SAND2011-7904C

Propagating Uncertainty from Simulation Parameters and Sampling Noise through Coupled Atomistic-to-Continuum Systems

Maher Salloum, Reese Jones, Khachik Sargsyan, *Bert Debusschere,* Helgi Adalsteinsson, Habib Najm

Sandia National Laboratories, Livermore, CA, USA
bjdebus@sandia.gov

Supported by the US Department of Energy, Office of Advanced Scientific Computing Research

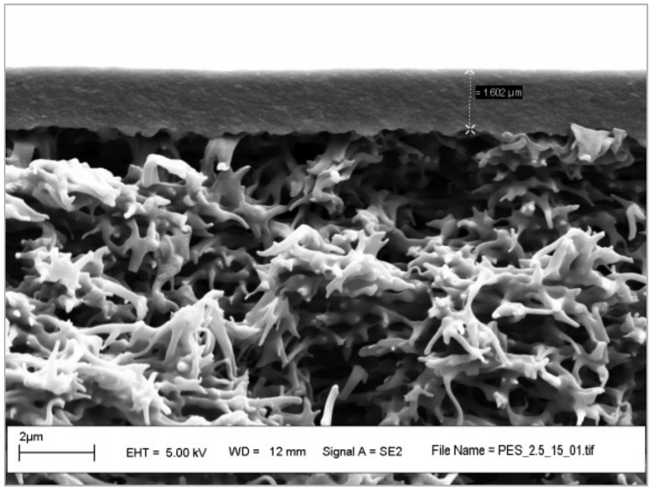
Quantifying prediction fidelity in multiscale multiphysics simulations: the overall team

- Sandia National Laboratories
  - Helgi Adalsteinsson
  - Bert Debusschere
  - Maher Salloum
  - Reese Jones
  - Khachik Sargsyan
- Johns Hopkins University
  - Omar Knio
  - Francesco Rizzi
- Texas Tech University
  - Kevin Long
  - Kaleb McKale
  - Jed Gohlke
  - Simon Rush
- Massachusetts Institute of Technology
  - Youssef Marzouk
  - Jinglai Li

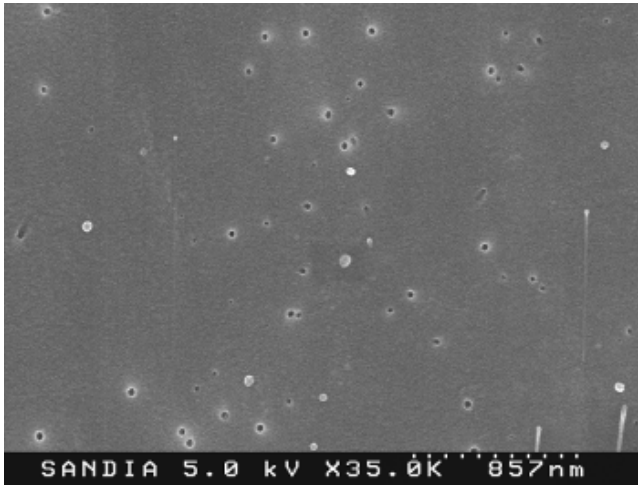
Overview

- Introduction and motivation
- Couette flow test case
- Multiscale coupling approach
  - Atomistic sampling noise
  - Parametric uncertainty
- Ongoing work
- Conclusions

Multiscale methods needed to account for phenomena coupled over wide ranges of time and length scales



Nanoporous polyaniline membrane

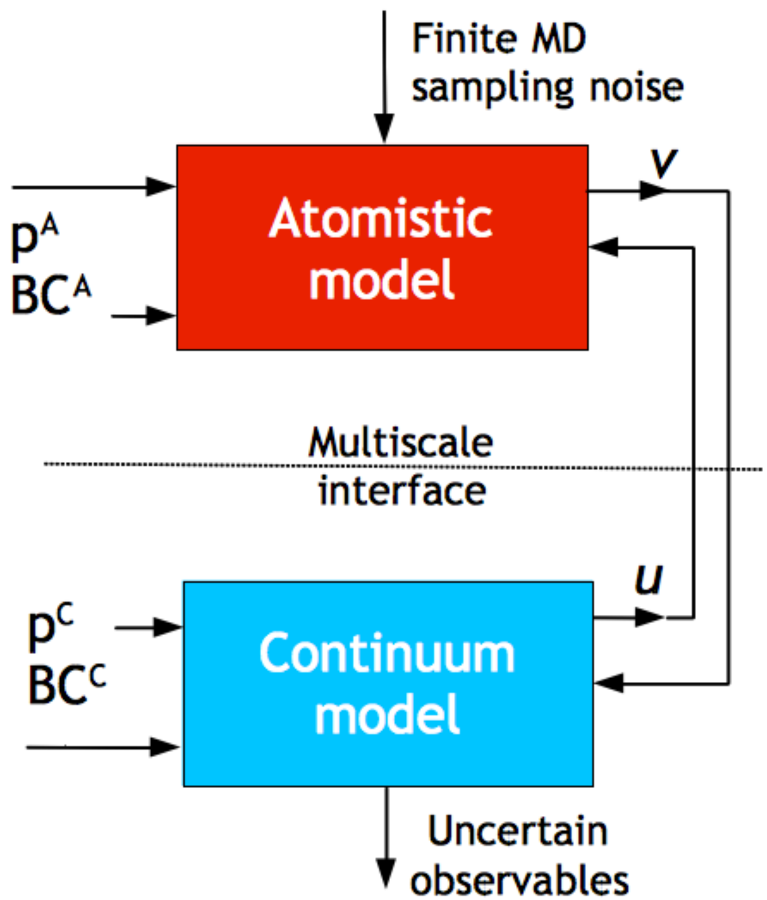


Polycarbonate track etched membrane

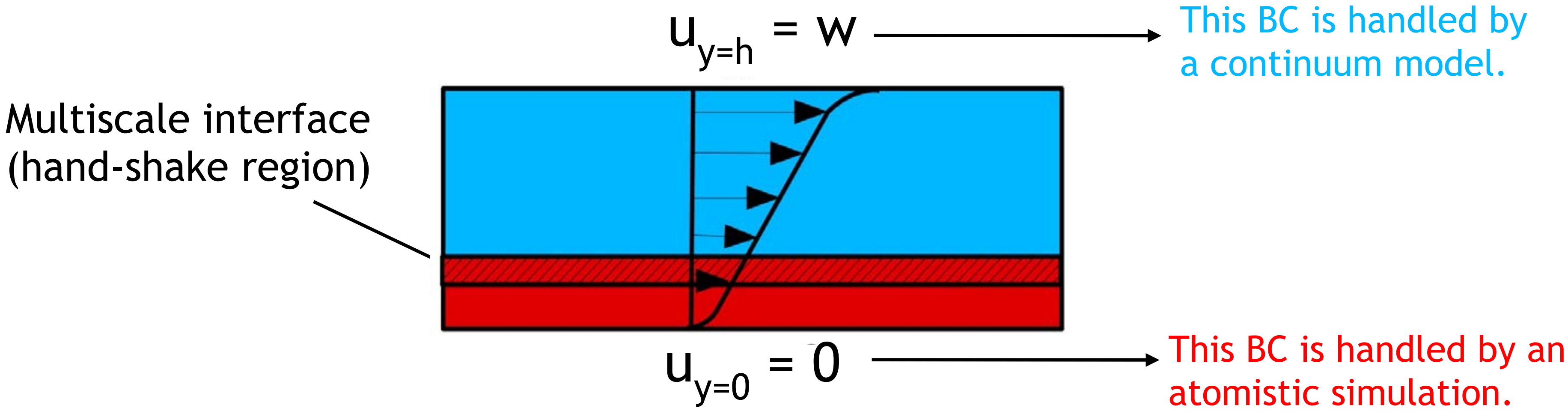
- Many key applications have macroscale behavior driven by microscale phenomena
  - E.g. ionic flux through nanopores in water desalination
- Multiscale simulations resolve key physics on different scales
  - Uncertainties on all scales
  - Uncertainties in coupling

