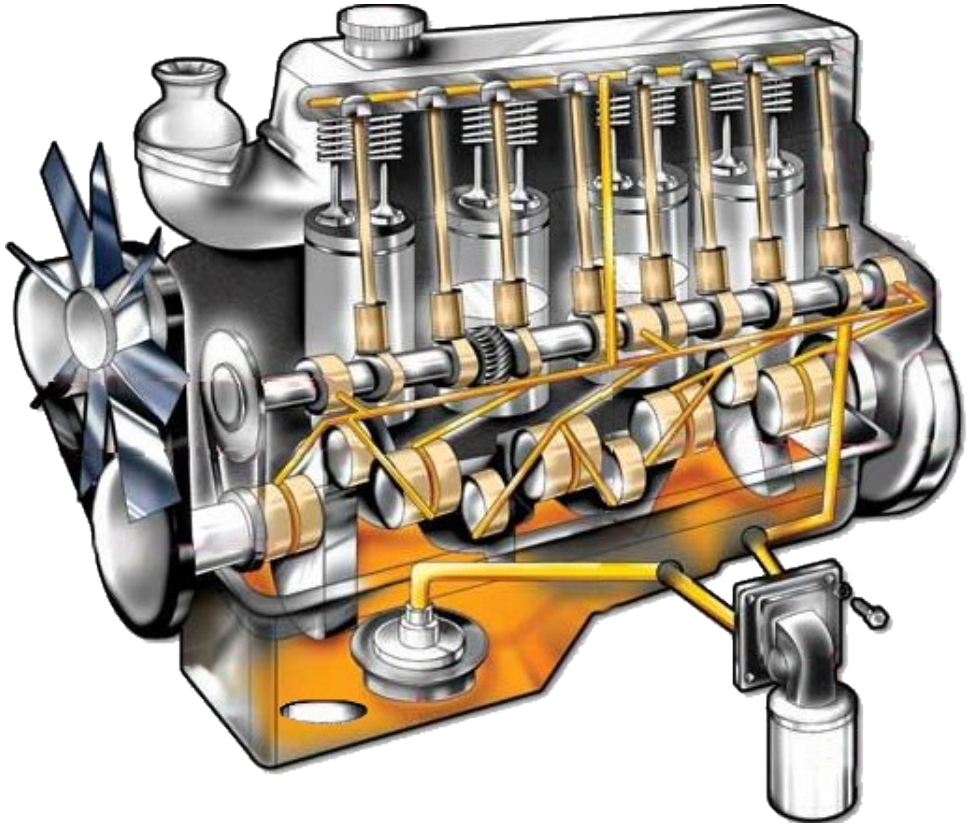
UNIVERSITY OF CAMBRIDGE

Climbing the greasy pole with machine learning: designing lubricants

Pavao Santak & **Gareth Conduit**

Theory of Condensed Matter group

Purpose of lubricants



- Reduce friction between surfaces
- Protect them from wear
- Transfer heat
- Remove dirt
- Prevent surface corrosion

Other requirements for lubricants

Stable

Not spontaneously combust

Liquid

Affordable

Ease of manufacture

Machine learning algorithm to

Train from sparse datasets

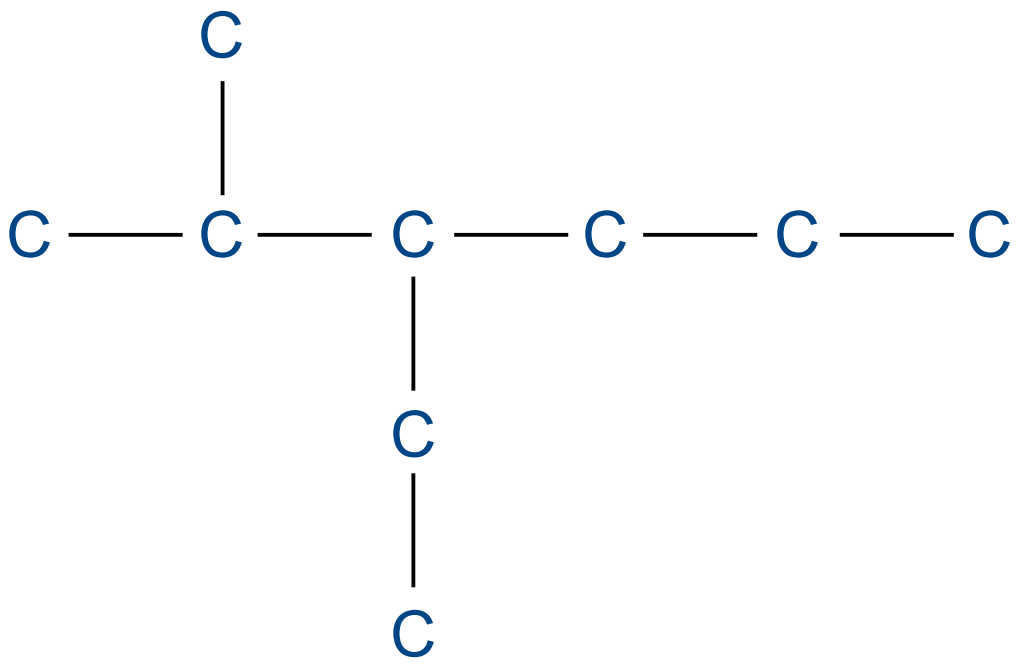
Merge simulations, physical laws, and experimental data

Reduce the need for expensive experimental development

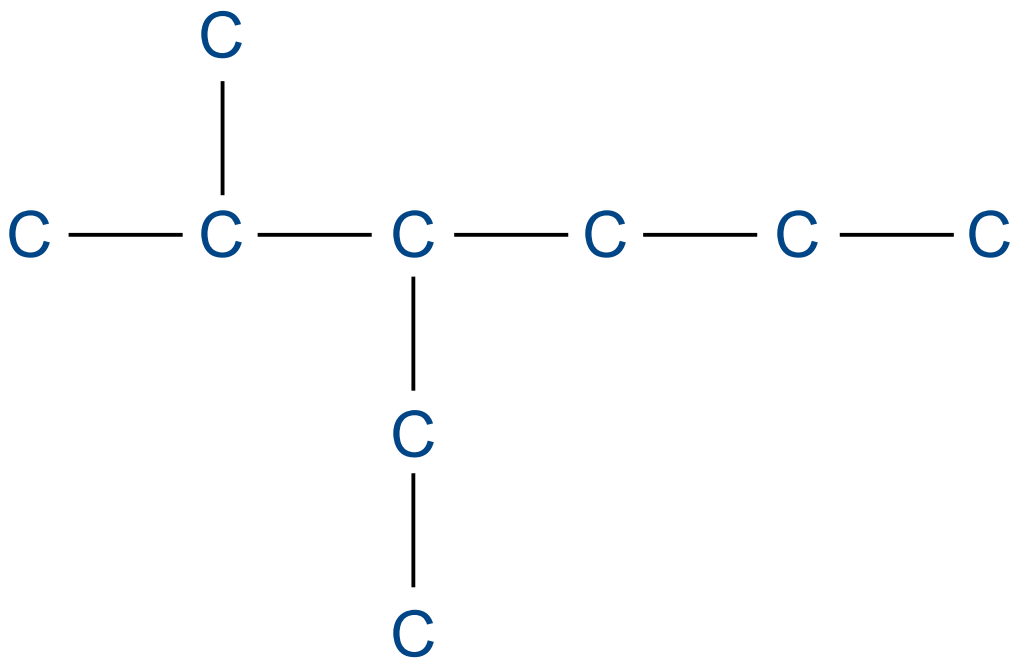
Accelerate discovery

Generic with proven applications in materials discovery and drug design

Typical lubricant hydrocarbon



Scope of lubricant study



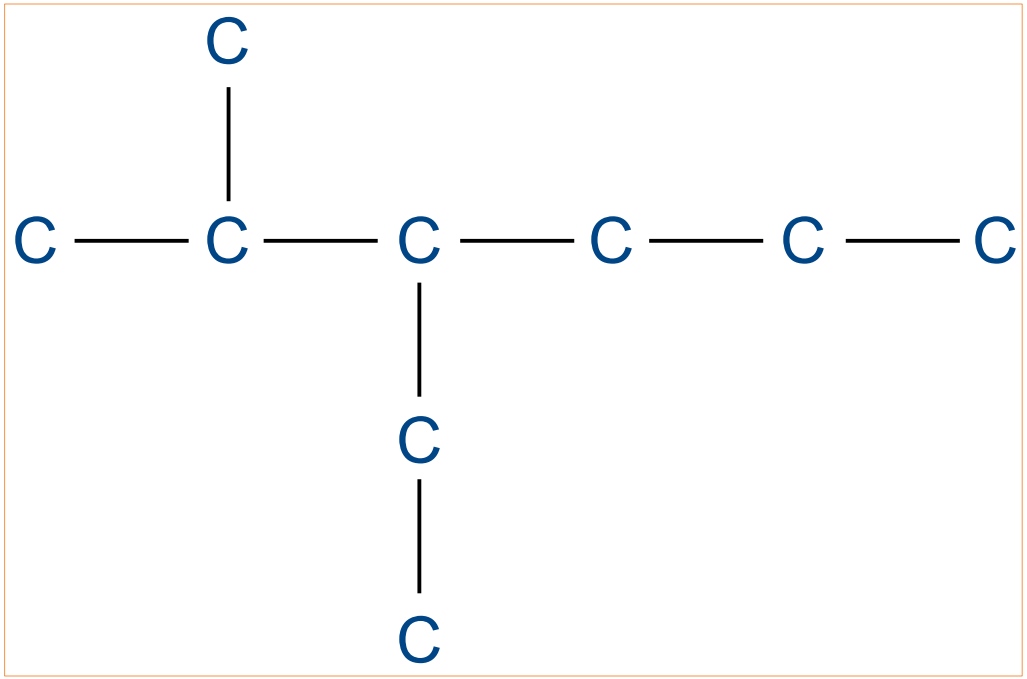
Comprise up to **thirty** carbon atoms

Consider up to **two** branches

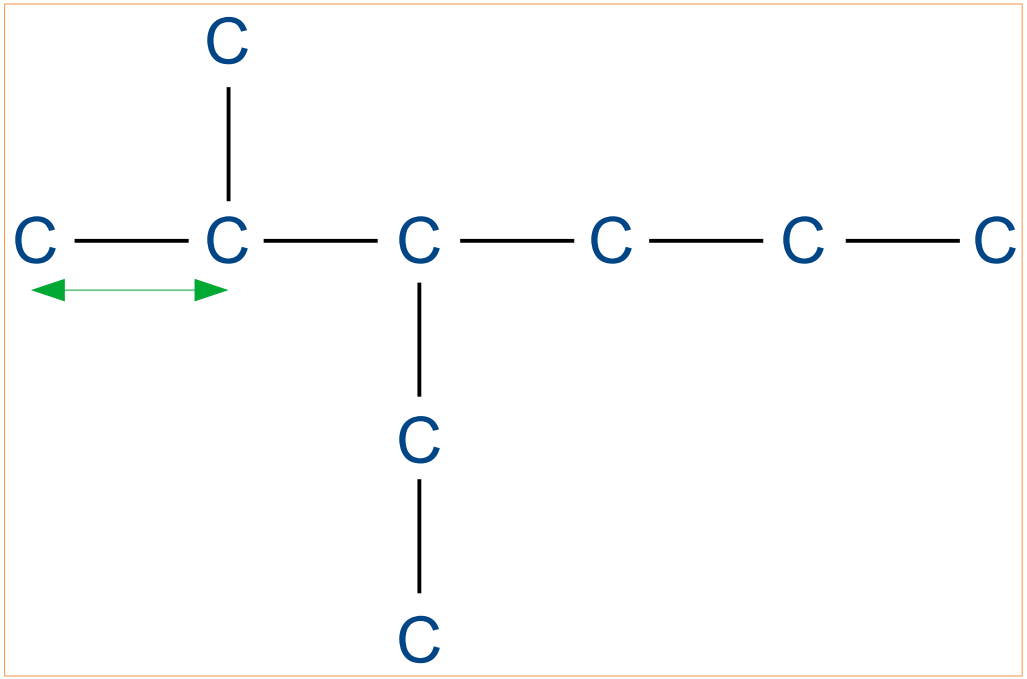
All **single** bonds, no additional elements

