

Group B

Mentors: James Taylor, Daichi Mukunoki

Improving through optimization
and increased efficiency of
multiscale simulation

Group members

☐ Ivonina Mariia

Background: Quantum Chemistry, Elongation Method

Affiliation: Kyushu University

☐ Nicholas Mills

Background: HPC I/O and Parallel File Systems

Affiliation: Clemson University

☐ Sun Qiwen

Background: Mathematics, Bayesian Inference

Affiliation: Nagoya University

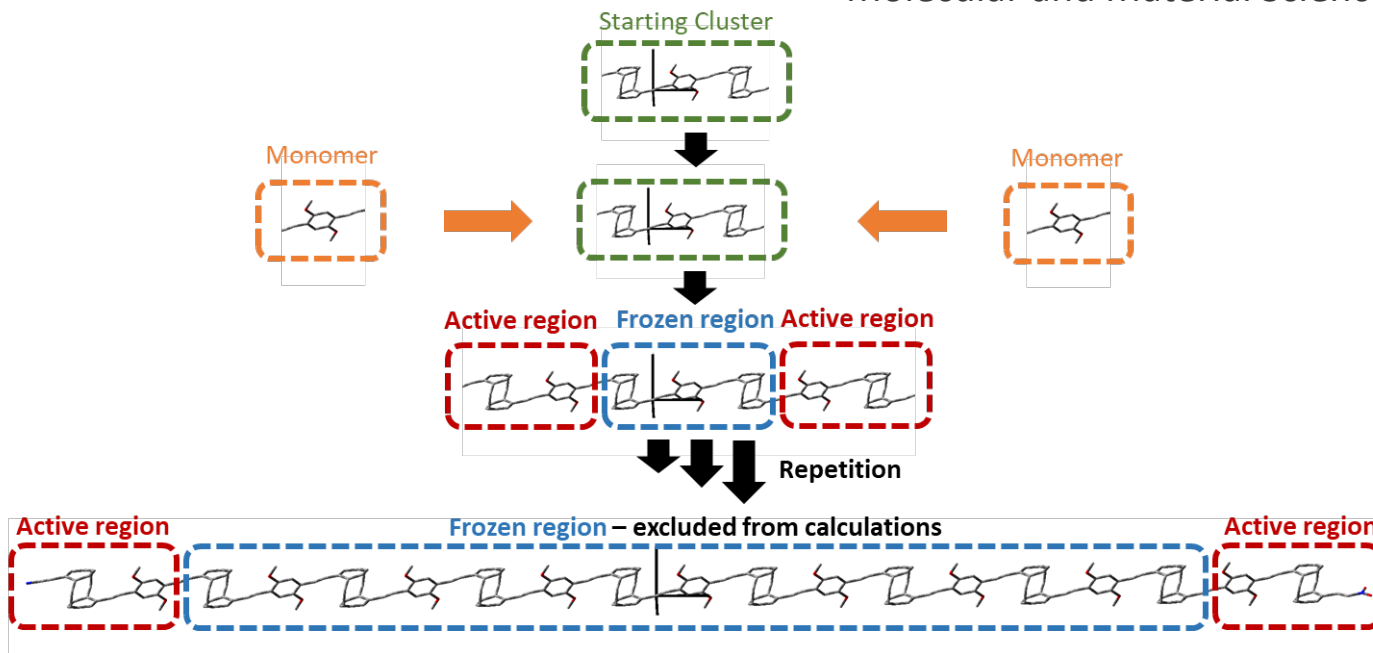
☐ Yuta Yamaguchi

Background: Physics, Molecular Dynamics simulation

Affiliation: University of Tokyo

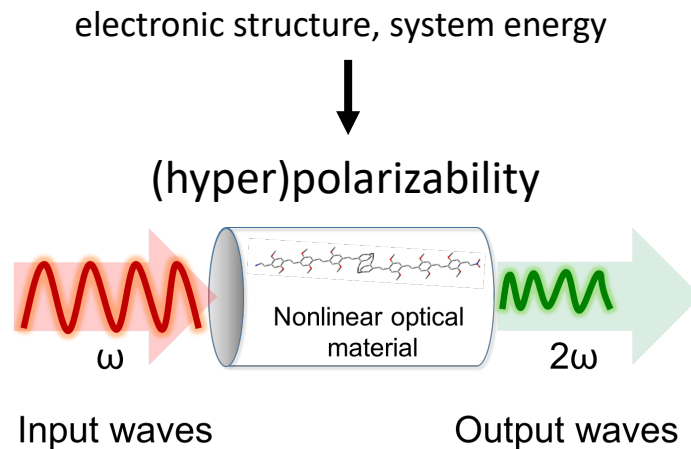
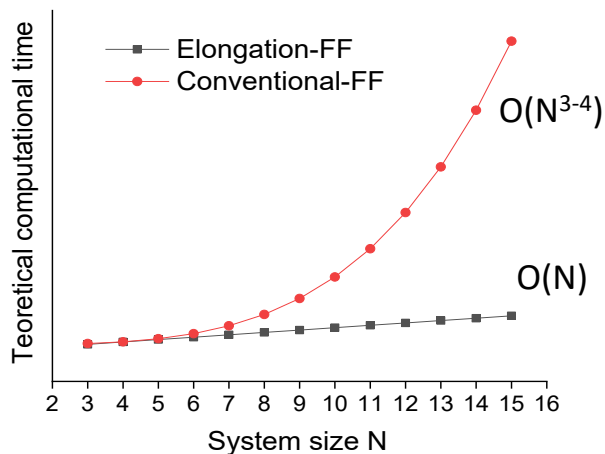
Elongation method for calculating nonlinear optical properties of long polymer chains

Ivonina Mariia,
Molecular and Material Science Department,
Kyushu U



Advantages

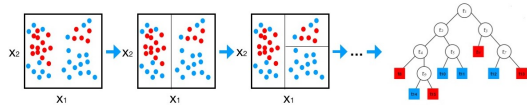
Applications



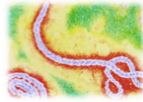
SUN Qiwen
 Nagoya University
 Graduate School of Mathematics D1
 R-CCS
 Data Assimilation Research Team JRA (2020 April~)

- Bayesian inference
- Probability
- Statistical learning
- Functional analysis

The Application of Tree-based Methods on the Analysis of MathSciNet Database



Group 1: *Ebola*



Question:
 How Can We Estimate
 the Size of Ebola Outbreak in Kenya?
