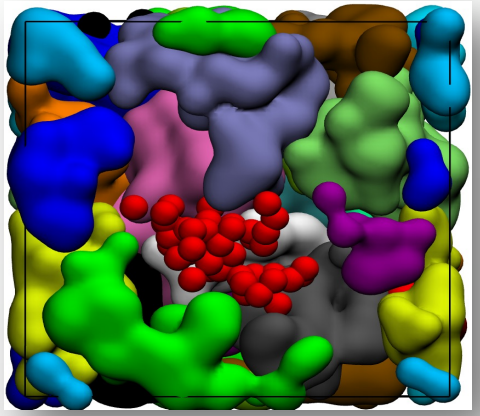
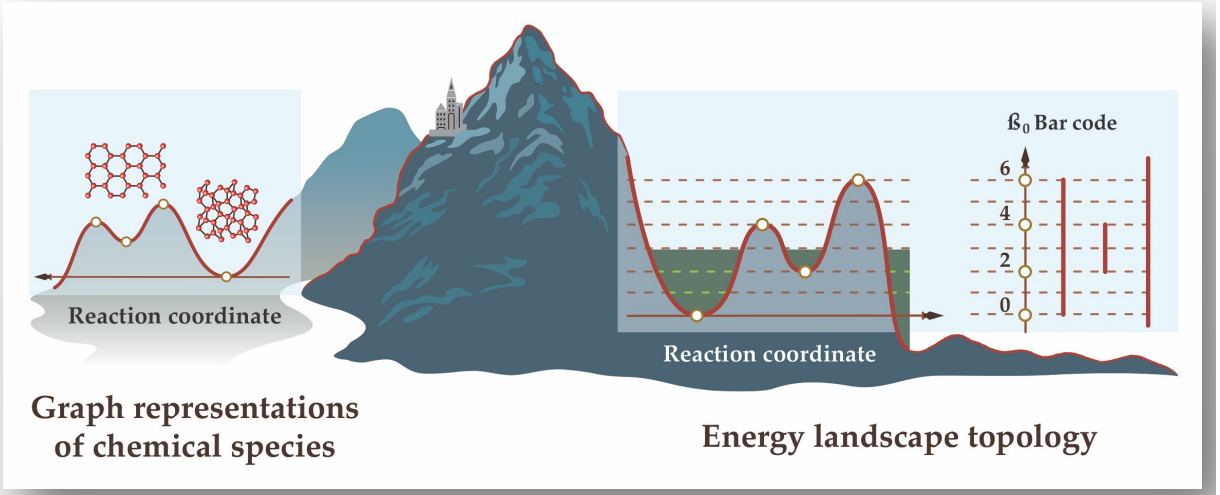
A Holistic, High Dimensional Perspective on Extracting and Encoding Chemical Data

**a problem for soft matter materials design





Aurora Clark

Department of Chemistry

University of Utah

Current Group members:

*Jackson Elowitt**
*Joshua Bilsky**
Simantini Paul
Yihui Wei
*Leonardo Liet**
Qing Guo
*Swarup Banerjee**

Former post-docs:

*Michael Servis**
*Abdullah Ozkanlar**
Catherine Moulder
Maxime Pouvreau
Lance Edens
*Zhou Liu**
Changyong Qing
Matthew Wander
David Semrouni
Jian Wu
Hung Le
Rafal Gorniak
*Biswajit Sadhu**
Andrew Launder
Mostafa Kuccukkal
Payal Parmar
Jadwiga Kuta

Former PhD students

Will Smith *Tiecheng Zhou** *Enrique Alvarado*** *Nitesh Kumar** *Alex Samuels*
Dan Pope *Yasaman Ghadar* *Nathan May ***** *Bojana Ginovska* *Ernesto Martinez*
*Chun-Hung Wang** *Sujala Bhattarai* *Morgan Kelley*

An Integrated Math + Chemistry Team

Henry Adams
Colo. State

Brittany Story
CSU/U Tenn (PhD stud)

Biswajit Sadhu
BARC (post-doc)

DELTA
Data Science + Topology = Chemistry

NSF

Harnessing the Data Revolution (HDR) Frameworks – Clark PI

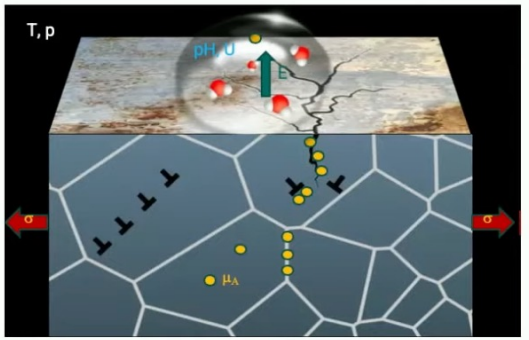
Complementarity With Prior Talks

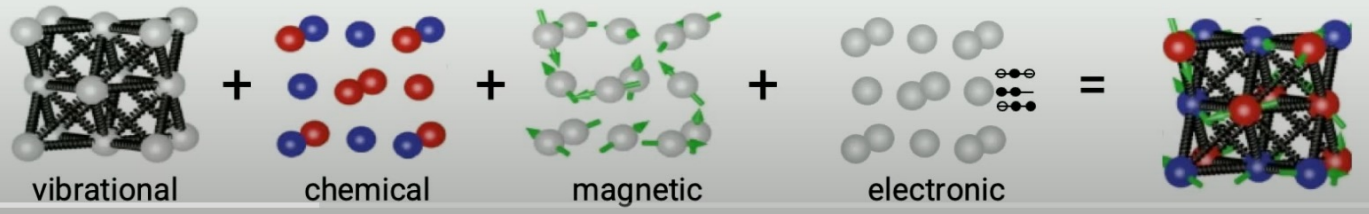
Jörg Neugebauer – Enabling next-generation materials science simulations by automated workflows

Methodological challenges

Exploring and navigating high-dimensional configuration spaces

$$Z(V,T,x,\alpha_{phase}) = \left\langle e^{-E^{QM}\left(\{\vec{R}_I,Z_I,\sigma_I,f_i,\ldots\}_V\right)/k_BT} \right\rangle_{V,T,x,\alpha_{phase}}$$





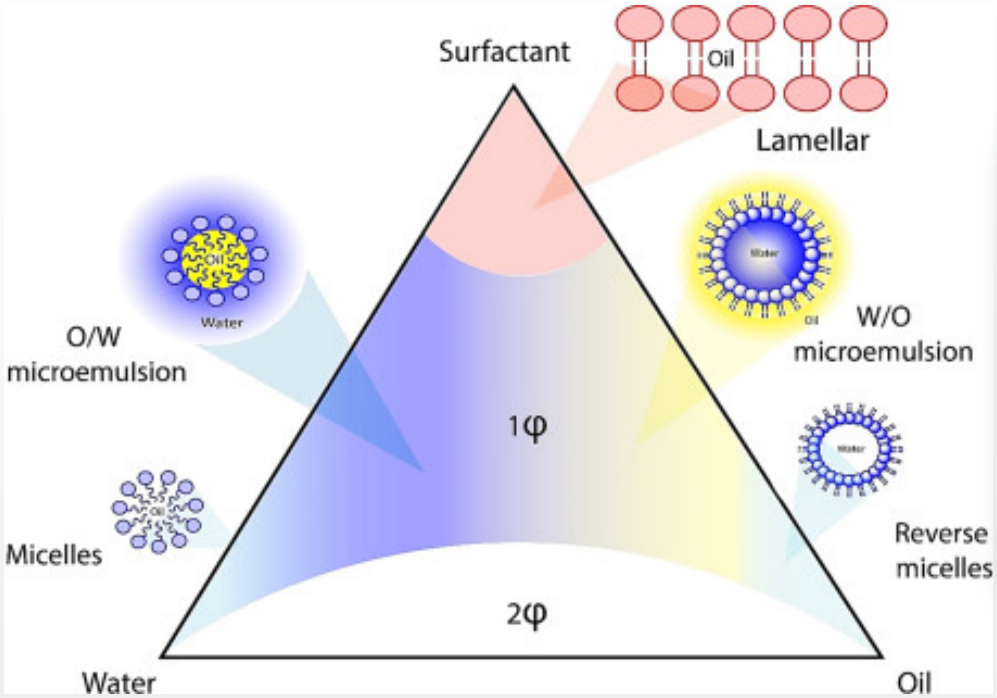
3:23 / 51:39

Partition Function Encodes State Dist. and Energy

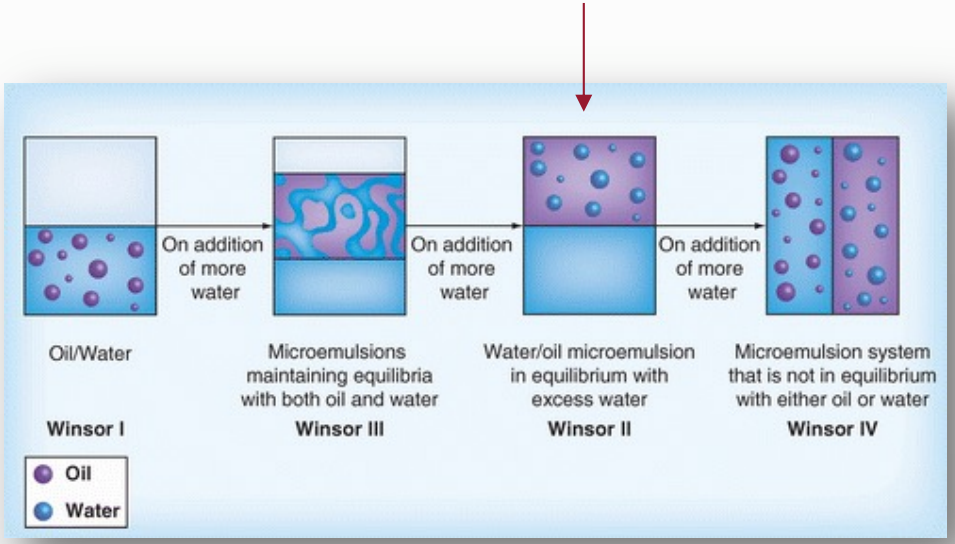
- In solid/crystalline phase – contributors to *Z* directly connect structure and energy distributions (vib, rot, elect, config contributions)
- What about soft matter? (ignoring liquid crystals and other weird things)
  - Liquids, solutions often very heterogeneous
  - Interfacial regions (L/L interfaces)
  - Gels
- Intrinsically disordered
  - Varying correlation lengths in space. (configuration states harder to define)
  - Correlation times – how easy to escape out of minima
- Often the dynamical phenomena are a key property of the material

Soft Matter Ensembles - Emulsions

- Very complex and rich multicomponent phase diagrams

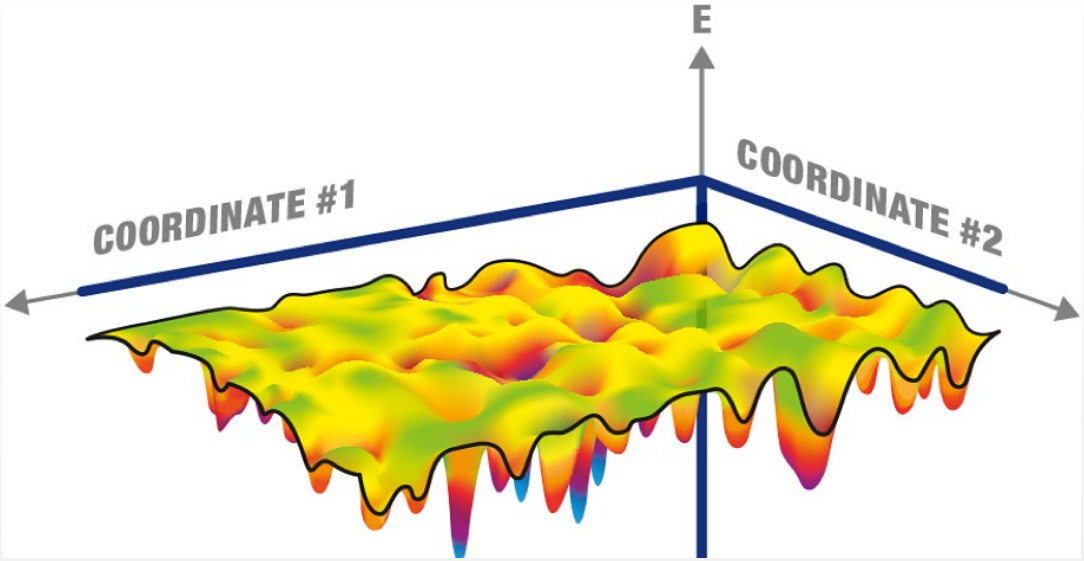


Transport key to separations science



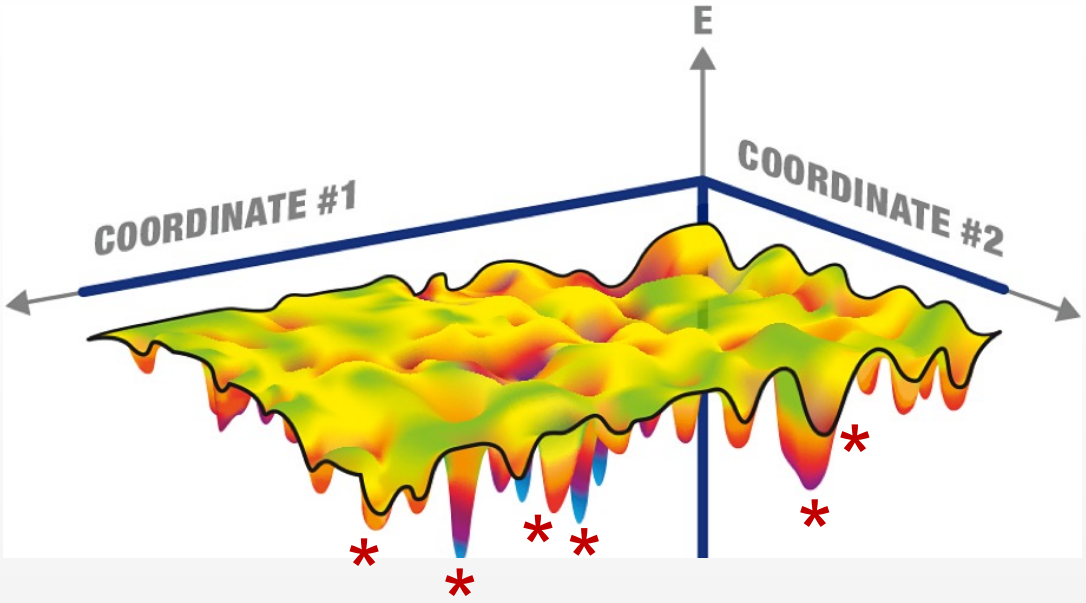
An Energy Landscape Perspective

***let's be agnostic to the type of EL at this point



Find the global minimum using some approximate Hamiltonian

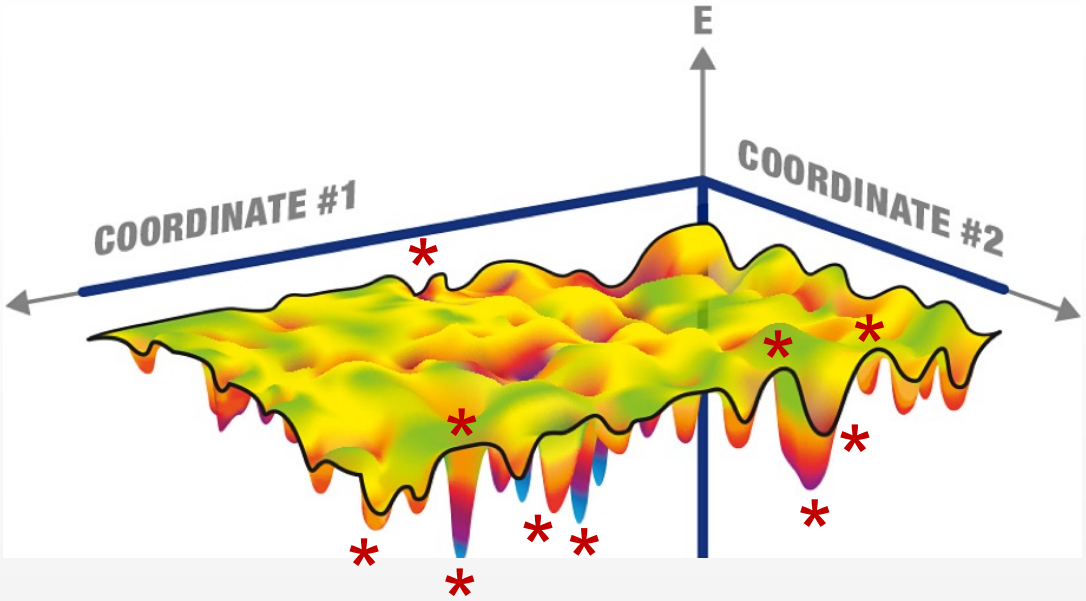
An Energy Landscape Perspective



Sampling configurational space - finding all relevant local minima on the EL

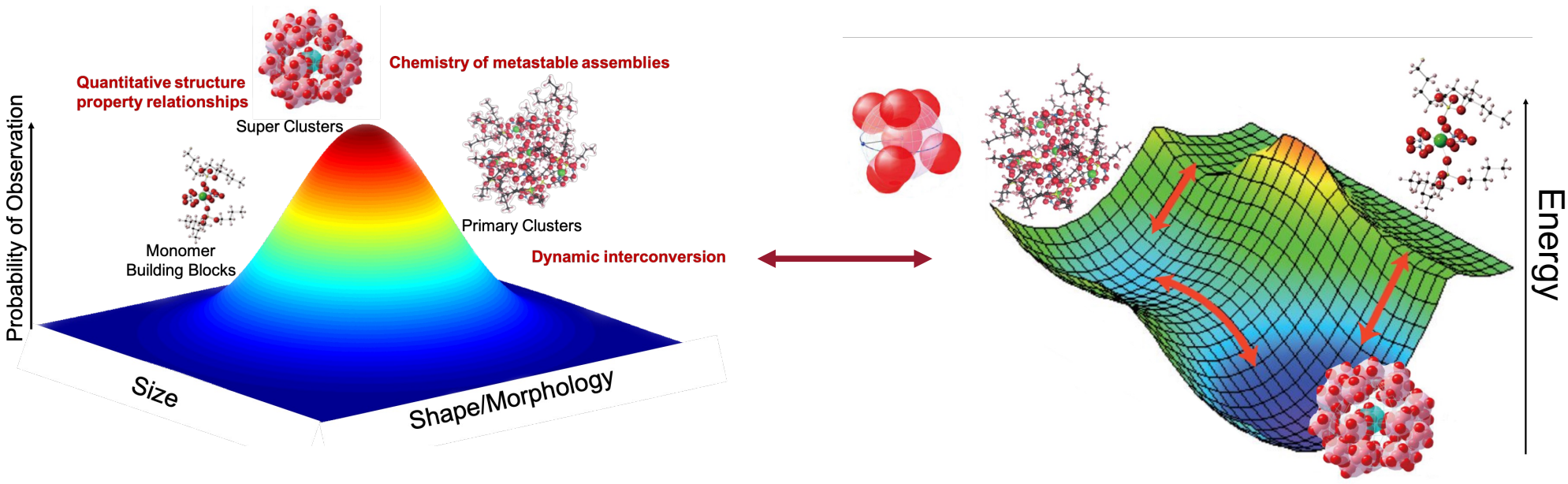
-index 0 critical points

An Energy Landscape Perspective



Sampling the fluctuations means sampling minima and maxima
-including index 1 CP's and higher order CP's for some processes

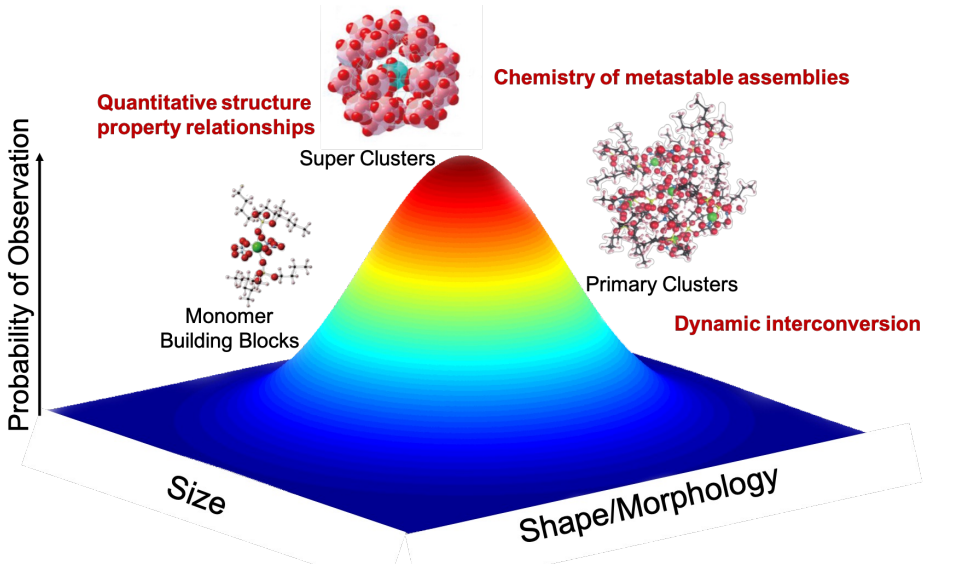
Key Challenges for Soft Matter Design

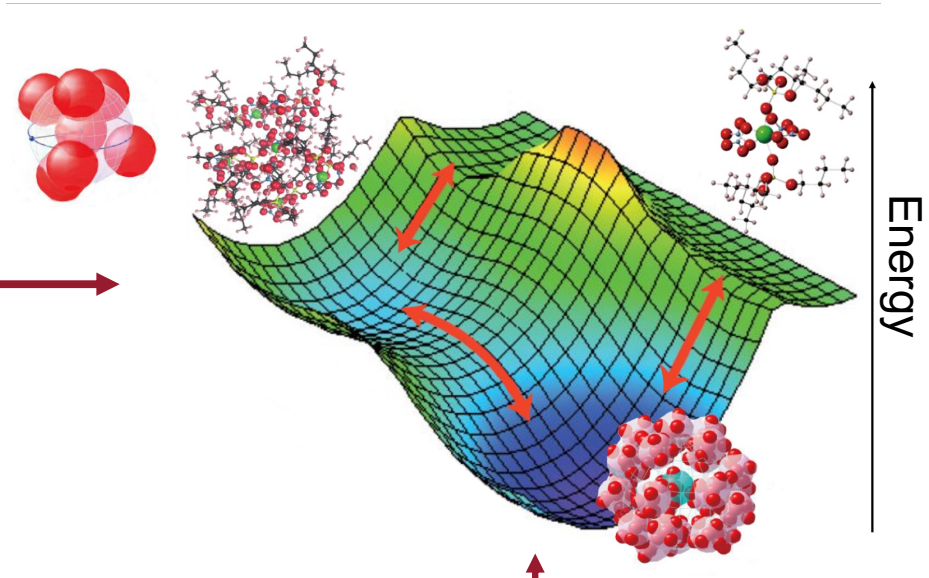


If we sample configuration space (using molecular dynamics)

- How to define states?
- How to relate to energy landscape? (choosing higher dimensional representations)

Key Challenges for Soft Matter Design





This is a lie

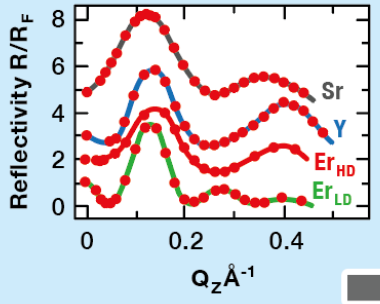
If we sample configuration space (using molecular dynamics)

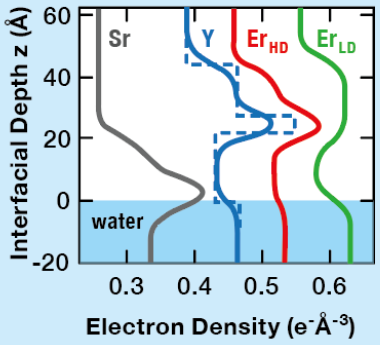
- How to define states?
- How to relate to energy landscape? (choosing higher dimensional representations)

Soft Matter Design Workflow

Simulations: atomistic & coarse grained

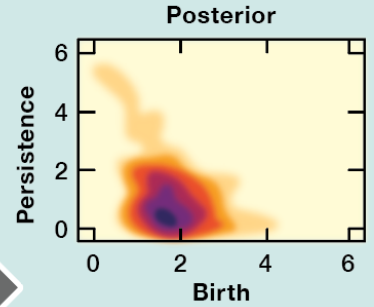
Synthesis and characterization





?

ML w/ UQ

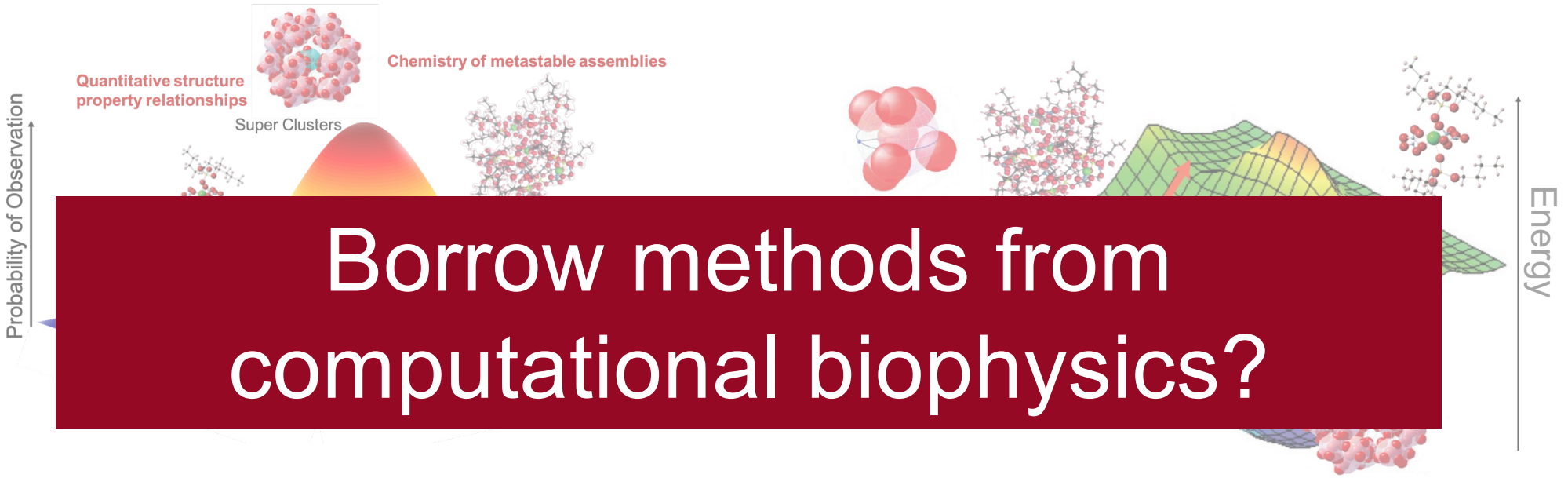


Compute the Bayes Factor

$$BF^{ij}(D) = \frac{p(D|D_{Y_i})}{p(D|D_{Y_j})}$$

DECISION

Key Challenges for Soft Matter Design

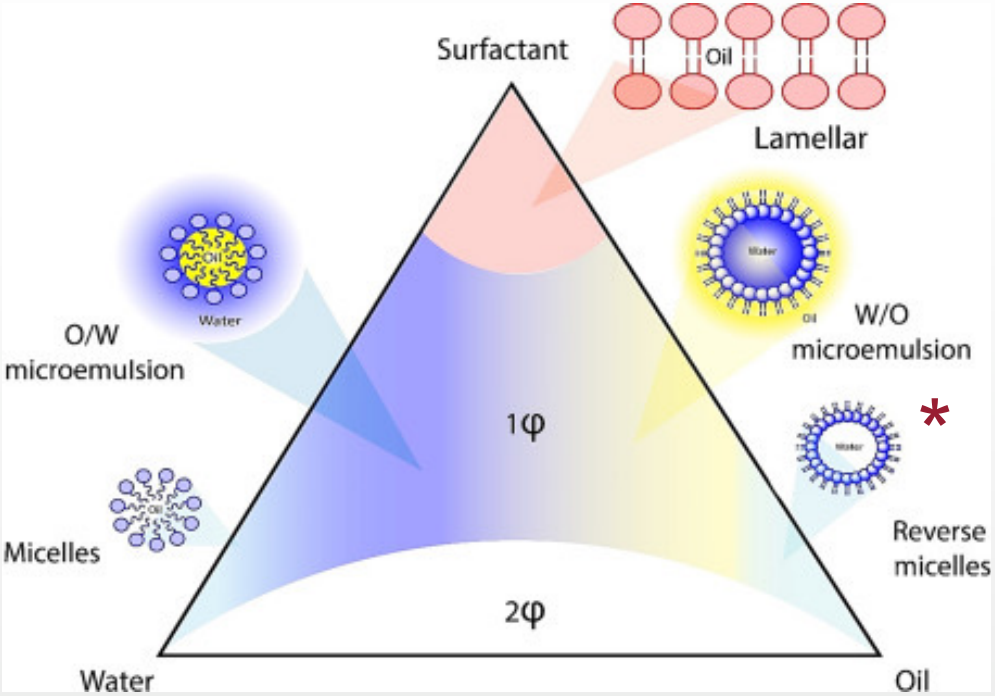


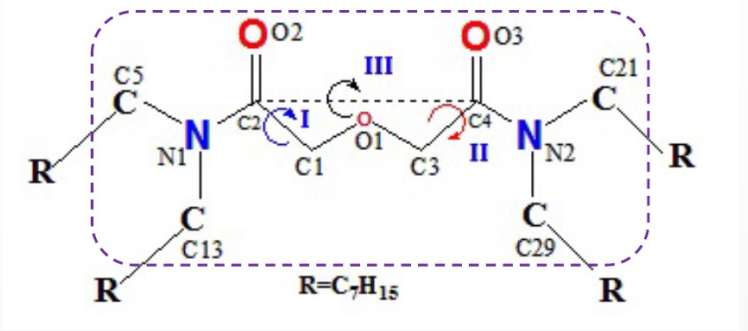
Borrow methods from computational biophysics?