First component of basis set

Total number of C atoms



Second component of basis set



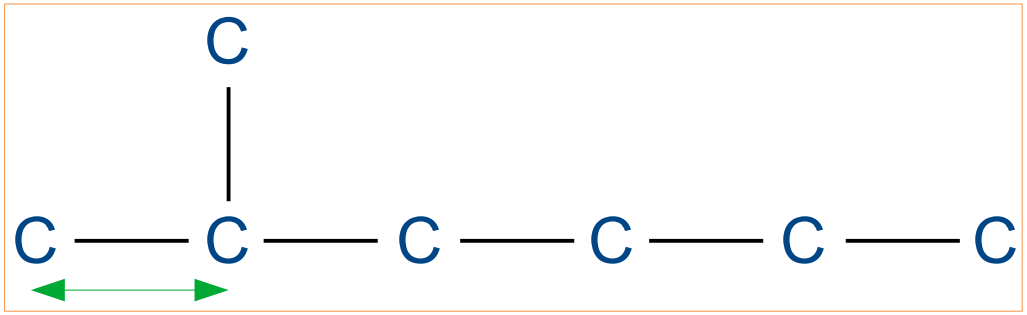
Total number of C atoms

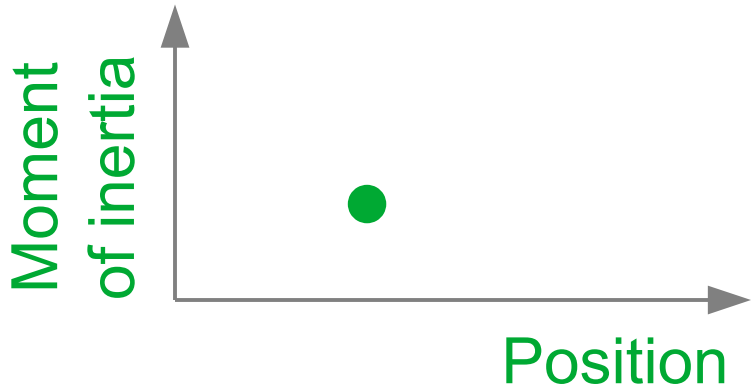
Distance of first branch from end

Basis set should be continuous

Total number of C atoms

Distance of first branch from end

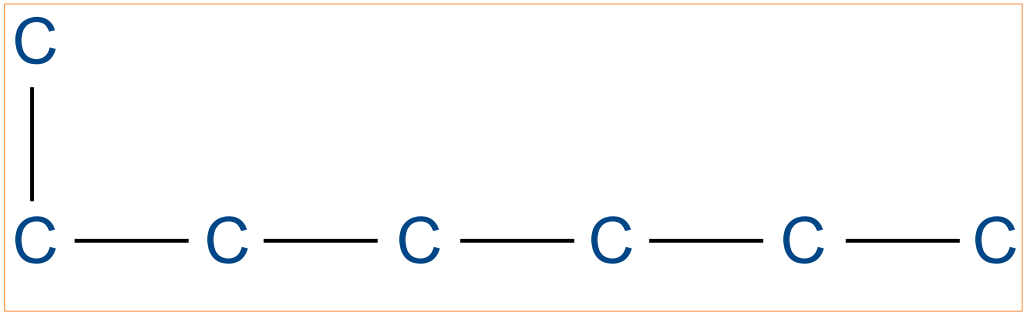


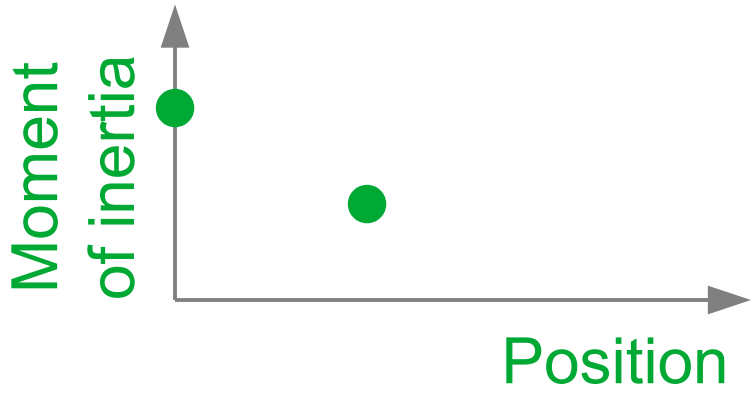


Zero equivalent to no branch

Total number of C atoms

Distance of first branch from end

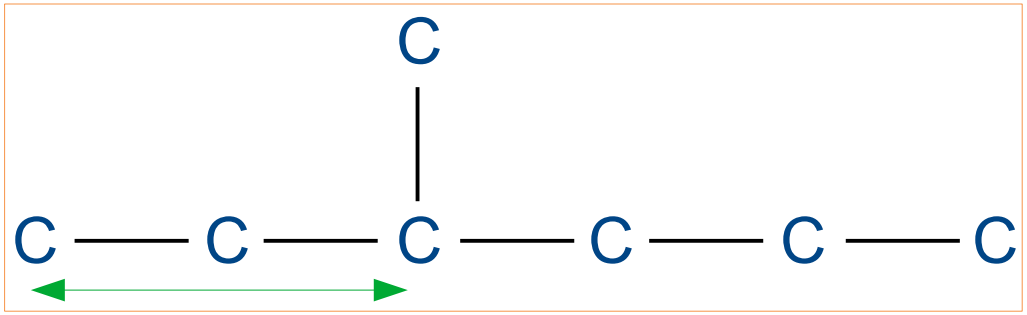




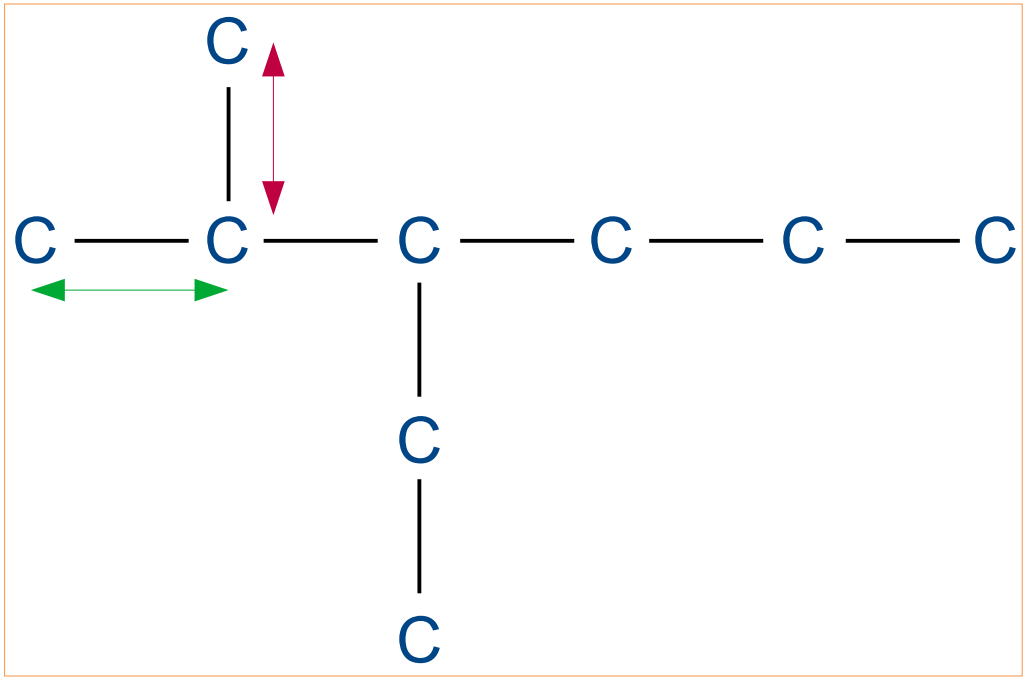
Two equivalent to center

Total number of C atoms

Distance of first branch from end



Third component of basis set

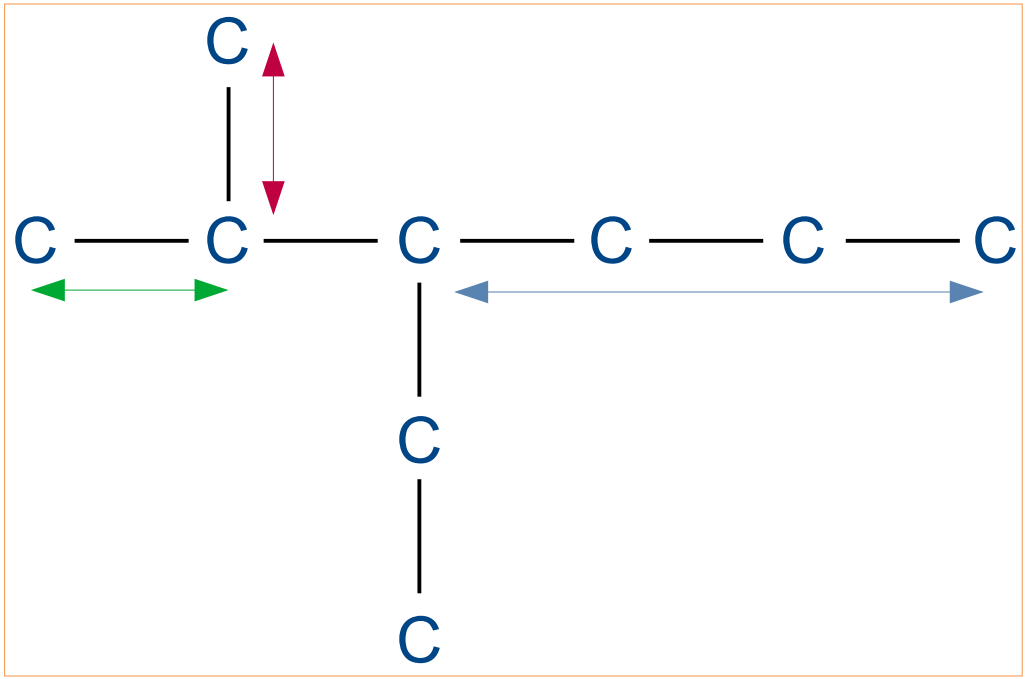


Total number of C atoms

Distance of first branch from end

Length of first branch

Fourth component of basis set



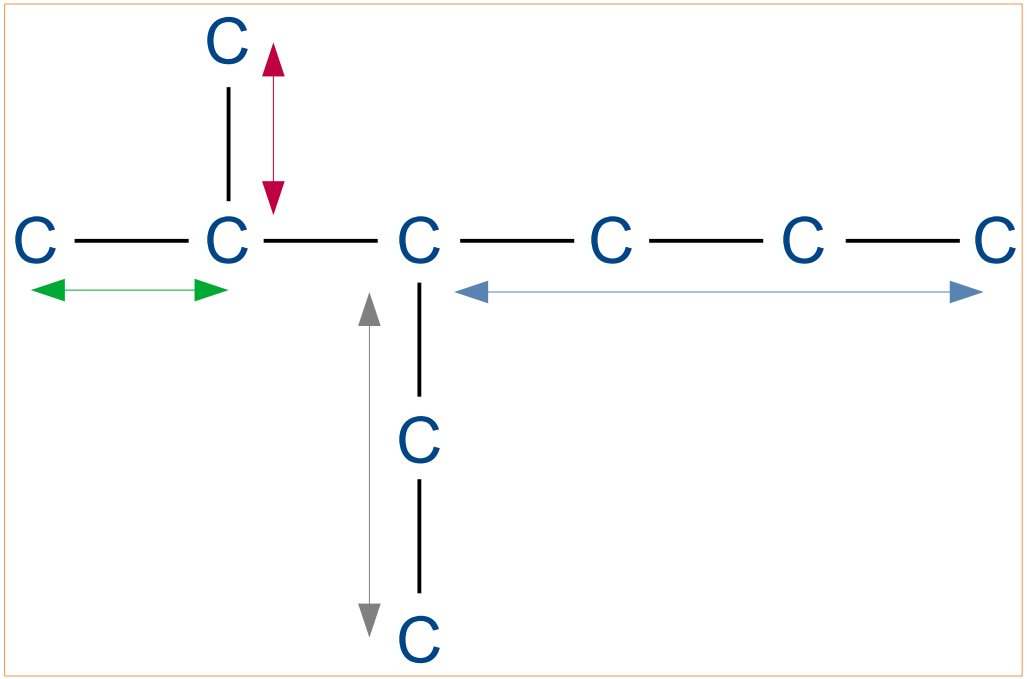
Total number of C atoms

Distance of first branch from end

Length of first branch

Distance of second branch from end

Fifth component of basis set



Total number of C atoms

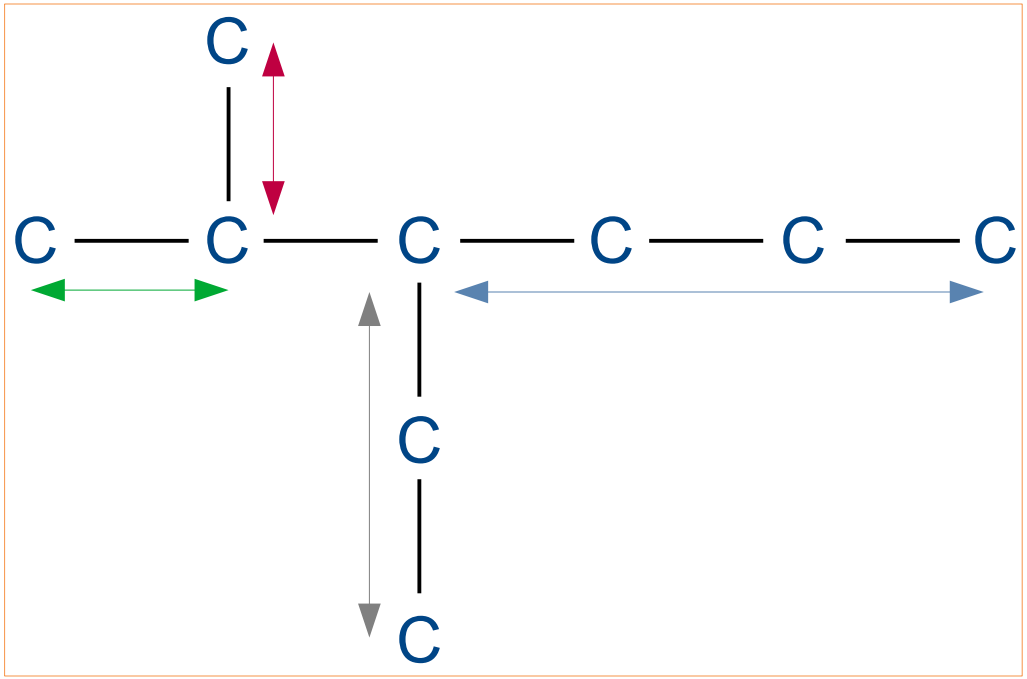
Distance of first branch from end

Length of first branch

Distance of second branch from end

Length of second branch

Full basis set



Total number of C atoms

Distance of first branch from end

Length of first branch

Distance of second branch from end

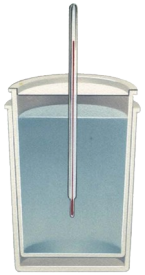
Length of second branch

(9,1,1,3,2)