 (with Quarantine Intervention)

Modeling the Effect of
 Vaccination for Children to Prevent
 Hand, Foot, and Mouth Disease

Bayesian Filtering, Predicting and Smoothing

Sun Qiwen in collaboration with Data Assimilation Research Team (RIKEN)
 Graduate School of Mathematics, Nagoya University, Japan
 m7040@math.nagoya-u.ac.jp

Background

- Unobserved signal $X_t \in \mathbb{R}^n$, $t \in \mathbb{N}$, and observations $Y_t \in \mathbb{R}^m$, $t \in \mathbb{N}$
- A probabilistic state space model is described as

$$\begin{aligned} &\text{initial distribution: } p(X_0) \\ &\text{dynamical system: } p(X_t | X_{t-1}), t \geq 1 \\ &\text{measurement system: } p(Y_t | X_t), t \geq 1 \end{aligned}$$
- with the assumptions of Markov properties, namely,

$$\begin{aligned} p(X_0 | X_{t-1}, Y_t, \beta) &= p(X_0 | X_{t-1}), & p(X_t | X_{t-1}, Y_t, \beta) &= p(X_t | X_{t-1}), T \geq 1, \\ p(X_t | X_{t-1}, Y_t, \beta) &= p(X_t | X_{t-1}), & p(Y_t | X_{t-1}, Y_t, \beta) &= p(Y_t | X_{t-1}), T \geq 1. \end{aligned}$$
- Question: find or approximate the following distributions

$$\begin{aligned} &\text{predicting distribution: } p(X_t | Y_{1:t-1}) \\ &\text{filtering distribution: } p(X_t | Y_{1:t}) \\ &\text{smoothing distribution: } p(X_t | Y_{1:T}), T \geq 1 \end{aligned}$$