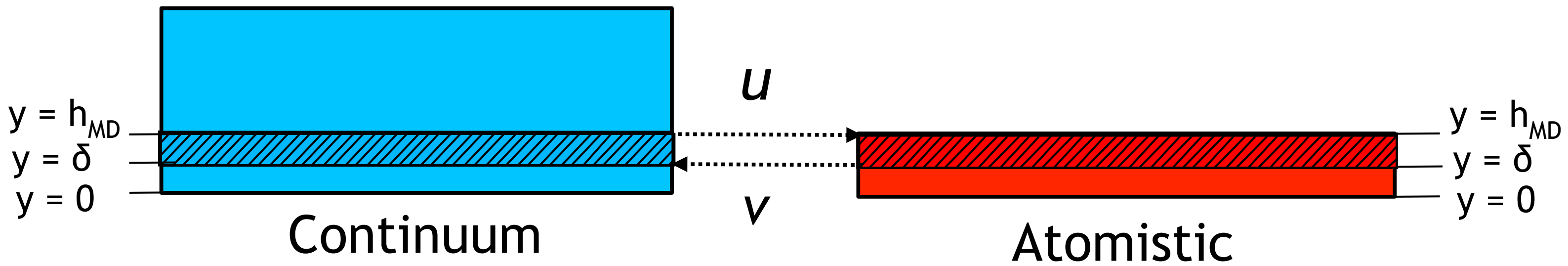
Predictive multiscale simulation requires quantification of the many sources of uncertainty

- This talk focuses on
  - Assessment of sampling noise on the atomistic level
  - Propagation of parametric uncertainty and sampling noise across scales
- Related work presented at this meeting
  - Forward propagation and inference of parametric uncertainty on atomistic level: Knio *et al.*, poster 1
  - Propagation of uncertainty through the continuum level: Long *et al.* poster 48



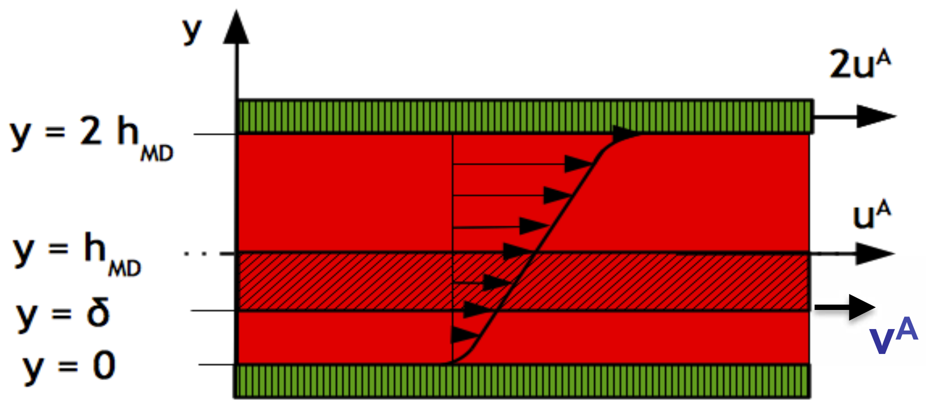
Canonical plane Couette flow is used as model problem for algorithm development

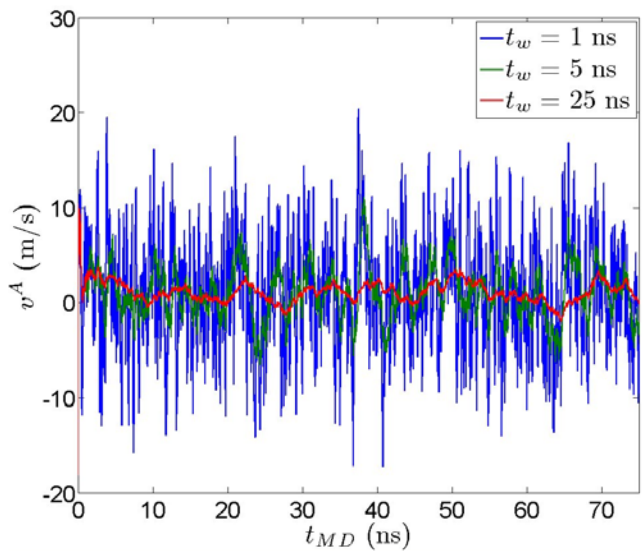




- Complex enough to illustrate key challenges in coupling atomistic to continuum with uncertainty
- Simple enough to make computations tractable

Finite sampling in extracting macroscale observables from atomistic simulations results in uncertainty

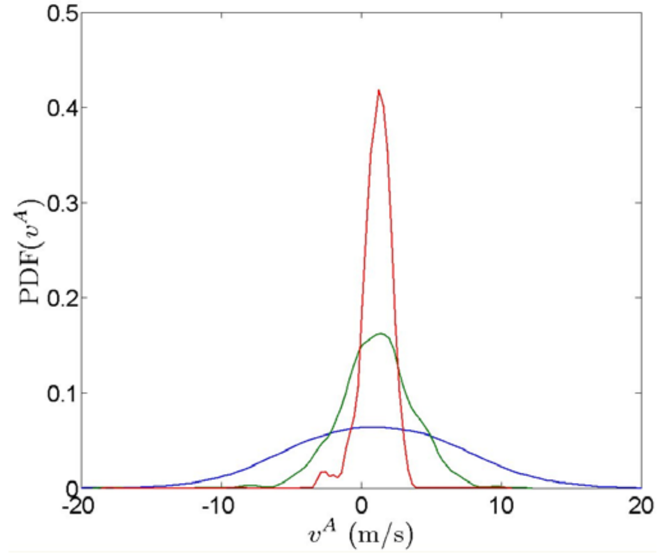




- Molecular dynamics deterministic
  - Lennard Jones
  - LAMMPS

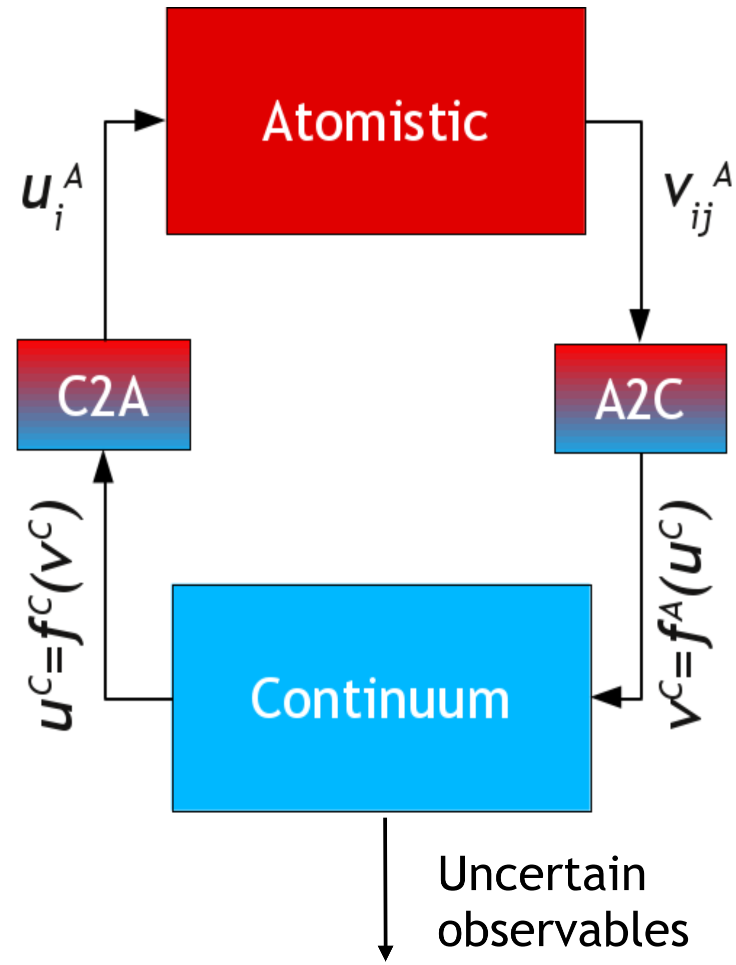
$$\phi_{ij} = 4\phi_0 \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$