Target properties

Melting point

Heat capacity



Boiling point

Vapor pressure

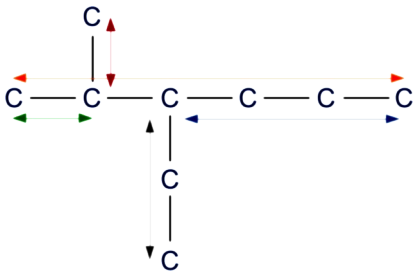
Density

Flash point

Viscosity index

Neural network for lubricant design

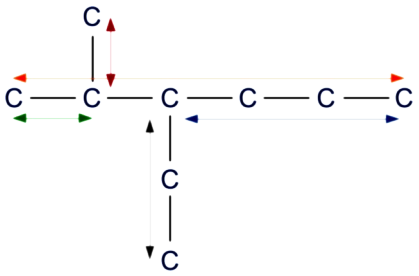
Structure



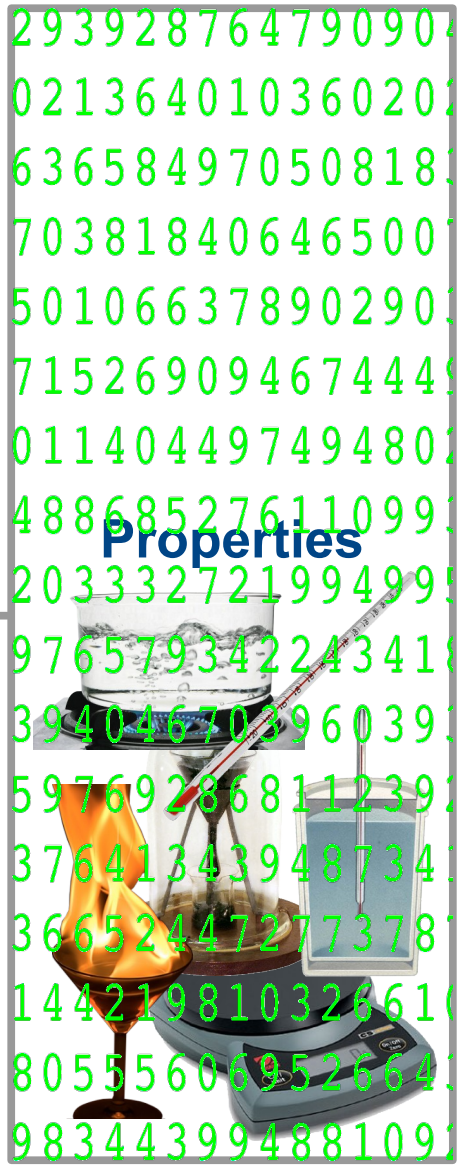
Properties

Train the neural network

Structure

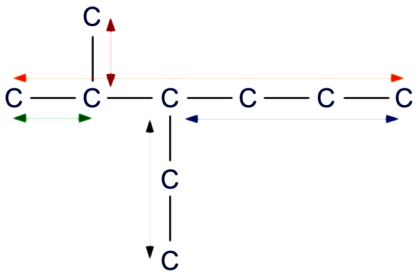


Properties



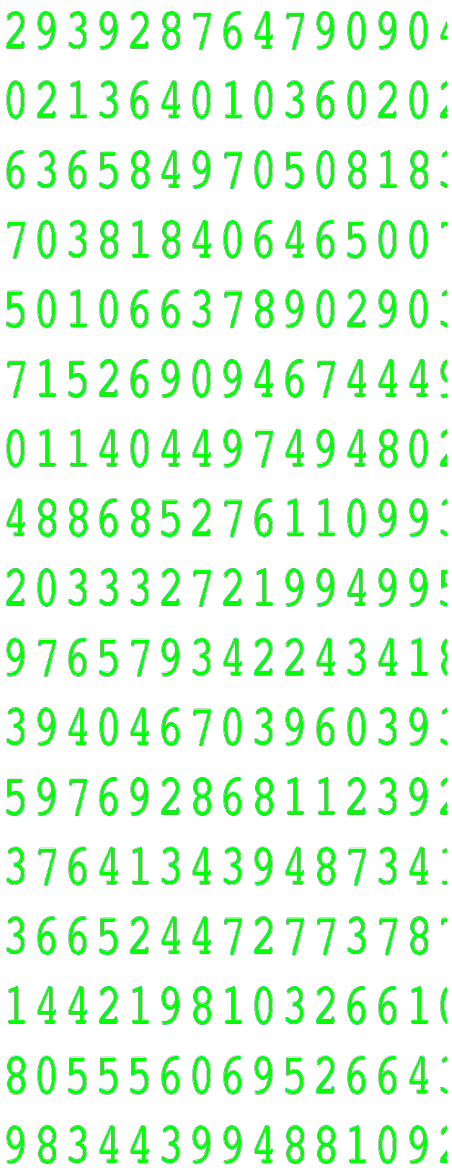
Predict using the neural network

Structure

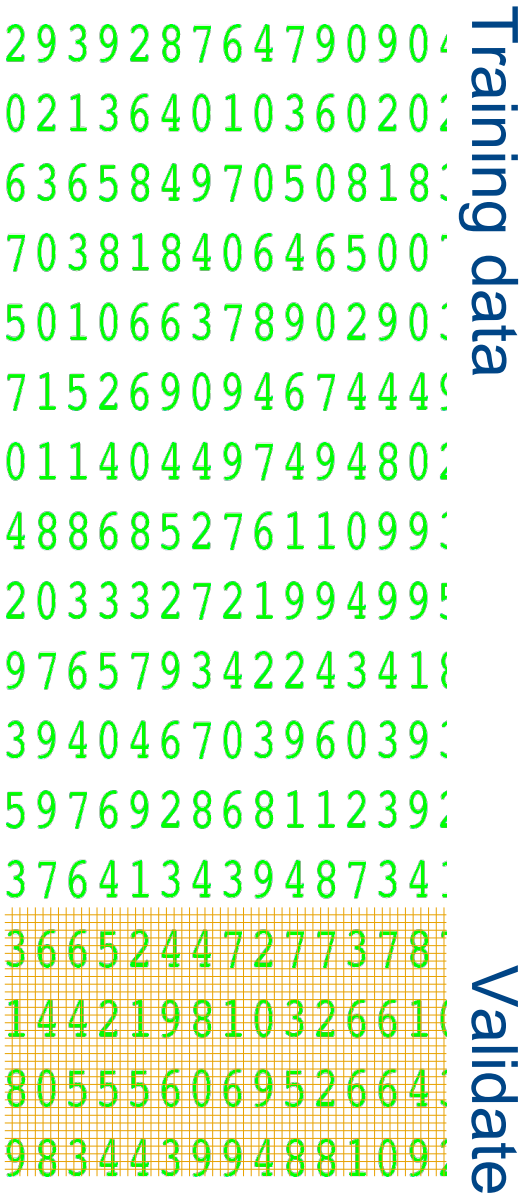


Properties

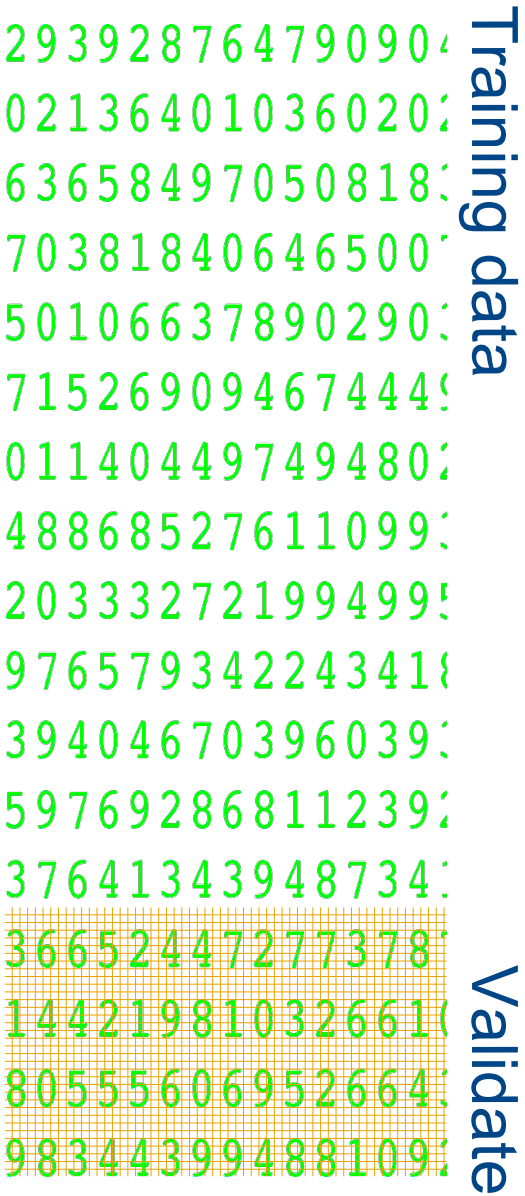
All data available for training

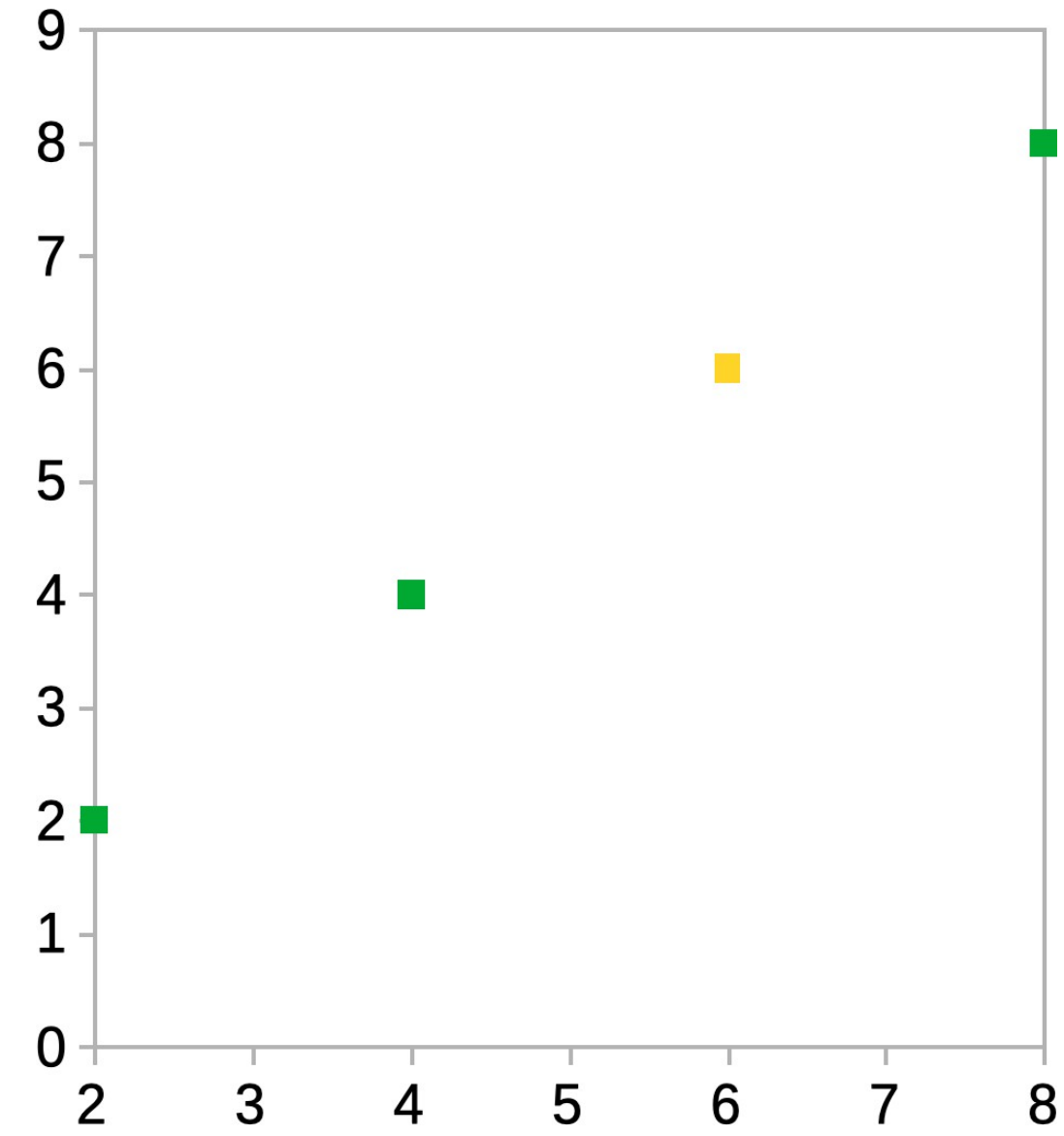


Split into training and hold out validation set

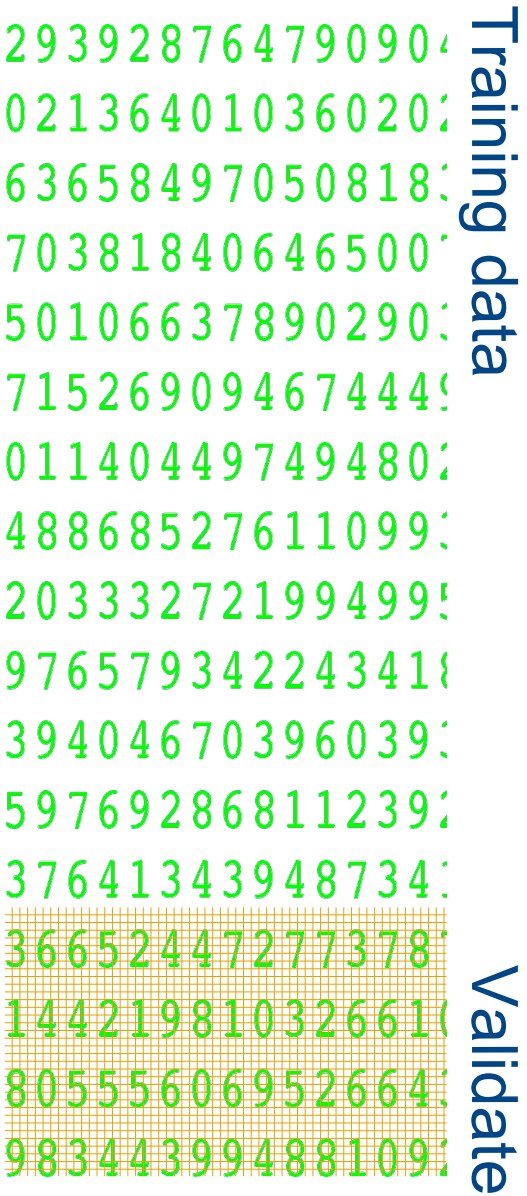


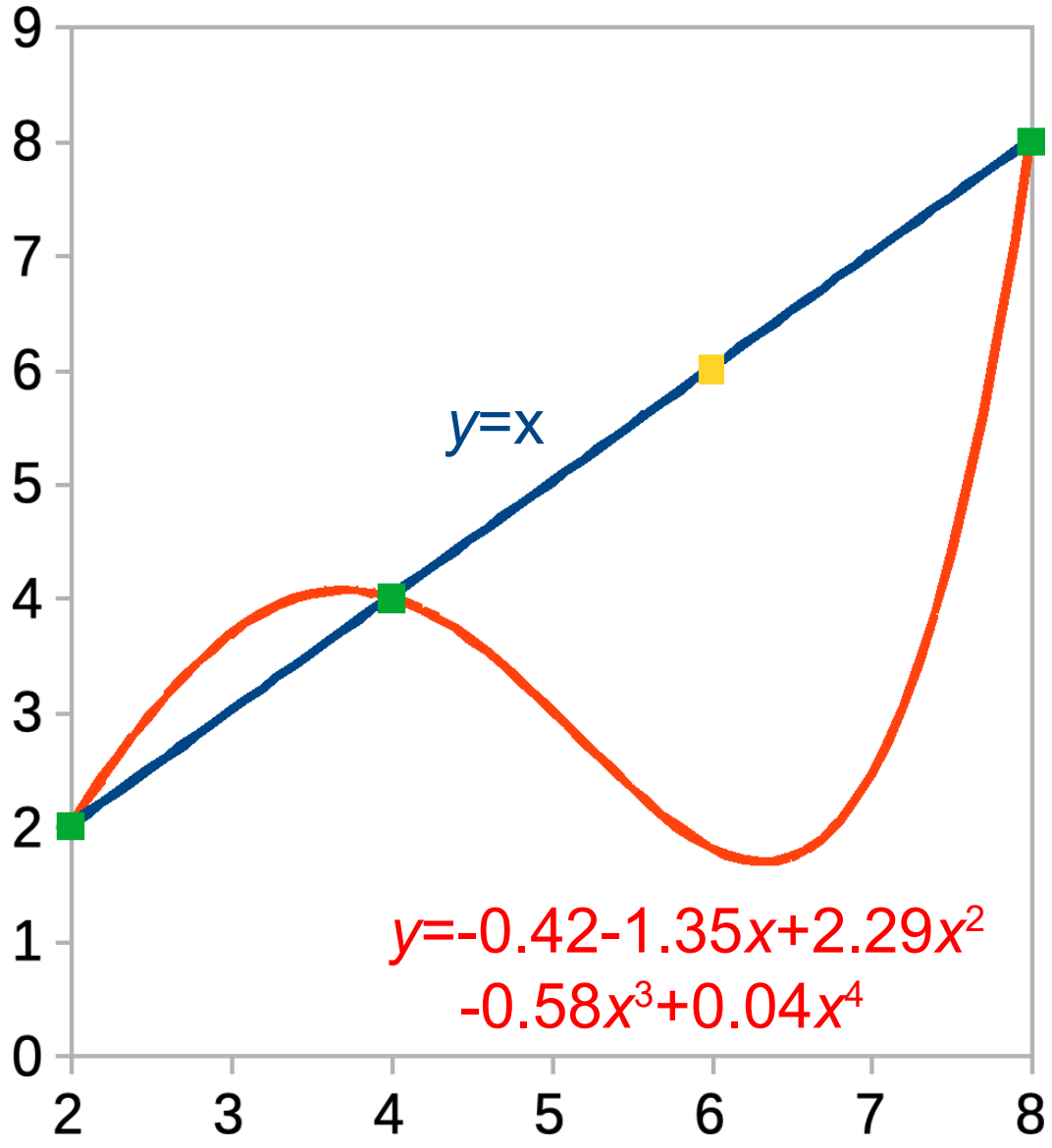
Split into training and hold out validation set