Application

• Cellular Automaton
 A plane with \mathbb{Z}^2 cells. The three states of cells are described by white, red and blue. The current state of each cell only depends on the states of cells around it (including itself) at the last time point. The signal is unknown, the observations are observed by randomly changing the color of some cells on the original signal and shading some of them. The patterns of incorrect and shaded cells are given. The analysis is the estimate of the signal by using the weighted summation of particles given by (3).

Weights are computed by comparing the corresponding cells in each particle and the observation. The comparison can be made between one cell in the particle and one cell at the same position in the observation. One can also compare a cell along with its neighbors with cells at the same positions in the observation. The distance between the interested cell and its neighbors is called the range.

• A study of the effect of range
 The size of cellular automaton is 50×50 . The number of particles is 200 ($\beta = 200$). The particle filter is used for each time step ($\beta \geq 0$). The observe error is set to be 20% or 50%. In the following picture, the vertical axis represents the analysis error which is the percentage of incorrect states between the analysis and the signal. The color of each curve represents the percentage of shaded cells.
 The effect of range ($\beta = 200$, $\beta \text{ cycle} = 1$, observe error = 20%)

The effect of range ($\beta = 200$, $\beta \text{ cycle} = 1$, observe error = 50%)

For observe error = 50 and range 0, the analysis error can be observed to lower than 20% over 90% of cells are shaded. For range 5, the decrease of analysis error is faster. One finds the most on each range by using the information of analysis error.

A special model with closed solution (Kalman Filter)

• Linear Gaussian Model

$$X_t = A_{t-1} X_{t-1} + q_{t-1}, \quad Y_t = H_t X_t + r_t$$
 where $q_{t-1} \in \mathcal{N}(0, Q_{t-1}), \quad r_t \in \mathcal{N}(0, R_t)$
 for all $t = 1, A_{t-1}$ and H_t are transition matrix and measurement matrix respectively. The initial distribution of X_0 is also Gaussian with mean m_0 and covariance matrix P_0 .

• Kalman Filter and RTS Smoother

$$\begin{aligned} p(X_t | Y_{1:t}) &= \mathcal{N}(m_t^f, P_t^f) & p(X_t | Y_{1:T}) &= \mathcal{N}(m_t^b, P_t^b) \\ p(X_{t-1} | Y_{1:t}) &= \mathcal{N}(m_{t-1}^f, P_{t-1}^f) & p(X_{t-1} | Y_{1:T}) &= \mathcal{N}(m_{t-1}^b, P_{t-1}^b) \end{aligned}$$
 where

$$\begin{aligned} m_t^f &= A_{t-1} m_{t-1}^f, & P_t^f &= A_{t-1} P_{t-1}^f A_{t-1}^T + Q_{t-1}, \\ m_t^b &= m_t^f + P_t^f H_t^T (Y_t - H_t m_t^f)^{-1} (Y_t - H_t m_{t-1}^b), & P_t^b &= P_t^f - K_t (Y_t - H_t m_{t-1}^b)^{-1} K_t^T, \\ m_t^b &= m_t^f + Q_{t-1} P_{t-1}^b (Y_t - H_t m_{t-1}^b)^{-1} (Y_t - H_t m_{t-1}^b), & P_t^b &= P_t^f + Q_{t-1} P_{t-1}^b (Y_t - H_t m_{t-1}^b)^{-1} (Y_t - H_t m_{t-1}^b)^T. \end{aligned}$$