If we sample configuration space (using molecular dynamics)

- How to define states?
- How to relate to energy landscape?

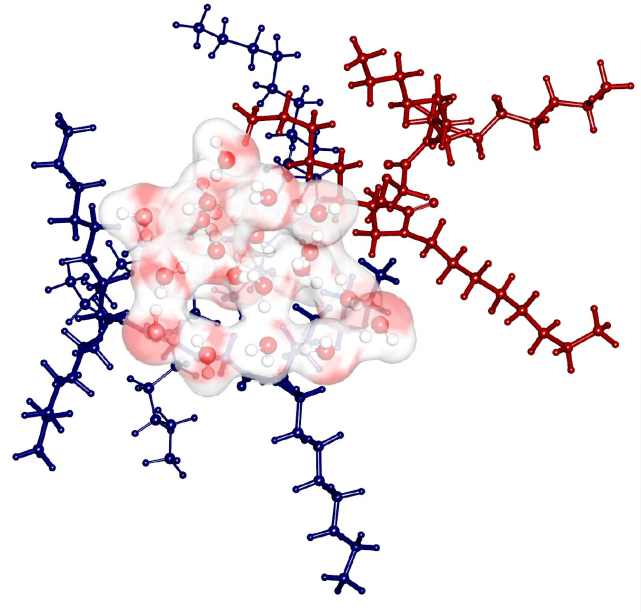
An Example Using Microemulsions



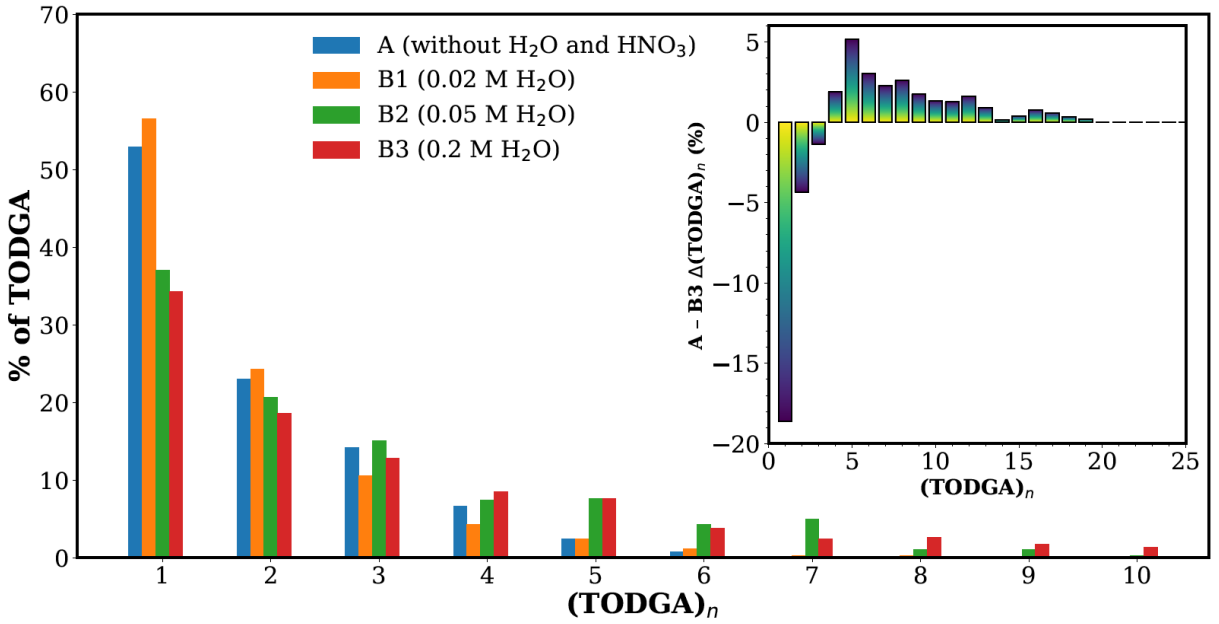


TODGA

Polar Cosolutes and Reverse Micelle Formation



- H_2O and HNO_3 increase TODGA micelle size



- What's molecular origin of this?
- Can we reduce dimension of the EL to be conceptually meaningful?

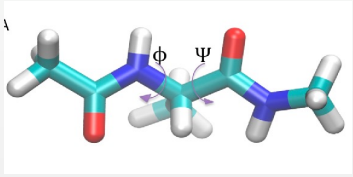
Markov Models

Finalizing Number of Microstates

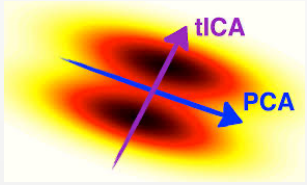
Coarse-graining of microstates into macrostates using PCCA++

Trajectory
- MD data

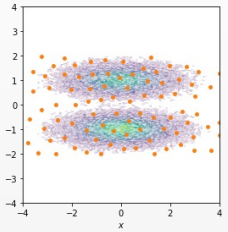
Featurization
- Order parameters



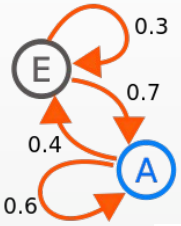
Dimension reduction
- TICA and/or PCA



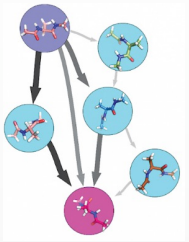
Discretization
- Clustering



Generate transition matrix for MSM



Analysis
- Thermodynamics



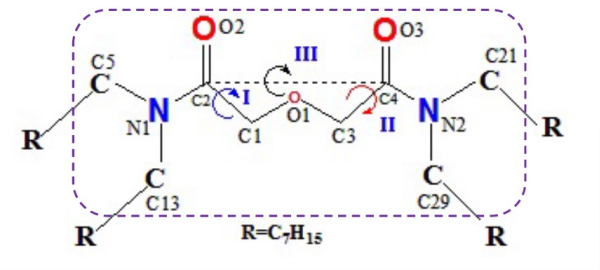
MSM BUILDER

http://msmbuilder.org/

http://emma-project.org/latest/

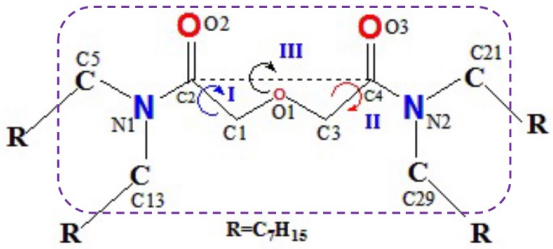
Sadhu & Clark, In Prep

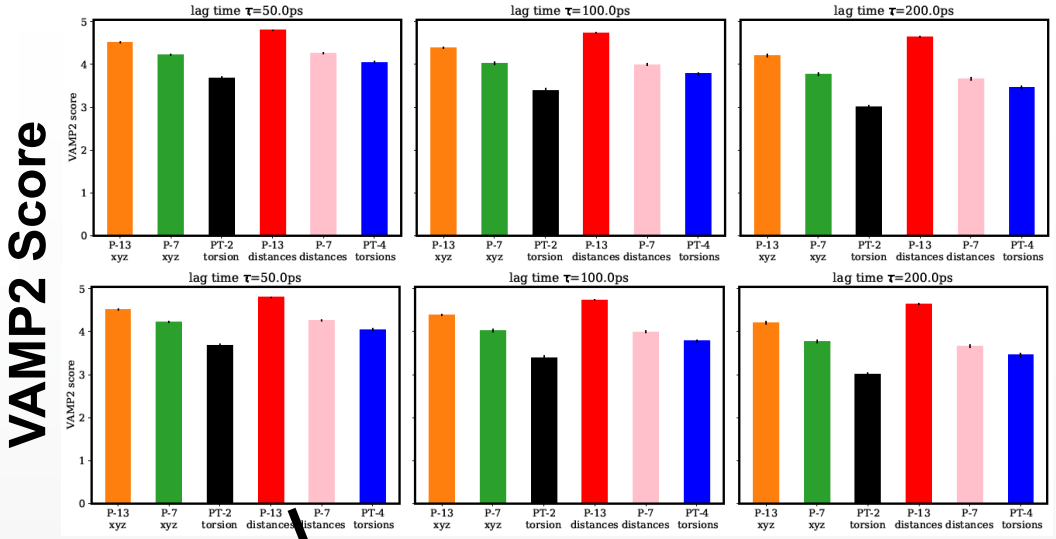
Feature Selection



Features	Feature descriptions	Features	Feature descriptions
P-13 xyz	xyz coordinates of 13 polar core atoms of TODGA	P-13 distances	Intramolecular distances involving 13 polar core atoms of TODGA.
P-7 xyz	xyz coordinates of 7 polar core atoms of TODGA	P-7 distances	Intramolecular distances involving 13 polar core atoms of TODGA.
PT-2 torsion	Two torsions	PT-4 torsions	Four torsions

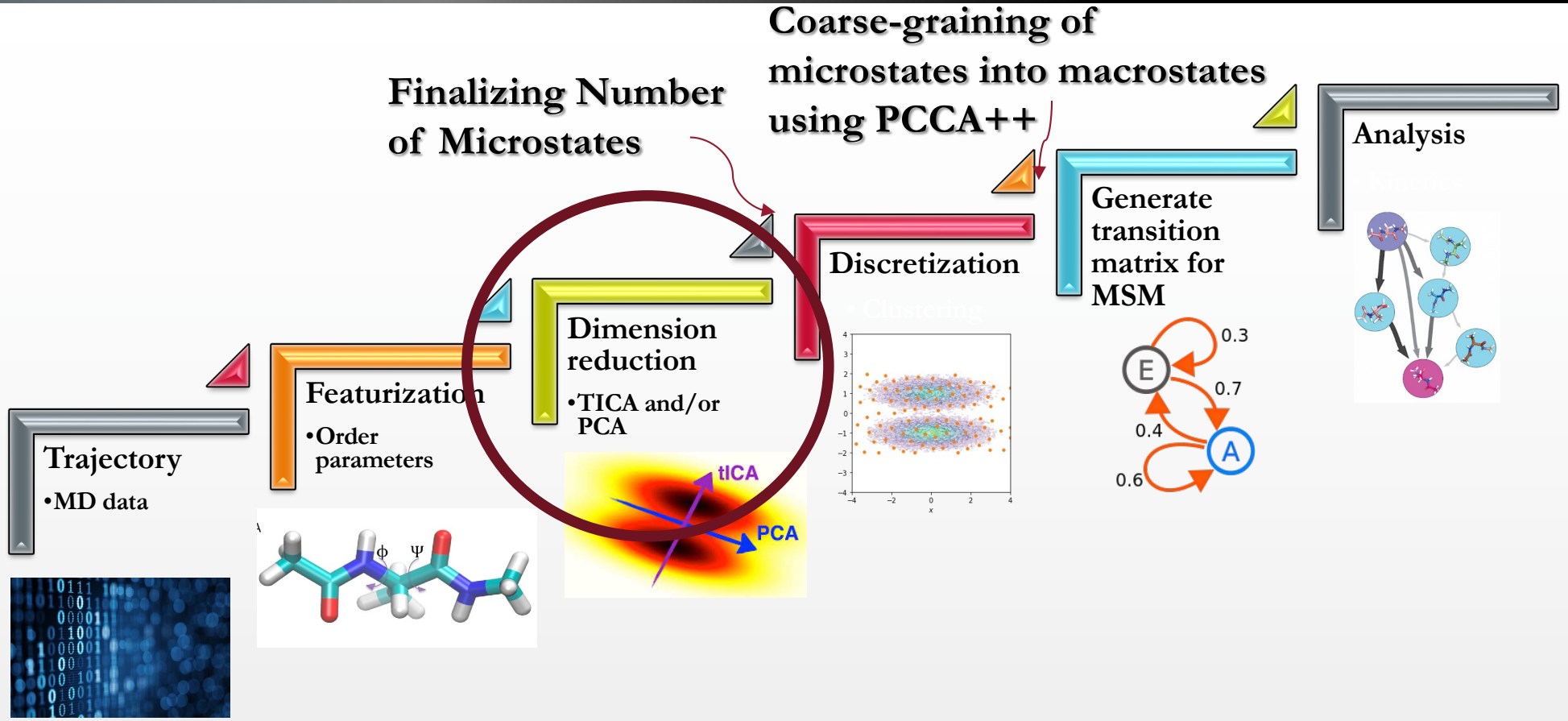
Feature Selection





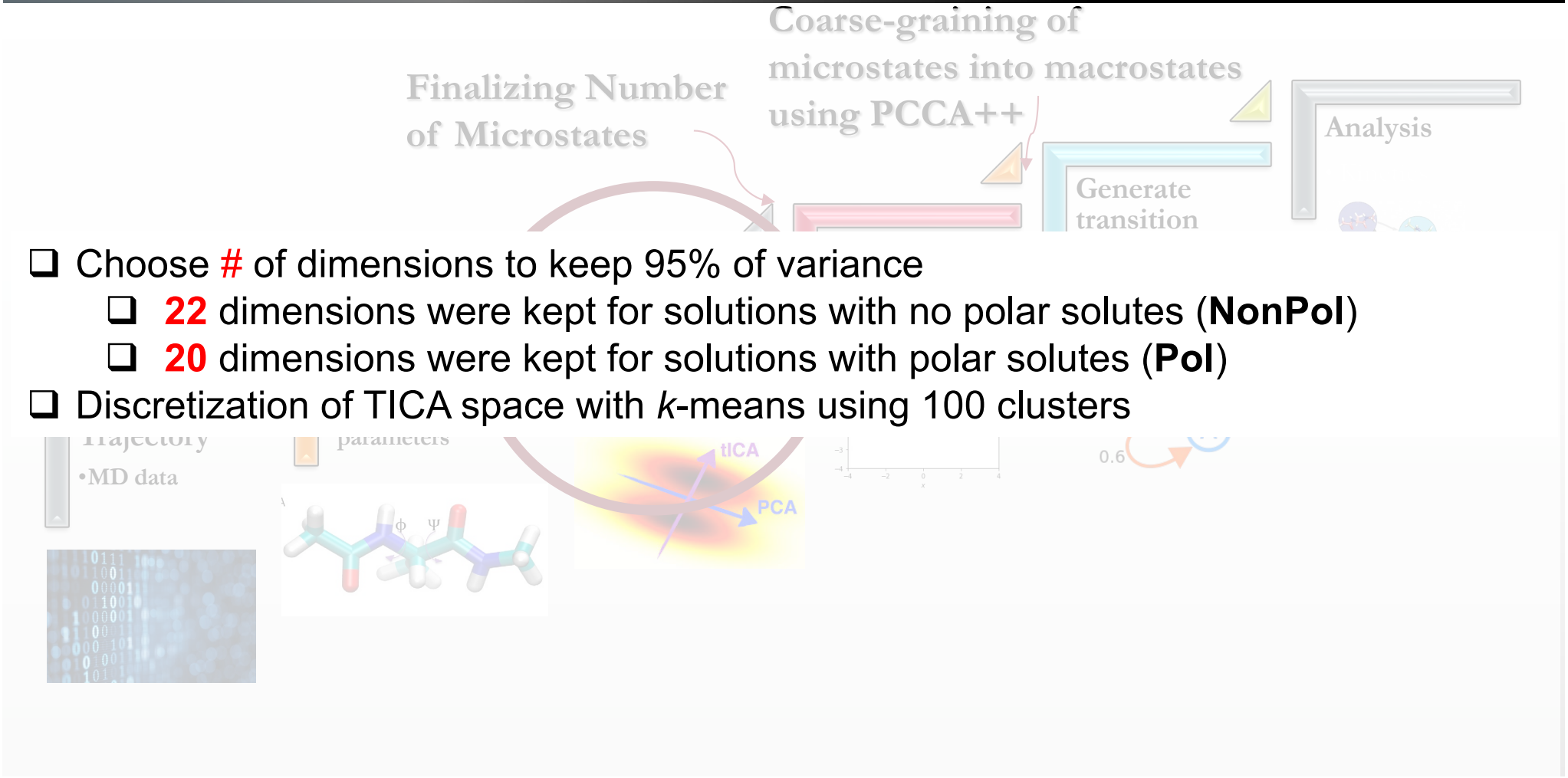
Features	Feature descriptions	Features	Feature descriptions
P-13 xyz	xyz coordinates of 13 polar core atoms of TODGA	*P-13 distances*	*Intramolecular distances involving 13 polar core atoms of TODGA.*
P-7 xyz	xyz coordinates of 7 polar core atoms of TODGA	P-7 distances	Intramolecular distances involving 13 polar core atoms of TODGA.
PT-2 torsion	Two torsions	PT-4 torsions	Four torsions

BMSM for Conformational Landscapes



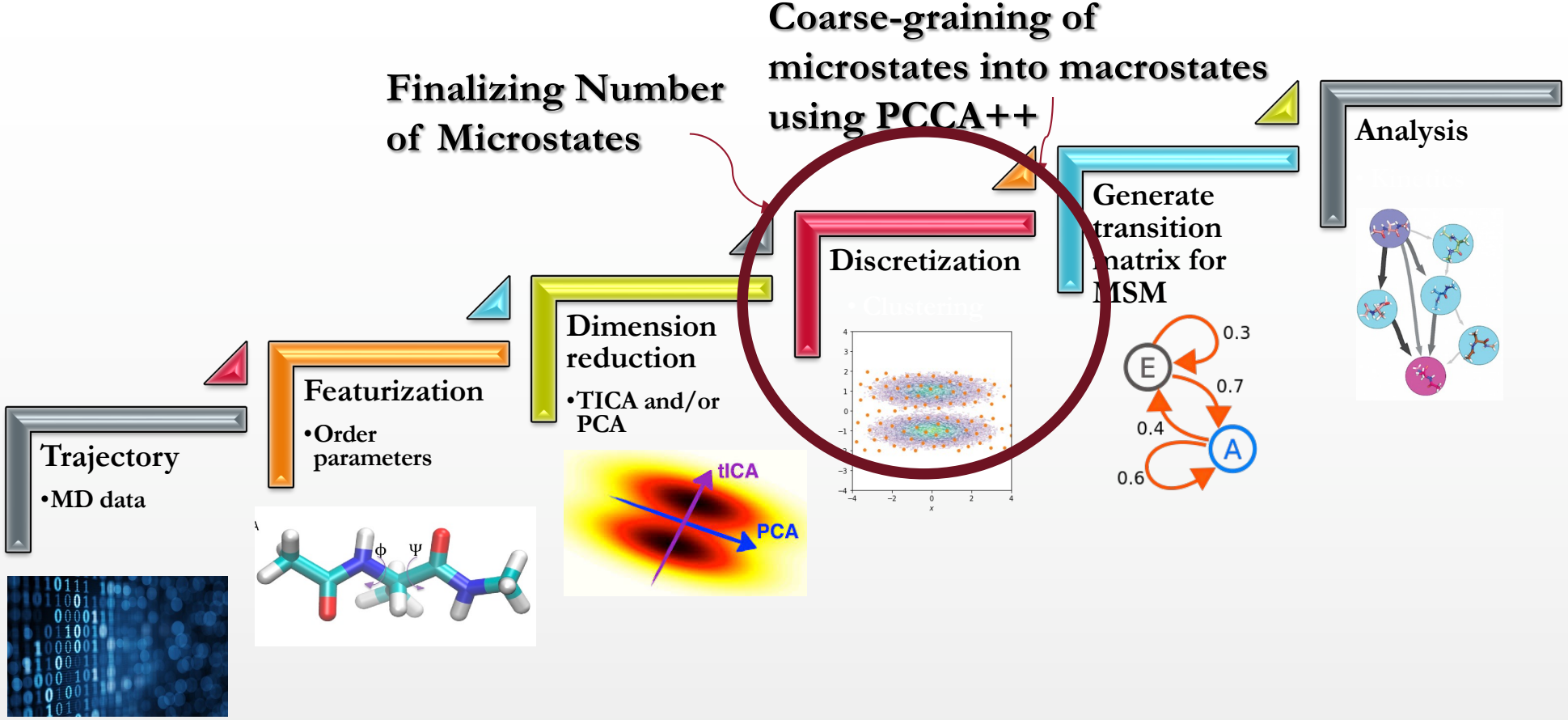
Sadhu & Clark, In Prep

BMSM for Conformational Landscapes



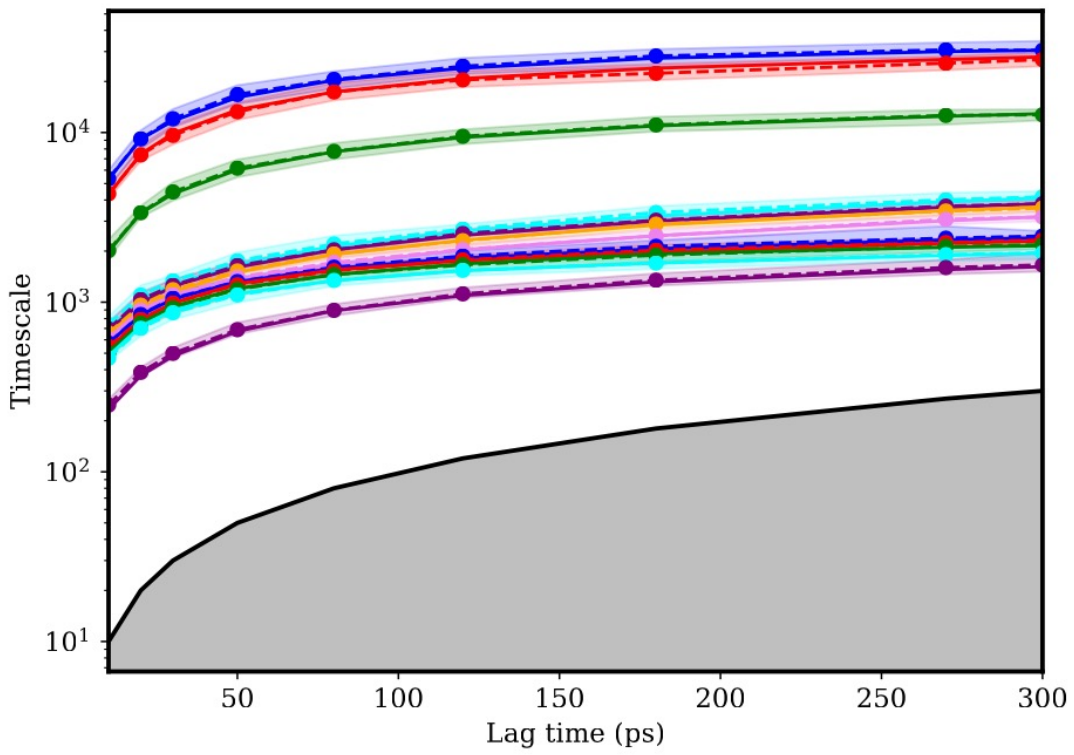
- Choose # of dimensions to keep 95% of variance
  - **22** dimensions were kept for solutions with no polar solutes (**NonPol**)
  - **20** dimensions were kept for solutions with polar solutes (**Pol**)
- Discretization of TICA space with *k*-means using 100 clusters

BMSM for Conformational Landscapes



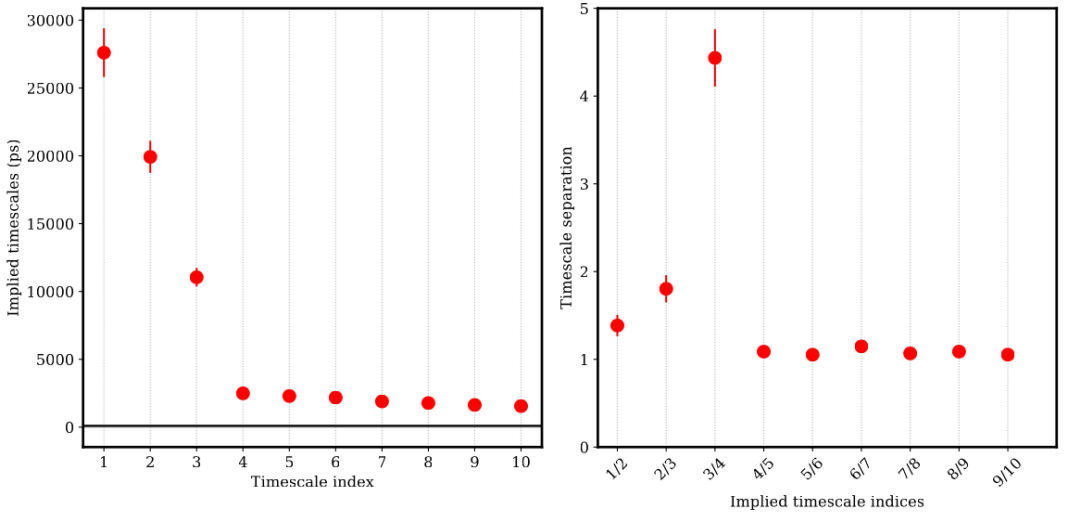
Sadhu & Clark, In Prep

BMSM for Conformational Landscapes



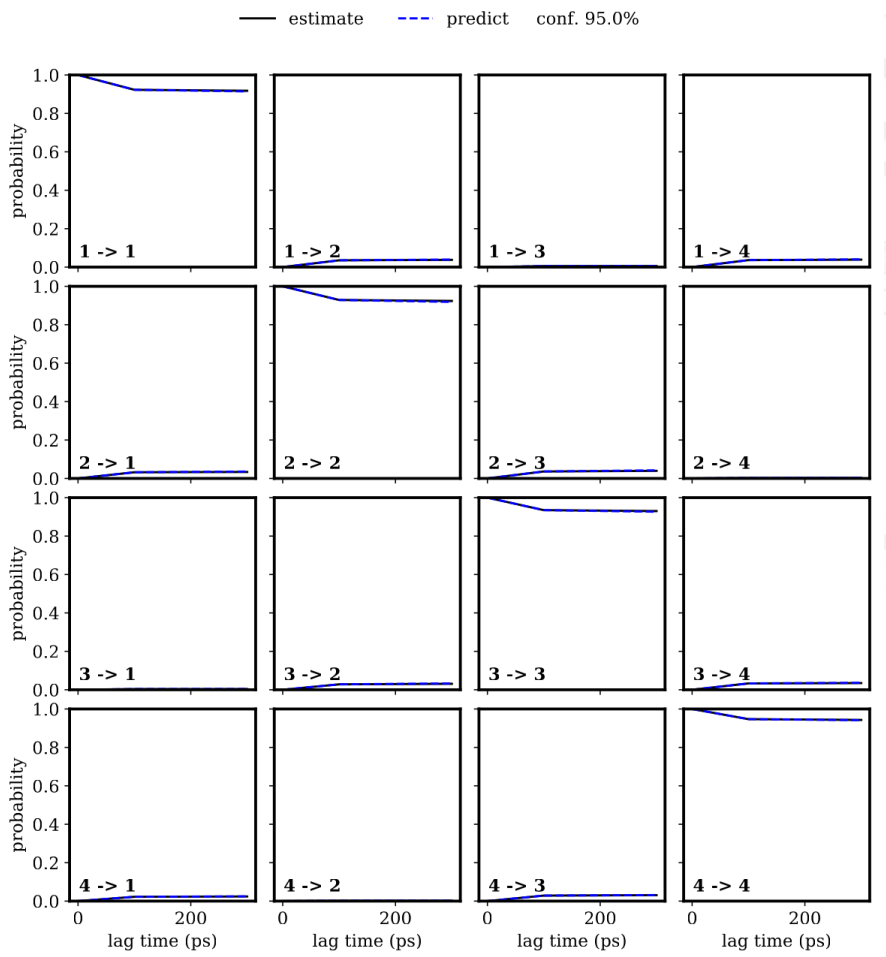
- Implied relaxation timescale used to validate lagtime in discretization
  - 50 ps for non pol
  - 100 ps for pol
- Separation of timescales
- Chapman-Kolmogorov test

