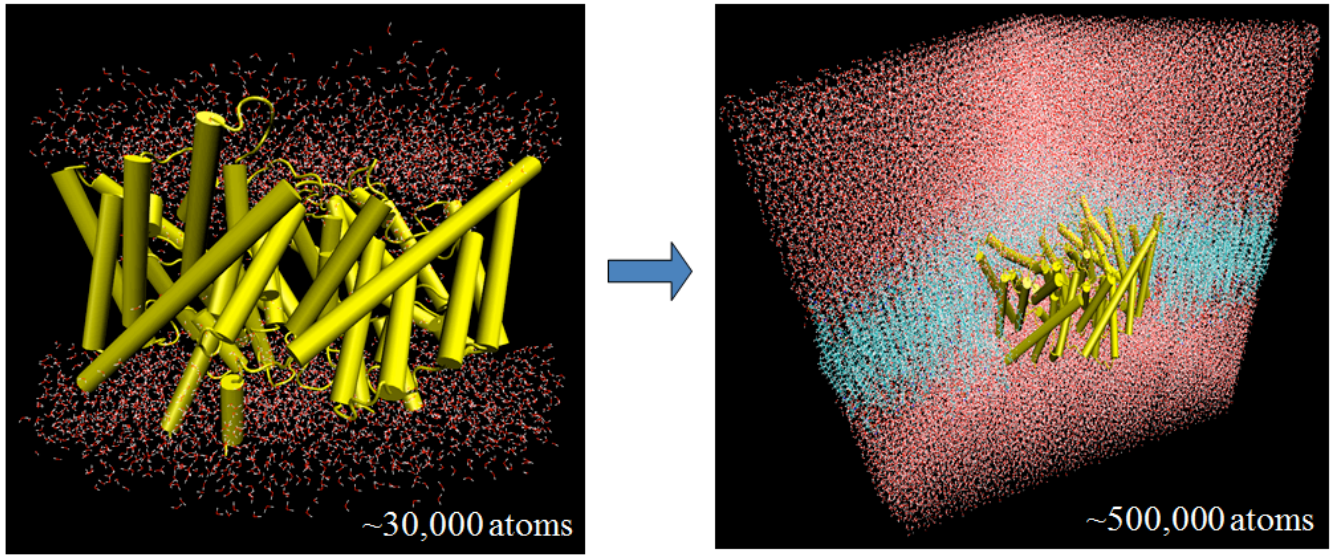
# Efficient and Accurate Multiscale Reactive Molecular Simulations




No lipid & restrained backbone

All atom membrane and protein

Chris Knight (ANL/CI)

PI: Gregory A. Voth (Univ. Chicago/ANL)

# A) Project Overview

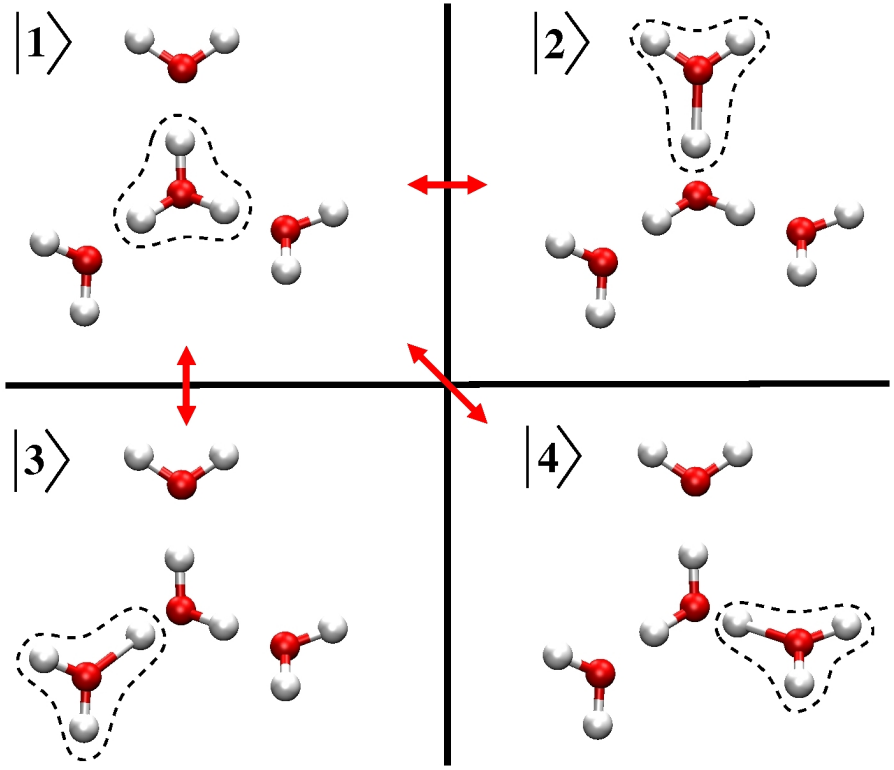
- The project
  - Implementation of multiple program algorithms for multistate simulations.
- Science goals
  - Fundamental understanding of charge transport processes and multiscale phenomena
- The participants, description of team
  - Graduate students and postdocs at U. Chicago and ANL.
- History
  - Reactive methodology in development since early 90's
  - Current code is about 3 years old.
- Goals
  - Scale a single simulation (~1 million atoms) to leadership scale