Particle Filters (PI)

For some special models, one may find closed solutions, but for more general models, closed solutions may not exist. For some function f , the problem of computing

$$E_{\text{opt}}[f(X)] = \int f(x) p(x) dx$$
 can be approximated by the perfect Monte Carlo approximation. The convergence is ensured by the Law of Large Numbers. If a sample $X_t^i, i = 1, 2, \dots, N$, called particles, can be drawn from $p(X_t | Y_{1:t})$ then

$$E_{\text{opt}}[f(X)] \approx \frac{1}{N} \sum_{i=1}^N f(X_t^i)$$

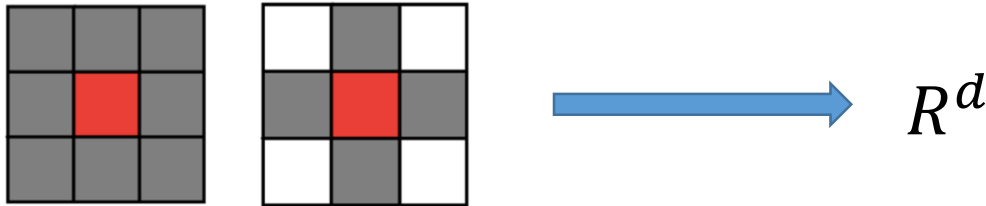
• Importance sampling
 Assume that there is an importance distribution $\pi(X_t | Y_{1:t})$ with support $\text{supp}(\pi(X_t | Y_{1:t})) \supset \text{supp}(p(X_t | Y_{1:t}))$ then

$$\begin{aligned} E_{\text{opt}}[f(X)] &= \int f(x) \frac{p(x | Y_{1:t})}{\pi(x | Y_{1:t})} \pi(x | Y_{1:t}) dx \\ &= \int f(x) w(x) \pi(x | Y_{1:t}) dx \end{aligned}$$
 where X_t^i is a sample from importance distribution π .

References

[1] J. Durbin, N. Friel, N. Ombao, Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, 2005.
 [2] C. Kirchner, Non-Gaussian State-Space Modeling of Volatility, Time Series Journal of the American Statistical Association, Vol. 121, No. 480 (Dec), 2012, pp. 1022-1041.
 [3] C. Kirchner, Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-Space Modeling, Journal of Computational and Graphical Statistics, Vol. 5, No. 1 (Mar), 1993, pp. 1-26.

The modeling and application of cellular automata theory



Moore neighborhood and Von Neumann neighborhood

Yimin Gong 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **242** 012106

Some applications:

- Traffic flow based on cellular automaton
- Simulation of complex land use system based on Neural Network-based cellular automaton
- Temperature field simulation

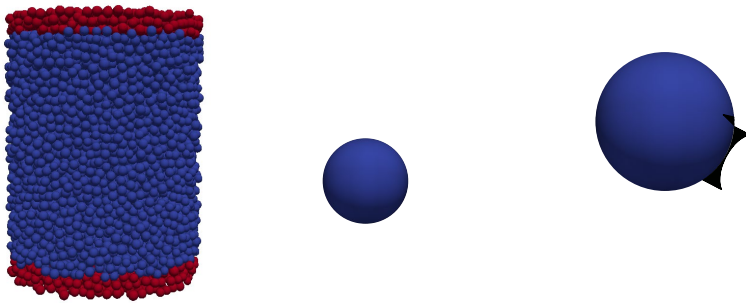
Questions :

- How should us simplify a three-dimensional system into cellular automata?
- Which kind of training method we can use to find potential parameters?