- Sampling noise due to
  - Finite size of domain
  - Finite size of averaging time window t_w
- For non-trivial MD simulations, we can not sample our way out of this

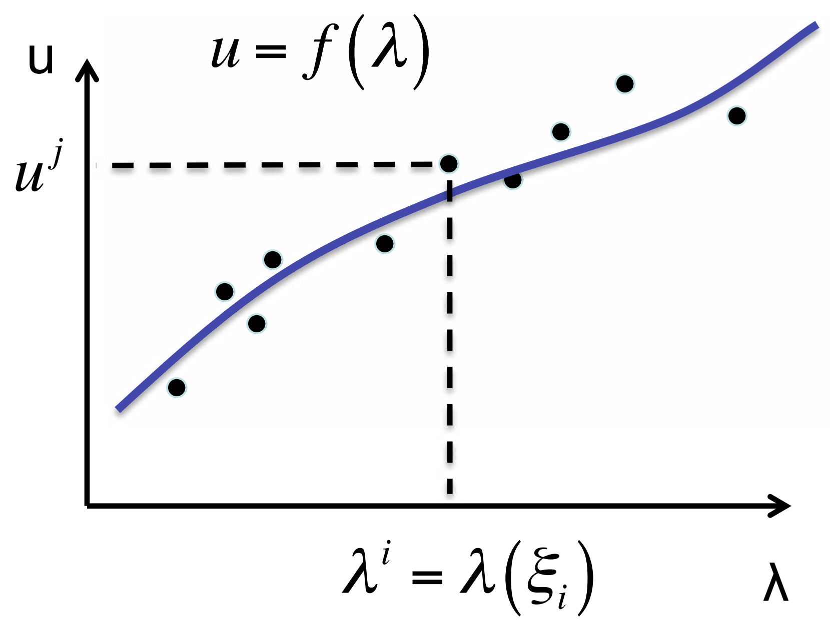


Building blocks in atomistic to continuum coupling

- Characterizing sampling noise in atomistic output
- Uncertain inputs and sampling noise in atomistic simulation
- Continuum simulation with uncertain inputs
- Coupled atomistic and continuum simulation
  - Sampling noise only
  - Sampling noise and uncertain inputs



Bayesian inference of response surfaces



$$\lambda = \sum_{k=0}^{P} \lambda_k \psi_k(\xi) \qquad u = \sum_{k=0}^{P} u_k \psi_k(\xi)$$

$$\mathbf{d} = \left\{u^j\right\}_{j=1}^{N} \qquad \mathbf{u} = \left\{u_k\right\}_{k=0}^{P}$$

$$u^j = f\left(\lambda^i\right) + s\,\eta_{ij} \qquad \eta_{ij} \sim \mathcal{N}(0,1)$$

$$P\left(\mathbf{u}, s^2 \middle| \mathbf{d}\right) \propto P\left(\mathbf{d} \middle| \mathbf{u}, s^2\right) P\left(\mathbf{u}, s^2\right)$$

$$\mathbf{u} \sim \mathcal{St}\left(\bar{\mathbf{u}}, S, \gamma\right)$$

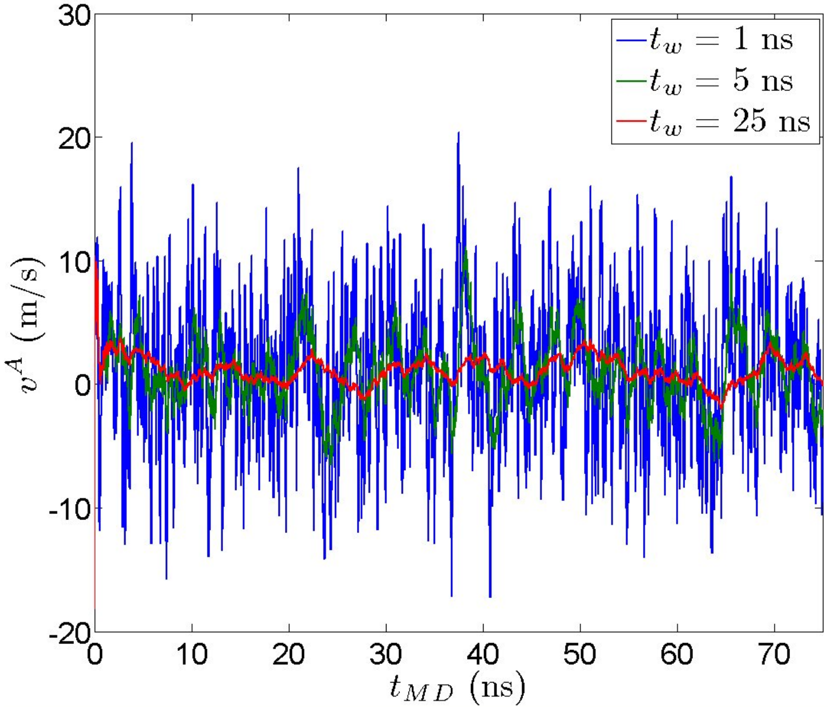
$$u = \psi(\xi)^T \bar{\mathbf{u}} + \zeta \sqrt{\psi(\xi)^T S \psi(\xi)}$$

$$\zeta \sim \mathcal{St}(0,1,\gamma)$$

Analytical expression for u as student-t process if:

- Gaussian noise model
- Infinite (improper) uniform prior on u_k
- Jeffreys prior for s^2
- Marginalize over s^2

Quantification of sampling noise in atomistic simulations



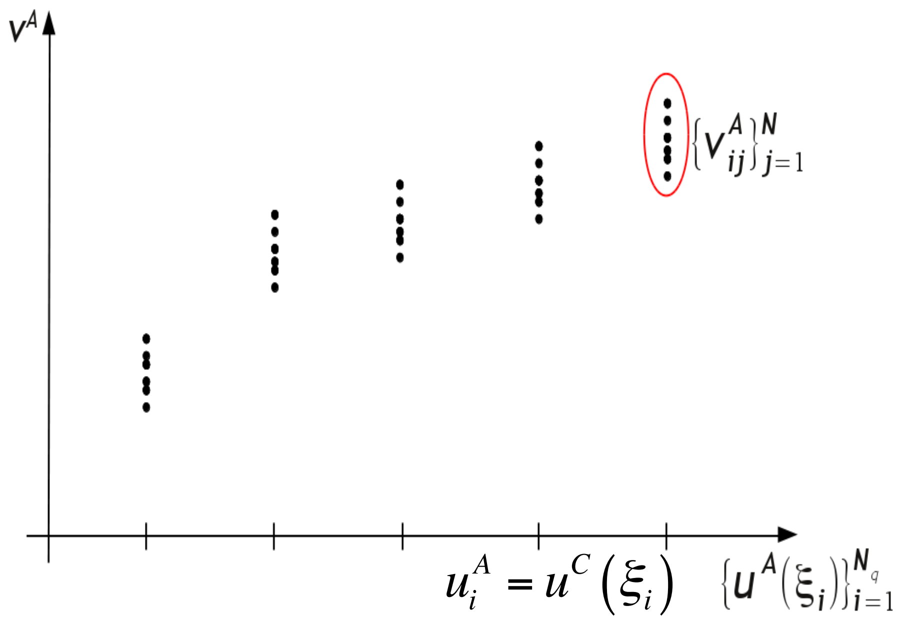
$$\mathbf{d} = \left\{ v_j^A \right\}_{j=1}^{N}$$

$$v_j^A = v^C + s\,\eta_j$$

$$v^C = \bar{v}^C + \sqrt{S}\,\zeta \qquad \zeta \sim St(0,1,\gamma)$$

- Infer v^C from N short-term averaged MD velocity samples v_j^A
  - Gaussian model for data noise due to Central Limit Theorem (CLT)
  - Analytical solution gives v^C as student-t random variable
- Averaging over longer time window or more data reduces sampling noise