BMSM for Conformational Landscapes



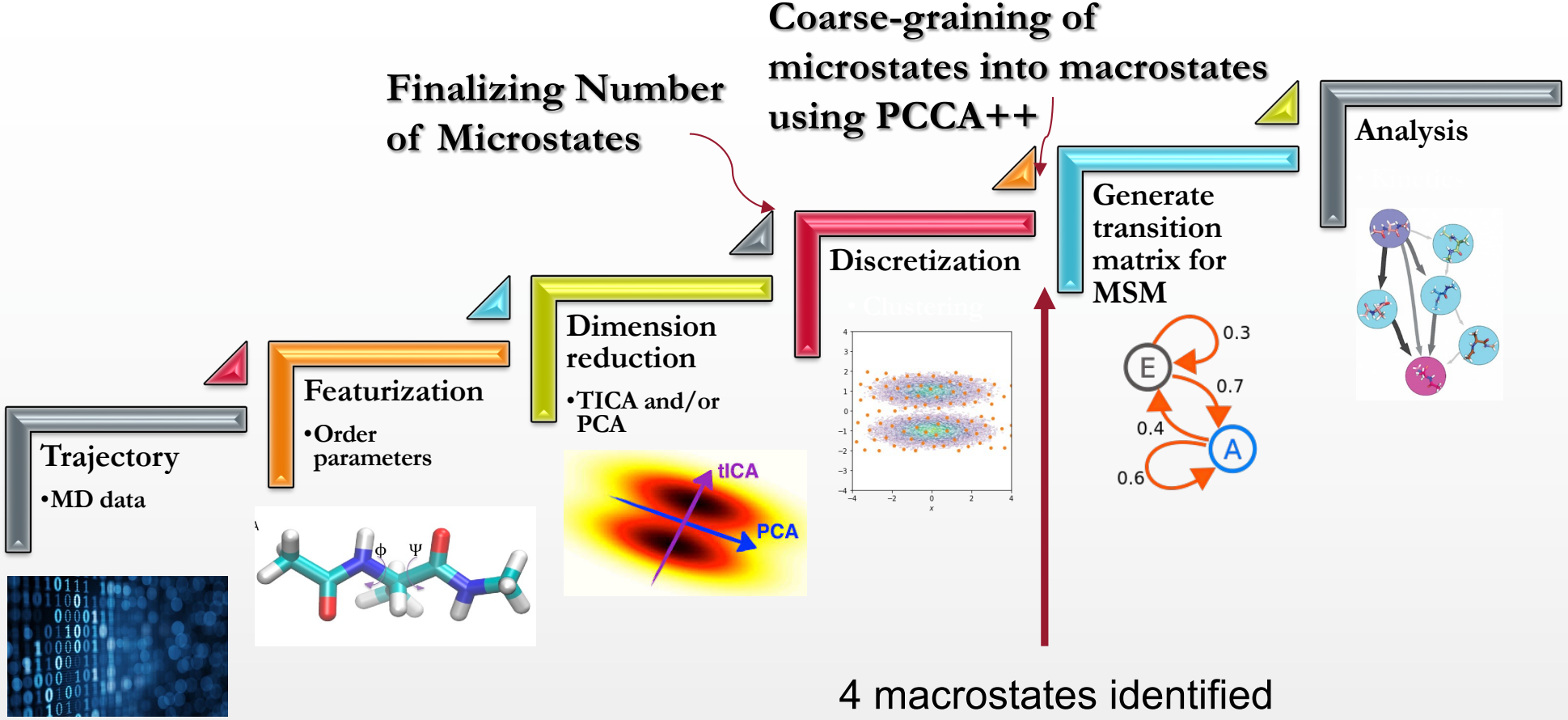
- Implied relaxation timescale used to validate lagtime in discretization
  - 50 ps for non pol
  - 100 ps for pol
- Separation of timescales
- Chapman-Kolmogorov test

BMSM for Conformational Landscapes



- Implied relaxation timescale used to validate lagtime in discretization
  - 50 ps for non pol
  - 100 ps for pol
- Separation of timescales
- Chapman-Kolmogorov test

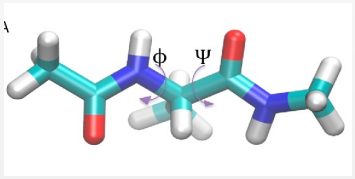
MSM for Conformational Landscapes



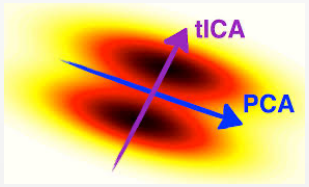
Sadhu & Clark, In Prep

MSM for Conformational Landscapes

Trajectory
- MD data

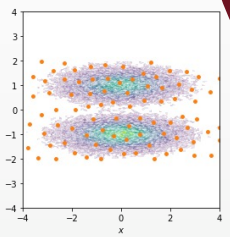


Featurization
- Order parameters



Dimension reduction
- TICA and/or PCA

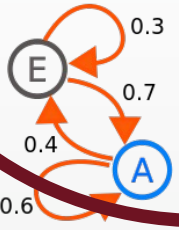
Finalizing Number of Microstates



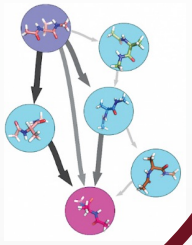
Discretization
- Clustering

Coarse-graining of microstates into macrostates using PCCA++

Generate transition matrix for MSM

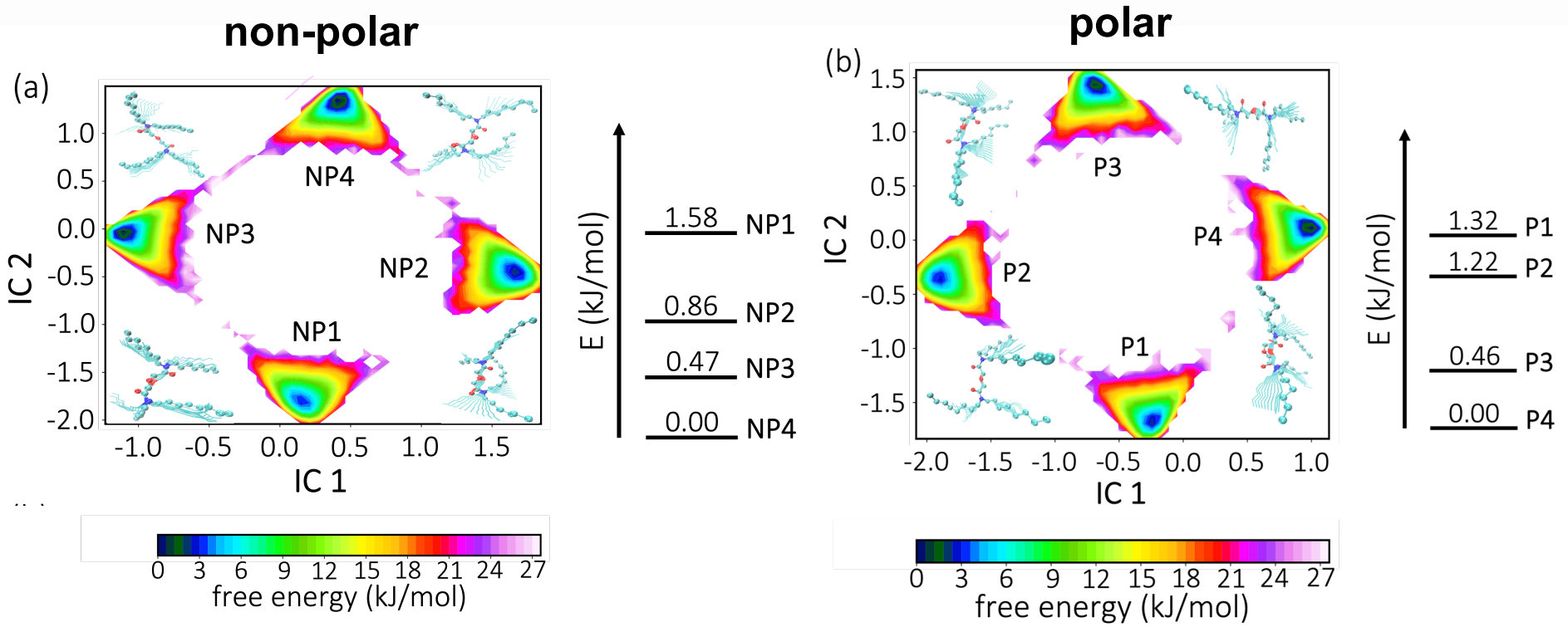


Analysis



Microstates and Transitions

- 4 coarse-grained macrostates observed in non-polar and polar conditions
- Free energy landscape in terms of independent components of PCA++

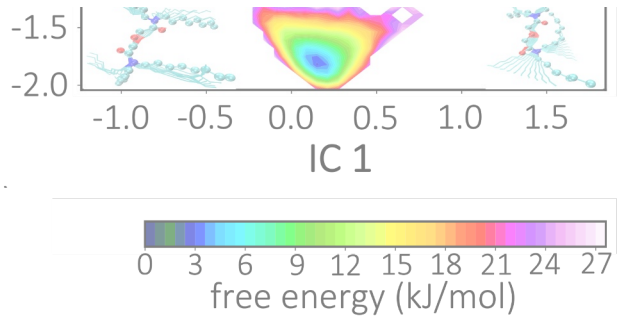


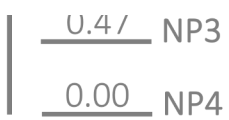
Microstates and Transitions

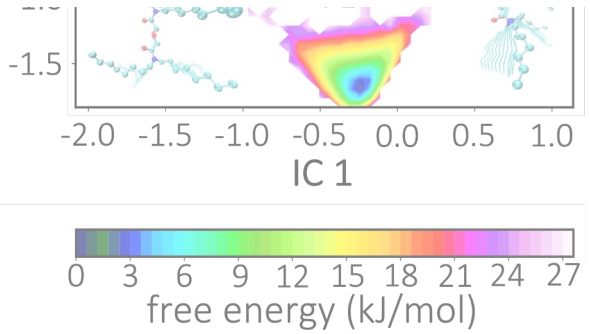
- 4 coarse-grained macrostates observed in non-polar and polar conditions
- Free energy landscape in terms of independent components of PCA++

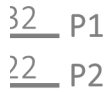
non-polar **polar**

- Free energy landscapes appear similar…but this is misleading.
  - What are the structures/configurations in the macrostates?





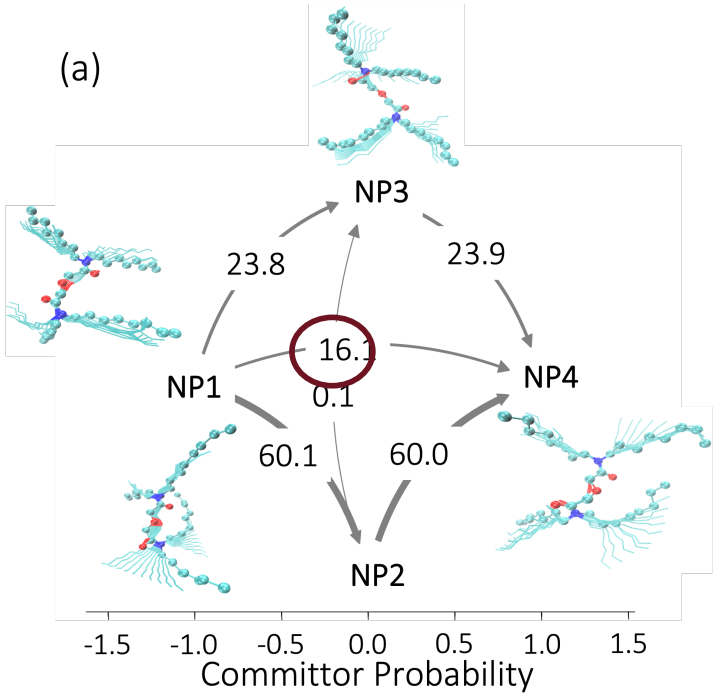


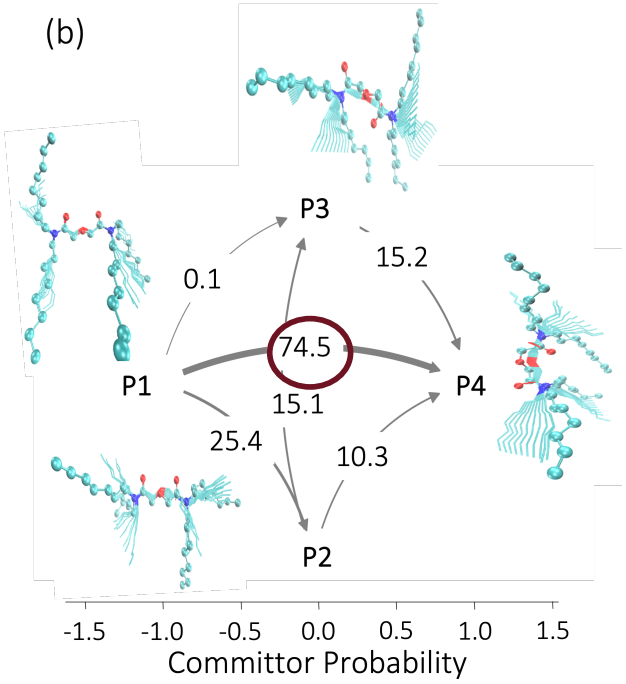


Microstates and Transitions

- Very different transition probabilities are observed, different conformational ensembles comprise a macrostate

non-polar **polar**



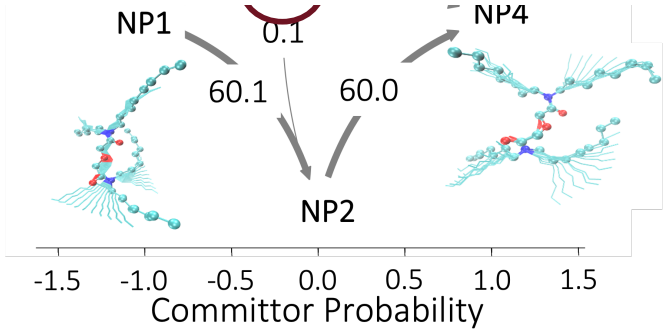


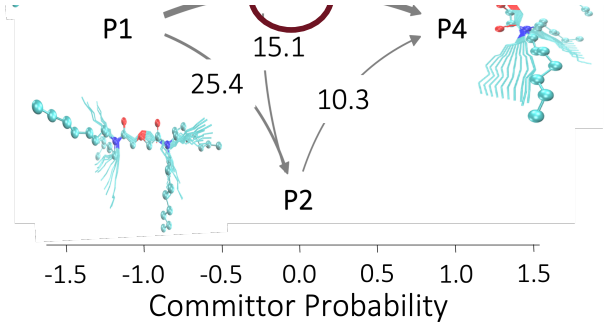
Microstates and Transitions

- Very different transition probabilities are observed, different conformational ensembles comprise a macrostate

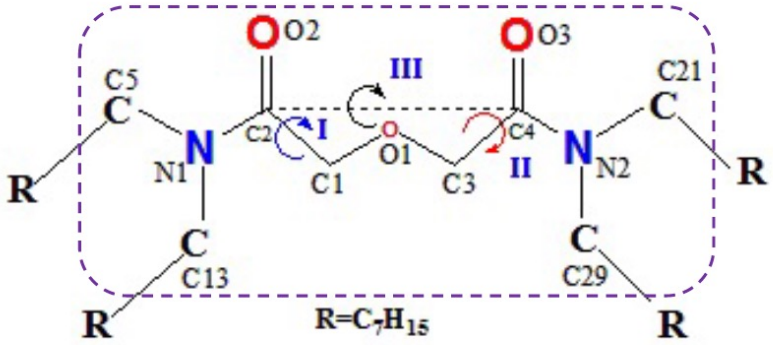
non-polar **polar**

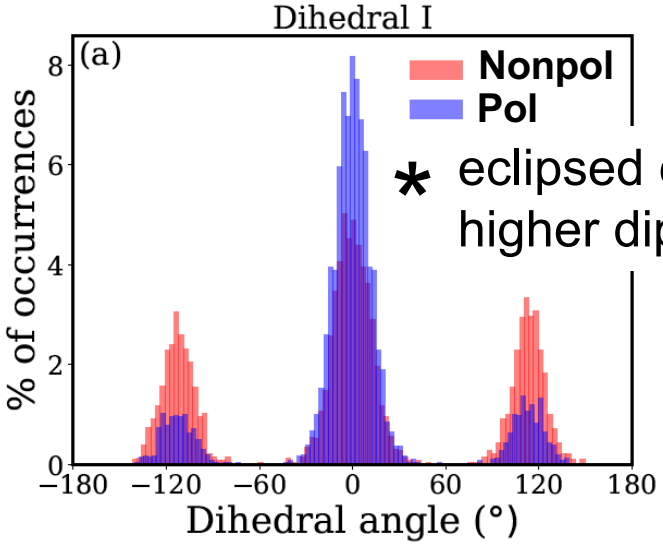
- Free energy landscapes appear similar…but this is misleading.
  - What are the structures/configurations in the macrostates?



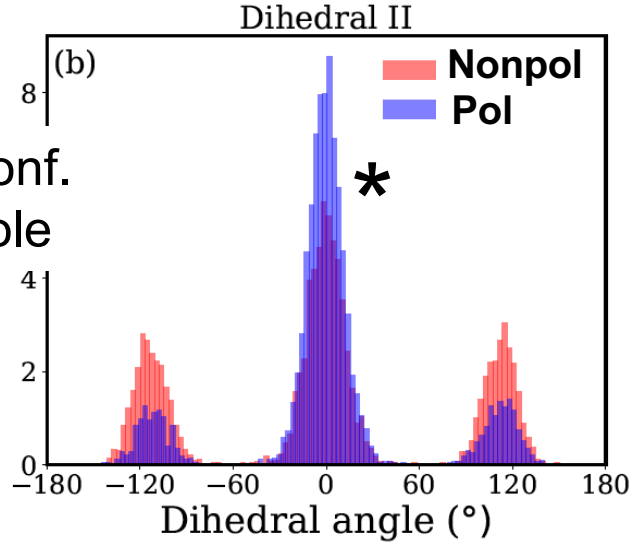


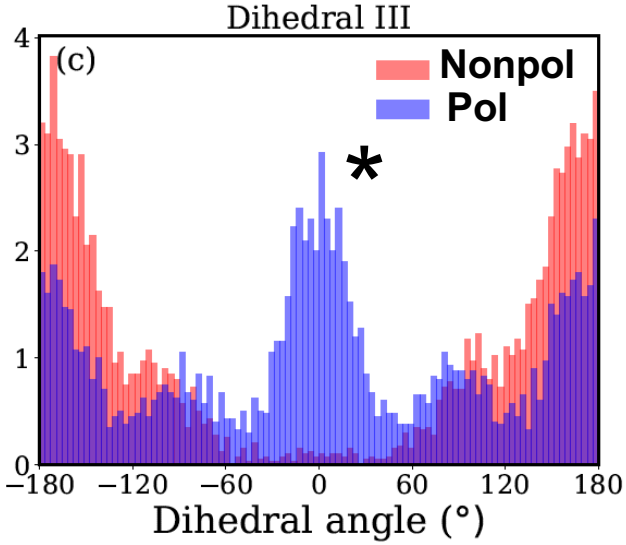
Manual Comparisons of Conformations





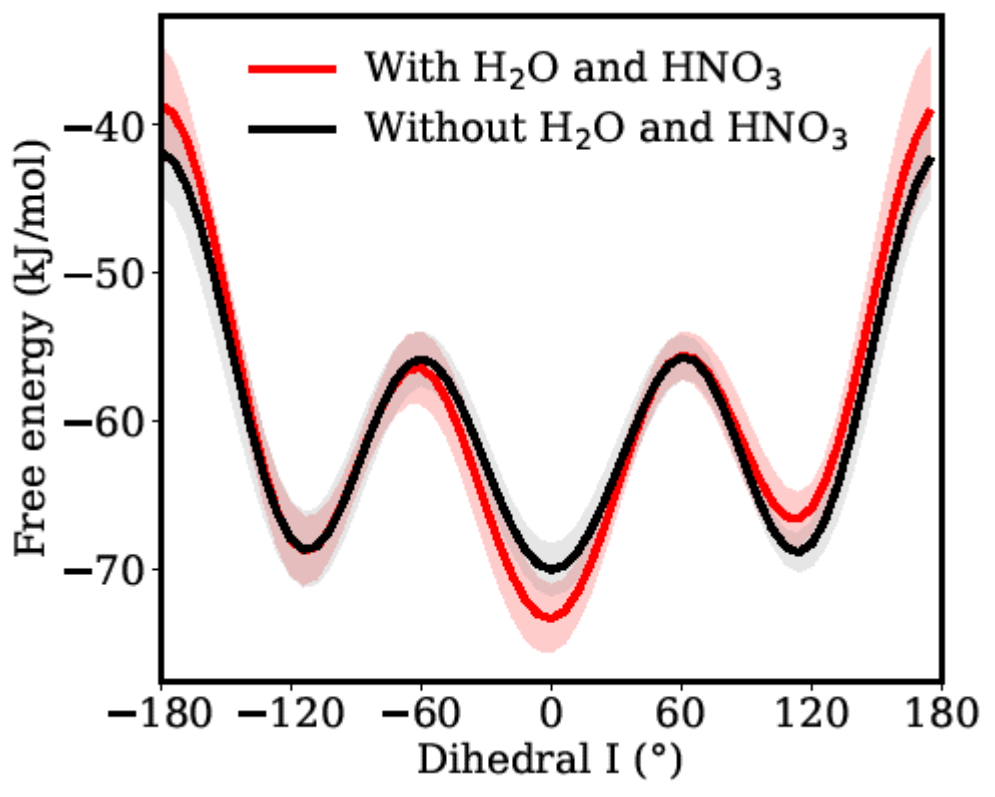
* eclipsed conf.
higher dipole





Polar Solute – Conformation Relationships

- Polar solute changes the free energy barrier of dihedral rotation



- Obtained from well-tempered metadynamics simulation

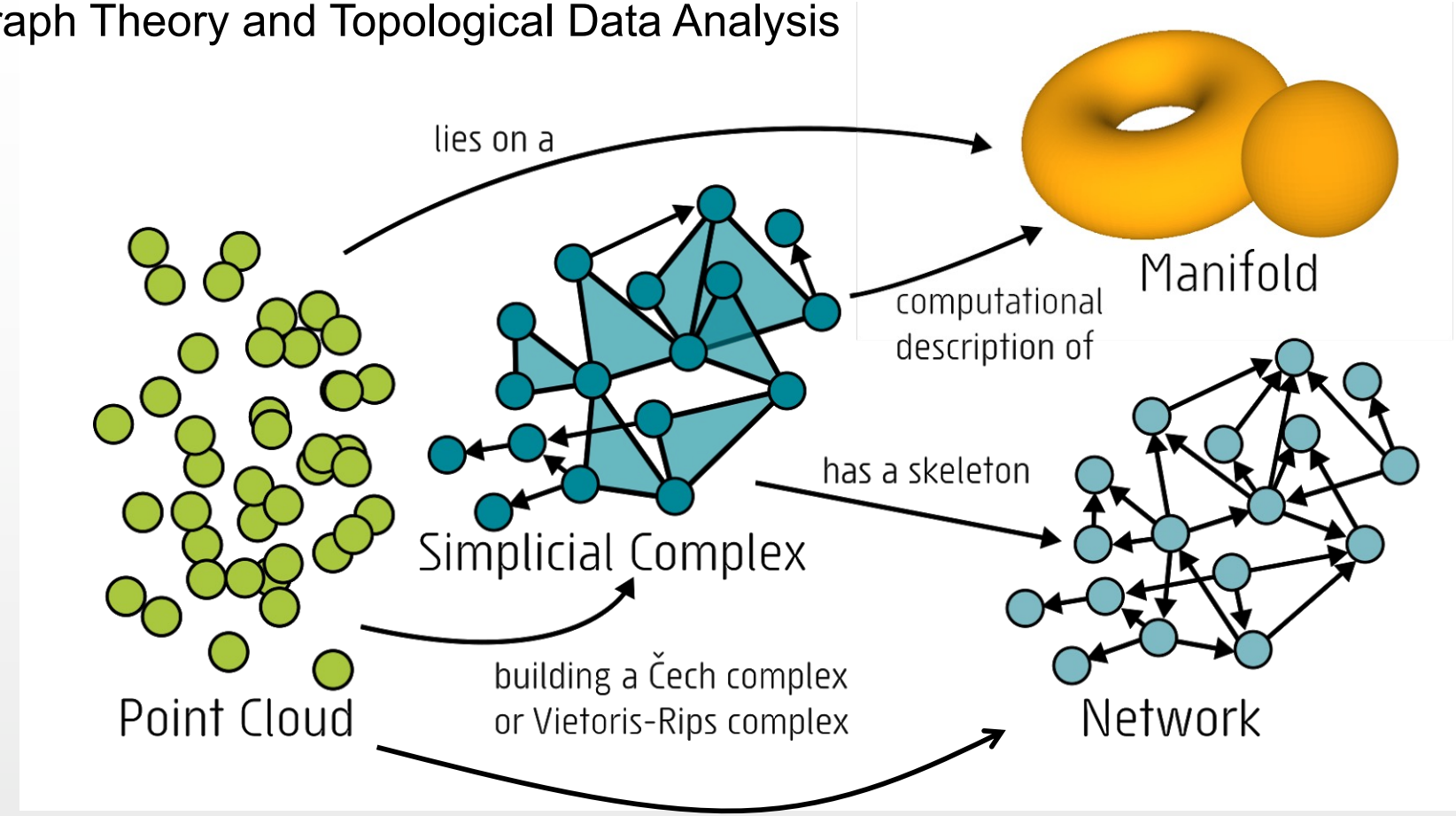
Polar Solute – Conformation Relationships

- Polar solute changes the free energy barrier of dihedral rotation

- This was very hard – it took a long time
- It does not easily scale across space and time

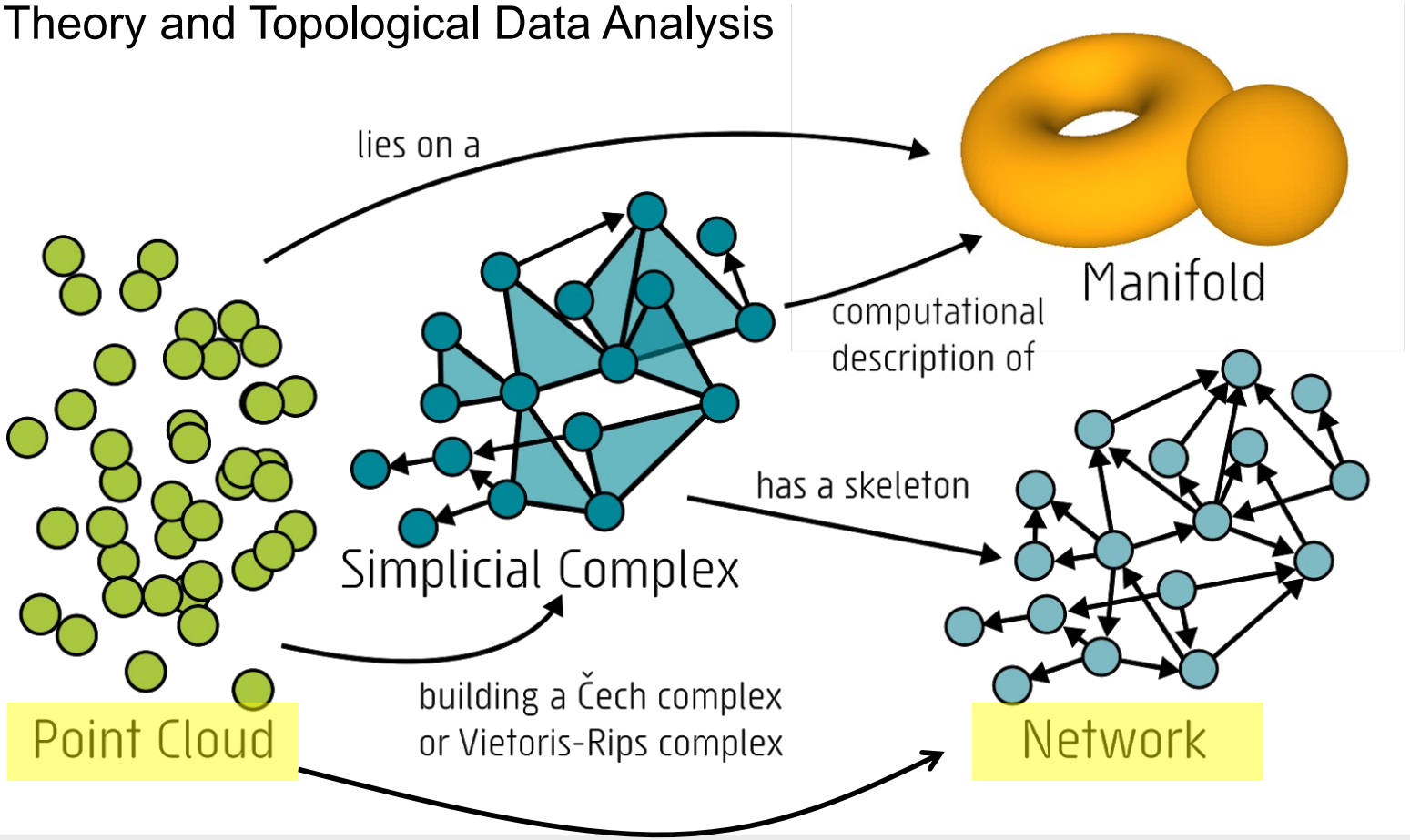
Alternative Approaches for Defining States

Graph Theory and Topological Data Analysis



Alternative Approaches for Defining States

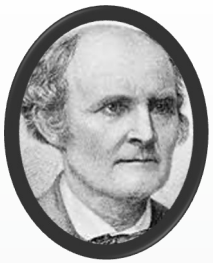
Graph Theory and Topological Data Analysis



Graph Theory Tools for Soft Matter Structure

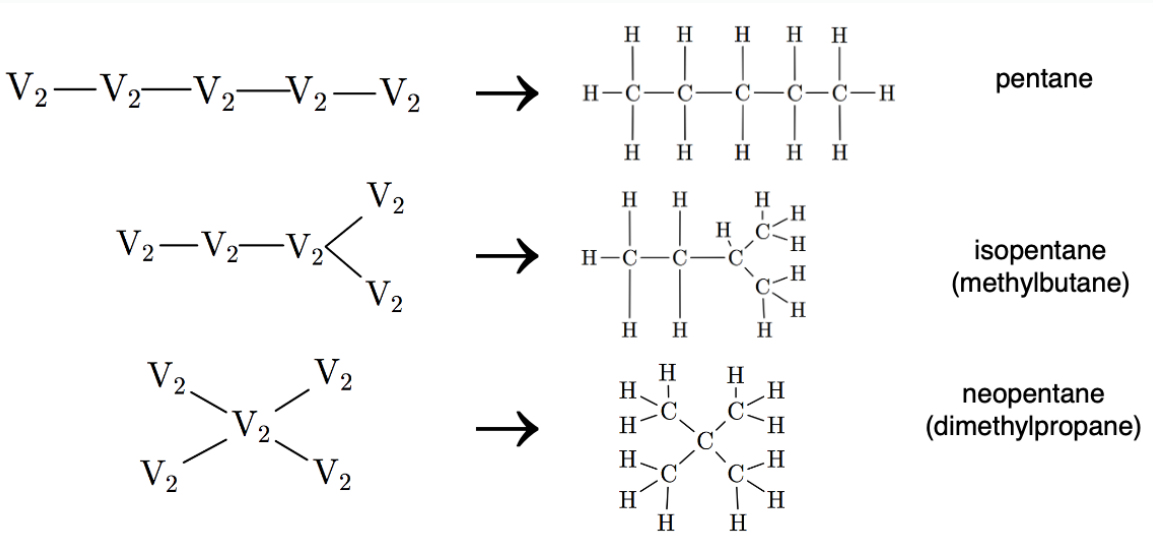
- Graph representations of chemical systems
  - Molecular graphs → date back to 1870's
  - intermolecular/interparticle interactions --> identify micelles
- Descriptors of graph connectivity & patterns (similar to molecular graphs)
  - Local and non-local, spectral GT approaches
- Hierarchical partitioning
  - Robust cluster or community analysis
- Time dependent evolution

Birth of Structure Theory (Aided by Graph Theory)



Arthur Cayley

- Derived generating functions to identify all possible isomers of tree alkanes with a given number of nodes



[Cay75a] A. Cayley, On the analytic forms called trees, with applications to the theory of chemical combinations, *Reports British Assoc. Adv. Sci.*, **45** (1875), 257-305 = *Math. Papers*, Vol. 9, pp. 427-460 (see p. 451).