Failure processes of cemented grains depending on packing fractions

Yuta Yamaguchi
University of Tokyo

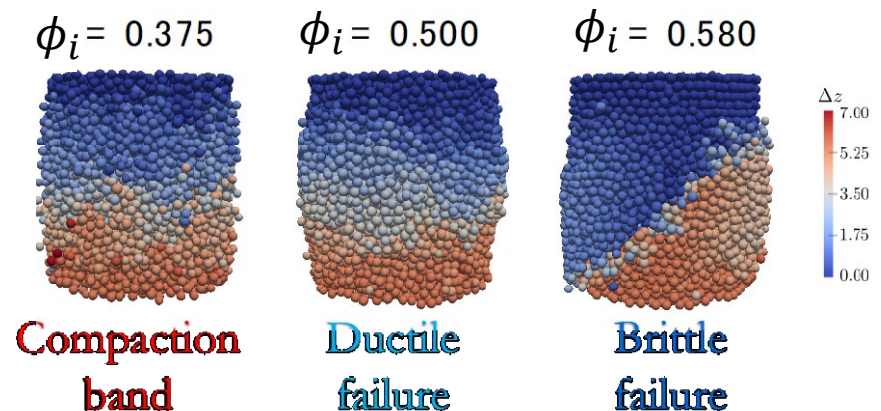
MODELLING



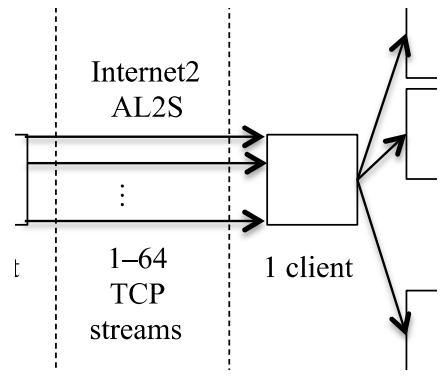
- Uniaxial compress simulations of cohesive porous media by means of DEM simulations.
- The material properties are firmly based on inputs from experiments.

PARAMETER STUDIES

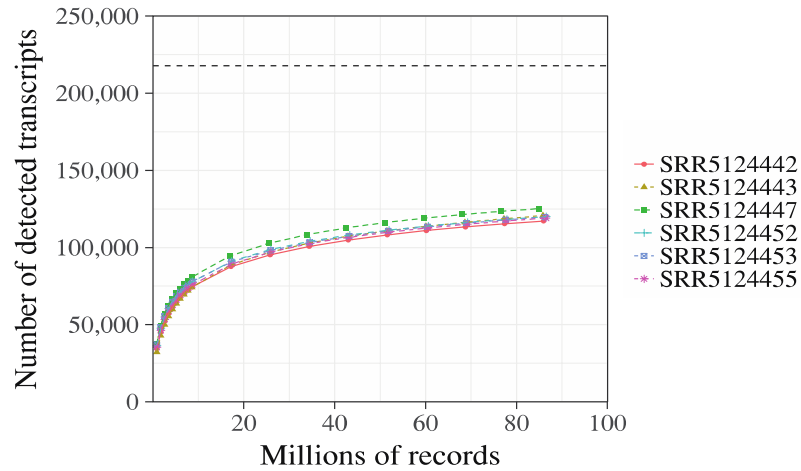
- Our parameter studies show three different modes of failure processes; brittle failure, ductile failure, and compaction bands.



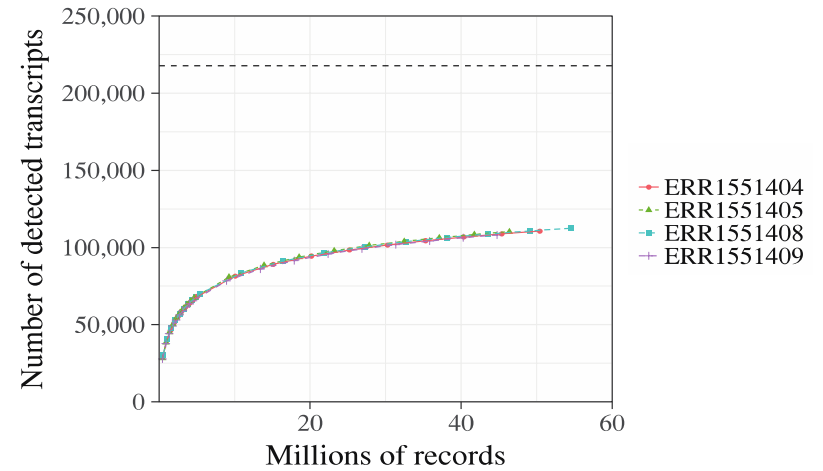
NICK Mills, Ph.D. Candidate
Clemson University
*Department of Electrical and Computer
Engineering*