Propagating parametric uncertainty and sampling noise through atomistic simulations



$$v_{ij}^A = v^C(\xi) + s\,\eta_{ij}$$

$$u^C = \sum_{k=0}^{P} u_k^C \psi_k(\xi)$$

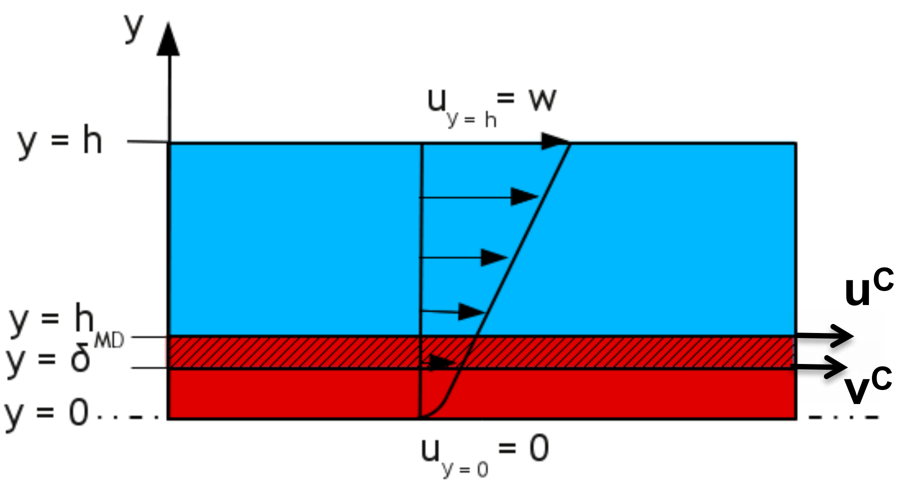
$$v^C = \sum_{k=0}^{P} v_k^C \psi_k(\xi)$$

$$v^C = \psi(\xi)^T \bar{\mathbf{v}}^C + \zeta \sqrt{\psi(\xi)^T S \psi(\xi)}$$

$$\zeta \sim St(0,1,\gamma)$$

- Infer v^C from N short-term averaged MD velocity samples v_{ij}^A
  - Sampled over a range of input velocities u_i^A
  - Gaussian model for data noise due to Central Limit Theorem (CLT)
  - Analytical solution gives v^C as student-t process over input velocity uncertainty

Propagating uncertainty through continuum



$$u^C = w - \frac{h - h_{MD}}{h - \delta}\left(w - v^C\right)$$

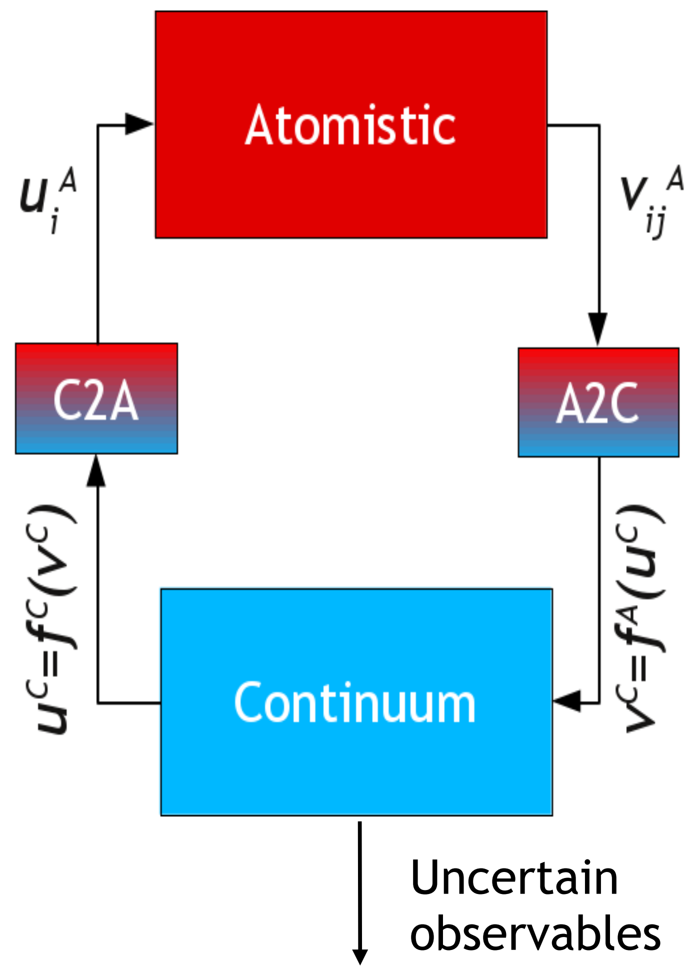
$$u^C = \sum_{k=0}^{P} u_k^C \psi_k(\xi) \quad v^C = \sum_{k=0}^{P} v_k^C \psi_k(\xi)$$

$$u_k^C = \begin{cases} w + \frac{(w - v_k^C)(h_{MD} - h)}{h - \delta}, & \text{for } k = 0 \\ \frac{-v_k^C (h_{MD} - h)}{h - \delta}, & \text{for } 0 < k \leq P \end{cases}$$

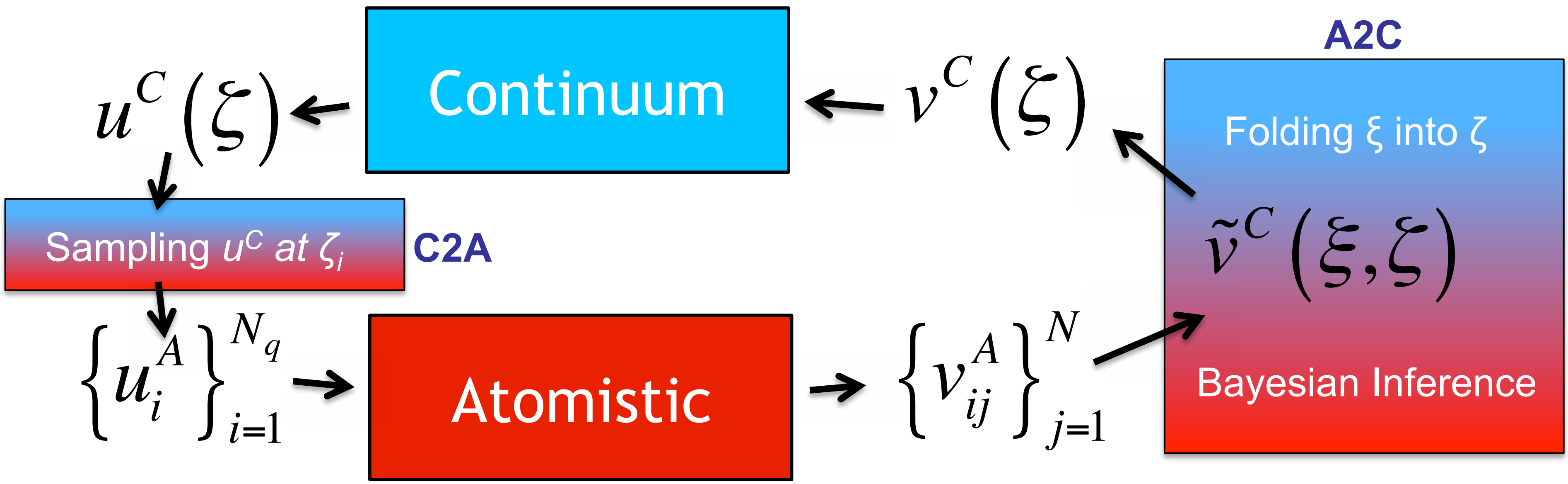
- Steady state, linear velocity profile
  - Allows analytical propagation of uncertainties