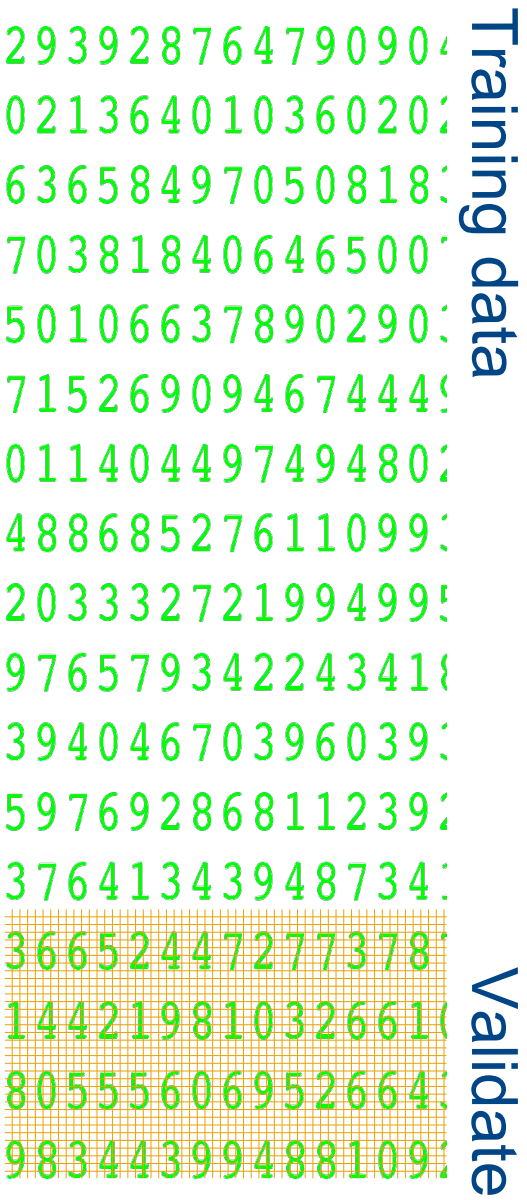


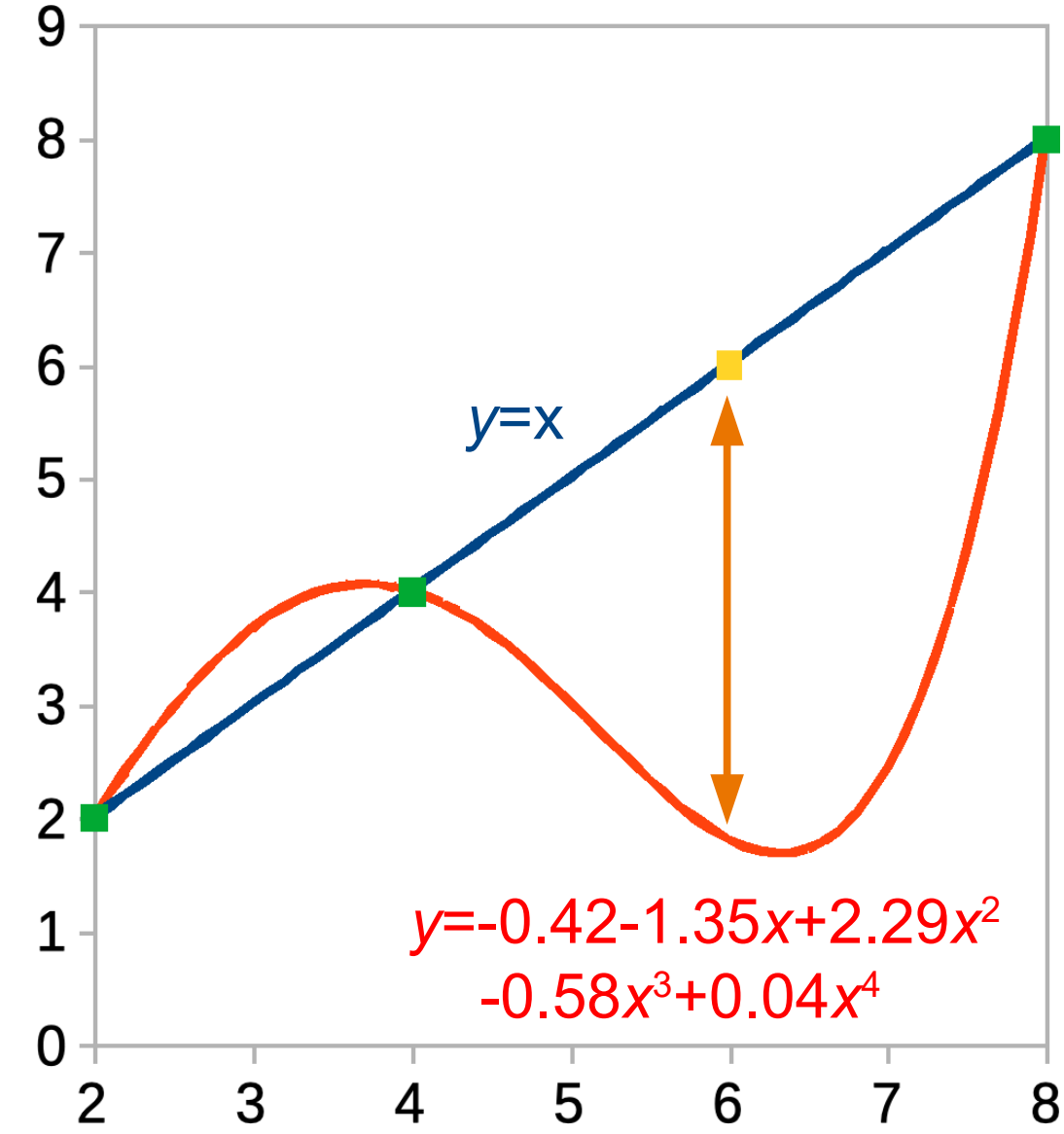
Two alternate fits to the data



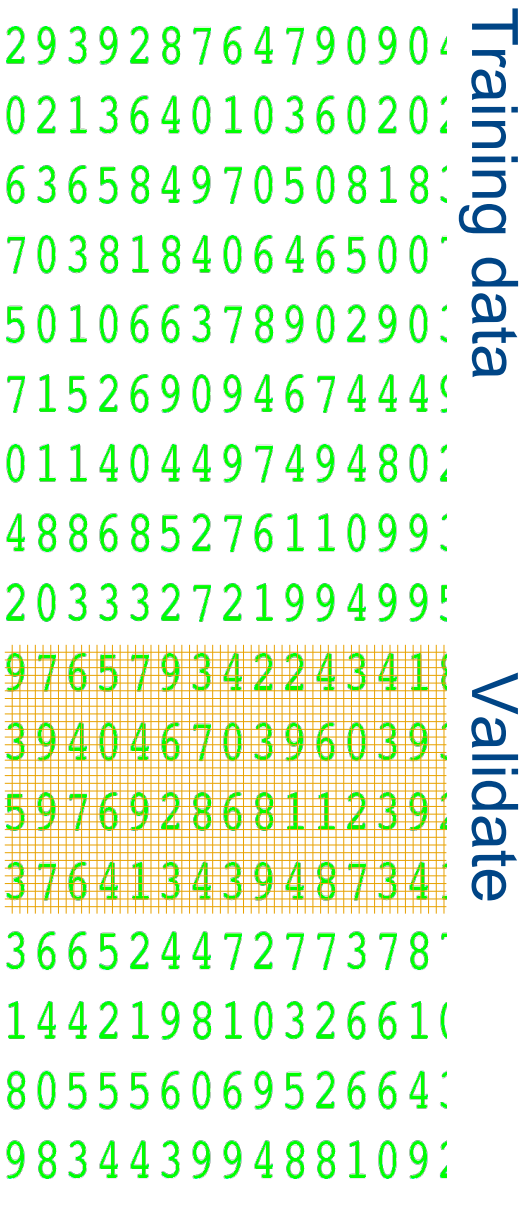


Penalize overfitted curve

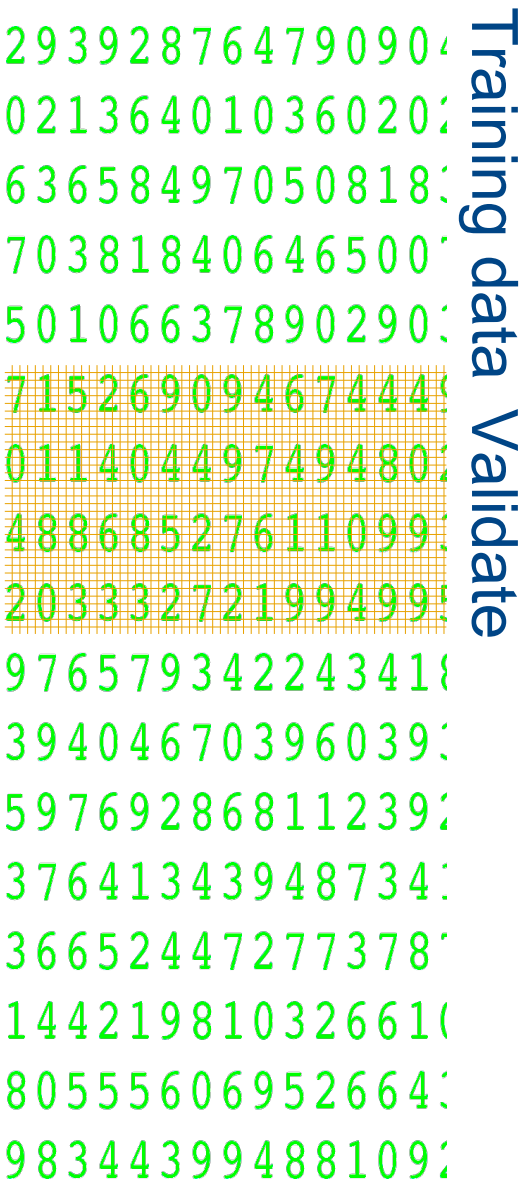




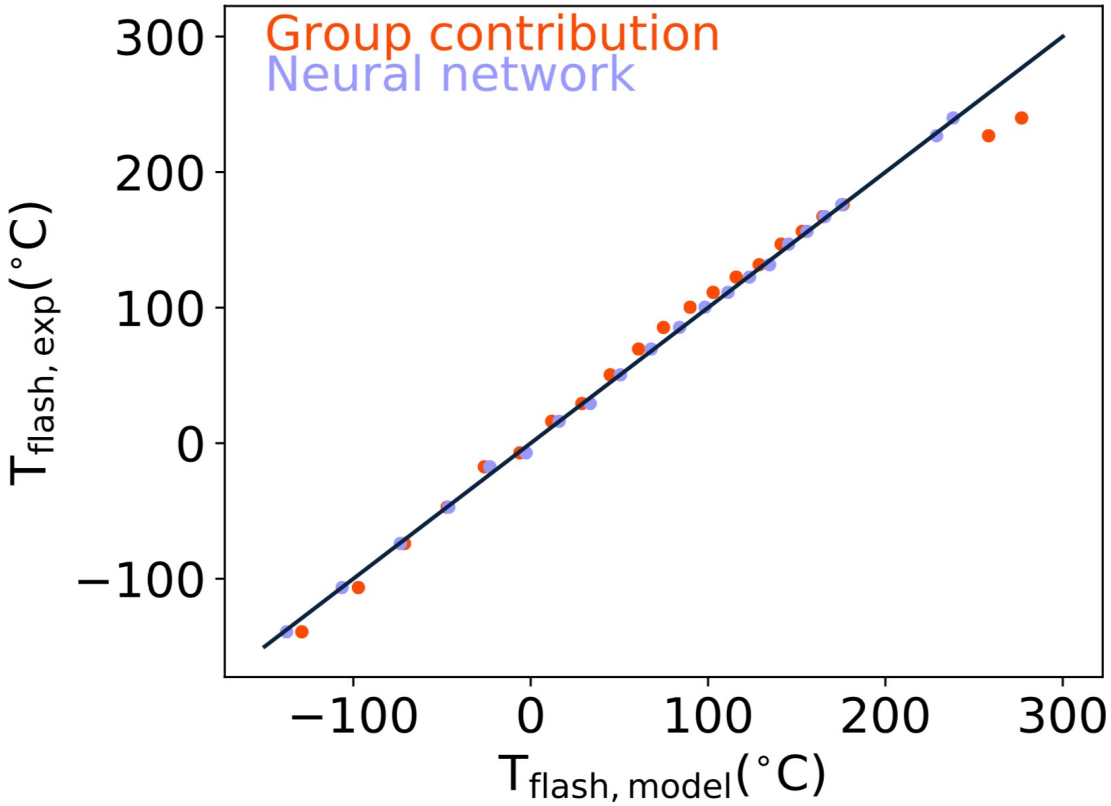
Crossover validation: second hold out set