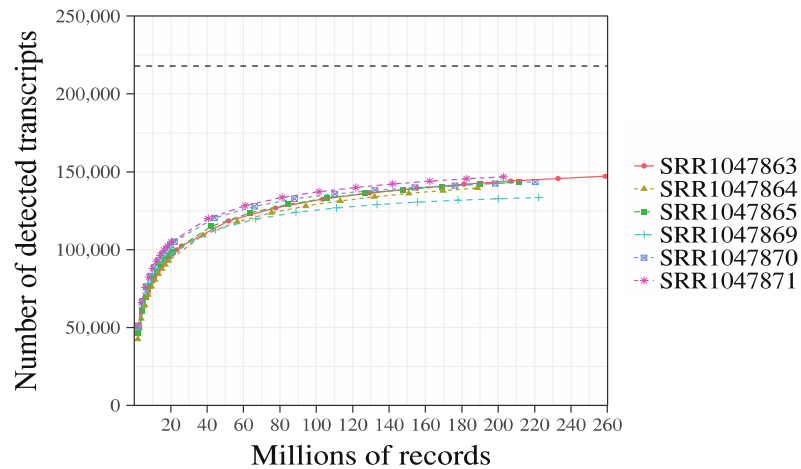
NICK Mills, Ph.D. Candidate
 Clemson University
 Department of Electrical and Computer
 Engineering