Coupled atomistic to continuum simulation accounting for sampling noise on atomistic level

- Sampling noise is only source of uncertainty
  - All external inputs deterministic
- Sampling noise in atomistic outputs propagated through coupling
  - Uncertain continuum simulation
  - Uncertain atomistic inputs
- Different approaches
  - Fixed point iteration on atomistic level
  - Intersecting sampled continuum response surfaces
  - Fixed point iteration on intersecting uncertain continuum response surfaces



Fixed point iteration on atomistic level



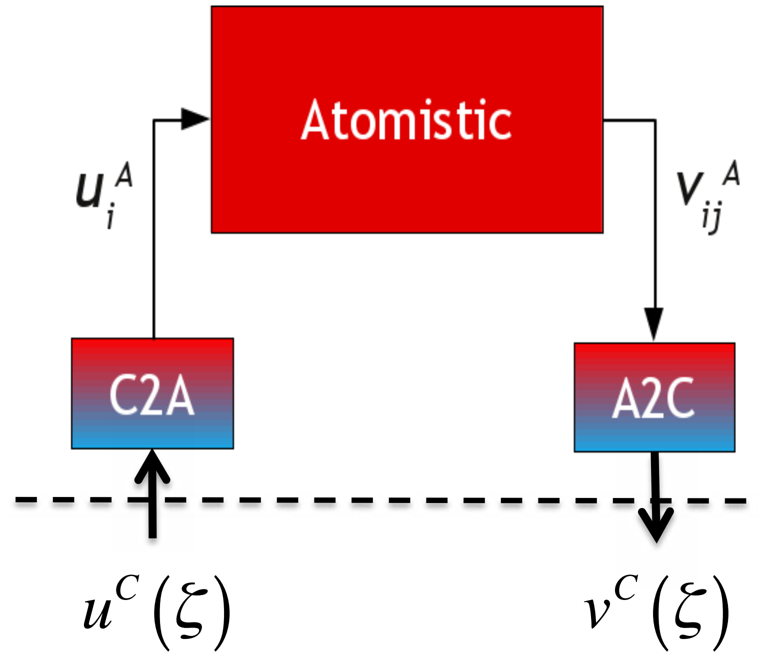
- Additional sampling noise introduced with every atomistic simulation
  - Merged with uncertainty present in atomistic input velocity
- Requires many atomistic simulations at nearby inputs
  - Expensive unless surrogate model used
- Salloum *et al.*, SIAM MMS, submitted 2011

Response surfaces make atomistic simulation available to the macroscale

$$v^C = \psi(\xi)^T \bar{\mathbf{v}}^C + \zeta \sqrt{\psi(\xi)^T S \psi(\xi)}$$

$$\zeta \sim \mathcal{St}(0,1,\gamma)$$

$$v^C = f^A(u^C, \zeta)$$

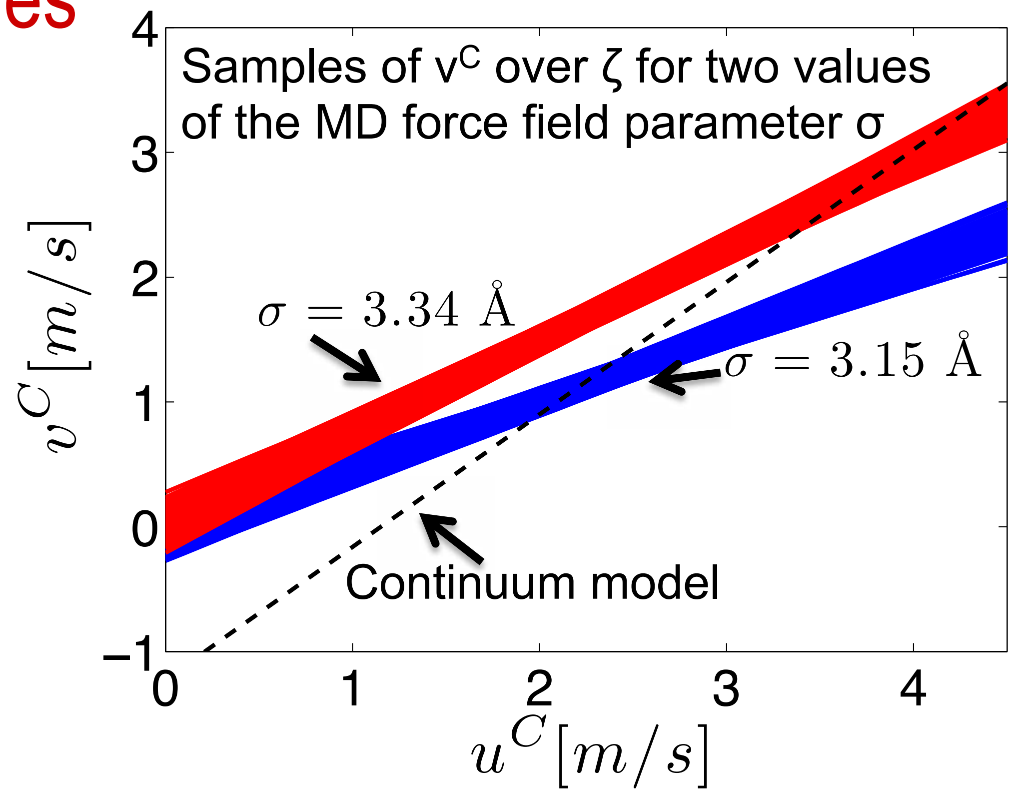


- Same formulation as propagation of parametric uncertainty
- Can be evaluated instead of running MD simulations
- Uncertainty due to sampling noise captured by ζ
  - Can be reduced by adding more MD data