Crossover validation: third hold out set



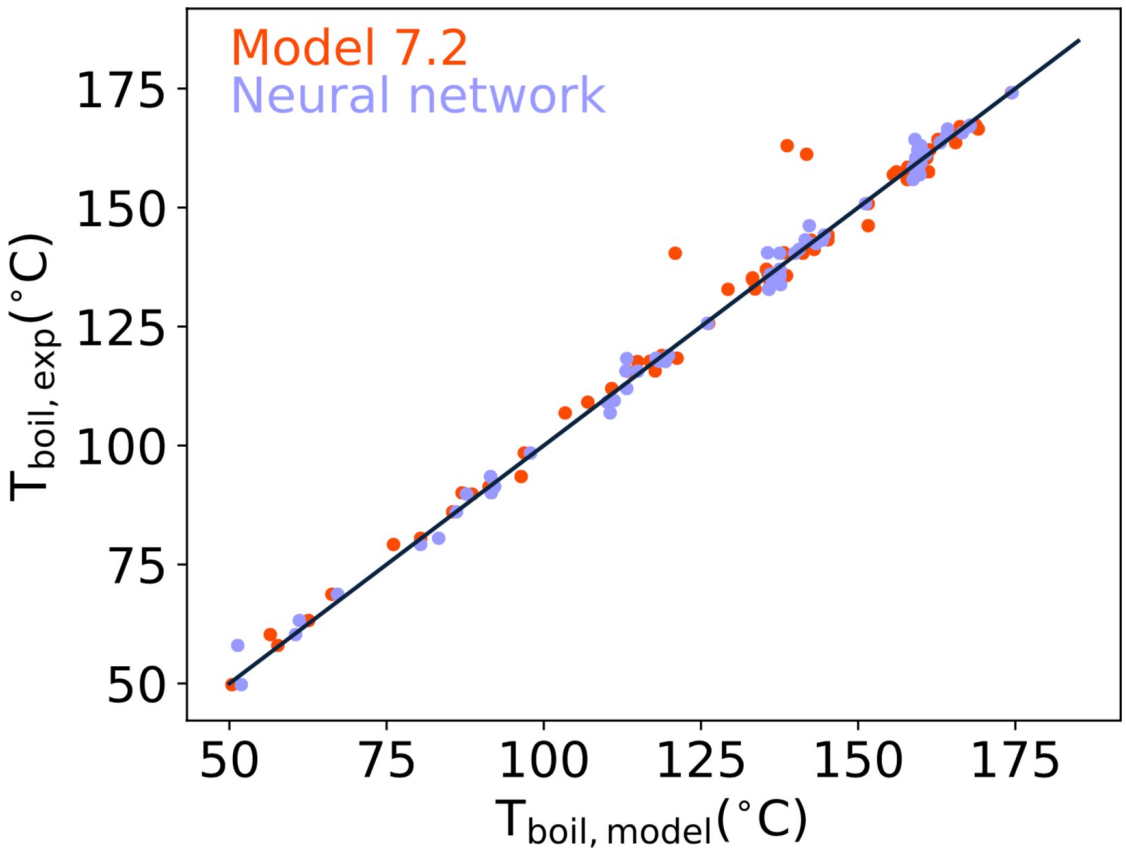
Flash point



Neural network R^2=0.997

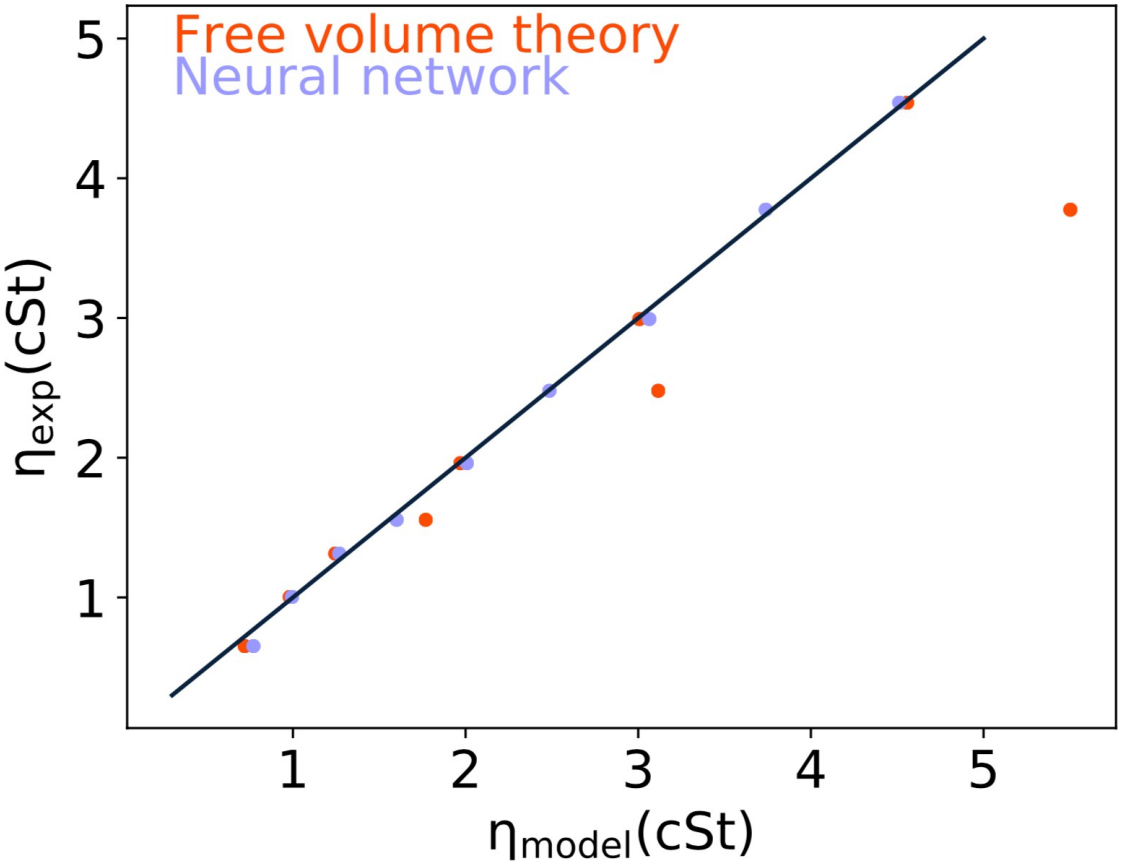
Group contribution R^2=0.971

Boiling point



Neural network	R^2=0.992
Model 7.2	R^2=0.977
Model 7.3	R^2=0.975

Kinematic viscosity



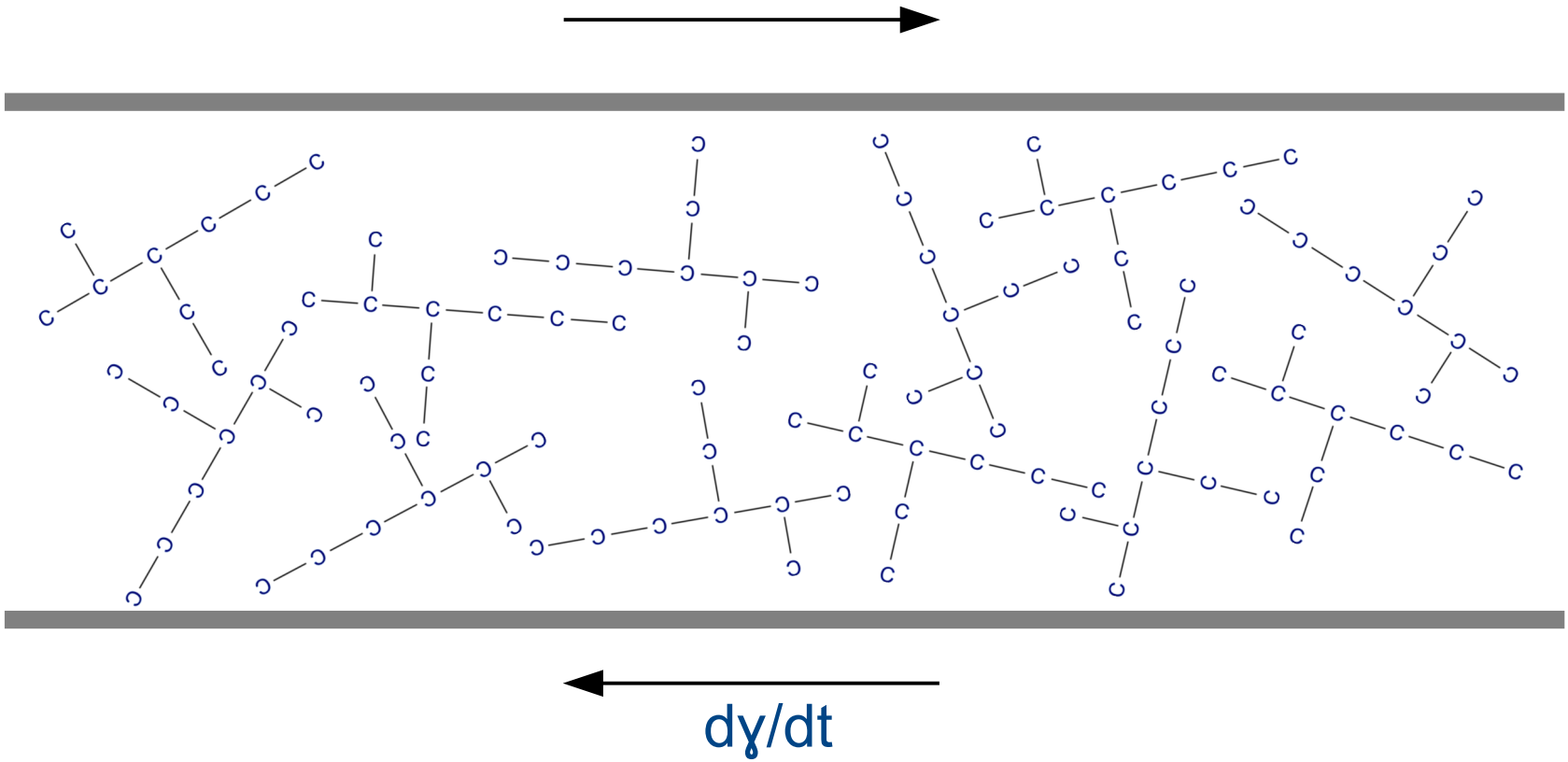
Neural network R^2=0.998

Free volume theory R^2=0.899

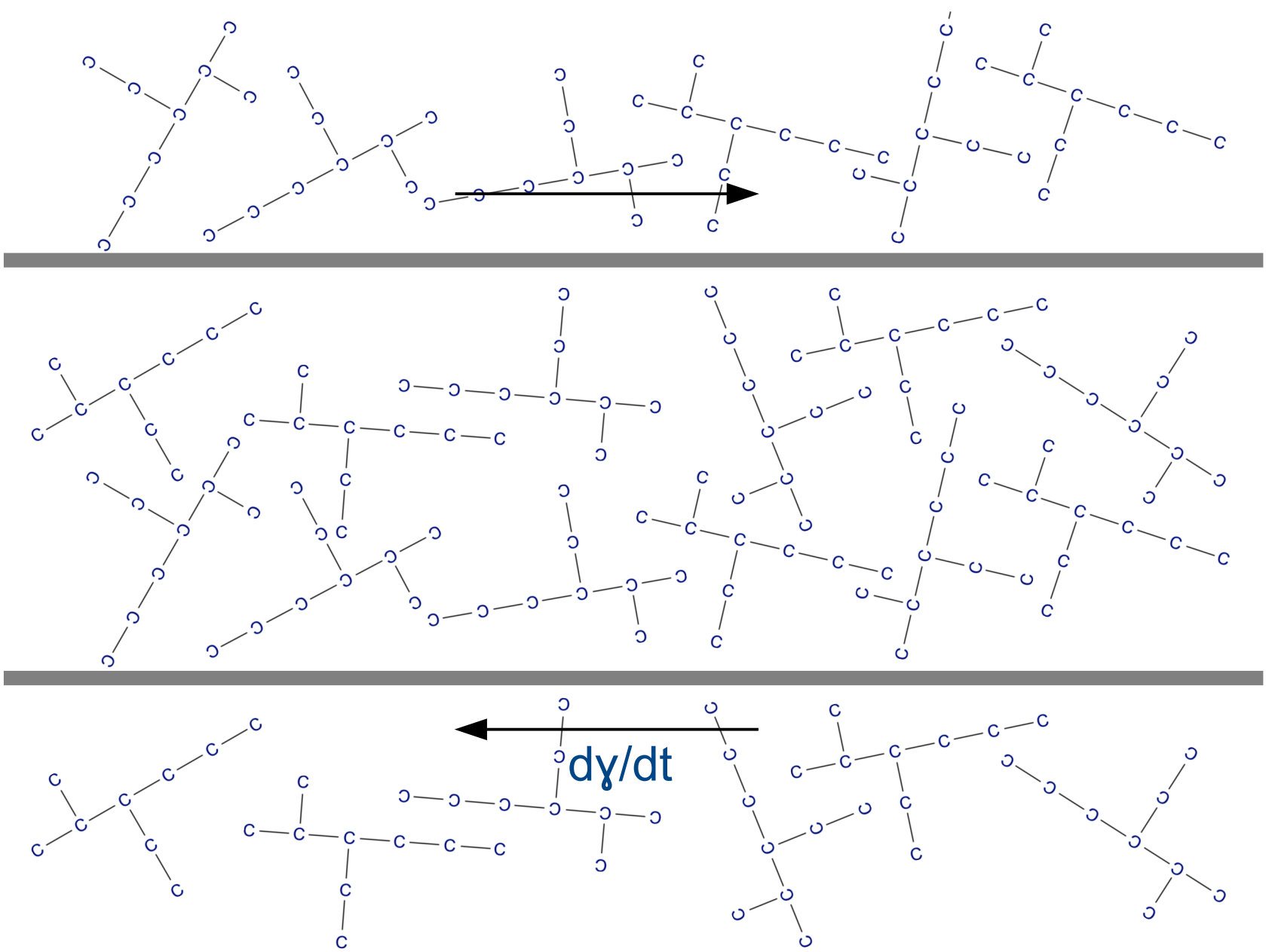
Summary of machine learning accuracies