# B) Science Lesson

- Construct linear combination of chemical bond topologies

$$\mathbf{H} = \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} \\ V_{12} & V_{22} & 0 & 0 \\ V_{13} & 0 & V_{33} & 0 \\ V_{14} & 0 & 0 & V_{44} \end{pmatrix}$$

$$\mathbf{F}_j = \sum_{mn} c_m c_n \mathbf{F}_j^{mn}$$



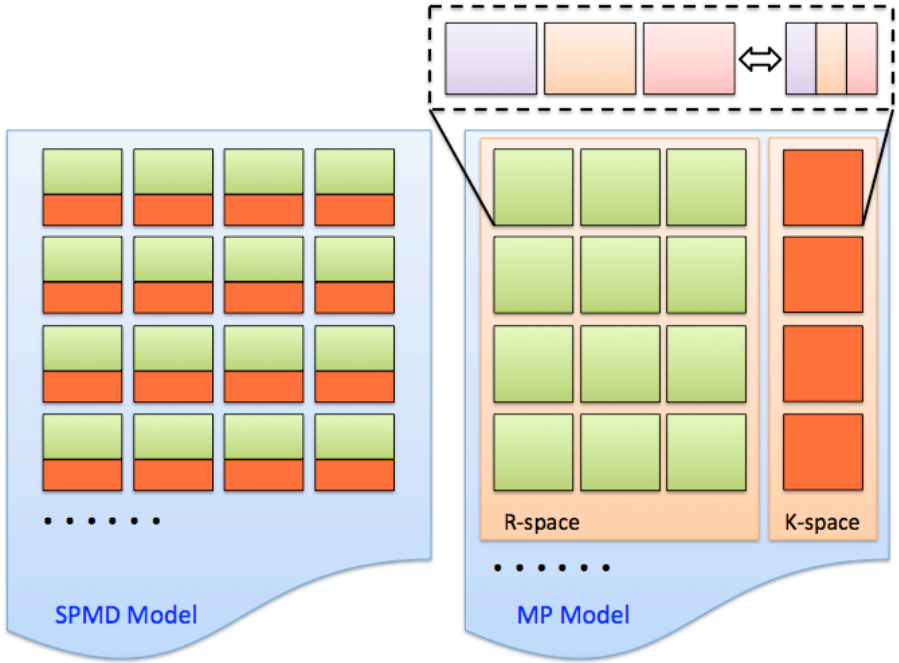

- Integrate Newton's EOM

# C) Parallel Programming Model

- LAMMPS & RAPTOR (developed in our group)
  - http://lammps.sandia.gov/
- Hybrid MPI + OpenMP (now w/ a flavor of multiple program)
  - Also supports GPU (Cuda and OpenCL)
- Highly portable C++ and only requires FFTW
- What platforms does the application currently run on?
  - All popular platforms.
- Current status & future plans for the programming model
  - Focusing heavily on multiple program parallelization

# D) Computational Methods

- Spatial domain-decomposition
- Rely heavily on FFTs for calculation of electrostatic interactions.
- We are currently working on extending to N > 2 partitions.
- Extensions to multiple reactive species (multiple Hamiltonian matrices).




$$\mathbf{H} = \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} \\ V_{12} & V_{22} & 0 & 0 \\ V_{13} & 0 & V_{33} & 0 \\ V_{14} & 0 & 0 & V_{44} \end{pmatrix}$$

# E) I/O Patterns and Strategy

- One set of output files per job
  - Everyone sends data to rank 0
- Approximate sizes of inputs and outputs
  - Inputs < 1 GB
  - Outputs ~100s GB in binary format (dcd, xtc,…)
- Checkpoint / Restart capabilities:
  - Coordinates, velocities, etc… (< 1 GB) in binary format
- Current status and future plans for I/O
  - Contemplating strategy for *n* ranks to write output (*n* < N)

# F) Visualization and Analysis

- How do you explore the data generated?
  - Molecular visualization typically w/ VMD (also serves as tool)
  - At larger scales, might need to switch to parallel visualization program
  - Reduce resolution (coarse-graining)
  - Analysis of 1-D and 2-D distribution functions

- Current status and future plans for your viz and analysis
  - Port analysis codes to GPUs

# G) Performance

- What tools do you use now to explore performance?
  - TAU and HPCToolkit
- What do you believe is your current bottleneck to better performance?
  - We know that the current bottleneck is FFTs and electrostatic calculations.
- What do you believe is your current bottleneck to better scaling?
  - Multiple Program strategy to simultaneously evaluate matrix elements
- Current status and future plans for improving performance
  - Multiple Program + OpenMP
  - approximate electrostatic algorithms (currently in develoment)

# H) Tools

- How do you debug your code?
  - gdb, totalview
- What other tools do you use?
  - VMD w/ Tcl scripts
- Current status and future plans for improved tool integration and support

# I) Status and Scalability

- How does your application scale now?
  - First time ever, we've recently run reactive simulations at 16K cores
- Where do you want to be in a year?
  - Multiple racks of Mira…
- What are your top 5 pains? (be specific)
  - 1-5: Cost of FFTs for accurate electrostatics
- What did you change to achieve current scalability? (replaced algorithms, data structures, math library, etc)
  - Multiple program, Multiple program, Multiple program
- Current status and future plans for improving scaling
  - ditto…

# J) Roadmap

- Where will your science take you over the next 2 years?
  - Efficient and accurate simulation of arbitrary chemical reactions in condensed phases over multiple time and length scales.

- What do you hope to learn / discover?
  - Fundamental understanding of charge transport processes and multiscale phenomena

: Exact Energies & Forces
: Approximate Off-Diagonal Couplings
: All matrix elements approximated.