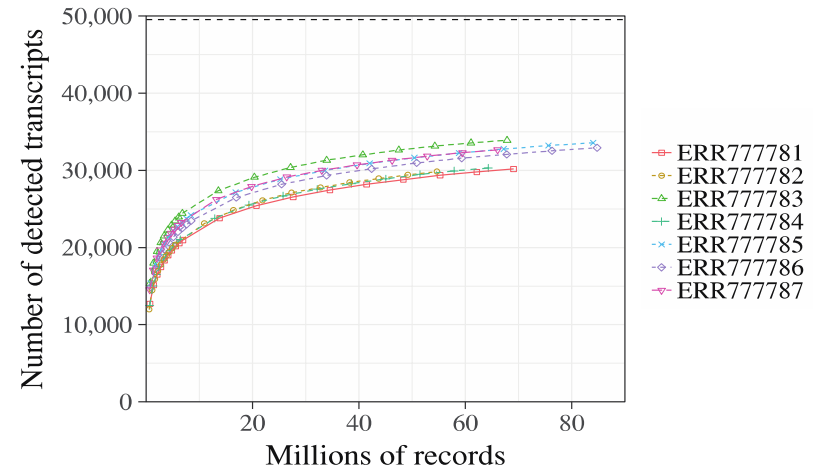
(a) bladder



(b) hypoxia

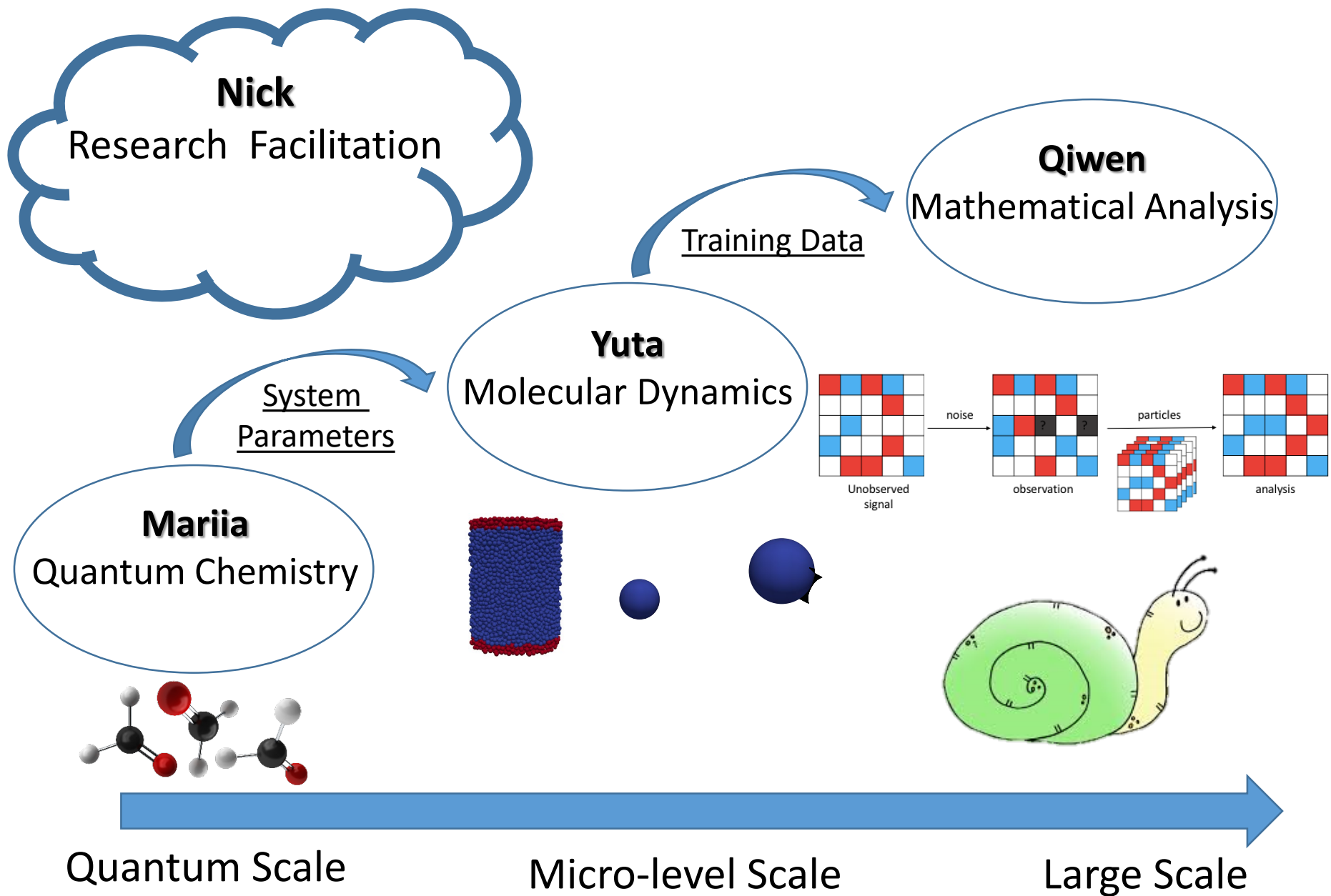


(c) nisc2



(d) oncopig

Optimization of Multiscale Collaboration



Objective

Quantum and **Micro** Scale – how to escape from the problem of system size limitation to more realistic reproduction of real systems properties?