Property	Machine learning	Alternate
Density	0.987	0.890
Heat capacity	0.996	0.995
Vapor pressure	0.962	-
Flash point	0.997	0.971
Melting point	0.998	0.991
Boiling point	0.992	0.977
Kinematic viscosity	0.998	0.899

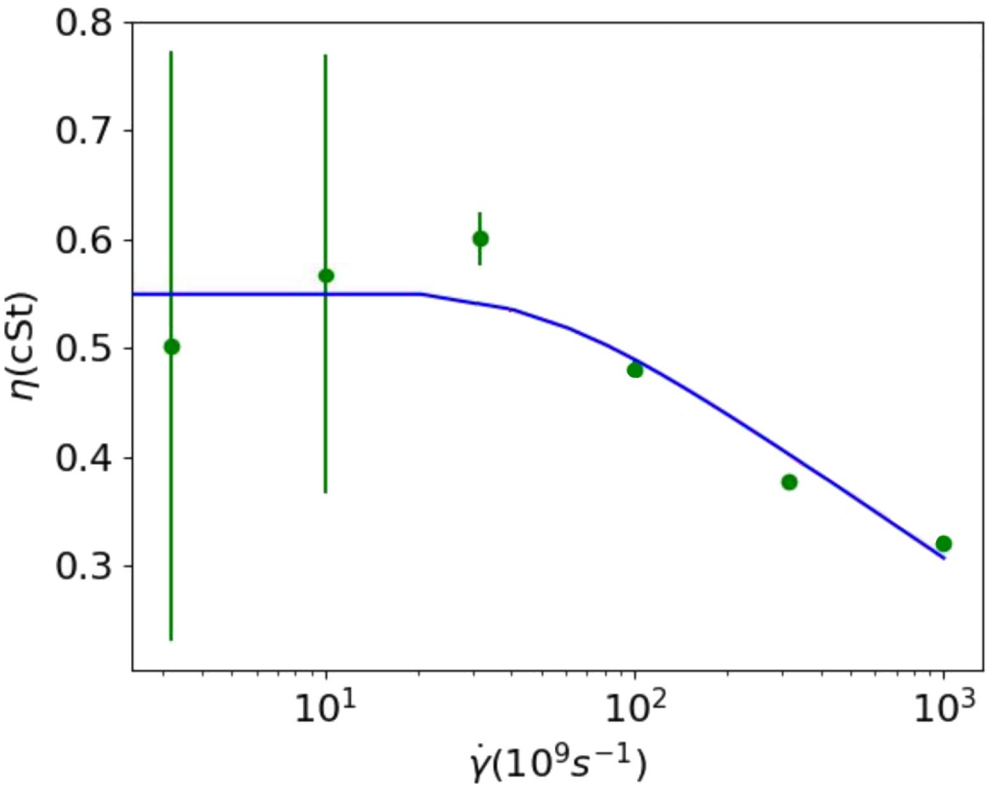
Non-equilibrium molecular dynamics



Periodic moving boundary conditions



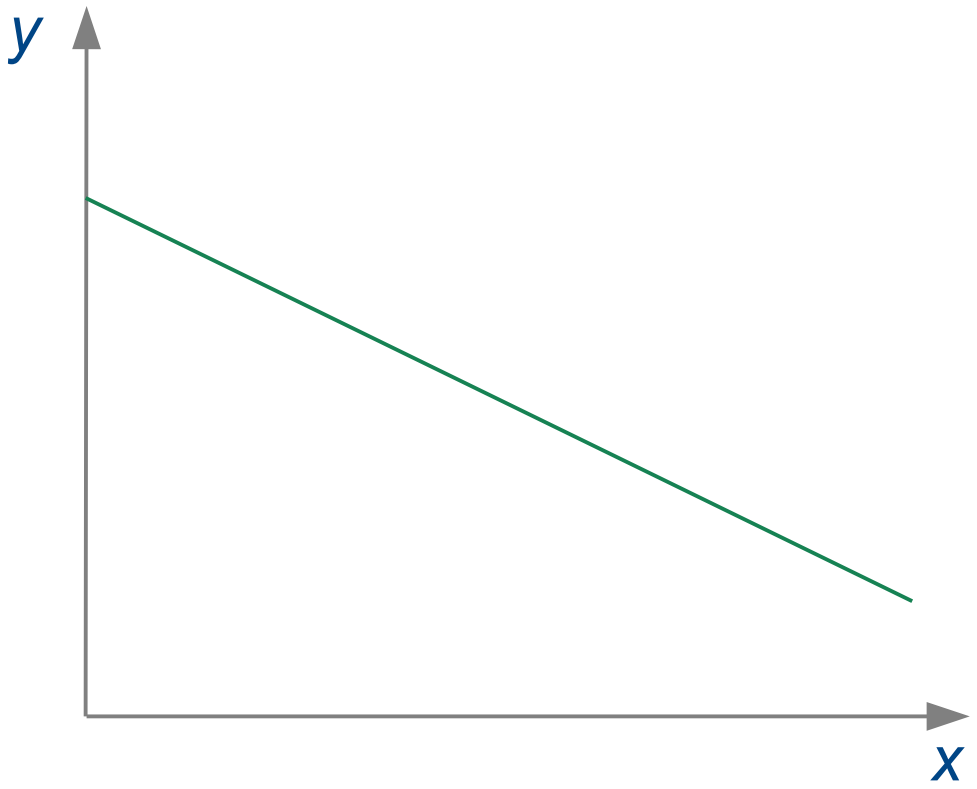
Kinematic viscosity of hexane



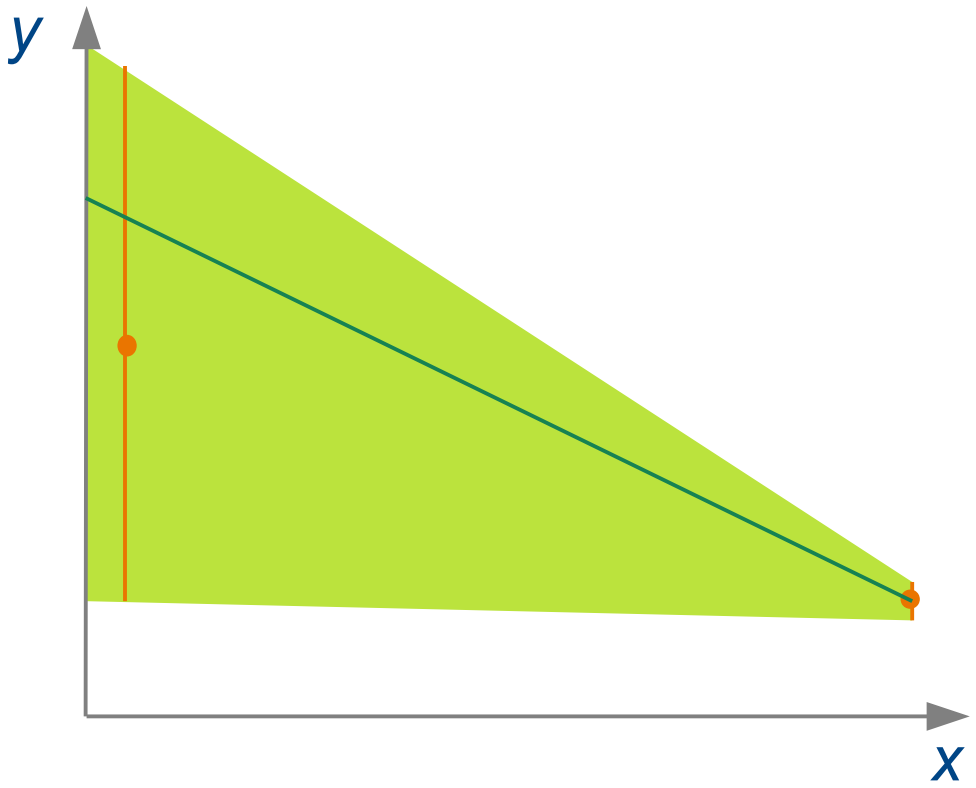
Increasing uncertainty on approaching zero shear rate