Intersecting sampled response surfaces readily provides coupling variables

$$v^C = f^A\left(u^C, \zeta\right)$$

$$u^C = f^C\left(v^C\right)$$

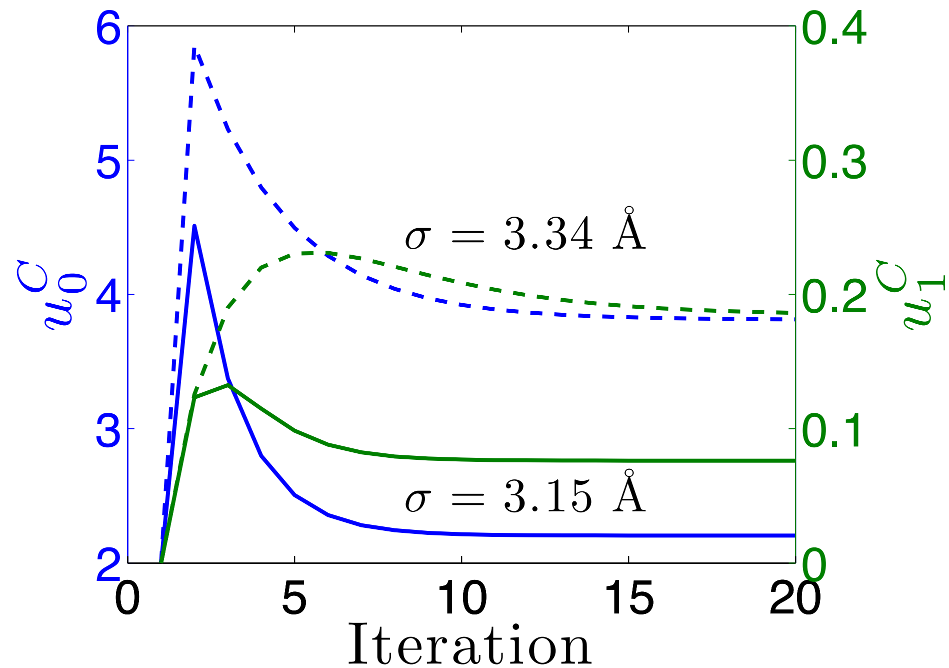


- Sample atomistic response surface over intrinsic variability ζ
  - Deterministic intersection with continuum response surface
- Project resulting samples of v^C onto PC basis
  - Mapping to PC random variables using inverse CDF of sampled v^C

Fixed point iteration on uncertain response surfaces

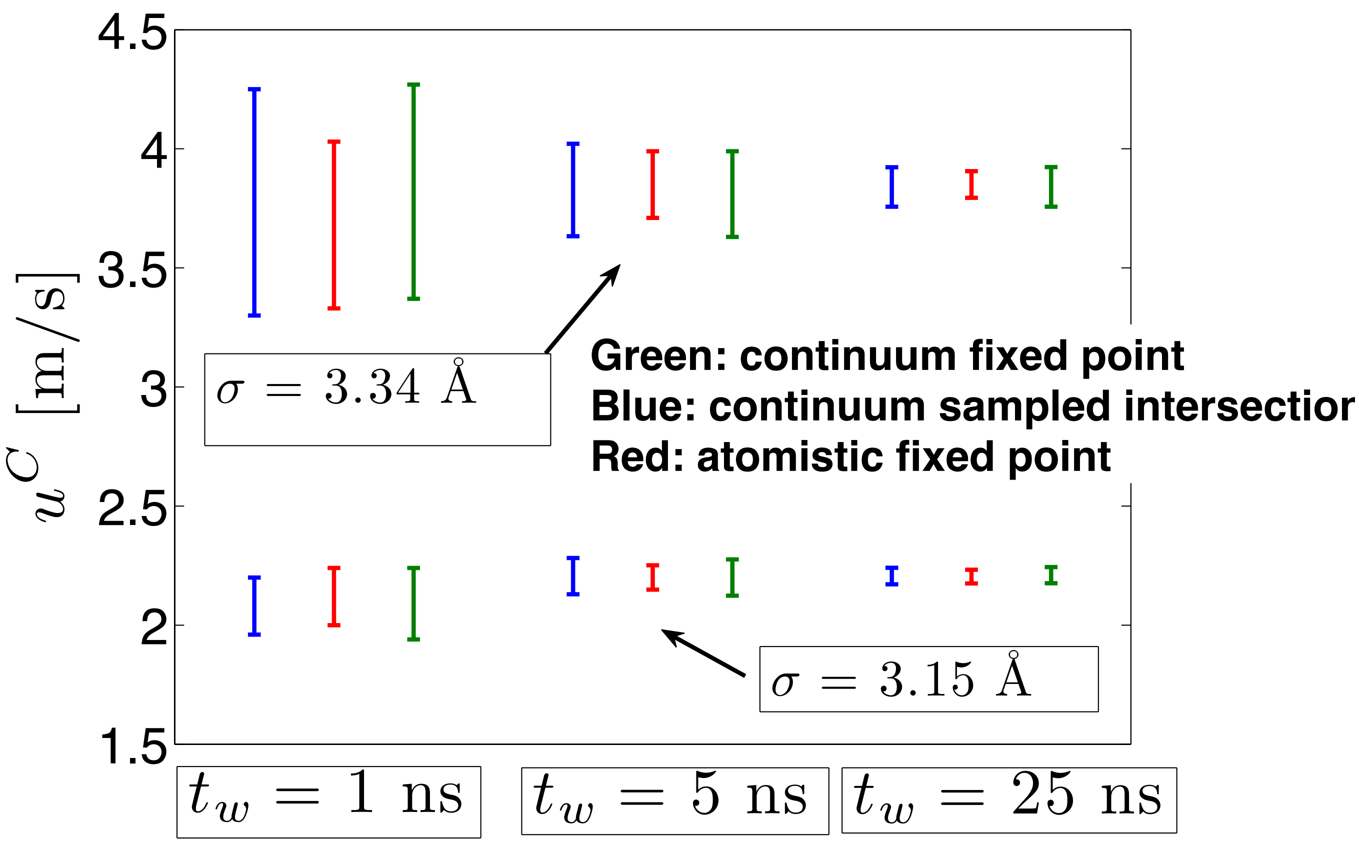
$$u^C = \sum_{k=0}^{P} u_k^C \psi_k(\zeta) \qquad v^C = \sum_{k=0}^{P} v_k^C \psi_k(\zeta)$$

$$v^C = f^A(u^C, \zeta) \qquad u^C = f^C(v^C)$$



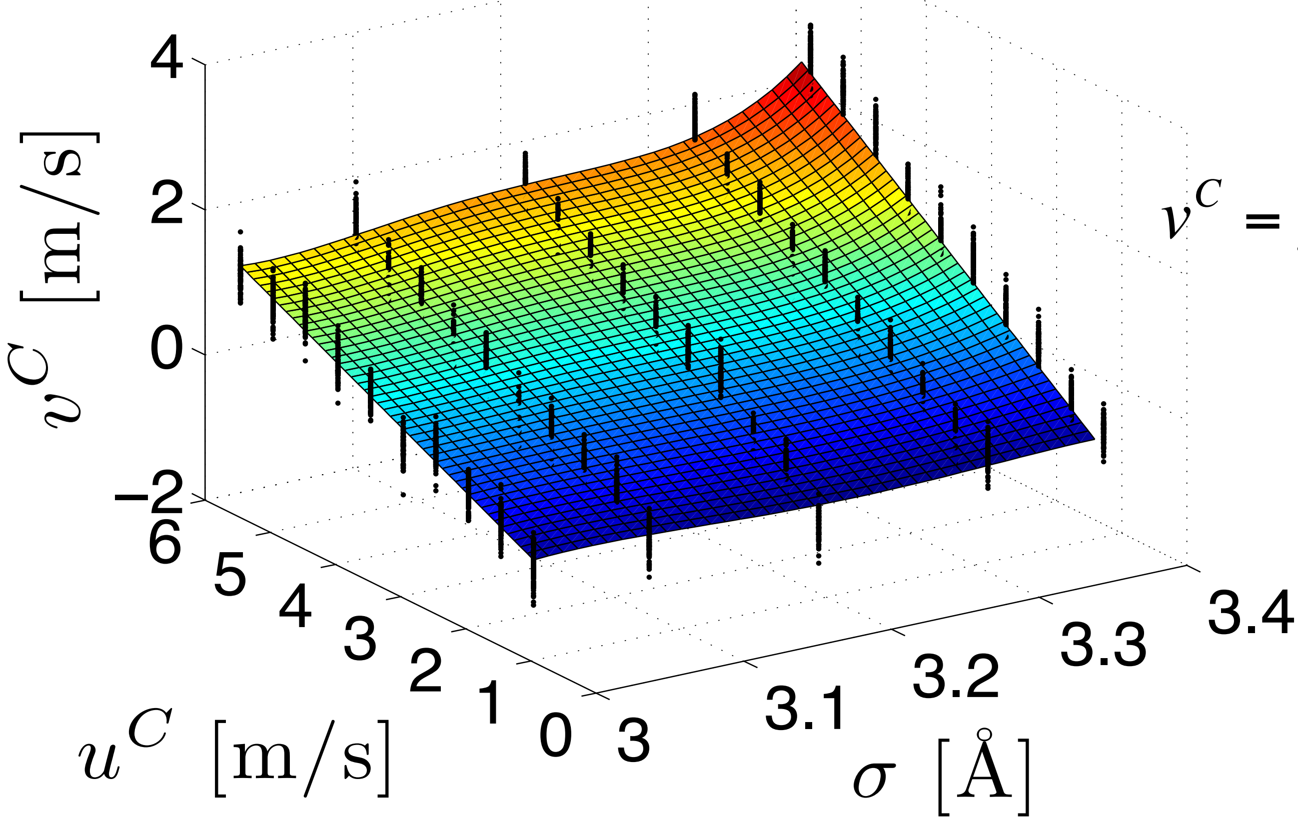
- Assume PC expansion for coupling variables to represent sampling noise
- Substitute PCEs into atomistic and continuum response surfaces
- Starting from an initial guess, iterate till convergence