GT and Chemistry Over Time (Some Highlights)

- Chemical reaction mechanisms based upon minimal edge changes and distances btw. adjacency matrices of reactants and products
- Principle of minimal structure change – Kolbe 1850

Definition 3.2. The *chemical distance* between two isomeric graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ with maximal common subgraph $G_1 \cap G_2 = (V_{12}, E_{12}, L_{12})$ is determined by

$$\mathrm{CD}(G_1, G_2) = |E_1| + |E_2| + |L_1| + |L_2| - 2|E_{12}| - 2|L_{12}|. \qquad (3.1)$$

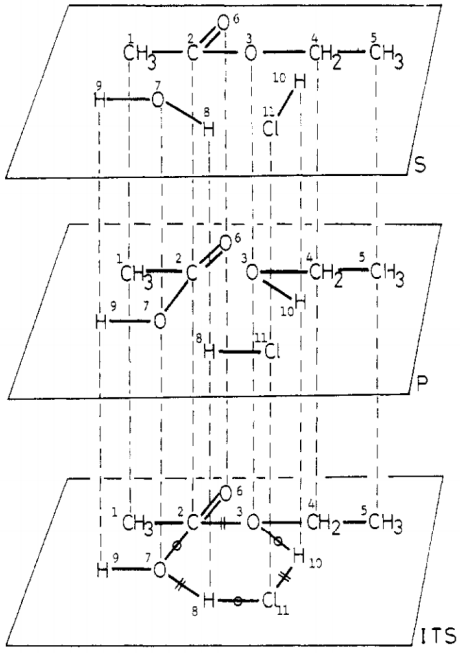


Figure 2. Constitution of imaginary transition structures (ITS's). This example is the hydrolysis of ethyl acetate catalyzed by hydrochloric acid. The abbreviations S and P correspond to the starting and product stages, respectively.

Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4, 205–212; Balaz, V. et. al. Disc. Appl. Math **1992**, *35*, 1-19.

GT and Chemistry Over Time (Some Highlights)

Quantum graphs – many different types here, though original formulation was for electron movement as in a wire along bonds (relevant to understanding magnetism, e- transport, and a wealth of condensed matter physics applications)

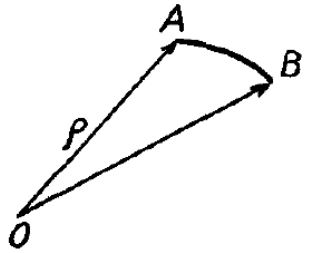


FIG. 1.

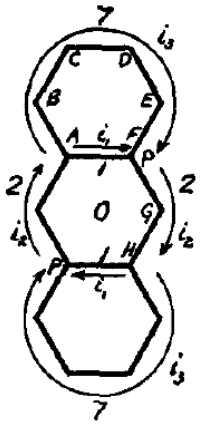


FIG. 2. Induced currents in anthracene.

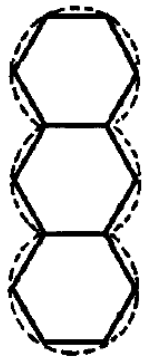


FIG. 3. Assumed path of induced currents in anthracene.

CHEMICAL REVIEWS

Review

Cite This: *Chem. Rev.* 2018, 118, 4887–4911

pubs.acs.org/CR

Quantum Interference, Graphs, Walks, and Polynomials

Yuta Tsuji,[†] Ernesto Estrada,[‡] Ramis Movassagh,[§] and Roald Hoffmann*[,⊥]

[†]Institute for Materials Chemistry and Engineering and IRCCS, Kyushu University, Nishi-ku, Fukuoka 819-0395, Japan

[‡]Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G11HX, United Kingdom

[§]IBM Research, MIT-IBM A.I. Lab, Cambridge, Massachusetts 02142, United States

[⊥]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, United States

Pauling, L. J. Chem. Phys. 4, 673 (1936); Ruedenberg, K.; Scherr, C. W. J. Chem. Phys. **21**, 1565 (1953)

Graph Theory Tools for Soft Matter Structure

- Graph representations of chemical systems
  - Molecular graphs → date back to 1870's
  - Intermolecular/interparticle interactions --> identify micelles
- Descriptors of graph connectivity & patterns (similar to molecular graphs)
  - Local and non-local, spectral GT approaches
- Hierarchical partitioning
  - Robust cluster or community analysis
- Time dependent evolution

Molecular Graphs and Partition Functions

Physica A 602 (2022) 127612

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Statistical–mechanical theory of topological indices

Ernesto Estrada

Institute for Cross-Disciplinary Physics and Complex Systems (IFISC, UIB-CSIC), Campus Universitat de les Illes Balears E-07122, Palma de Mallorca, Spain

***also some work relating Shannon Entropy to thermodynamic entropy, etc.

Graph Theory Tools for Soft Matter Structure

- Graph representations of chemical systems
  - Molecular graphs → date back to 1870's
  - Intermolecular/interparticle interactions --> identify micelles
- Descriptors of graph connectivity & patterns (similar to molecular graphs)
  - Local and non-local, spectral GT approaches
- Hierarchical partitioning in soft matter
  - Robust cluster or community analysis
- Time dependent evolution

Our Contributions to Descriptors

Spectral GT side:

- Determining Polyhedral Arrangements of Atoms Using PageRank, *Journal of Mathematical Chemistry,* **2012**, *50*, 2342.
- Novel Analysis of Cation Solvation Using Graph Theoretic Approaches, *Journal of Physical Chemistry B*, **2012**, *116*, 4263.

----many applications ensued---

- PageRank as a Collective Variable to Study Complex Chemical Transformations and Their Energy Landscapes, *Journal of Chemical Physics*, **2019**, *150*, 134102.

Centrality Measures:

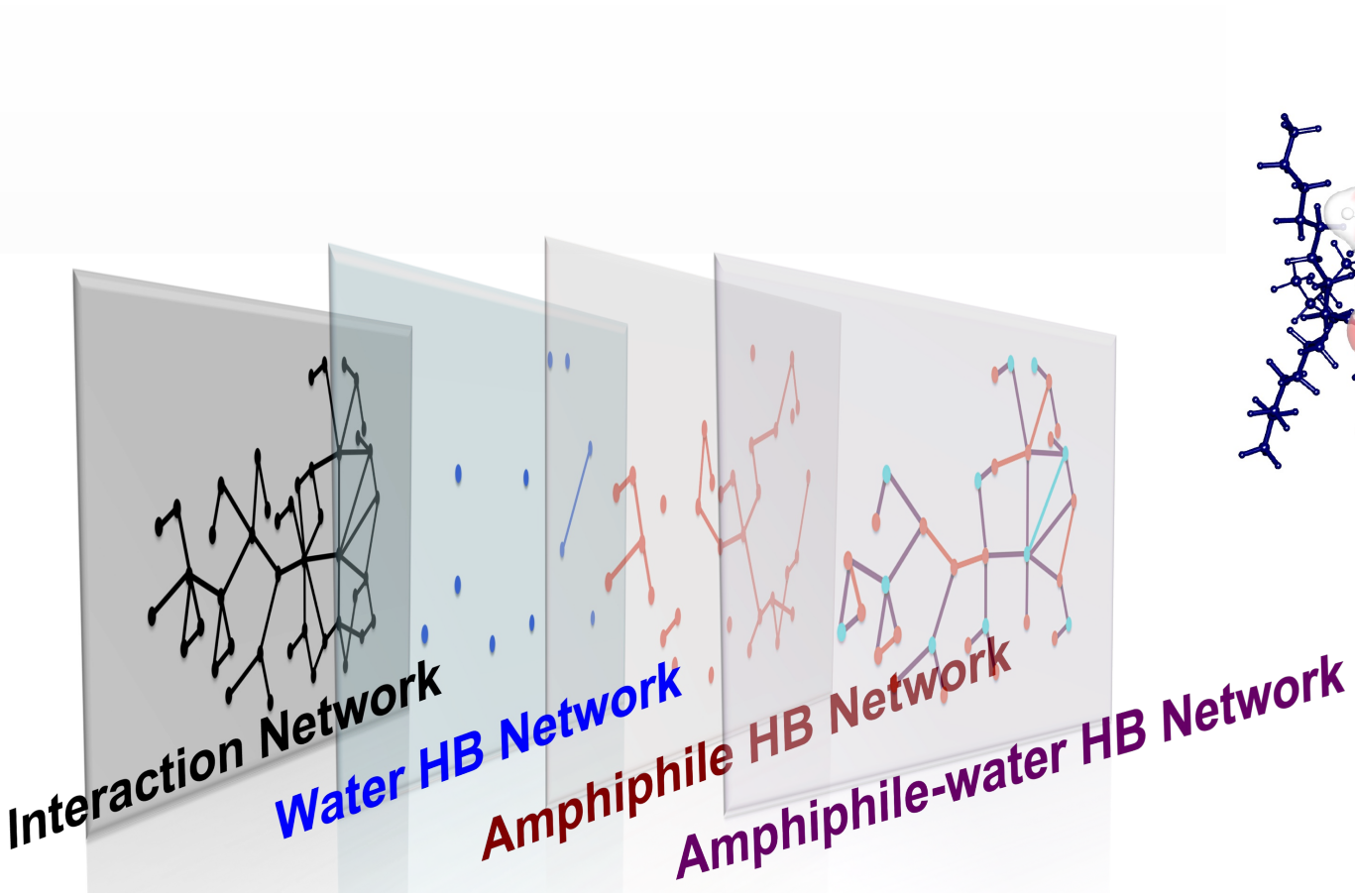
- Deconstructing the Confinement Effect upon the Organization and Dynamics of Water in Hydrophobic Nanoporous Materials: Lessons Learned from Zeolites, *Journal of Physical Chemistry C,* **2017**, *121*, 22015

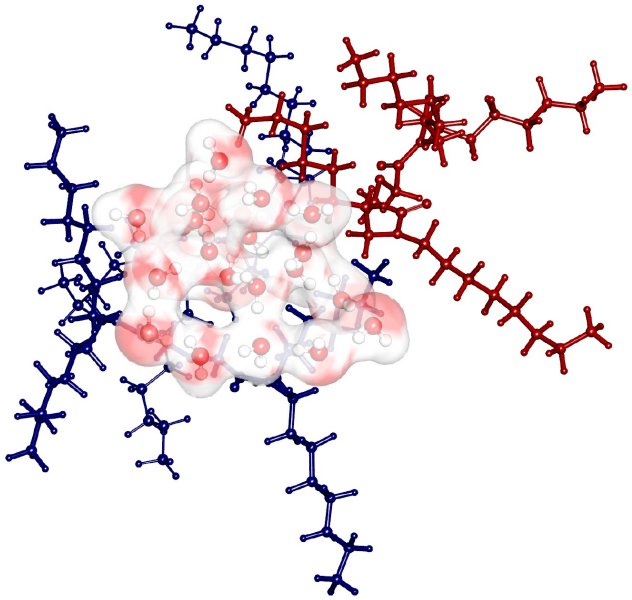
---many applications using different measures to understand speciation---

Graph Theory Tools for Soft Matter Structure

- Graph representations of chemical systems
  - Molecular graphs → date back to 1870's
  - Intermolecular/interparticle interactions --> identify micelles
- Descriptors of graph connectivity & patterns (similar to molecular graphs)
  - Local and non-local, spectral GT approaches
- Hierarchical partitioning in soft matter
  - Robust cluster or community analysis
- Time dependent evolution

From Molecules to Micelles





Hierarchical Graph Partitioning

- Modularity optimization helps identifying cluster in a generalized manner
- Find cluster partition that maximizes Q

$$Q = \frac{1}{N}\sum_{ij}^{N}(A_{ij} - \frac{k_i k_j}{n})\delta(c_i, c_j).$$

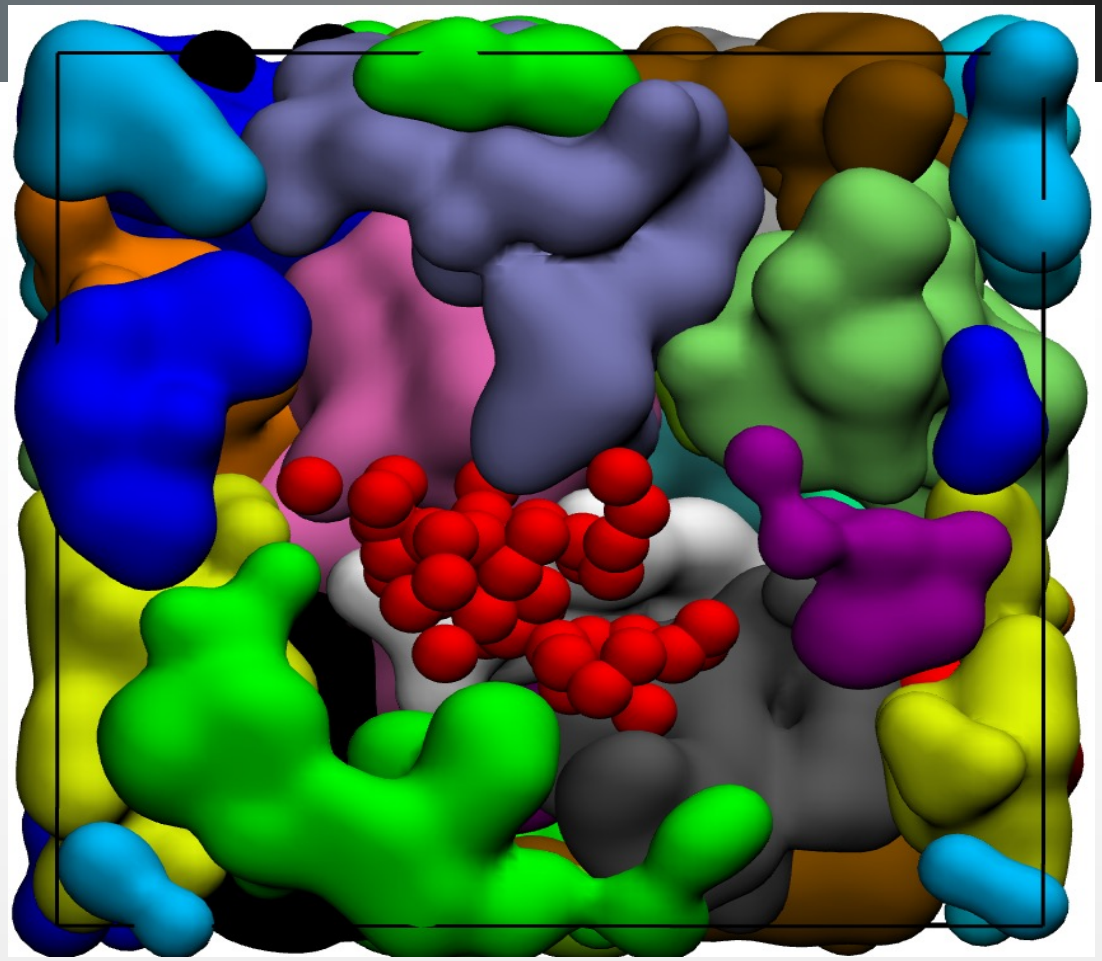
Edge weight between pair of nodes

0 if *i* and *j* in different clusters, 1 otherwise

k_i total of all of the edges weights to node *i*
n total of all of the edges weights in A

Total number of nodes (may also be 'n")

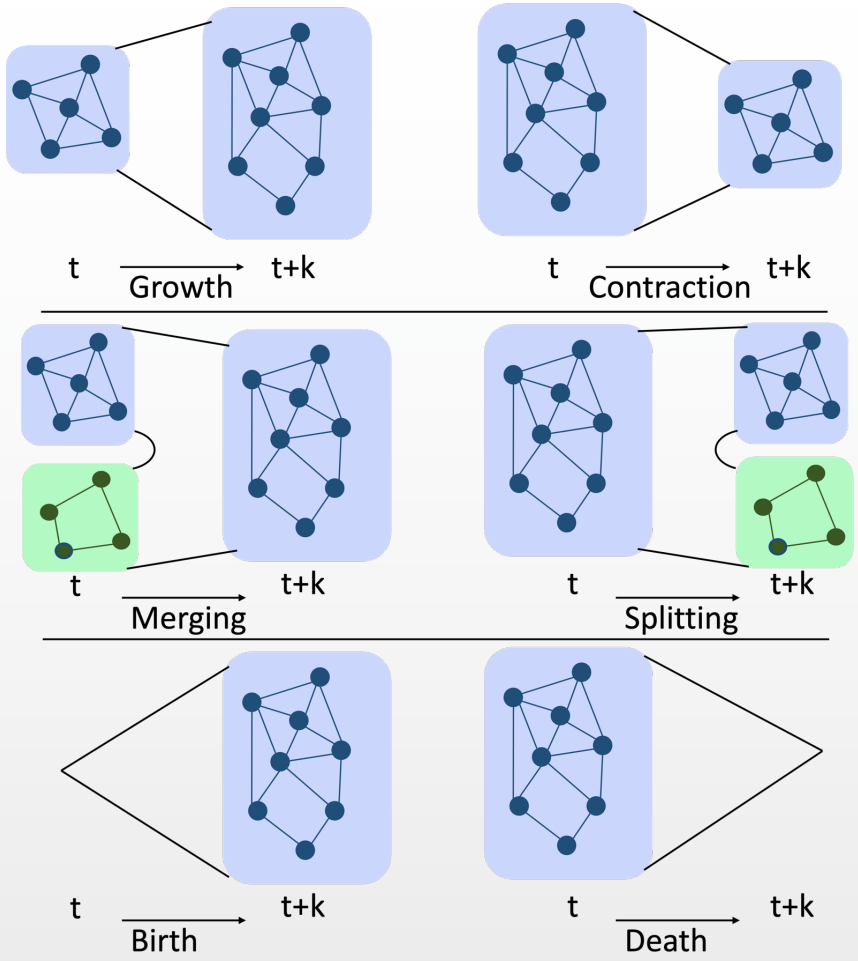
Newman, M. E. *PNAS* ***2006****, 113,* 8577.*;* Blondel, et. al. *J. Stat. Mech.* **2008** P1008.



- Do in multiple passes to coarse-grain and obtain hierarchical partitioning

Journal of Physical Chemistry B, **2021**, *125*, 3986-3993

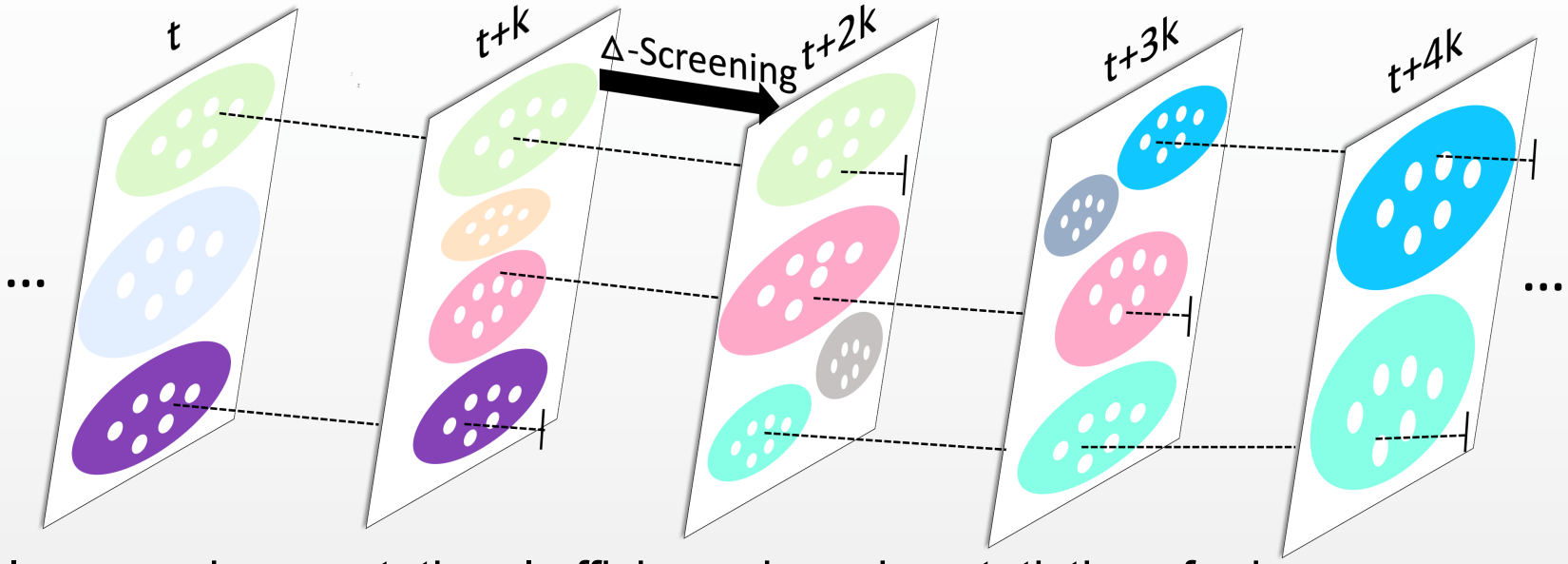
Modularity Optimization in Time



Journal of Chemical Theory and Computation, **2022**, *18*, 7043 – 7051

Modularity Optimization in Time

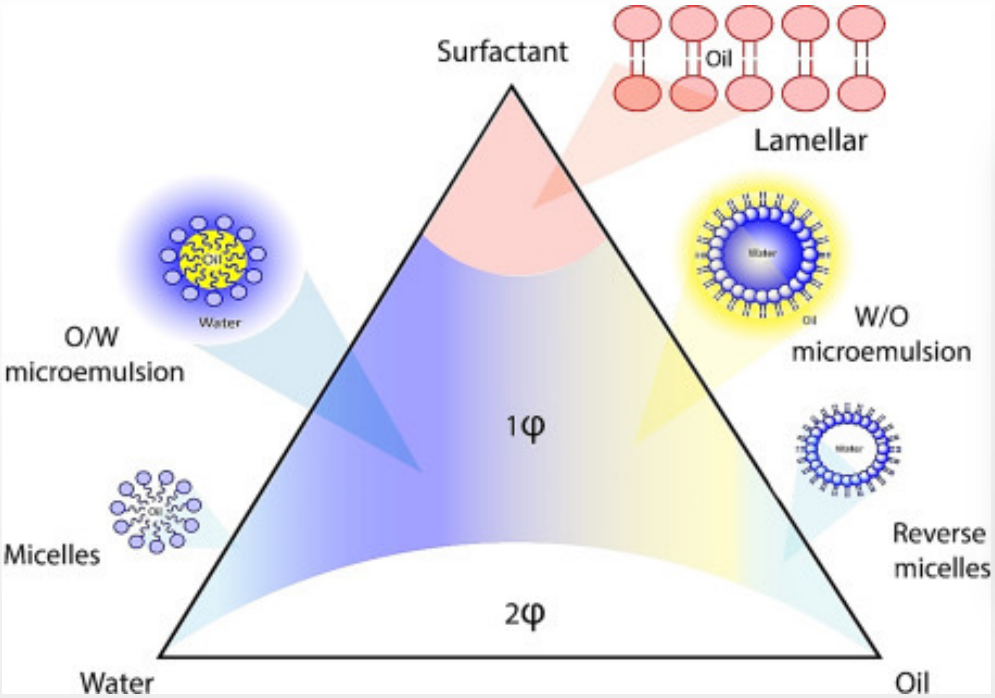
- Define “temporal” communities – with user defined constraints upon changes to node composition



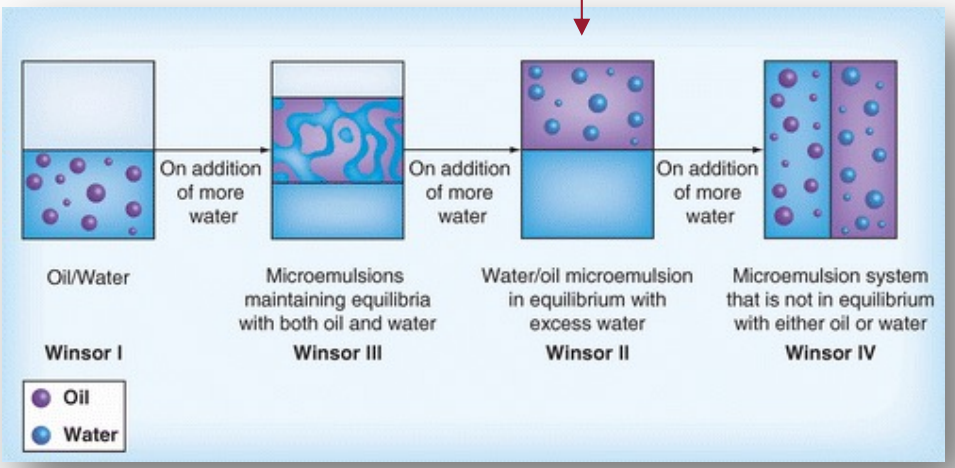
- Increased computational efficiency by using statistics of edge addition/deletion in time