Carreau model guides extrapolation to zero shear rate

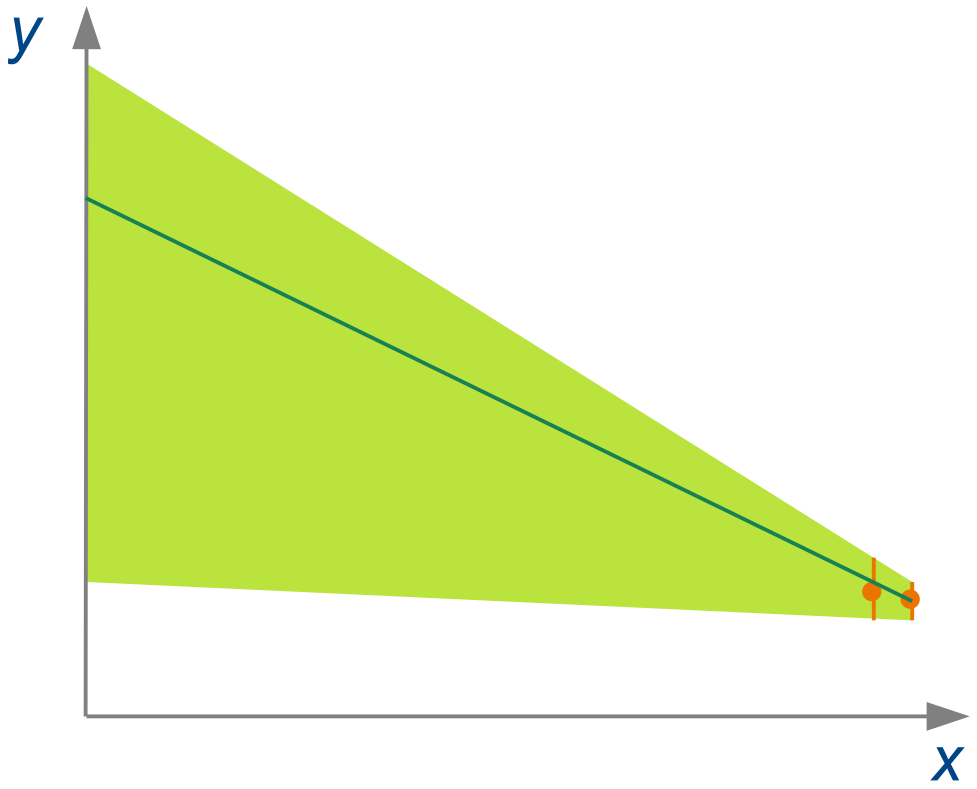
Extrapolating with a straight line to $x \rightarrow 0$



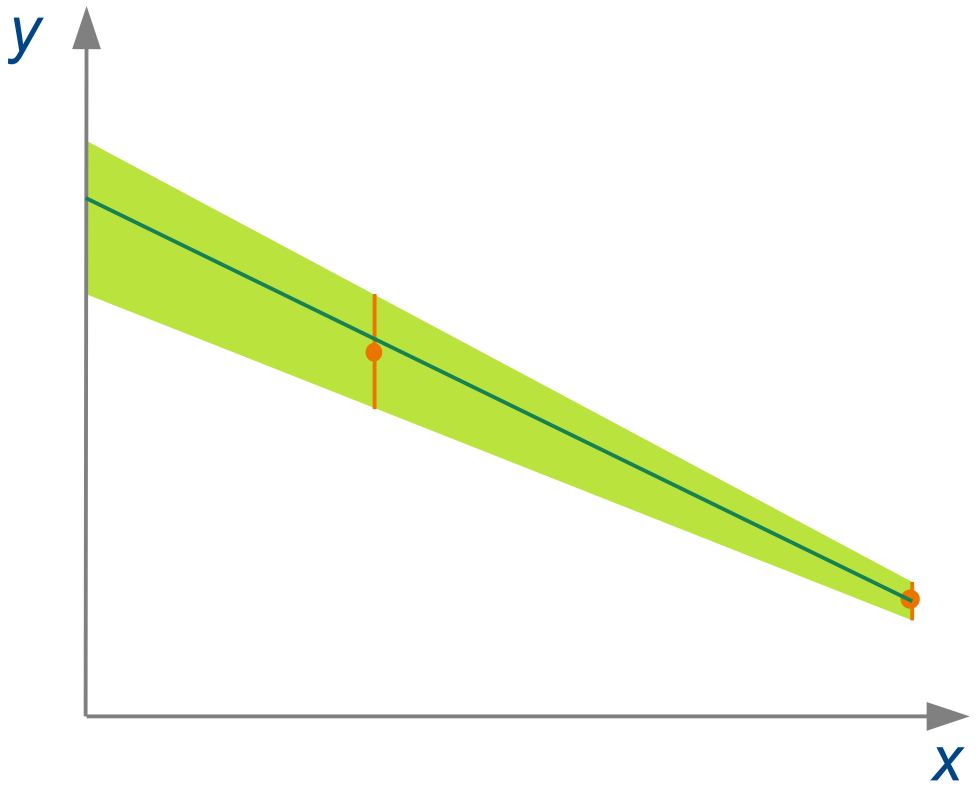
Use calculations far apart



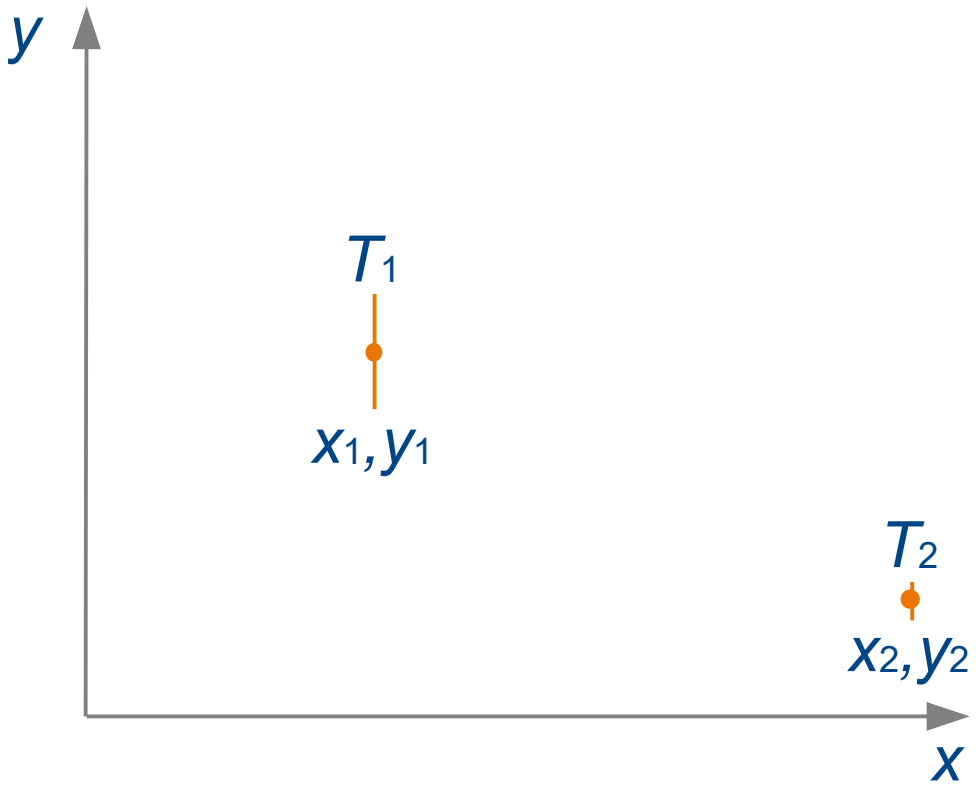
Use calculations near together



Intermediate spacing of calculations

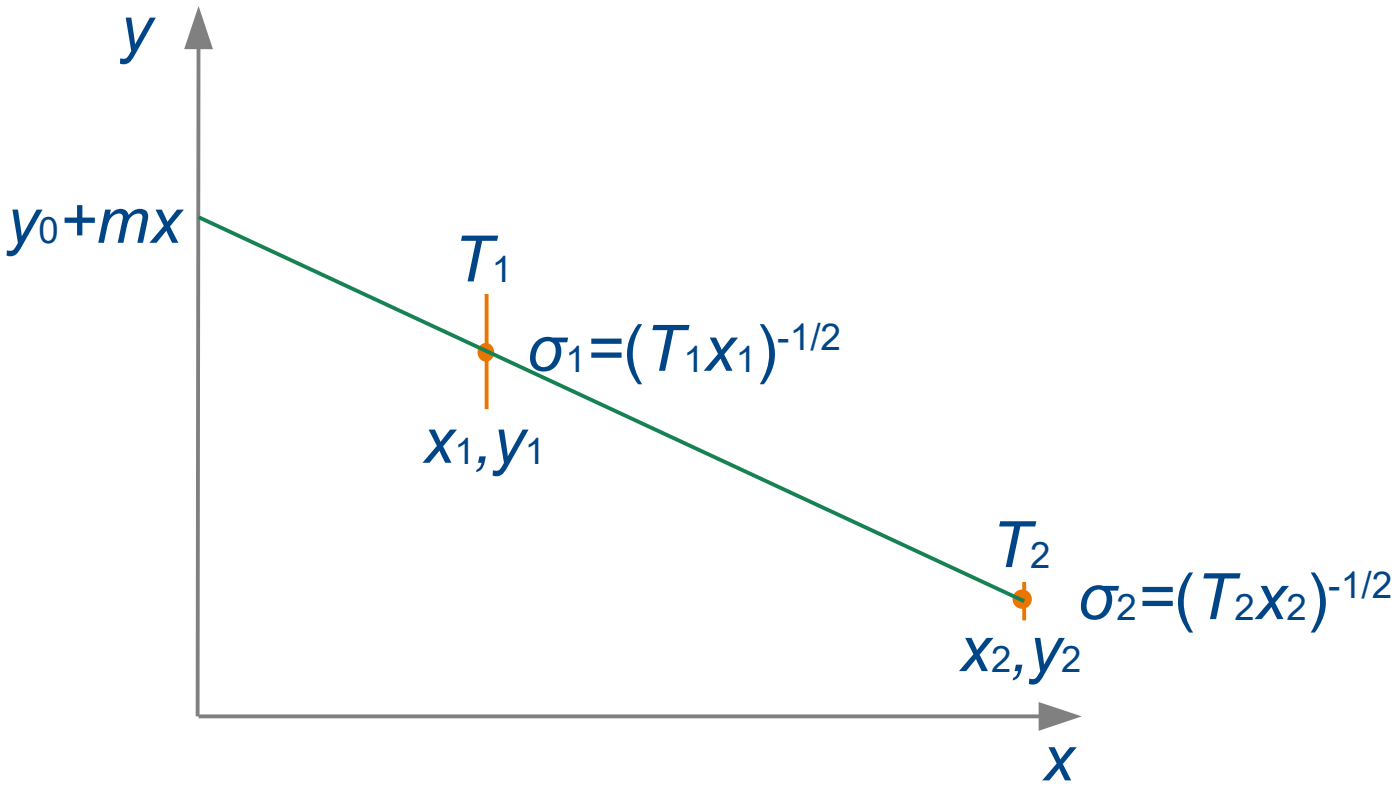


Optimize position of calculations



Setting up optimization problem

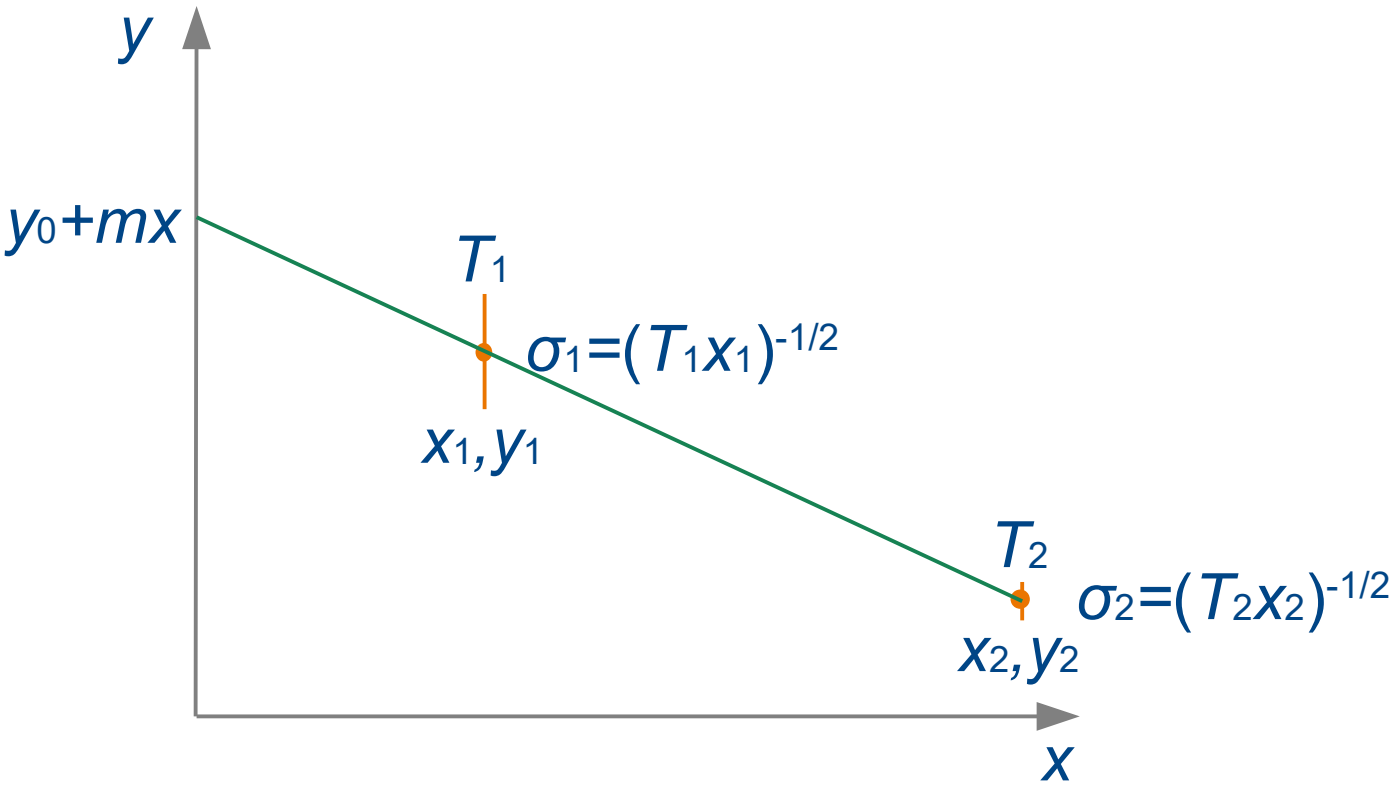
$$\chi^2=(y_0+mx_1-y_1)^2/\sigma_1^2+(y_0+mx_2-y_2)^2/\sigma_2^2$$