Large scale –

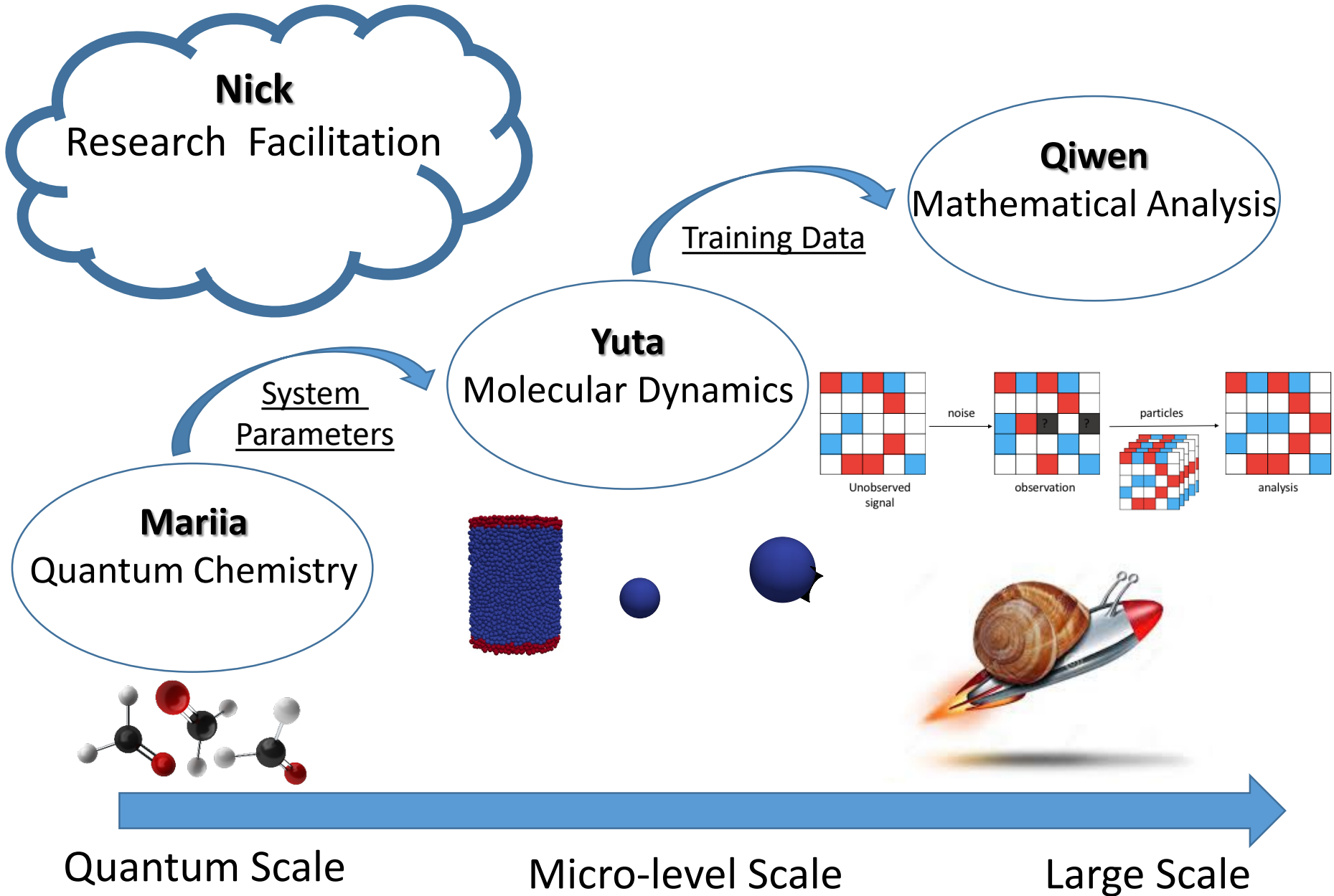
- how to select only useful information from the huge output data and transform it efficiently?
- how can we train the data properly?
- how can we select observations to reach an acceptable convergence level?

Solution

Quantum and **Micro** Scale – identification and parallelization of frequently executed regions of code (optimization)

Large scale – efficient transfer of observation and output data by identifying the useful information

Optimization of Multiscale Collaboration



Thank all of you!
(special thanks for James and Daichi)