Journal of Chemical Theory and Computation, **2022**, *18,* 7043 – 7051

Temporal Communities for Separations

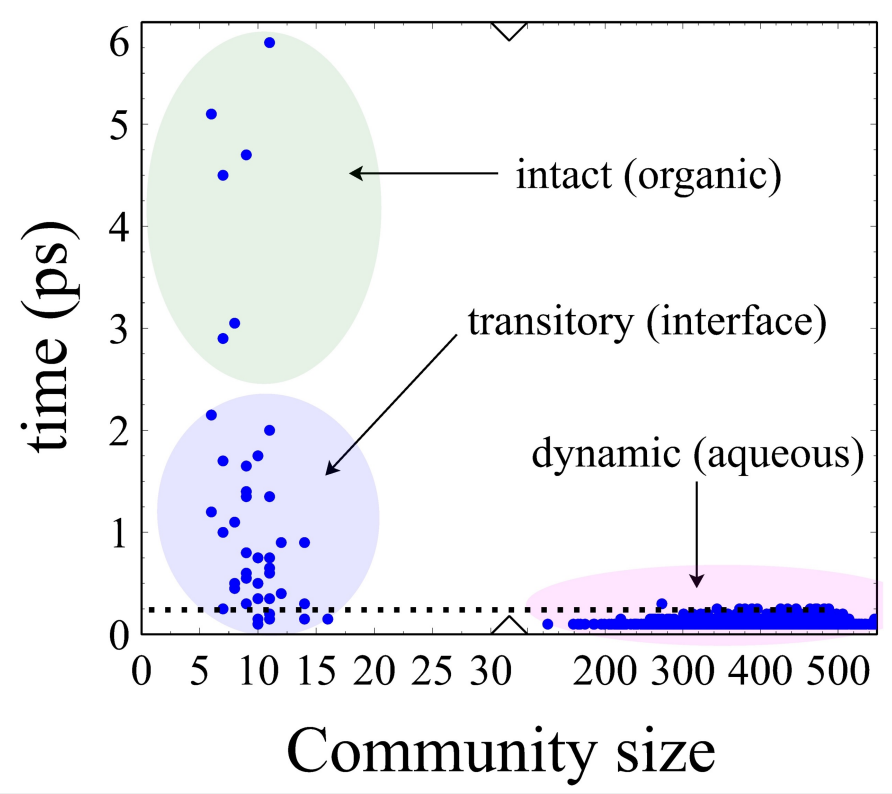


Transport key to separations science



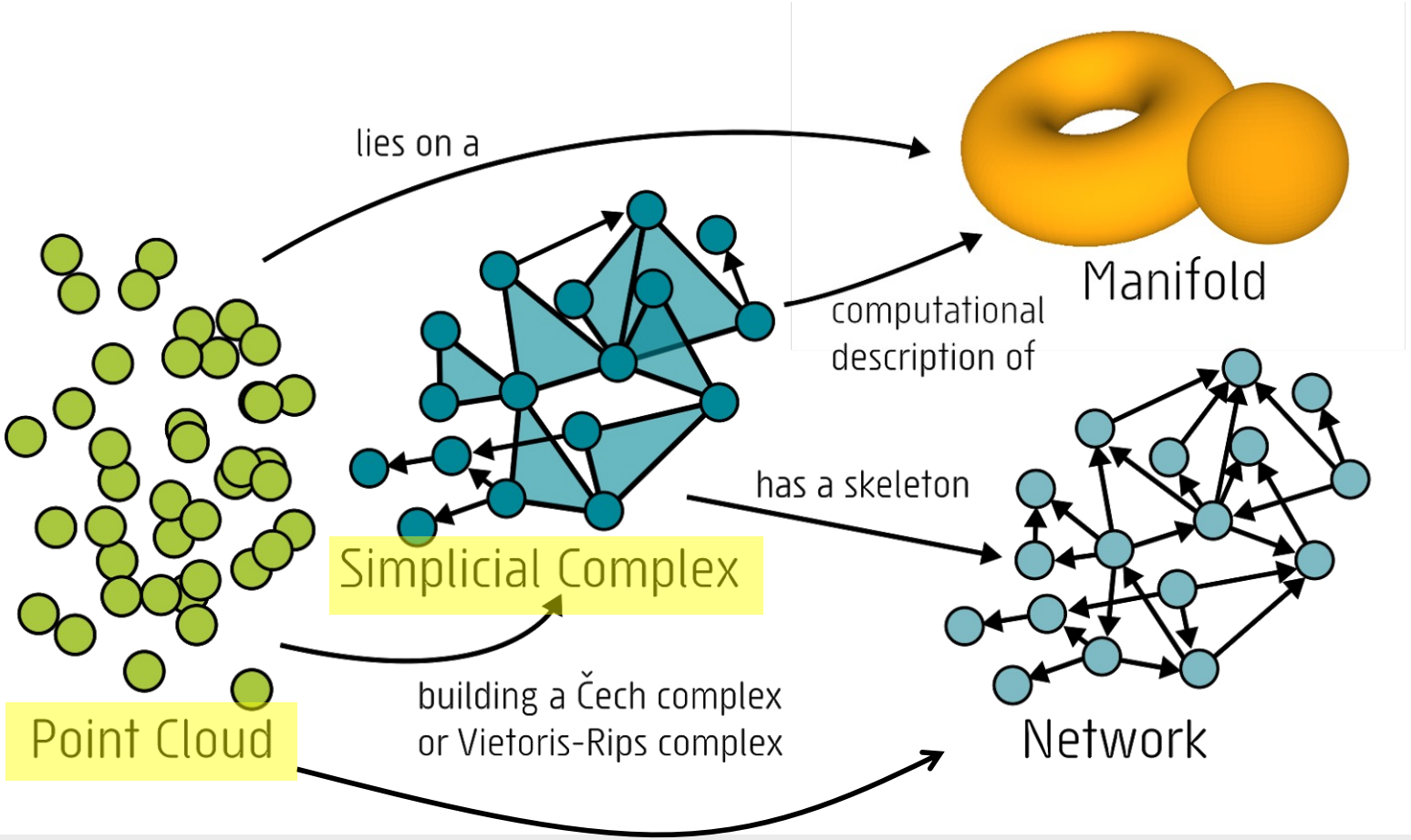
Temporal Graph Theory Analyses

Molecular dynamics of an actively transporting oil/water interface

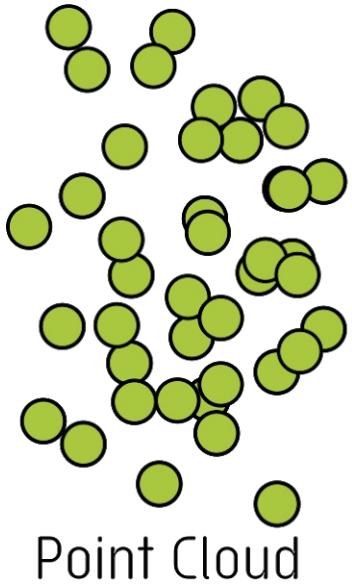


Journal of Chemical Theory and Computation, **2022**, *18,* 7043 – 7051

Graph Theory and Comp. Topology as a Framework

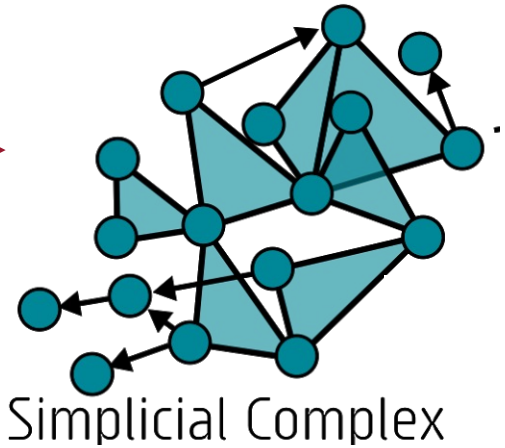


Graph Theory and Comp. Topology as a Framework



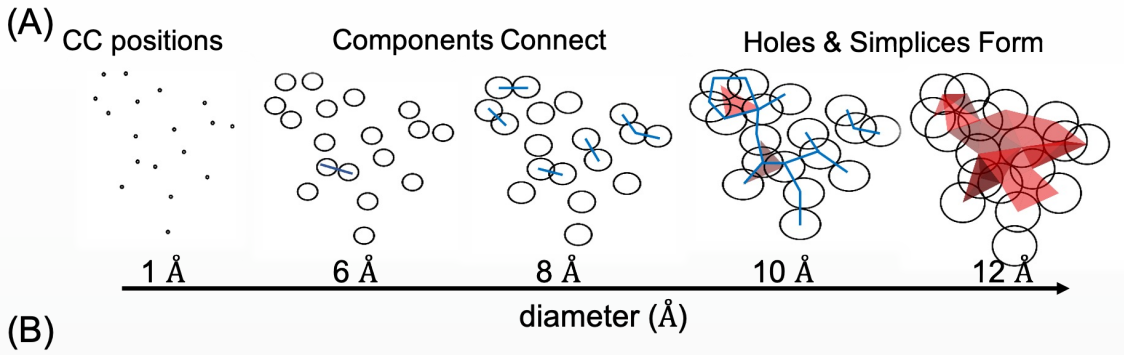
1. Choose a ball radius that forms the complex

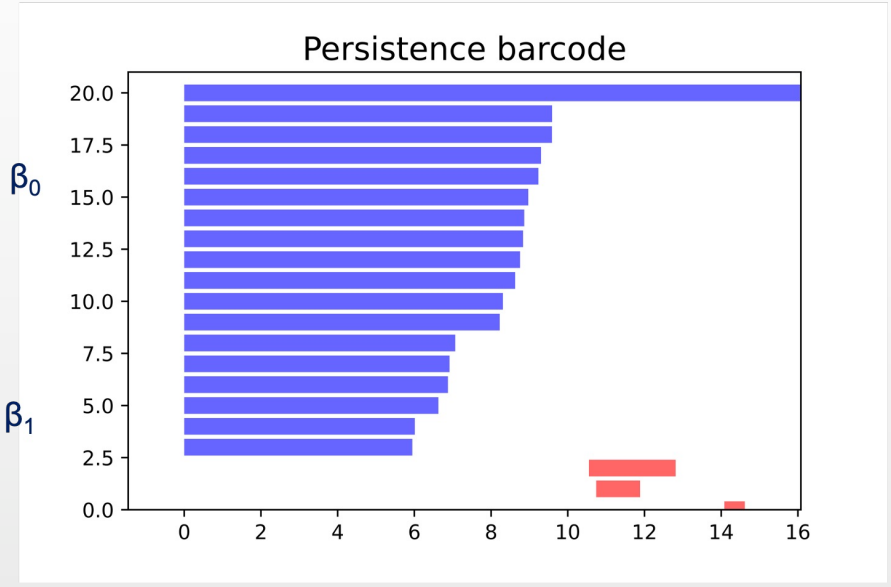
2. Vary ball radius and track how components change



Persistent Homology – Distance Filtered Graphs

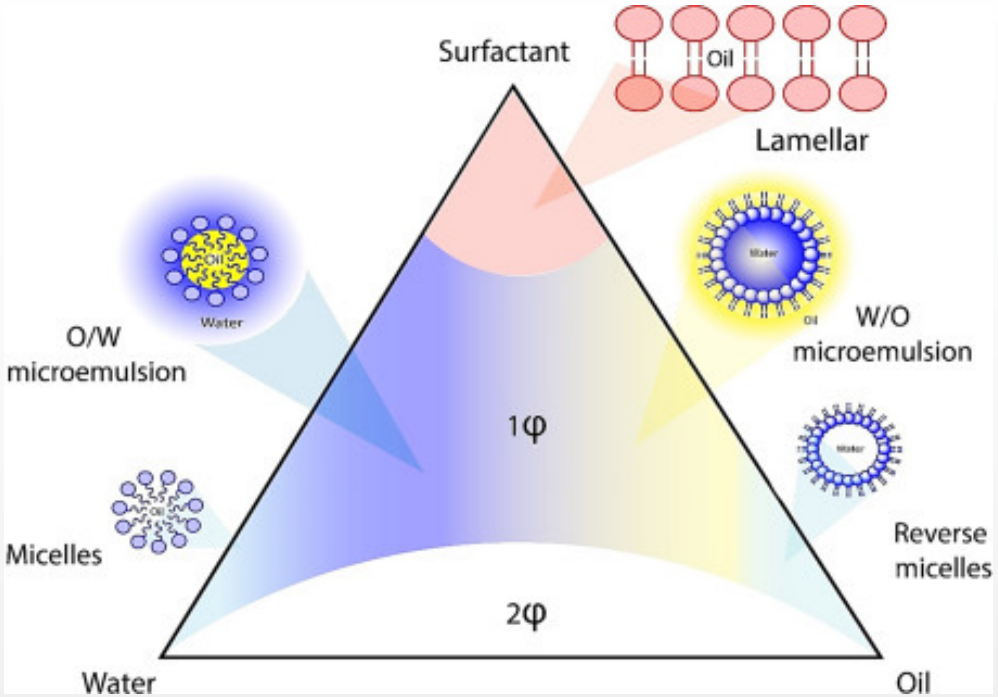
- Graphs formed by different distance filtrations of point cloud data

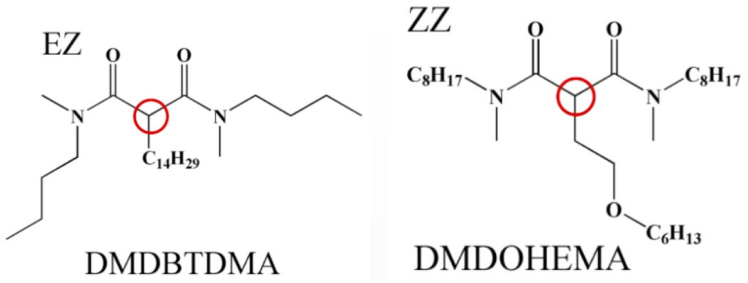


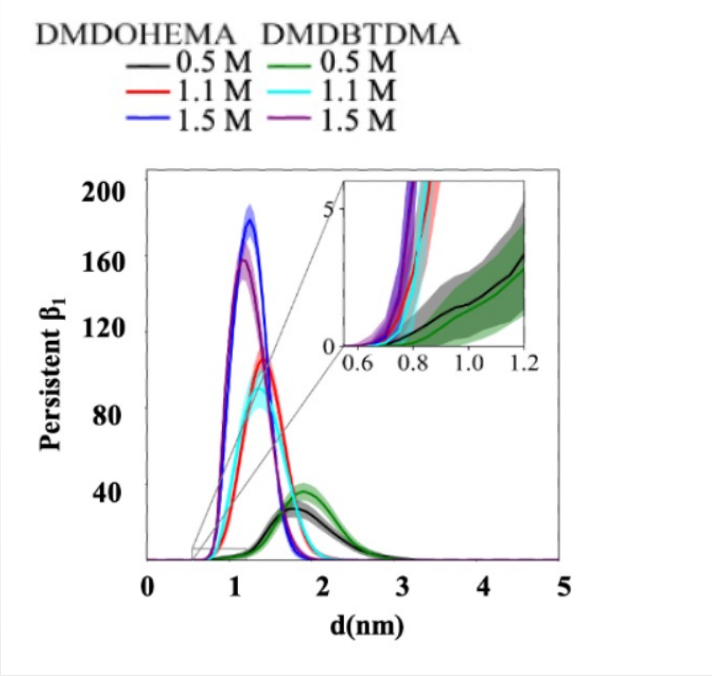


- Recording the number of components of the simplicial complex (zero dimensional information, β_0)
- Recording the number of holes in the simplicial complex (1-dimensional information, β_1)

Structures of Surfactant Aggregates



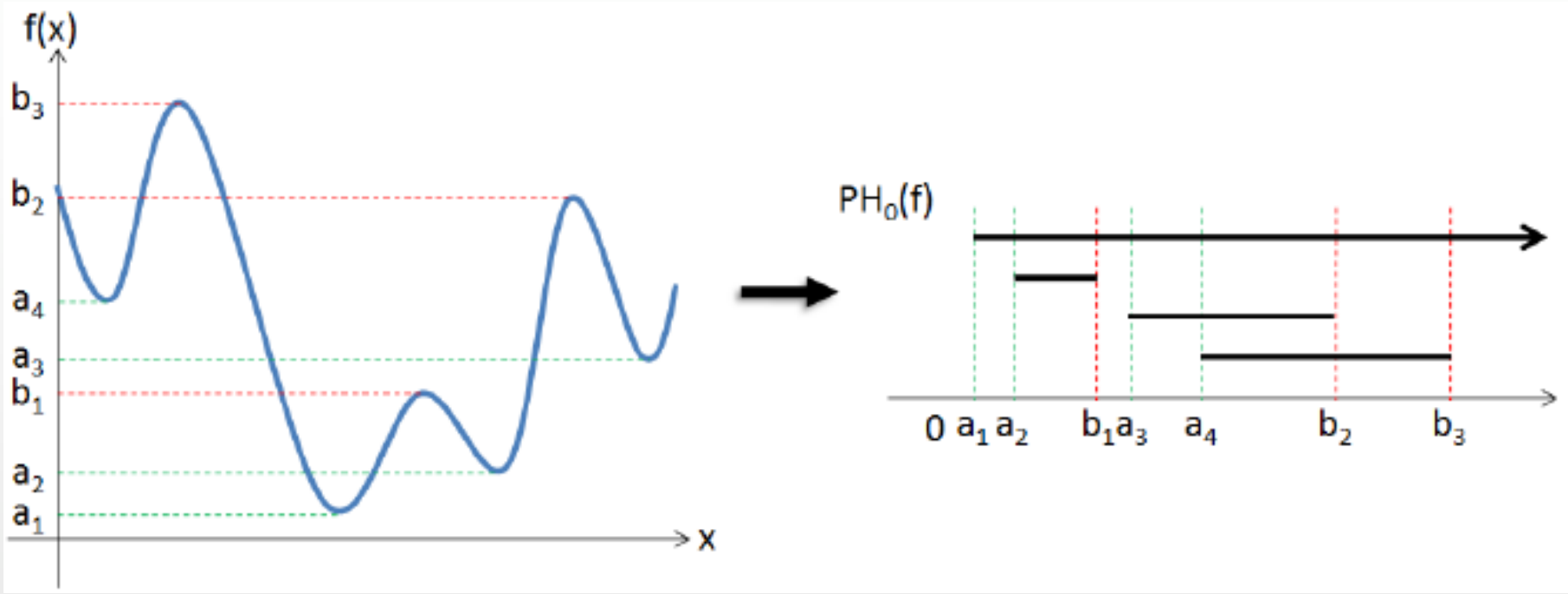


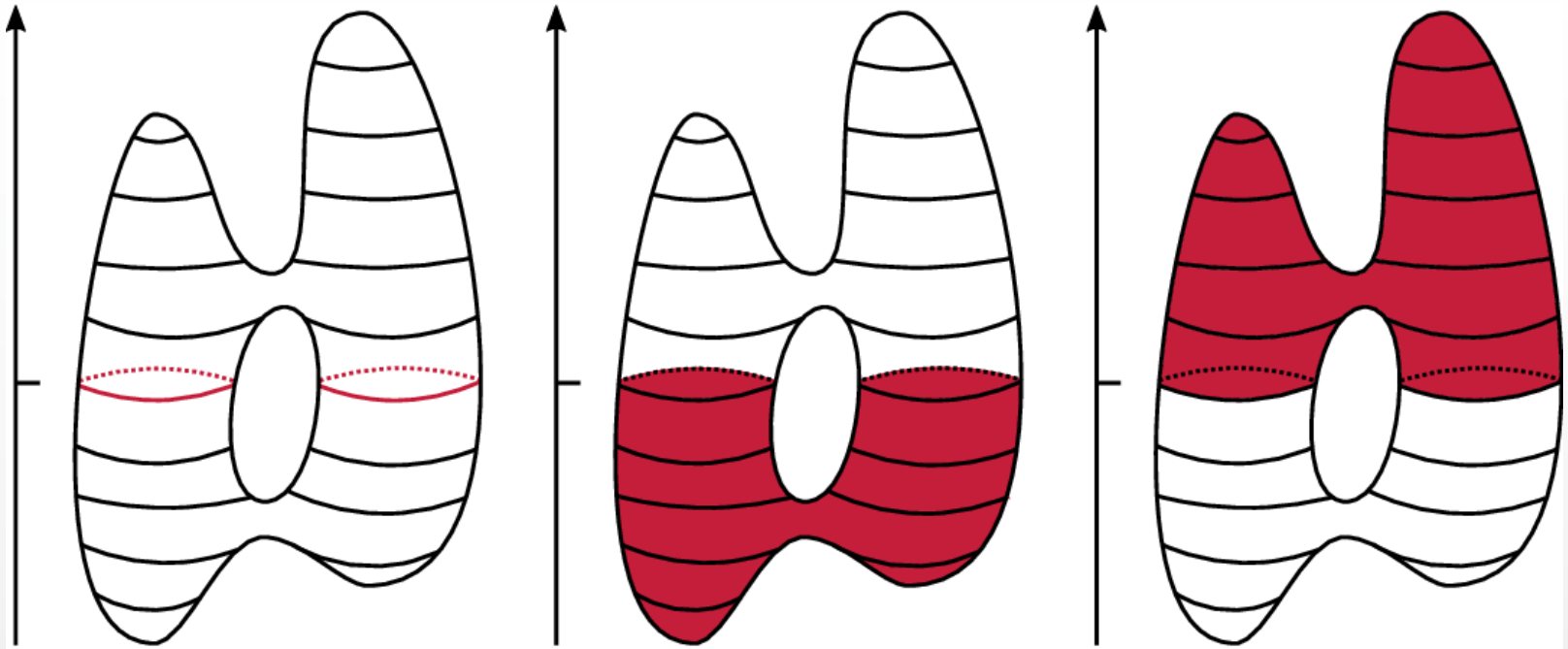


Journal of Molecular Liquids, **2022**, *345*, 117743.;
Journal of Physical Chemistry B, **2020**, *124,* 10822.

Sub- or Superlevelset Persistent Homology

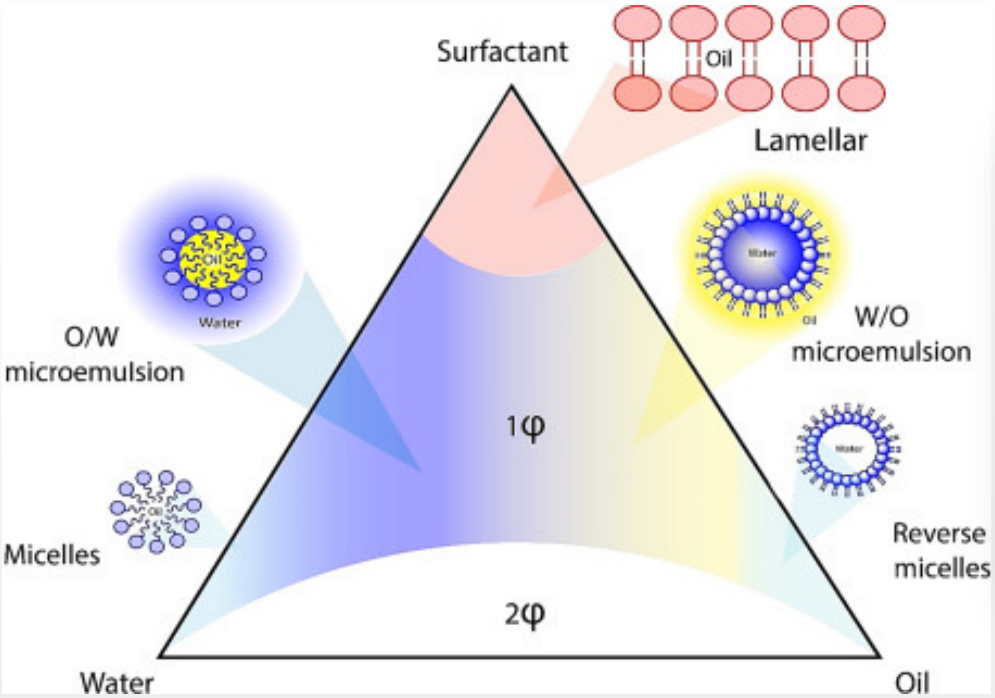
- Not only for point cloud data, but also functions
- Filtration starts from the lowest to highest – sublevelset
- Filtration starts from highest to lowest - superlevelset



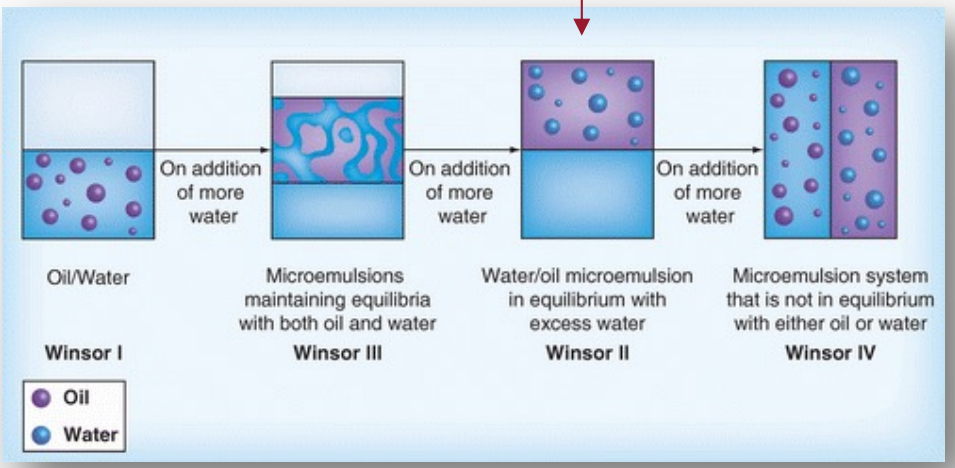


A level set $\mathcal{L}_y(f)$ A sublevel set $\mathcal{L}_y^-(f, y)$ A superlevel set $\mathcal{L}_y^+(f, y)$

Sublevelset PH for Separations



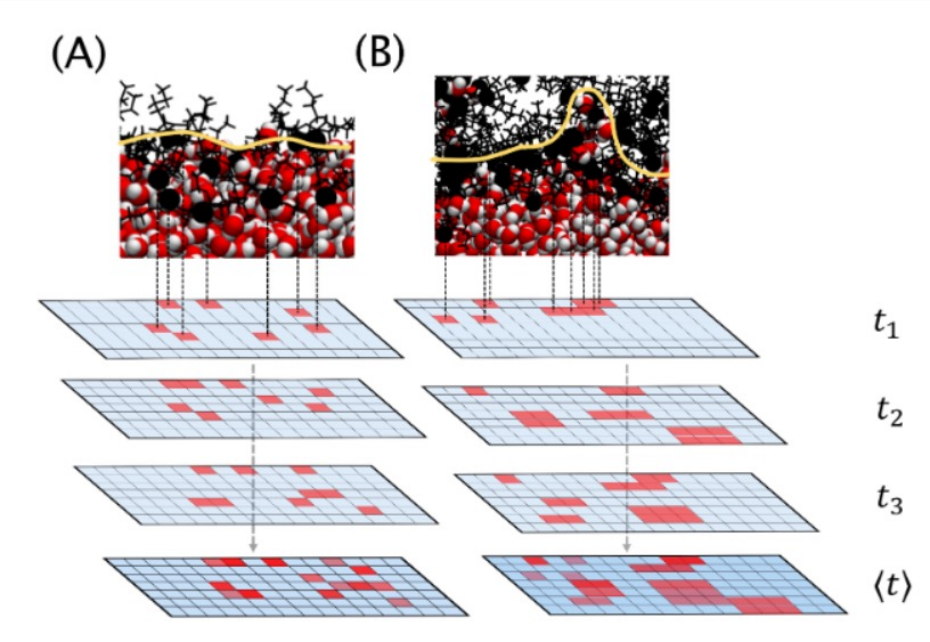
Transport key to separations science

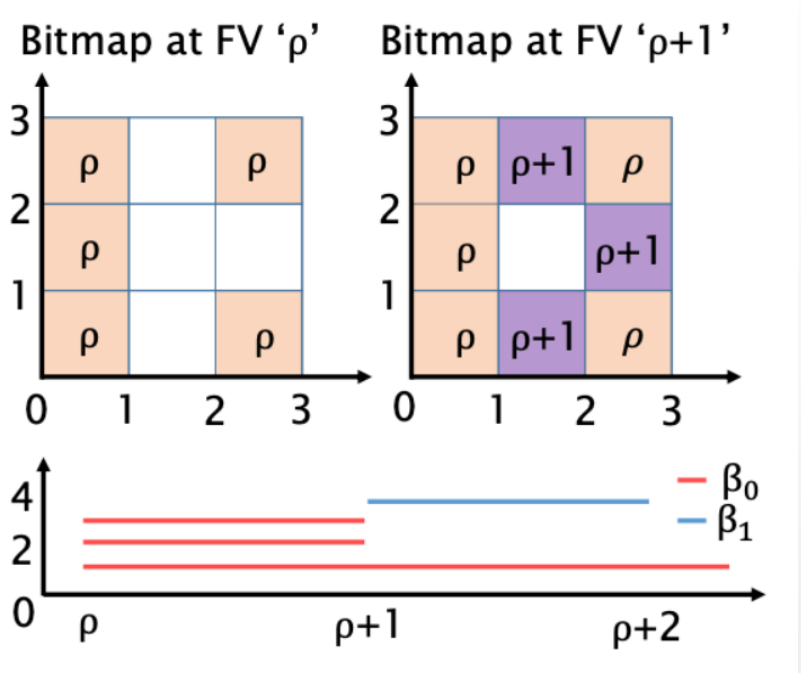


Sublevelset PH

- Molecular dynamics of surfactants at oil/water interfaces

Create time-dependent density images





Journal of Chemical Theory and Computation, **2023**, *In Press*
https://doi.org/10.26434/chemrxiv-2023-vwrxj

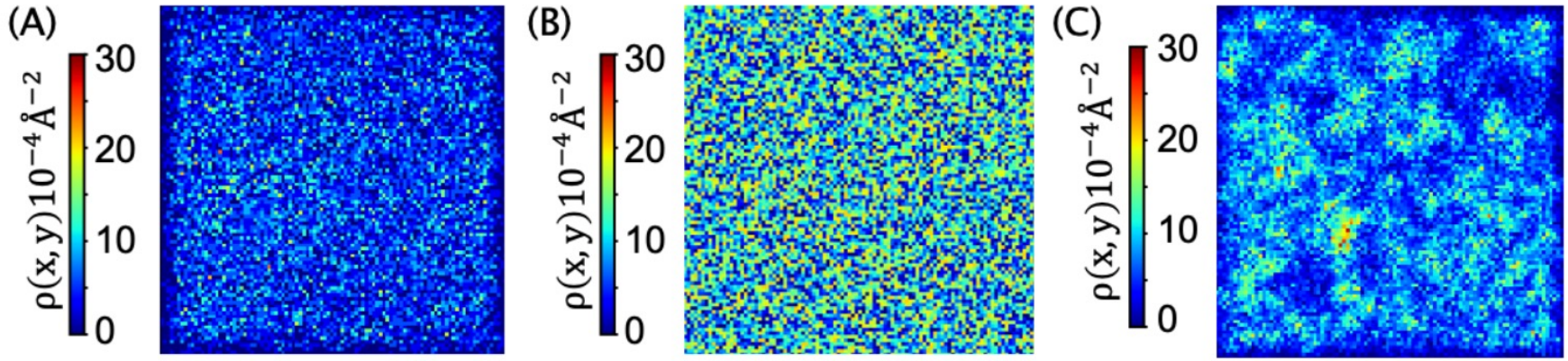


FIG. 3. Representative density surfaces analyzed in this study. (A) Adsorption density surface of ~ 72 TBP adsorbed at the water/vapor interface ($x = y = 60$ Å). (B) Stochastic distribution model surface representing ~ 72 TBP on a $x = y = 52$ Å surface, (C), and an average of ~ 79 TBP adsorbed at $LiNO_{3(aq)}$/hexane interface ($x = y = 52$ Å).