Fit the straight line

$\chi^2=(y_0+mx_1-y_1)^2/\sigma_1^2+(y_0+mx_2-y_2)^2/\sigma_2^2$

Solve $d\chi^2/dy_0=0$ and $d\chi^2/dm=0$ to deliver y_0, m, and uncertainty in y_0

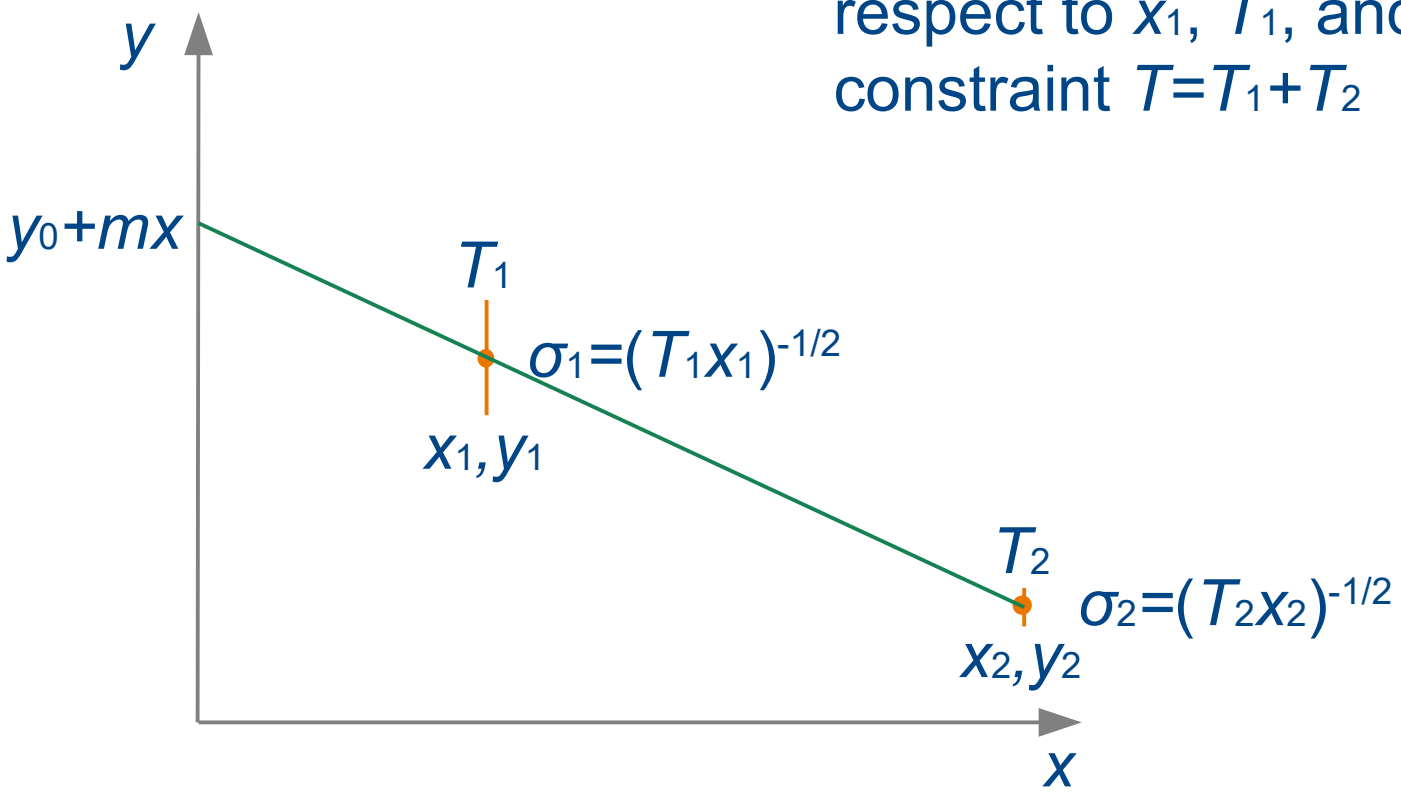


Solving optimization problem

$\chi^2=(y_0+mx_1-y_1)^2/\sigma_1^2+(y_0+mx_2-y_2)^2/\sigma_2^2$

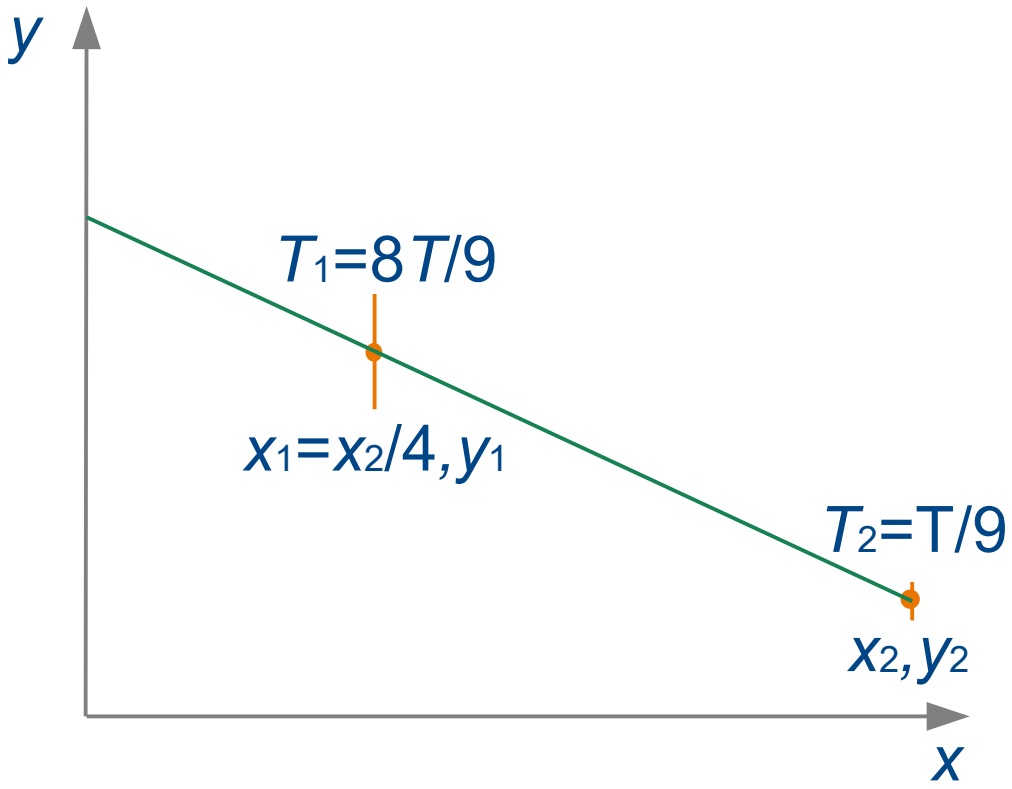
Solve $d\chi^2/dy_0=0$ and $d\chi^2/dm=0$ to deliver y_0, m, and uncertainty in y_0

Minimize uncertainty in y_0 with respect to x_1, T_1, and T_2 with constraint $T=T_1+T_2$

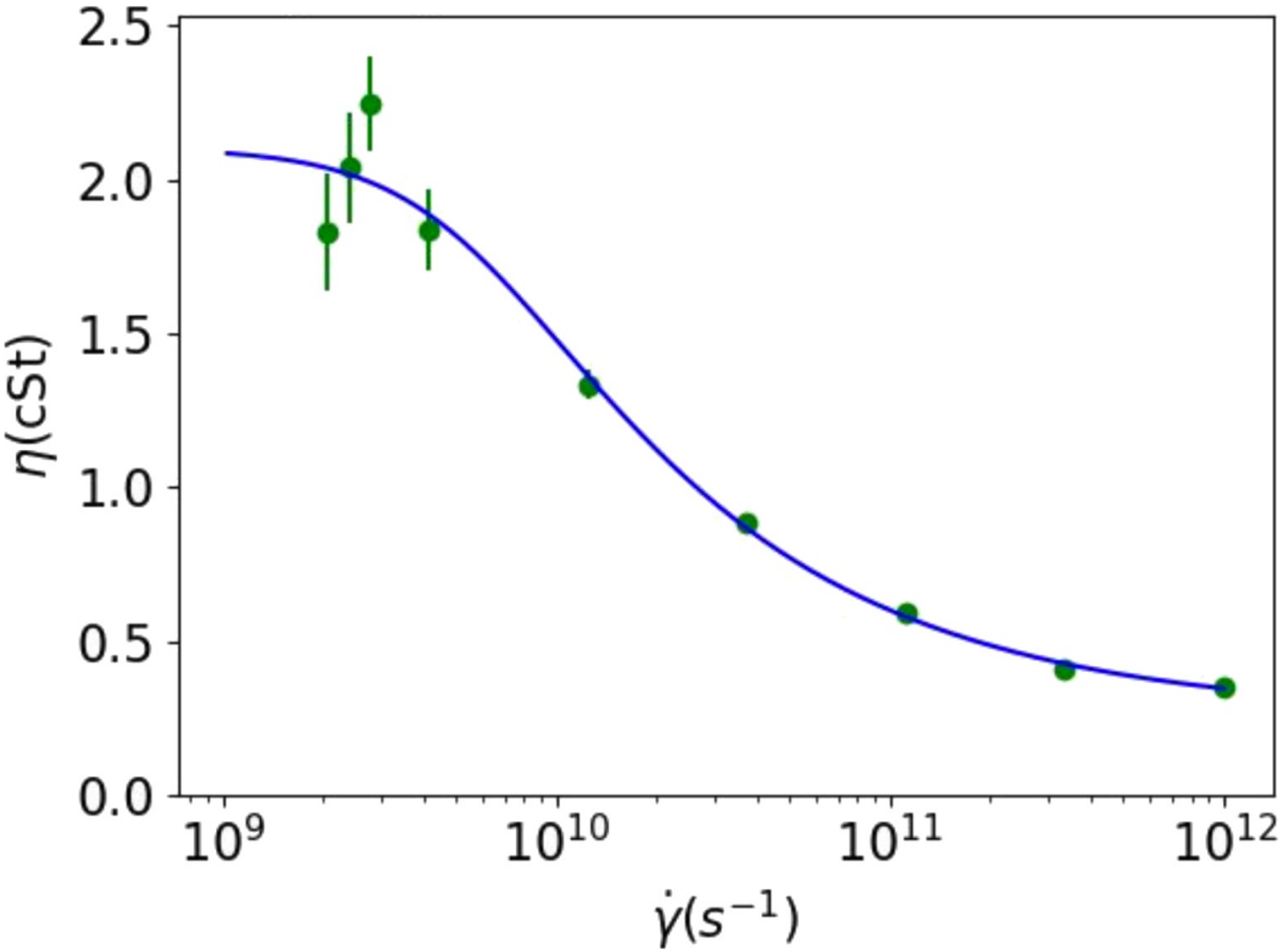


Optimal strategy

Developed for Diffusion Monte Carlo in
R.M. Lee, G.J. Conduit, N. Nemec,
P. López Ríos & N.D. Drummond
Phys. Rev. E 83, 066706 (2011)



Viscosity of hexadecane at 60°C



Viscosity of straight chain alkanes

Hydrocarbon	20°C	40°C	60°C	70°C	100°C
Hexane	0.52 (0.56)	0.37 (0.37)	0.33 (0.34)	-	-
Decane	1.23 (1.27)	1.01 (1.00)	0.83	0.61	0.52
Tetradecane	2.93 (3.06)	2.04	1.51	1.27	1.02
Hexadecane	4.49 (4.51)	2.94 (2.92)	2.19 (2.06)	-	1.45 (1.41)

Comparison to other approaches

Hydrocarbon	20°C	40°C	60°C	70°C	100°C
Hexane	0.52 (0.56)	0.37 (0.37)	0.33 (0.34)	-	-
Decane	1.23 (1.27)	1.01 (1.00)	0.83	0.61	0.52
Tetradecane	2.93 (3.06)	2.04	1.51	1.27	1.02
Hexadecane	4.49 (4.51)	2.94 (2.92)	2.19 (2.06)	-	1.45 (1.41)

Our molecular dynamics R^2=0.95

Other molecular dynamics R^2=0.69

Summary

Model many physically relevant properties of lubricants with machine learning

Optimal strategy for non-equilibrium molecular dynamics simulations of viscosity

P. Santak & G.J. Conduit Fluid Phase Equilibria 501, 112259 (2019)