The three coupling approaches are in agreement



- Larger σ gives more uncertainty
- Noise can be reduced through longer time averaging

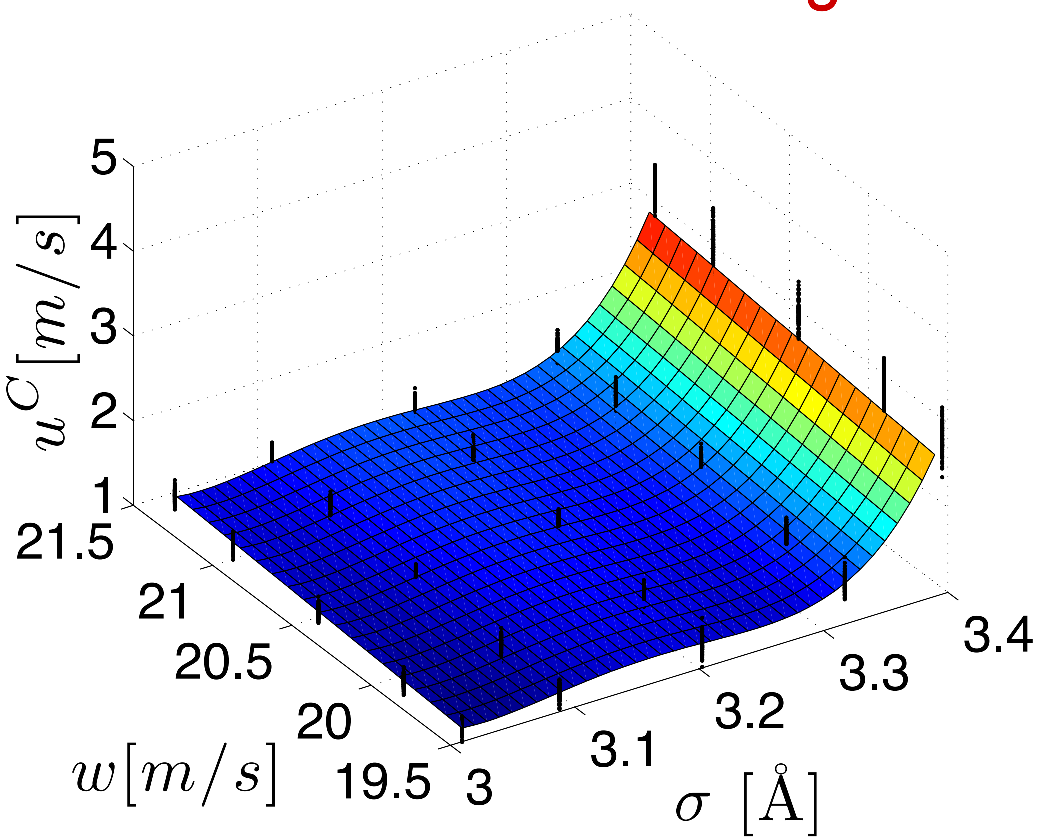
Coupled atomistic to continuum simulation with sampling noise and parametric uncertainty



$$v^C = f^A\left(u^C, \sigma, \zeta\right)$$

- Response surface f^A as function of input parameters
  - Generalization of case with sampling noise only
  - Inferred from MD data at sampled parameter values
  - Sampling noise represented as student-t process

Response surface intersection through sampling



- Intersect response surfaces at specific parameter values
  - Using previously discussed approaches
- Infer polynomial surface through those points

Intersection through fixed point iteration on uncertain continuum response surfaces

$$w = \sum_{k=0}^{P} w_k \psi_k(\xi)$$

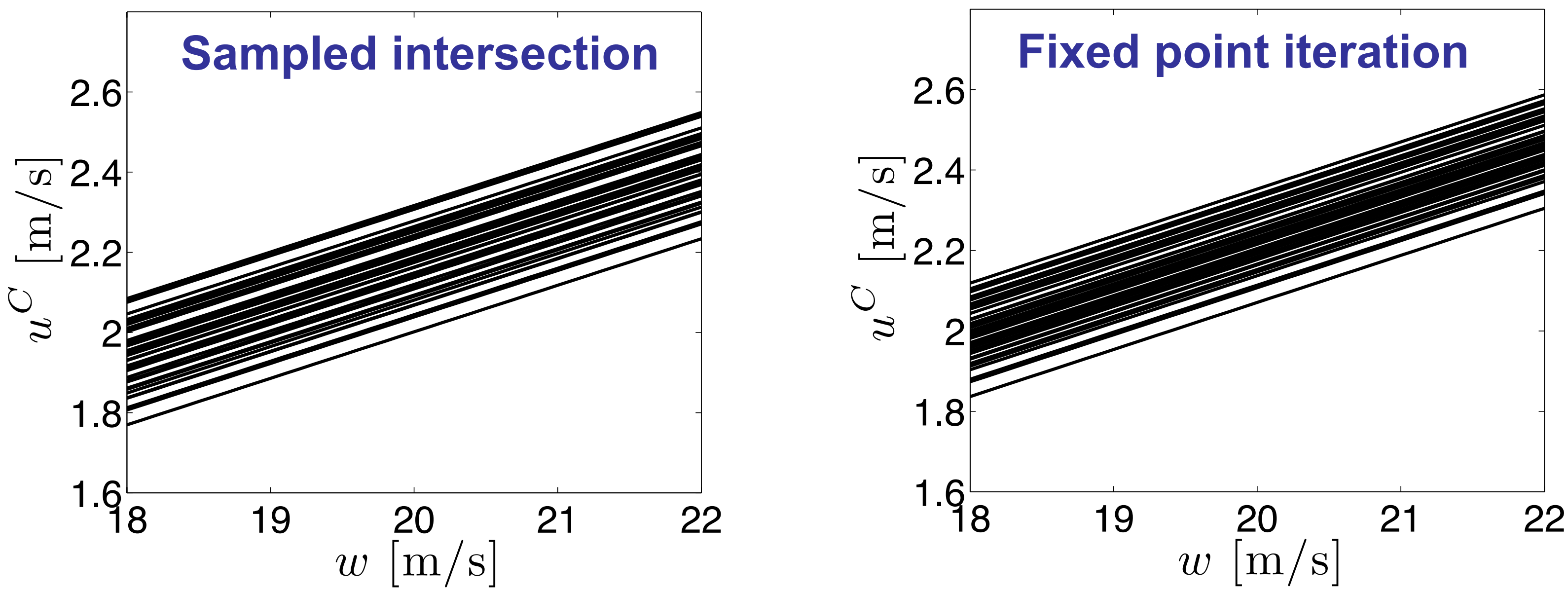
$$\sigma = \sum_{k=0}^{P} \sigma_k \psi_k(\xi)$$