PH Descriptors

- Goal – capture organization at surface by surfactants
- Excess Betti Curves
  - Compares the real betti curves with those from purely random images (at same density)

$$\tilde{\beta}_n(\rho) = \beta_n(\rho) - \beta_n'(\rho).$$

- Persistent entropy – measures variations in persistent lifetimes of the components

$$PE = \sum_i -\frac{l_i}{L} \log \frac{l_i}{L},$$

- Non-ideality index for surfaces – if surfactants don't interact we know # of components, variations indicate non-ideal interactions

$$I = \frac{\text{number of observed cycles}}{\text{average number of adsorbates}}$$

Journal of Chemical Theory and Computation, **2023**, *In Press*
https://doi.org/10.26434/chemrxiv-2023-vwrxj

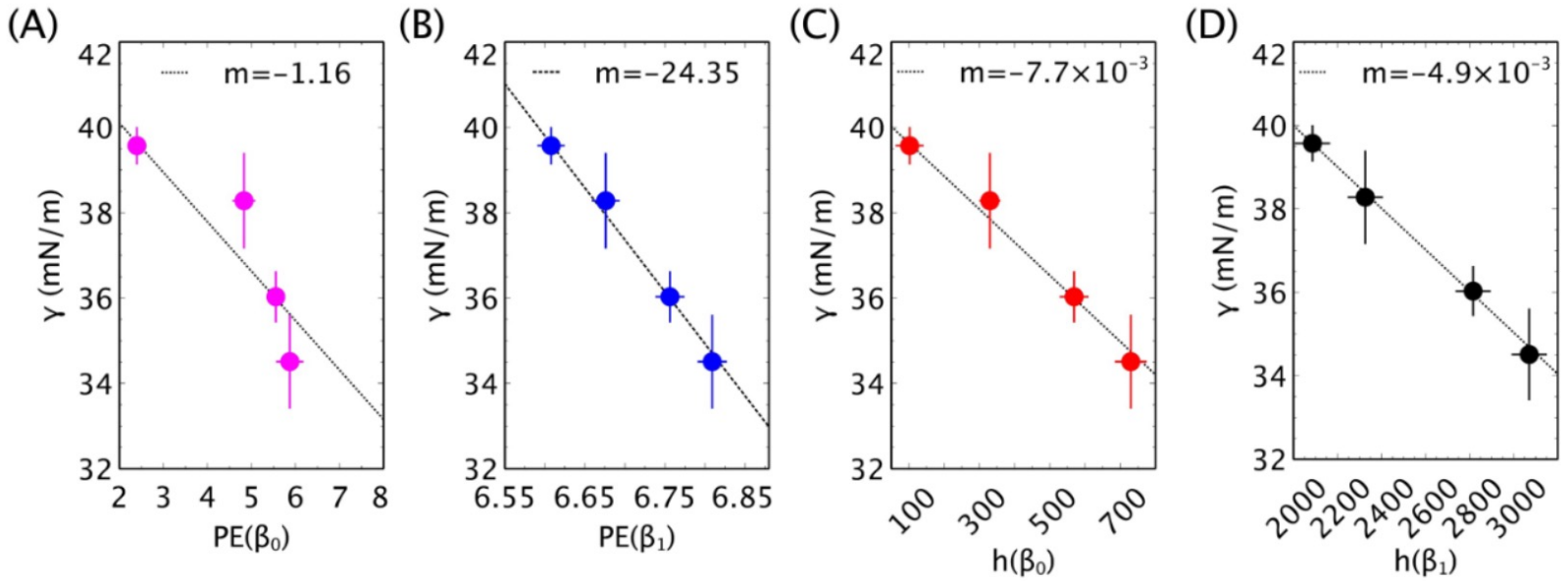
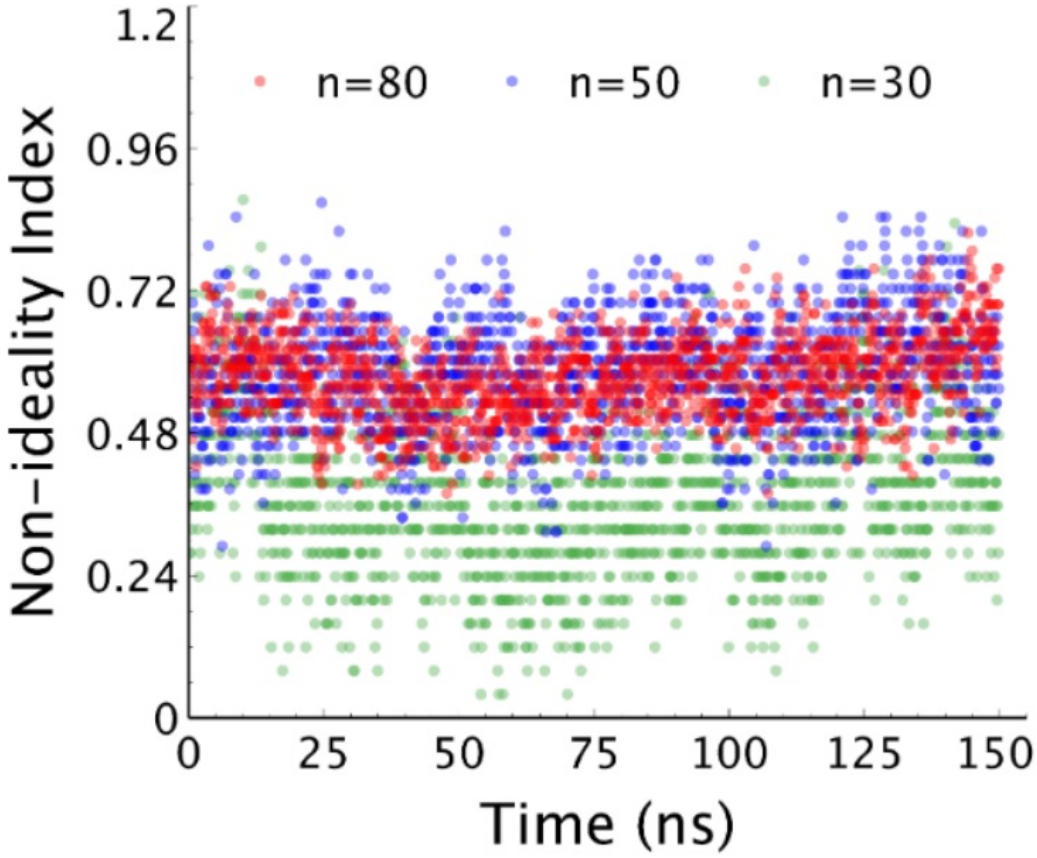


FIG. 7. The correlation between the interfacial tension (γ) with (A) 0- and (B) 1-d persistent entropy and (C) 0- and (D) 1-d Betti-index for the systems with 72, 96, 120 and 144 TBP per interface at the water/vapor surface. The data is fitted to a linear regression model to obtain the slopes provided in the inset.



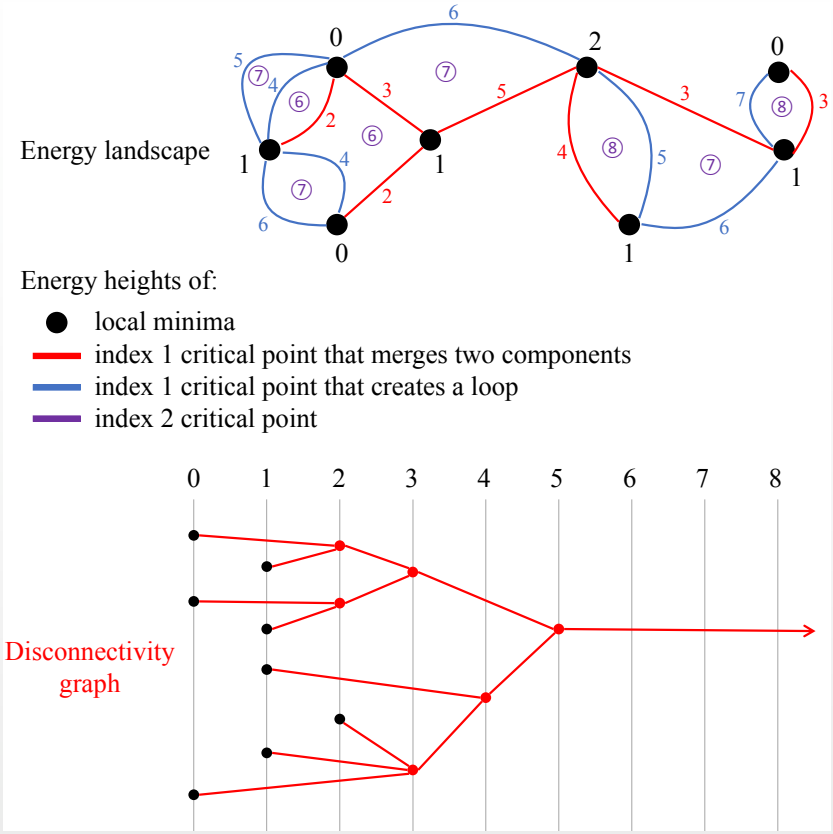
Non-transporting interfaces

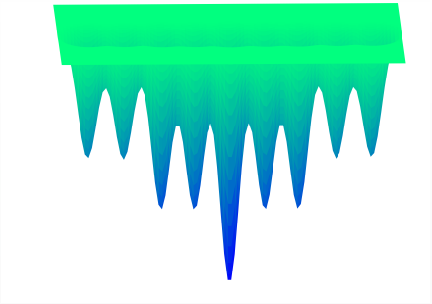
Transporting interfaces

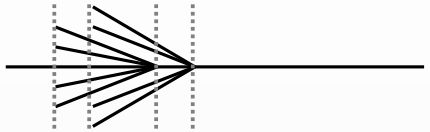
What have we learned?

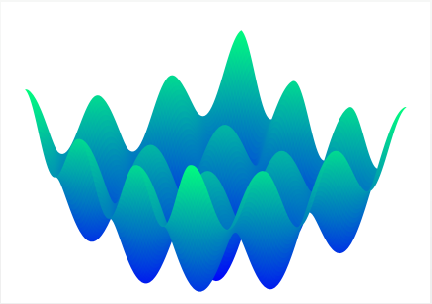
- Huge number of tools available for creating new representations of “states” in soft matter systems
- Across lengthscales (locally, globally)
- Capturing dynamical phenomena
- Now…how do we connect to energy landscapes?
  - Brute force mapping along collective variables (like PR)
  - Use dynamical information to obtain energetics (Markov)
  - Encode topological information
    - Use to predict EL?

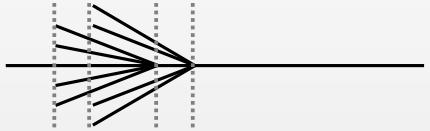
Compact Representations of PELs



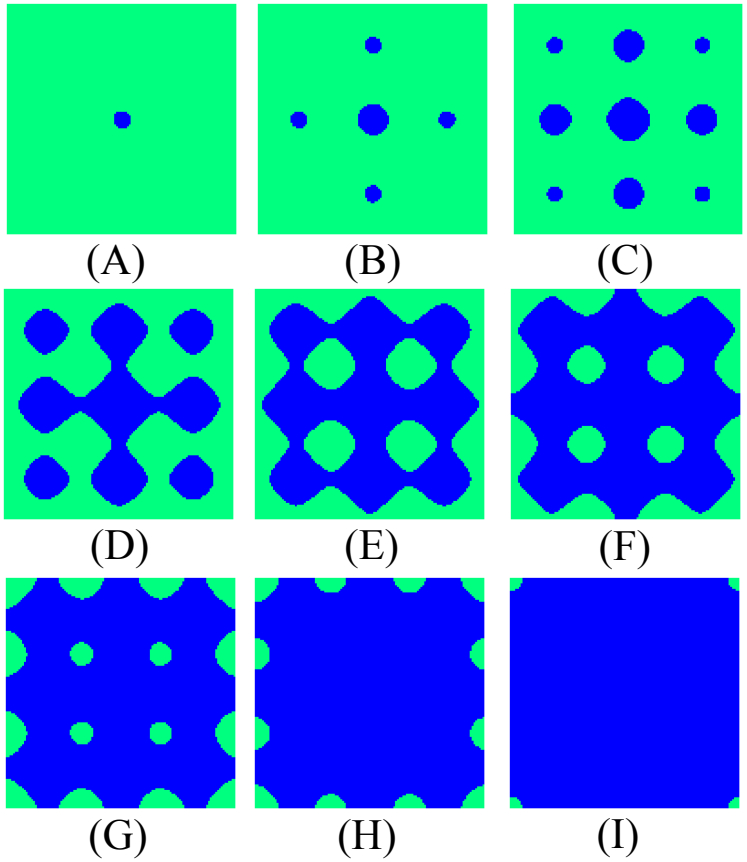


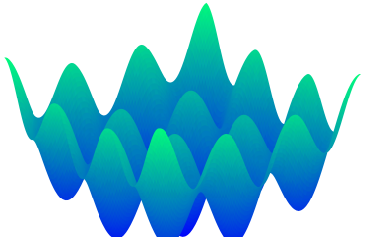


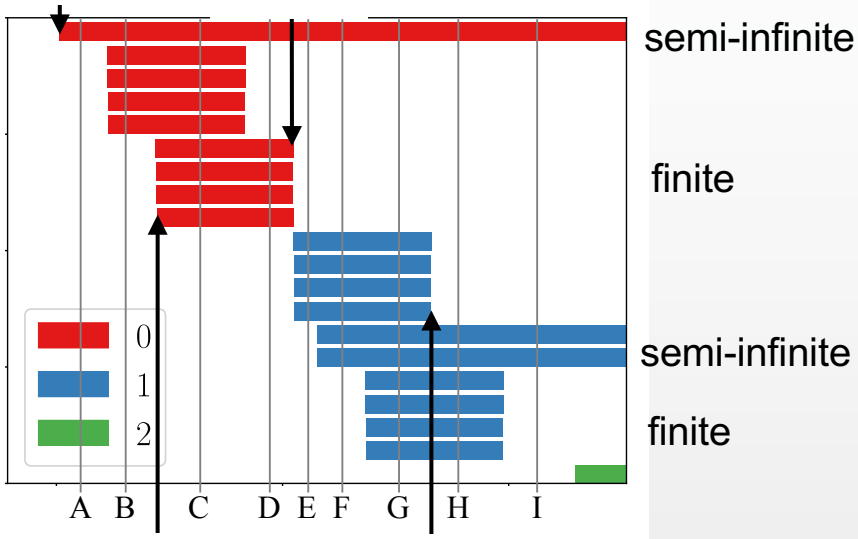




Sublevelset Persistent Homology of PEL

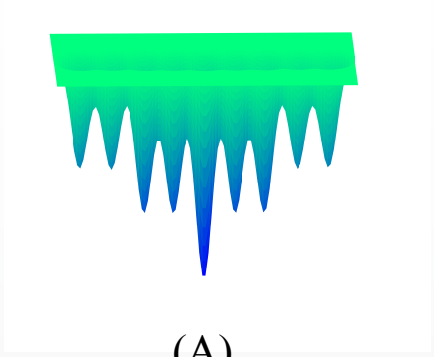




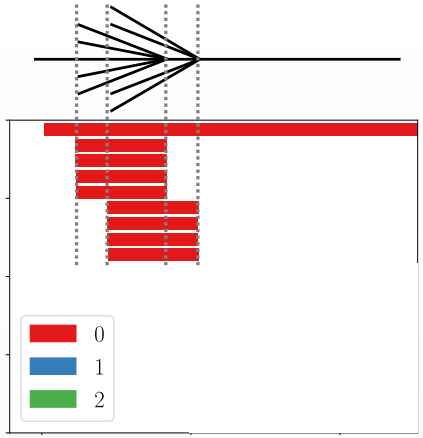


Journal of Chemical Physics, **2021**, *154,* 114114

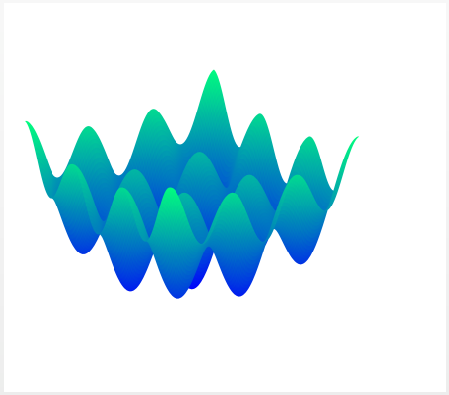
Higher Dimensionality Information is Impt!



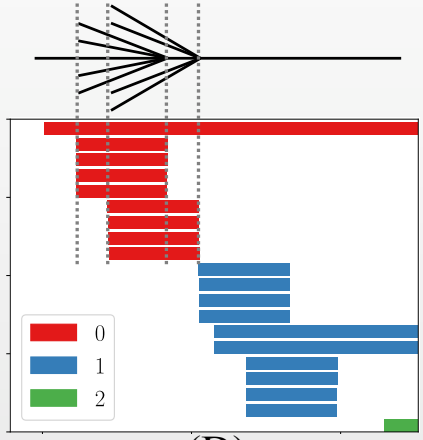
(A)



(B)

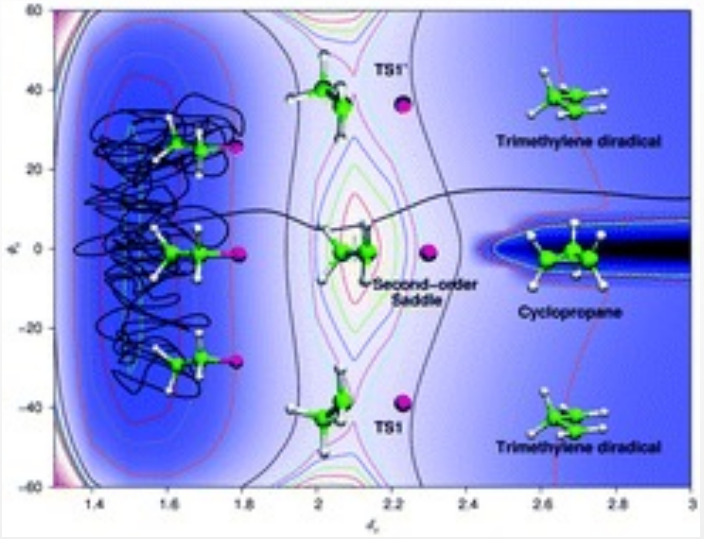


(C)



(D)

- Distinguishing between PEL
- Understanding chemical dynamics
- Relevance of 2nd order pathways



Phys. Chem. Chem. Phys., **2019**, *21*, 12837-12842

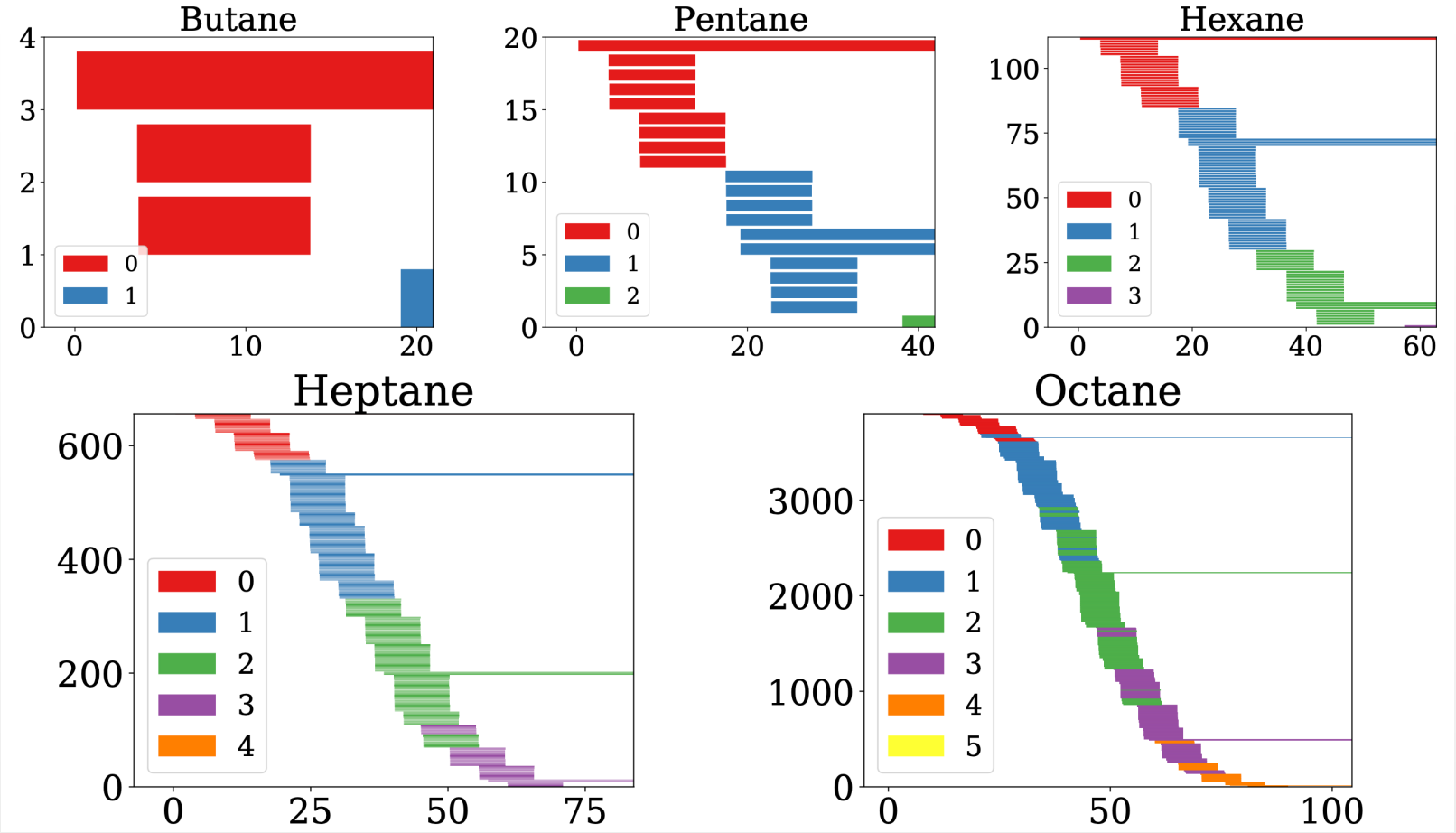
Sublevelset PH of PEL

- Visualization of higher-dimensions of PEL (no dim. red. needed!)
- Capturing more information about PEL topology
  - Bar lengths are barrier heights
  - # of bars are # of barriers
  - Include all indices of critical points (all dimensions of PEL)

- Create a platform for quantifying differences in PEL in high-dimensional space?

- What patterns emerge and can they be used for predicting PEL?

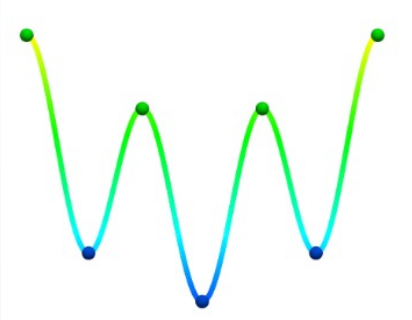
Sublevelset PH of Conformational PEL

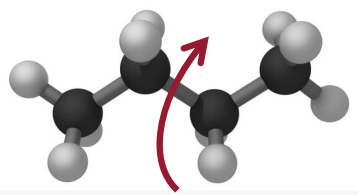


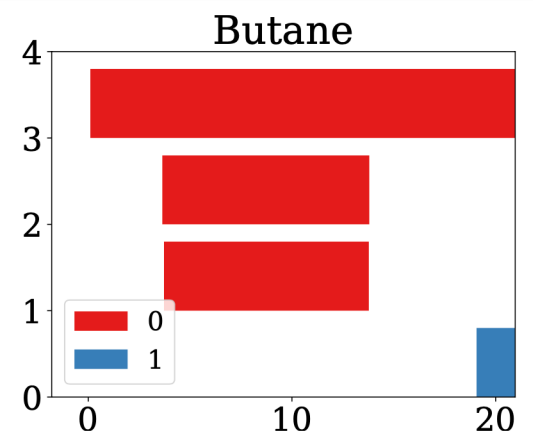
Journal of Chemical Physics, **2021**, *154,* 114114

Sublevelset PH of Conformational PEL

- What's notable about these barcodes?
- # of n-dimensional bars

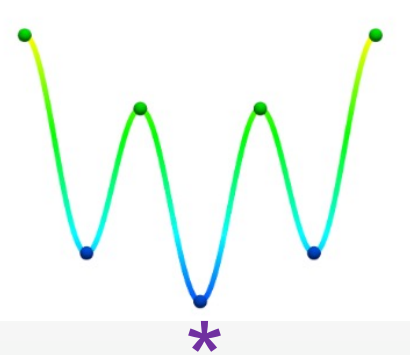


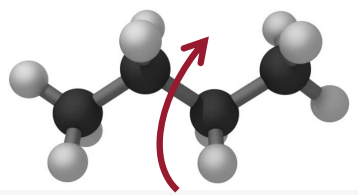


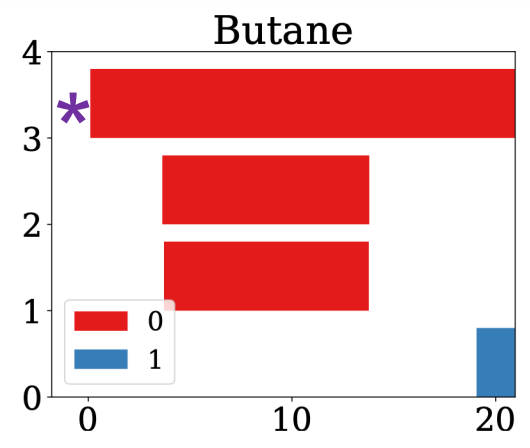


Sublevelset PH of Conformational PEL

- What's notable about these barcodes?
- # of n-dimensional bars



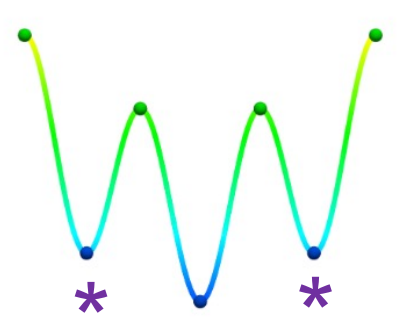


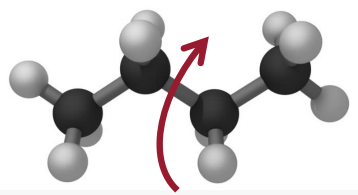


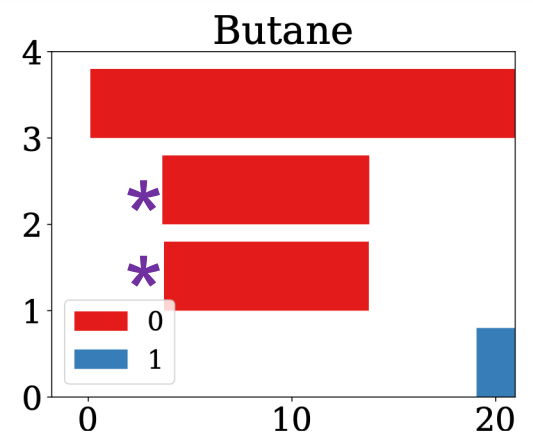
- Birth and death times of features
  - Relative energy differences of minima and maxima

Sublevelset PH of Conformational PEL

- What's notable about these barcodes?
- # of n-dimensional bars



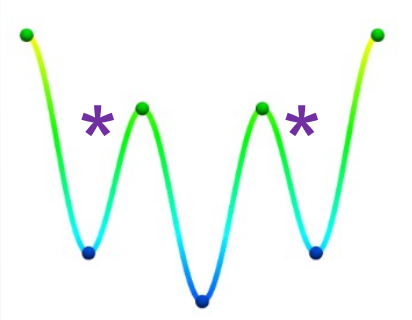


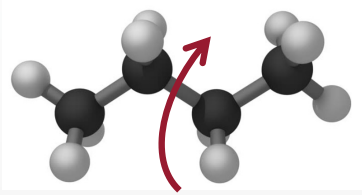


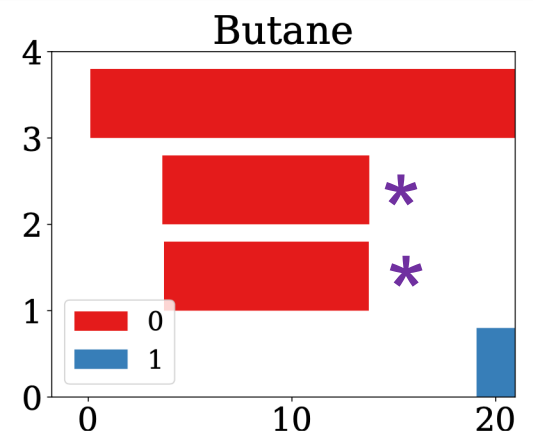
- Birth and death times of features
  - Relative energy differences of minima and maxima

Sublevelset PH of Conformational PEL

- What's notable about these barcodes?
- # of n-dimensional bars

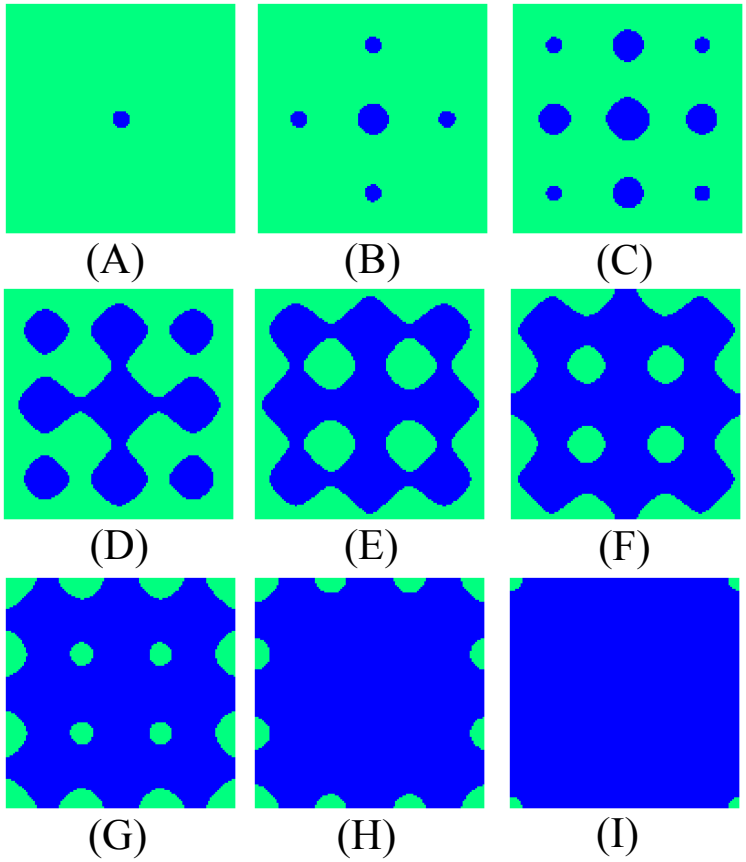


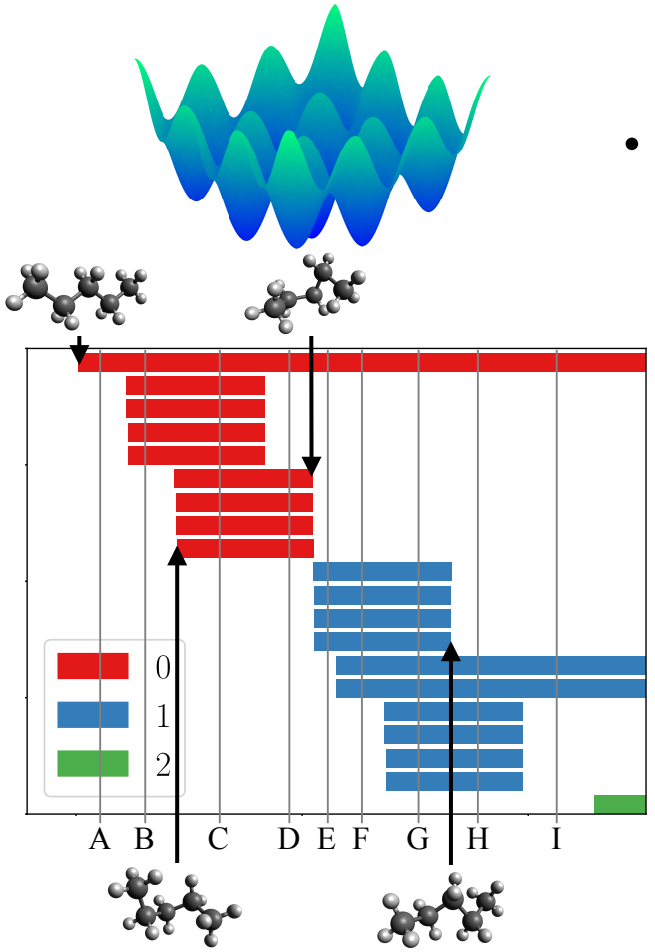




- Birth and death times of features
  - Relative energy differences of minima and maxima

Sublevelset PH of Conformational PEL

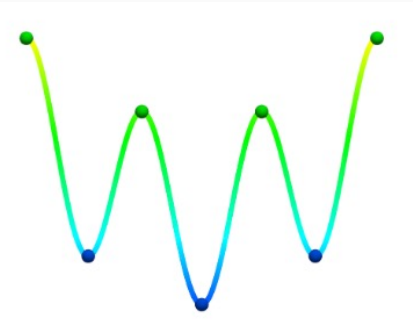


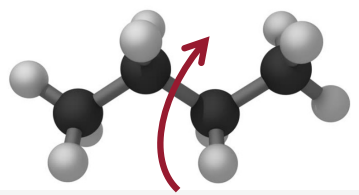


- Relate configuration to energy, local topology

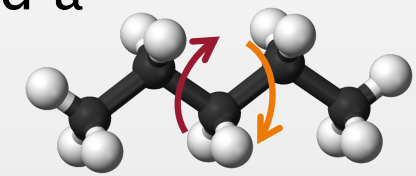
Patterns in the barcodes for n-alkanes

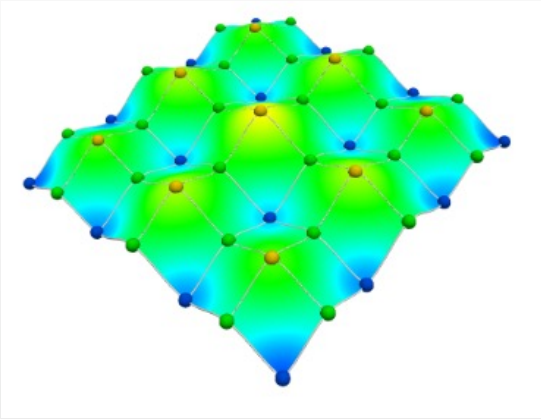
- The rotation about 1 dihedral has a PEL that can be plotted on the circle (1-dim torus)
- 6 critical points for each dihedral PEL (3 maxima and 3 minima)



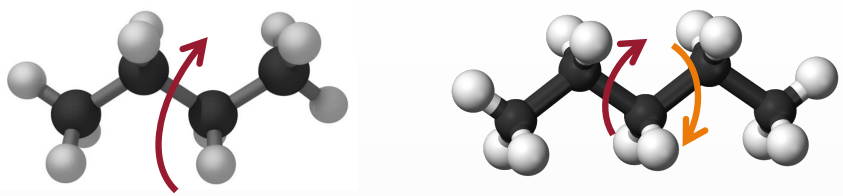


- For every added C-atom, add a
- dihedral degree of freedom





Patterns in the barcodes for n-alkanes



- The rotations of butane are 2x in pentane, all rotations are translated copies
- True as *n* C-atoms are added to alkane PEL
- PEL are nested functions if they are constructed additively
- Mathematically prove the # of bars in *k*-dimensions, birth and death times for any *n*-alkane

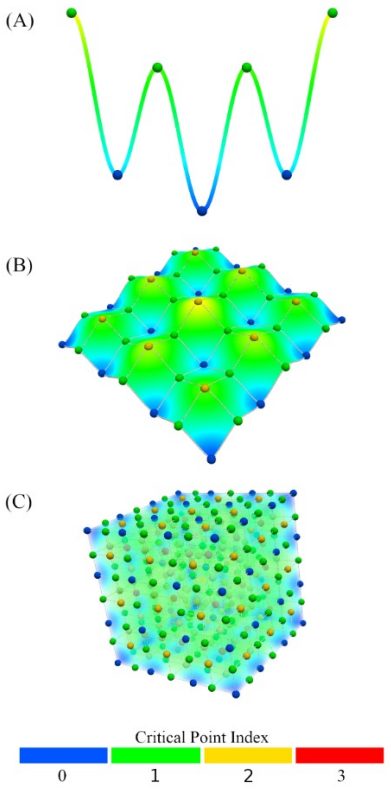
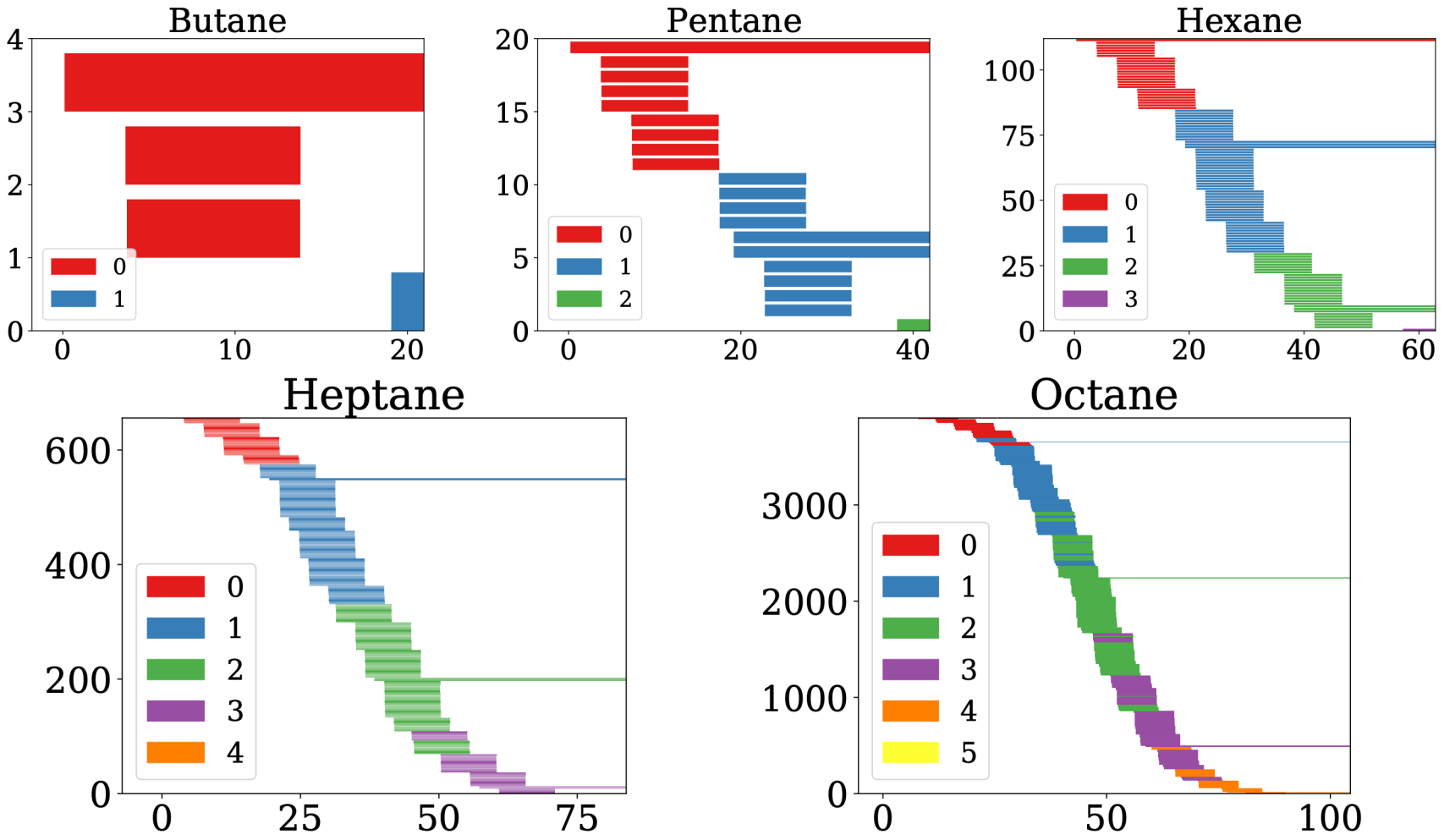


FIG. 3. Morse–Smale complexes for the analytical PEL for (A) butane, (B) pentane, and (C) hexane. The critical points are indicated and are colored by their index. For pentane and hexane, the unique flows between critical points whose indices differ by one are also indicated; see Appendix D. For (A) and (B), the energy scales are indicated by the vertical direction.

Journal of Chemical Physics, **2021**, *154,* 114114

Predicted Sublevelset PH



Extensions to Any Additive Intramolecular Potential

- Energy landscape $V: X \to \mathbb{R}$
- is a function over a product spac$X = X_1 \times \ldots \times X_n$.
- These are composed of building block functions

$$g_i: X_i \to \mathbb{R}$$

- Thus

$$V(x_1, \ldots, x_n) = g_1(x_1) + \ldots + g_n(x_n) = \sum_{i=1}^{n} g_i(x_i)$$

- Building blocks could be on the circle (dihedrals), HO's, etc.

J. Chem. Phys. **2023** *158*, 164104

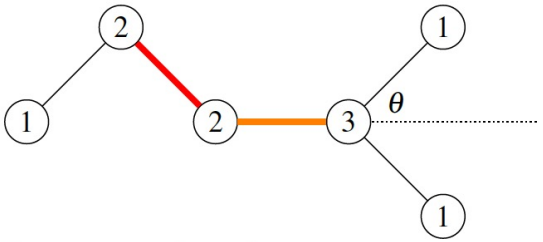
Extensions to Any Additive Intramolecular Potential

$$V : S^1 \times S^1 \to \mathbb{R}$$

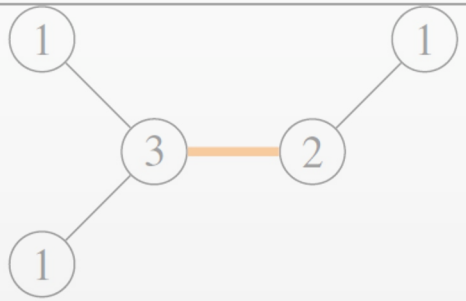
$$V(x_1, x_2) := g_1(x_1) + g_2(x_2)$$

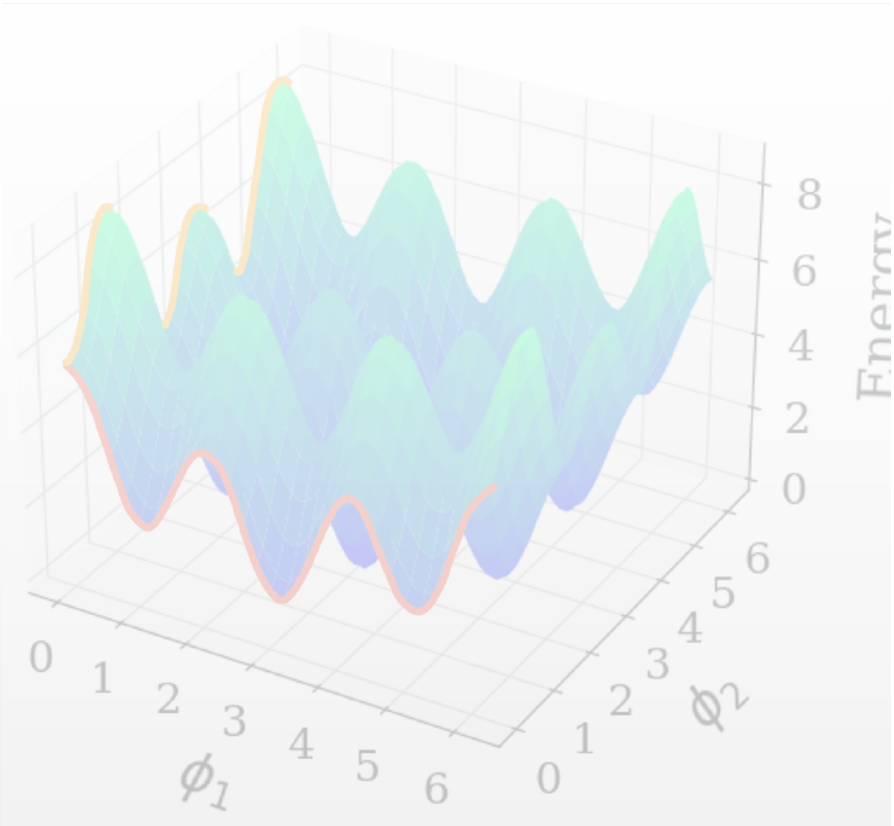
$$g_1 : S^1 \to \mathbb{R}$$

2-methylpentane



1-3-2-1 (isopentane)



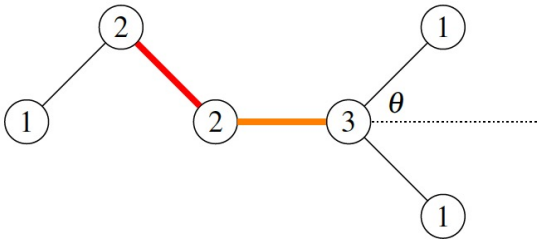


J. Chem. Phys. **2023** *158*, 164104

Extensions to Any Additive Intramolecular Potential

$$V : S^1 \times S^1 \to \mathbb{R}$$

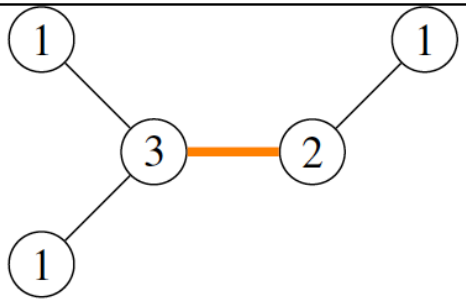
2-methylpentane

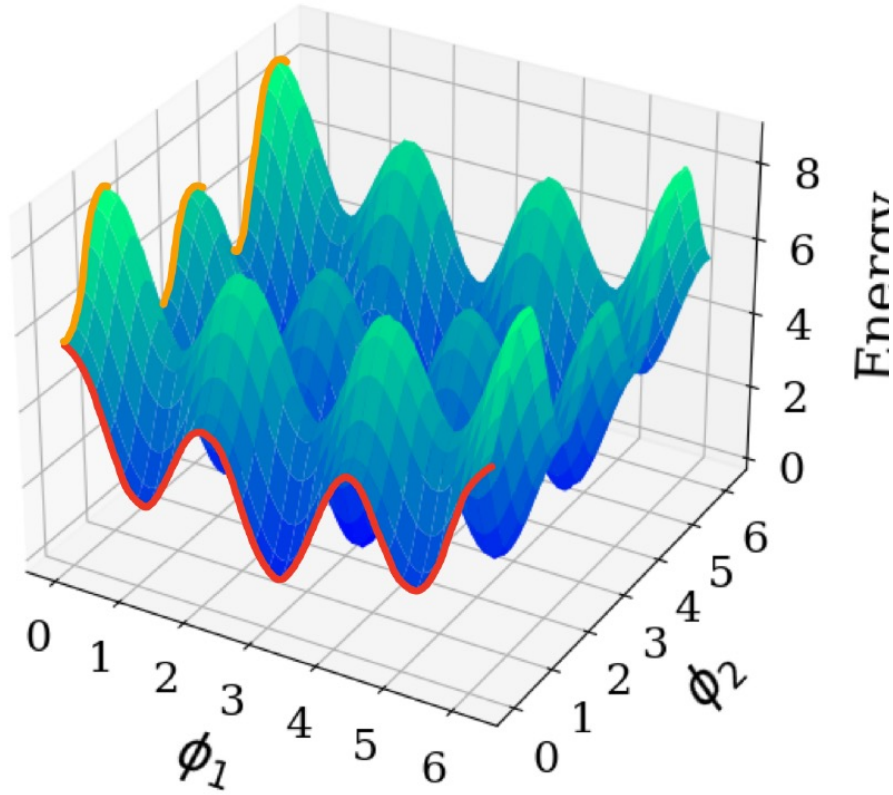


$$V(x_1, x_2) := g_1(x_1) + g_2(x_2)$$

$$g_1 : S^1 \to \mathbb{R}$$

1-3-2-1 (isopentane)



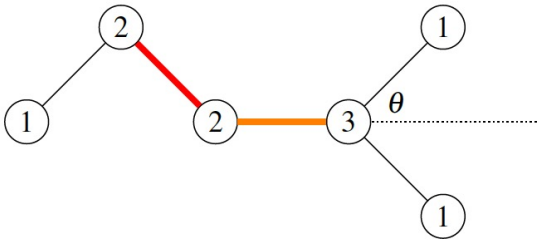


J. Chem. Phys. **2023** *158*, 164104

Extensions to Any Additive Intramolecular Potential

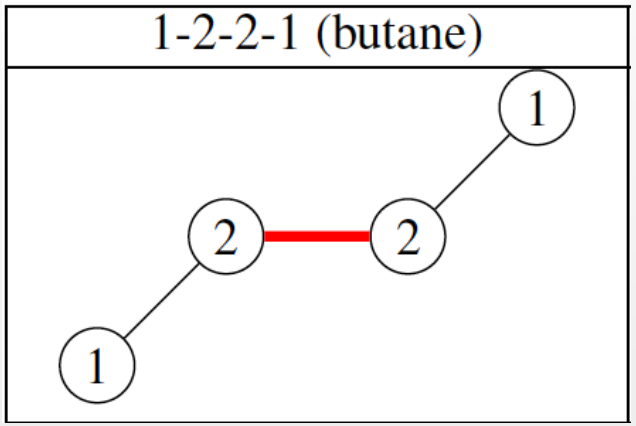
$$V: S^1 \times S^1 \to \mathbb{R}$$

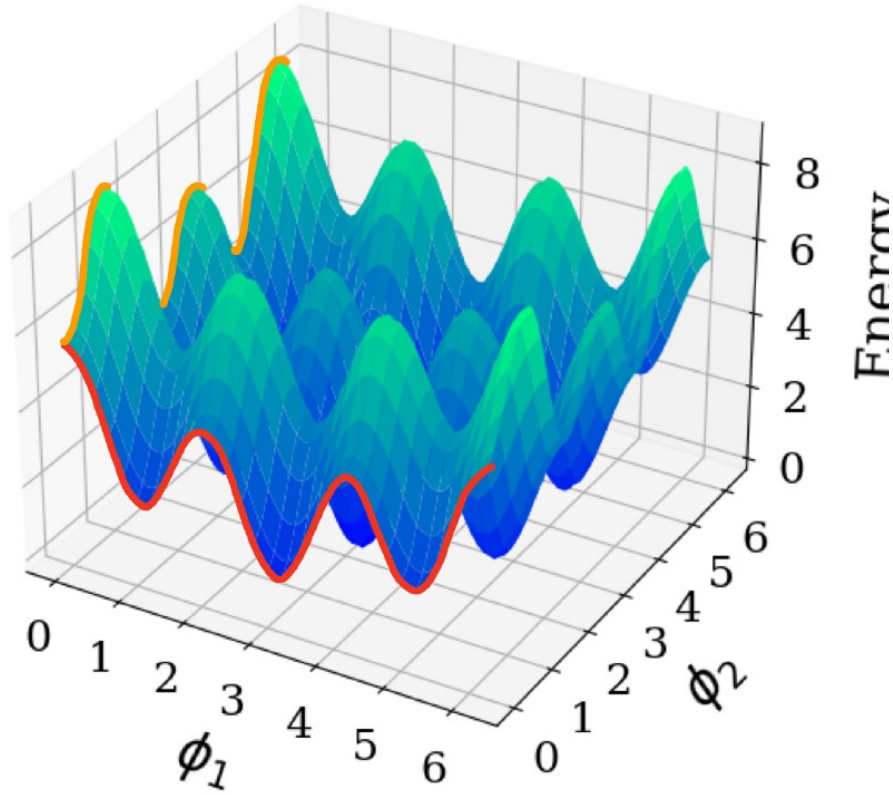
2-methylpentane



$$V(x_1, x_2) := g_1(x_1) + g_2(x_2)$$

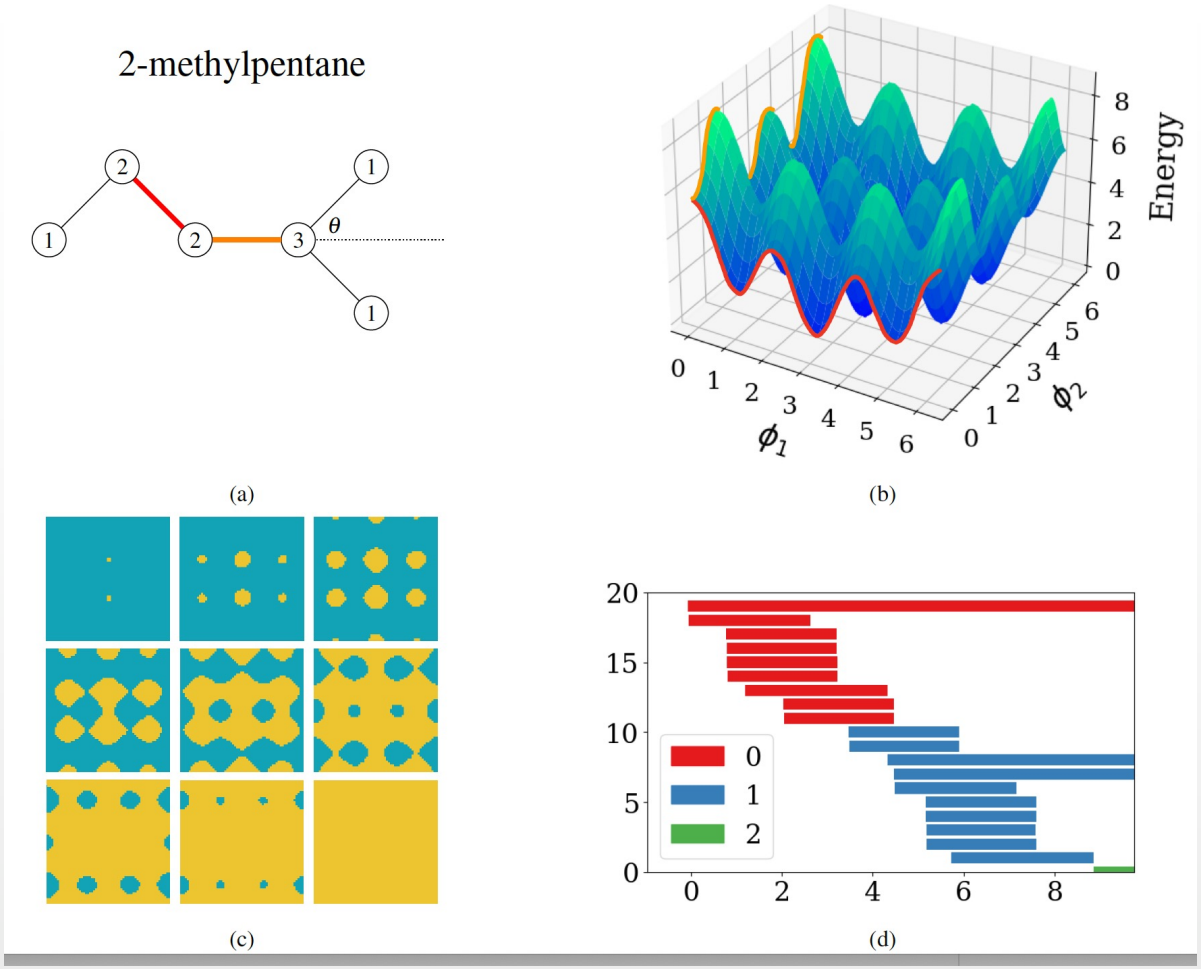
$$g_2: S^1 \to \mathbb{R}$$





J. Chem. Phys. **2023** *158*, 164104

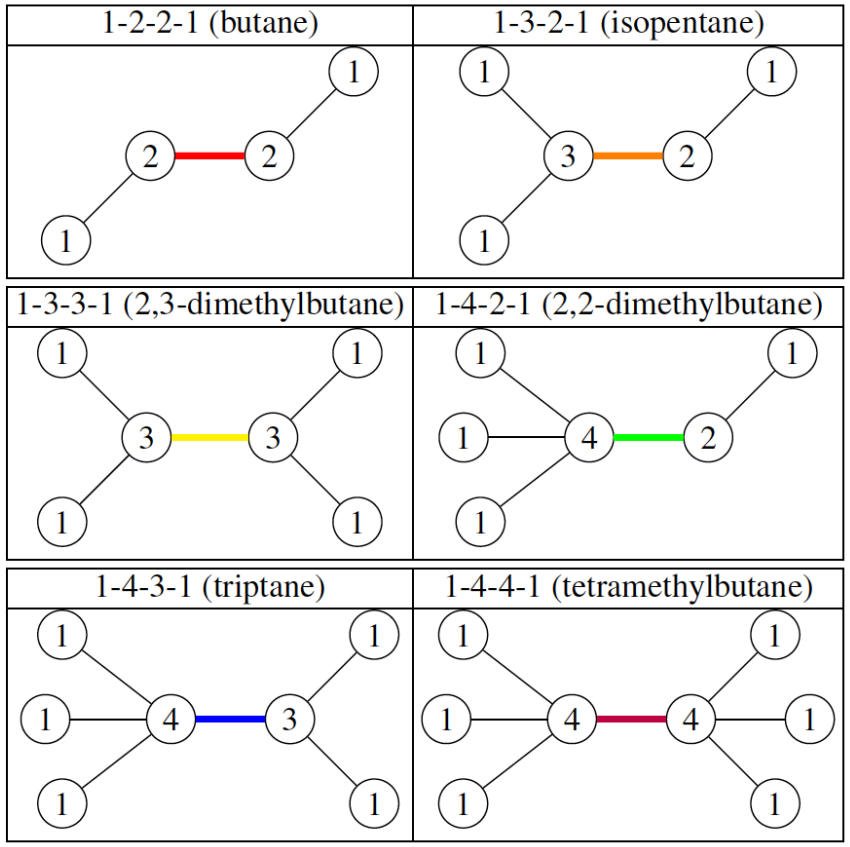
Extensions to Any Additive Intramolecular Potential



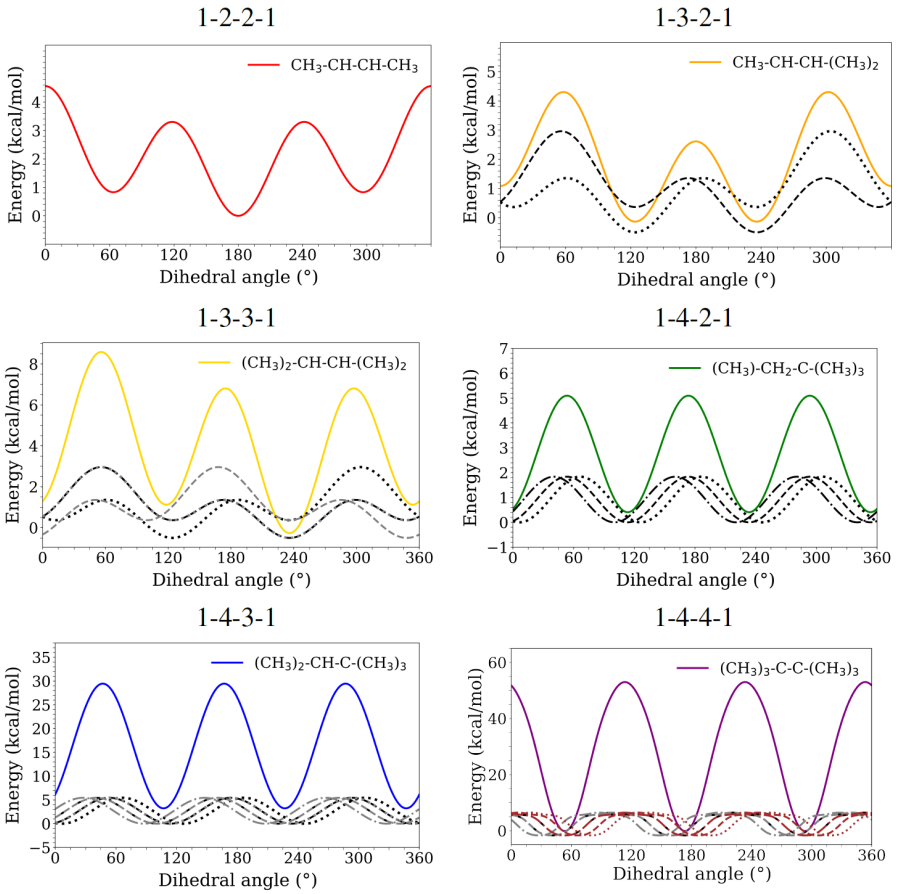
- The sublevel set filtration for V is a “tensor product” of the sublevelset filtrations of g_n

Branched Alkane Examples

Building block units

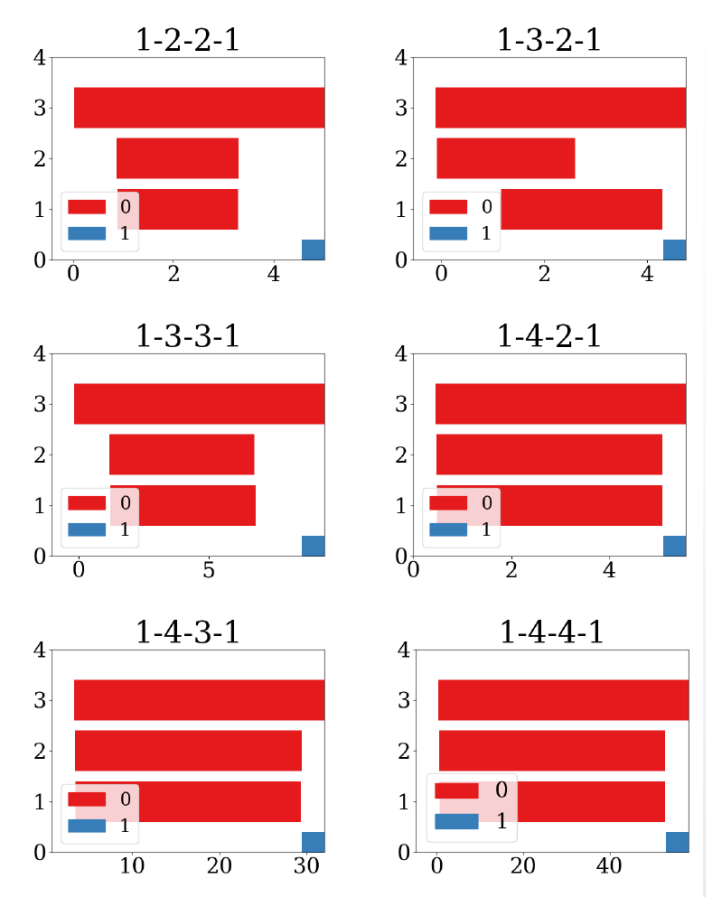


Building block functions

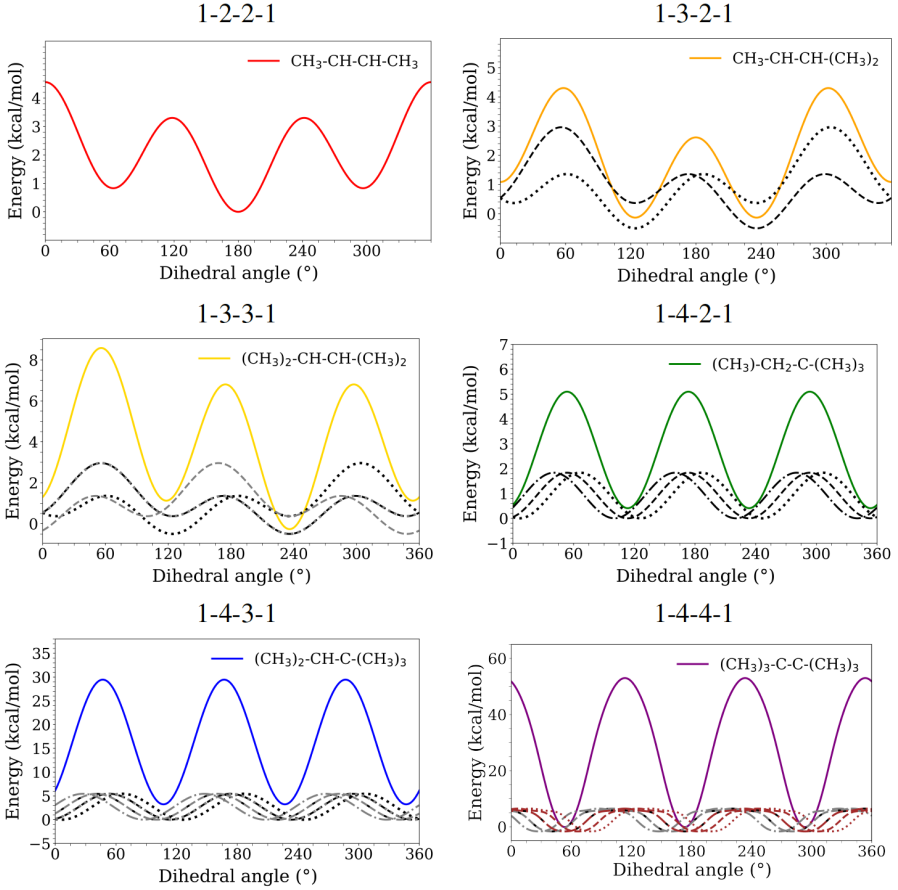


Branched Alkane Examples

Building block sublevel set PH

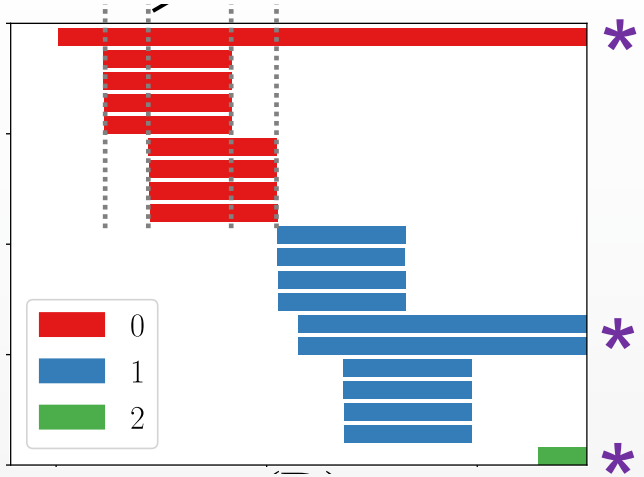


Building block functions



Mathematical Construction of the PH

- Generalizing the prior work
  - *n* building block functions on the circle
  - Domain of *V* is $(S^1)^n$.
  - *n*-dimensional torus has k-dimensional homology of rank $\binom{n}{k}$
  - The semi-infinite bars in a filtration space recover the homology of the space - thus



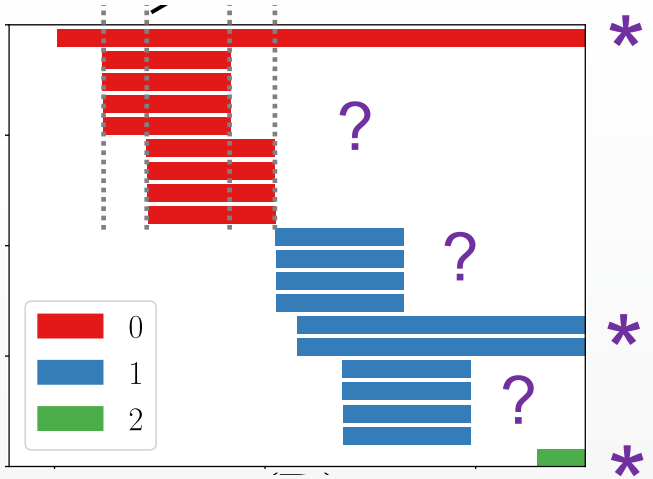
Lemma 1. *The energy landscape* $V : (S^1)^n \to \mathbb{R}$ *for any branched alkane built from n building block functions has* $\binom{n}{k}$ *semi-infinite bars in dimension k.*

Mathematical Construction of the PH

- **How to know total # of bars?**
  - semi-infinite + finite bars

$\binom{n}{k}$ semi-infinite bars in dimension k, for each $0 \leq k \leq n$

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \text{ semi-infinite bars}$$



- Every building block function is a Morse function, hence every V is a Morse function
- Birth and death times correspond directly to CP's (p_i) of V

2^n belong to semi-infinite bars

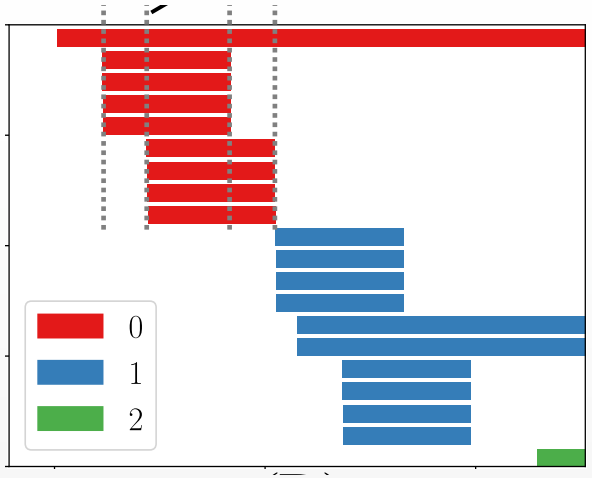
$\prod_{i=1}^{n} p_i - 2^n$ must be split between the birth and death times of the finite bars

- # of CP's is even on the circle – thus $\prod_{i=1}^{n} p_i - 2^n$ is even

Mathematical Construction of the PH

- # of finite bars is then

$$\frac{1}{2}\left(\prod_{i=1}^{n} p_i - 2^n\right)$$

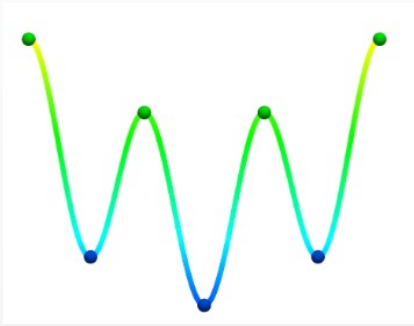


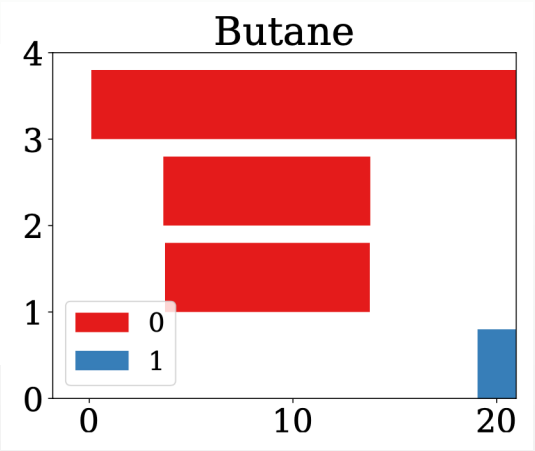
Theorem 1. *Suppose the additive energy function* $V : (S^1)^n \to \mathbb{R}$ *is given by* $V(\phi_1, \ldots, \phi_n) = g_1(\phi_1) + \ldots + g_n(\phi_n)$ *with each Morse function* $g_i : S^1 \to \mathbb{R}$ *having* p_i *critical points. The total number of bars in the sublevelset persistent homology of* V *is*

$$(2^n) \text{ semi-infinite bars} + \frac{1}{2}\left(\prod_{i=1}^{n} p_i - 2^n\right) \text{ finite bars.}$$

Mathematical Construction of the PH

- How many bars in each dimension?
  - Use the # of bars in *n* to identify # of bars in *n*+1 (nested aspect of V_n)
  - Begin with any building block function
  - Example dihedral angle (on the circle) – 6 CPs, 3 k=0 bars





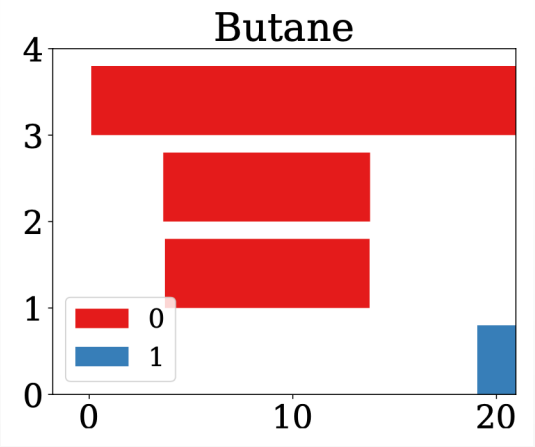
3 bars of dimension 0

1 bar of dimension 1

❑ Suppose these come from the following:

$$\binom{n}{k} + (3^n - 1)\binom{n-1}{k}$$

Mathematical Construction of the PH



$$\binom{1}{0} + (3^1 - 1)\binom{1-1}{0} = 3 \text{ bars of dimension } 0$$

$$\binom{1}{1} + (3^1 - 1)\binom{1-1}{1} = 1 \text{ bar of dimension } 1$$

- Does this equation hold as *n* (# building blocks) is increased?

$$V_n : (S^1)^n \to \mathbb{R} \longrightarrow V_{n+1} : (S^1)^{n+1} \to \mathbb{R}.$$

where

$$V_{n+1}(\phi_1, \ldots, \phi_{n+1}) = V_n(\phi_1, \ldots, \phi_n) + g_{n+1}(\phi_{n+1})$$

$$g_{n+1} : S^1 \to \mathbb{R}$$

Mathematical Construction of the PH

- **If you know the # of bars of each dimension of the building block, then there is a finite # of combinations of how to add them (Kunneth formula)**
  - Find each k-dimensional bar in V_{n+1} in terms of the sublevelset persistence bars of V_n and g_{n+1}

(i) combining k-dimensional bars of V_n with 0-dimensional bars of g_{n+1},

(ii) combining $(k-1)$-dimensional bars of V_n with 1-dimensional bars of g_{n+1}, or

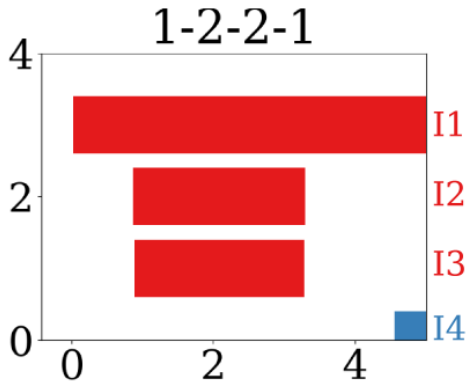
(iii) combining $(k-1)$-dimensional bars of V_n with 0-dimensional bars of g_{n+1} via torsion.

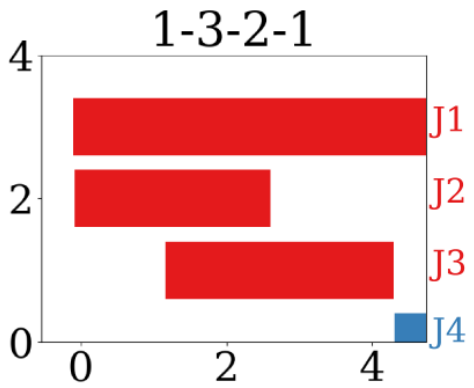
Mathematical Construction of the PH

$$
\begin{aligned}
&3\left[\binom{n}{k}+(3^n-1)\binom{n-1}{k}\right]+\left[\binom{n}{k-1}+(3^n-1)\binom{n-1}{k-1}\right]+2\left[(3^n-1)\binom{n-1}{k-1}\right]\\
=&3\binom{n}{k}+3(3^n-1)\binom{n-1}{k}+\binom{n}{k-1}+3(3^n-1)\binom{n-1}{k-1}\\
=&3\binom{n}{k}+3^{n+1}\binom{n-1}{k}-3\binom{n-1}{k}+\binom{n}{k-1}+3^{n+1}\binom{n-1}{k-1}-3\binom{n-1}{k-1}\\
=&3\binom{n}{k}+\binom{n}{k-1}+3^{n+1}\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right]-3\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right]\\
=&3\binom{n}{k}+\binom{n}{k-1}+3^{n+1}\binom{n}{k}-3\binom{n}{k}\\
=&\binom{n}{k-1}+3^{n+1}\binom{n}{k}\\
=&\binom{n}{k-1}+\binom{n}{k}+(3^{n+1}-1)\binom{n}{k}\\
=&\binom{n+1}{k}+(3^{n+1}-1)\binom{n}{k},
\end{aligned}
$$

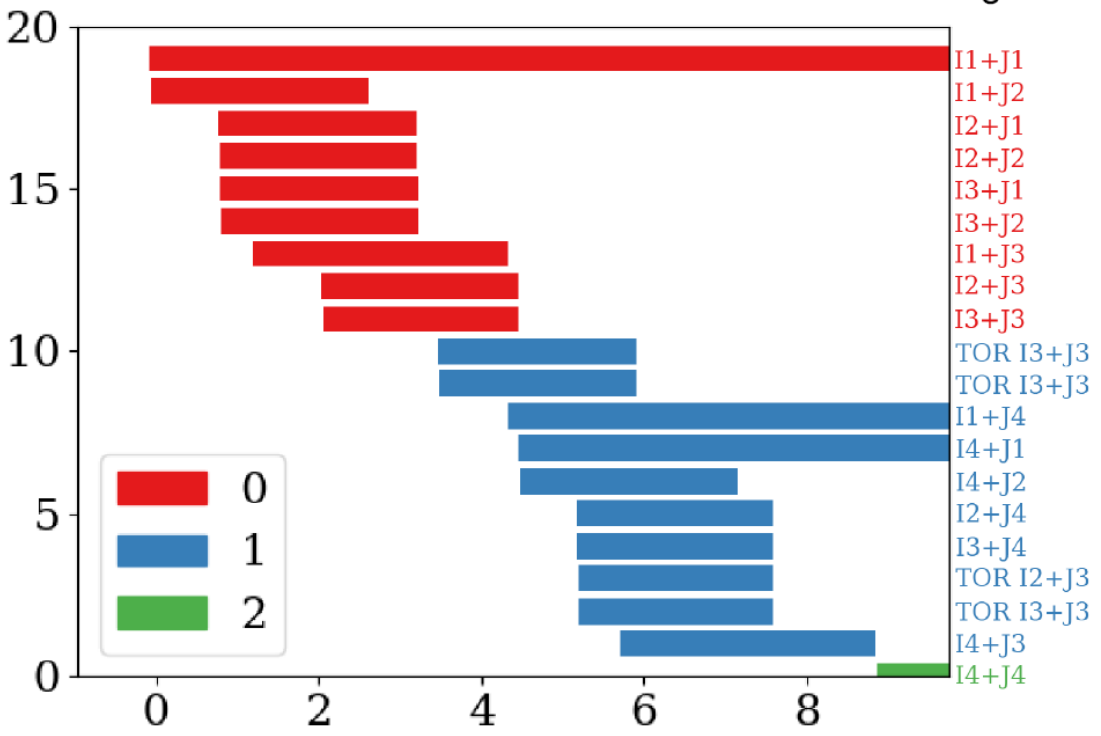
# of bars of each dimension k for a given *n* building blocks

Sublevel set Persistence of 2,2-dimethylpentane



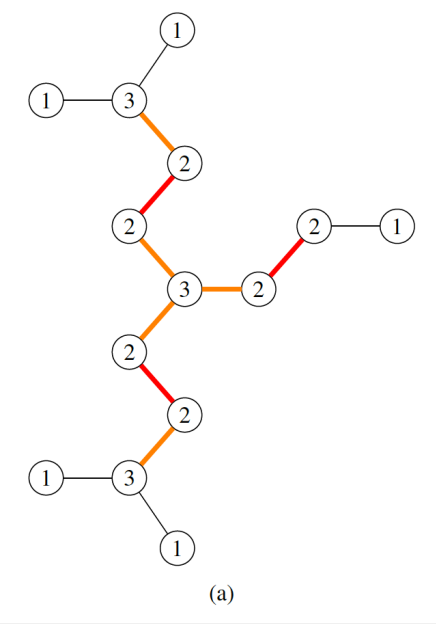


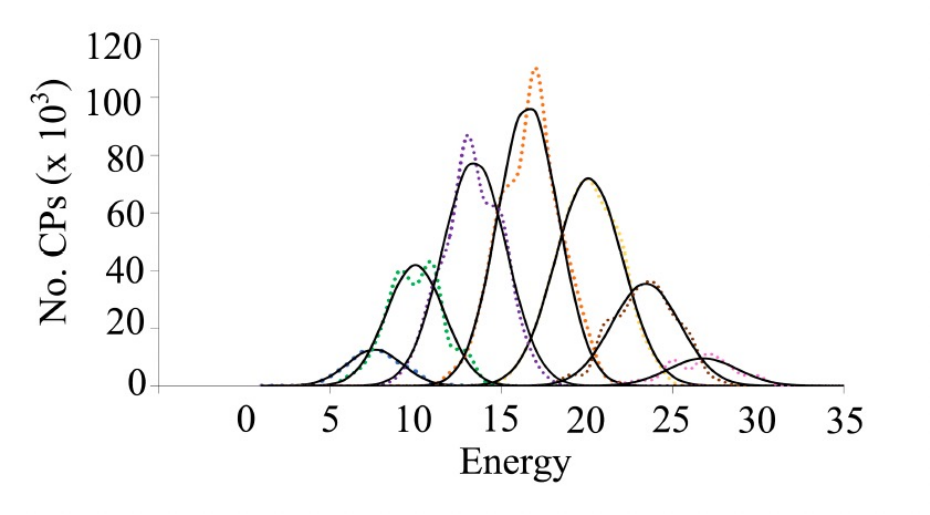
(a)



(b)

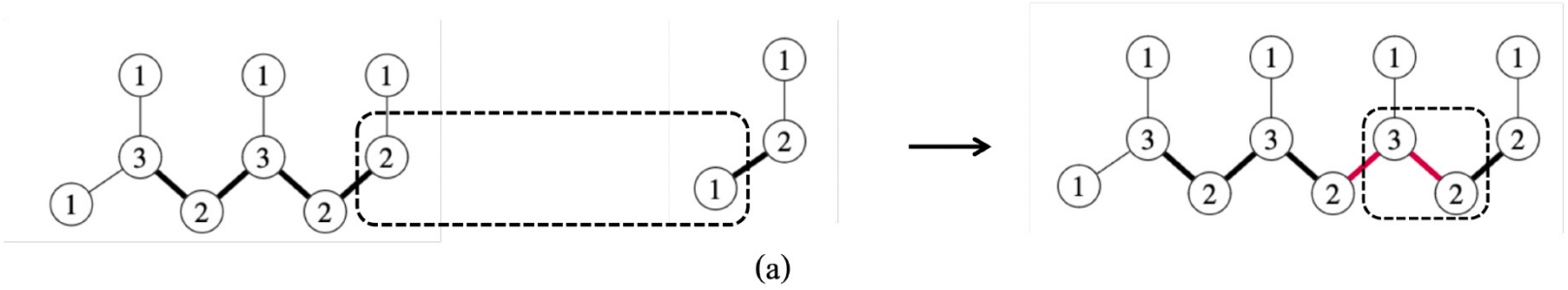
Topology of 2,2-dimethylpentane PEL

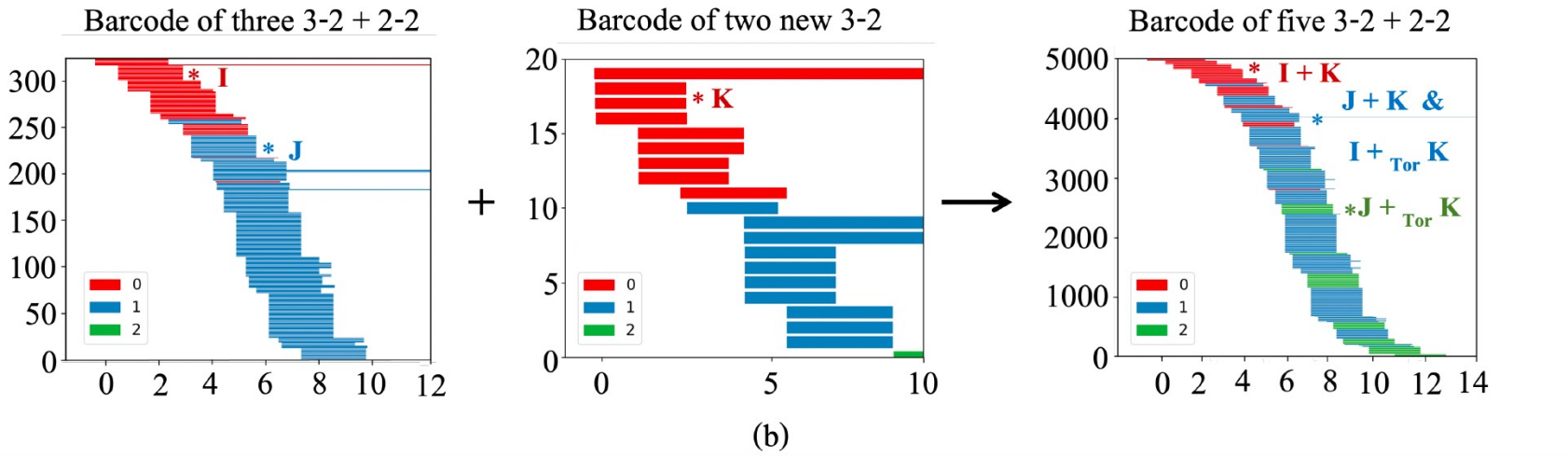




- 7-dim dihedral angle PEL
- 840,000 total CP's
- 32 global minima
- 186 index-0 local minima under 1 kcal/mol
- 1,200 index-0 CP before you overcome any index-1 barriers at 2.75 kcal/mol

Topology of change to PEL w/bond addition





New Opportunities

- Direct connection between structure and location on a PEL

- Learn about patterns in EL barriers and minima related to structure and changes to composition of the system
  - Experimental design strategies → stabilizing specific structure or dynamic phenomena for the application of interest

- Employ in enhanced sampling of EL → accelerate computational workflows

***other great into on applied topology can be found on the AATRN YouTube (Henry Adams)

Soft Matter Design Workflow