$$u^C = \sum_{k=0}^{P} u_k^C \psi_k(\xi,\zeta) \qquad v^C = \sum_{k=0}^{P} v_k^C \psi_k(\xi,\zeta)$$

$$v^C = f^A\left(u^C,\sigma,\zeta\right) \qquad u^C = f^C\left(v^C,w\right)$$

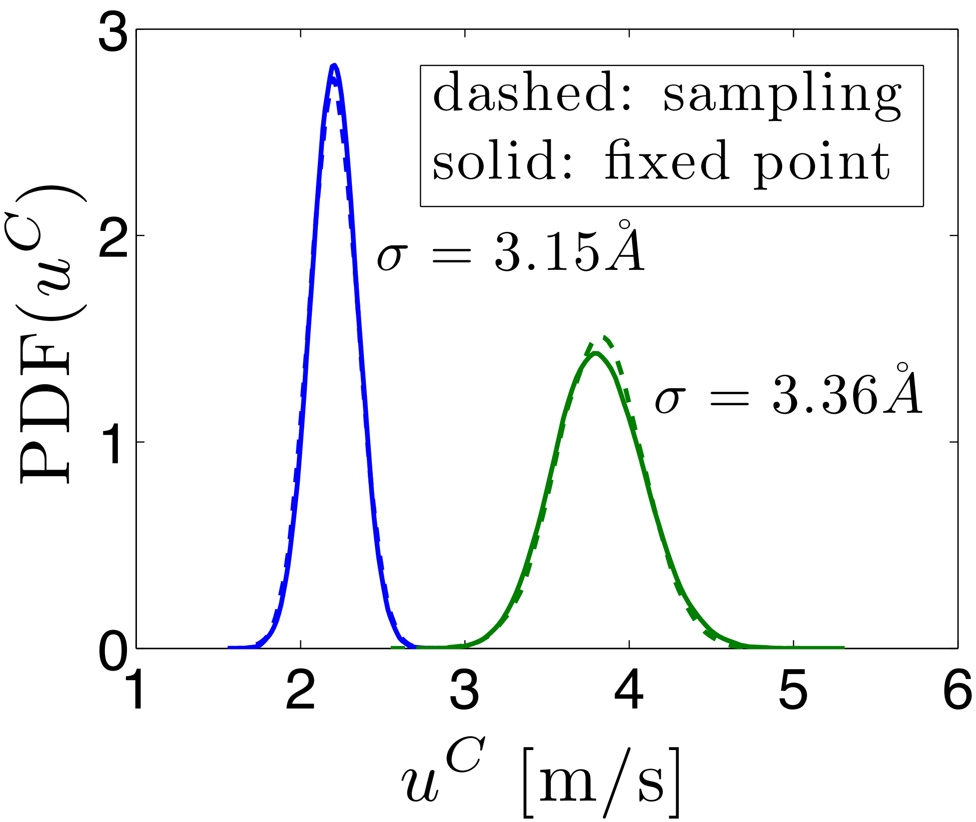
- Assume known uncertainties in w and σ
- Substitute PC expansions into response surfaces
- Iterate on expansions for u^C and v^C

The sampled intersection and fixed point iteration approaches agree well



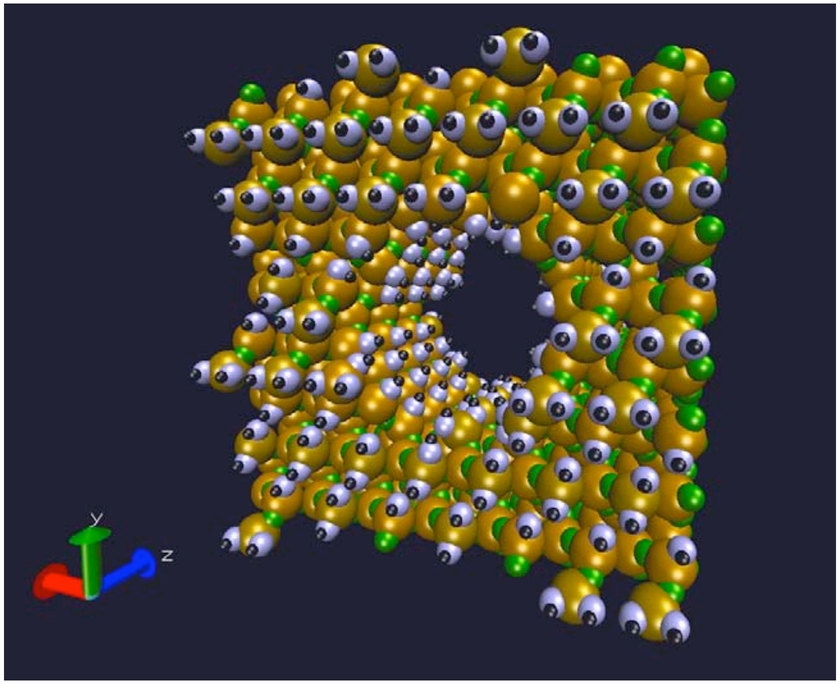
- Implemented for linear dependence on w at fixed σ
  - Good agreement with sampled intersection approach
- General non-linear case being implemented

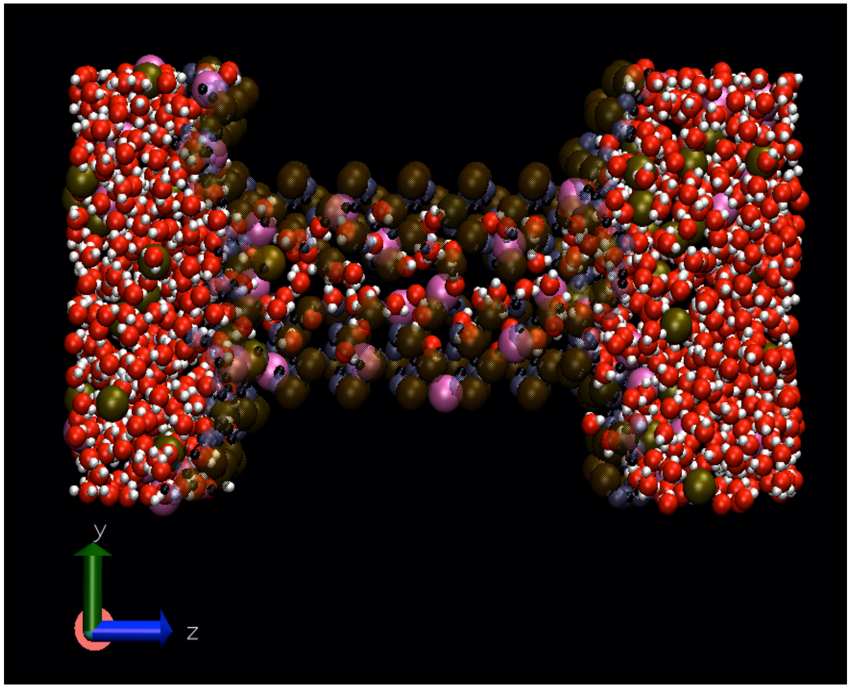
Forward propagation of uncertainty in driving velocity



- Assume uncertain driving velocity w
  - Gaussian with mean 20 m/s and standard deviation 1 m/s
- Uncertainty from w and MD sampling noise propagated into coupling variables

Application to more challenging multiscale problems





- Ionic fluxes (NaCl) through Silica nanopores
- MD concentration boundary conditions set by continuum
- Continuum flux boundary conditions set by MD
- Work in progress

Conclusions

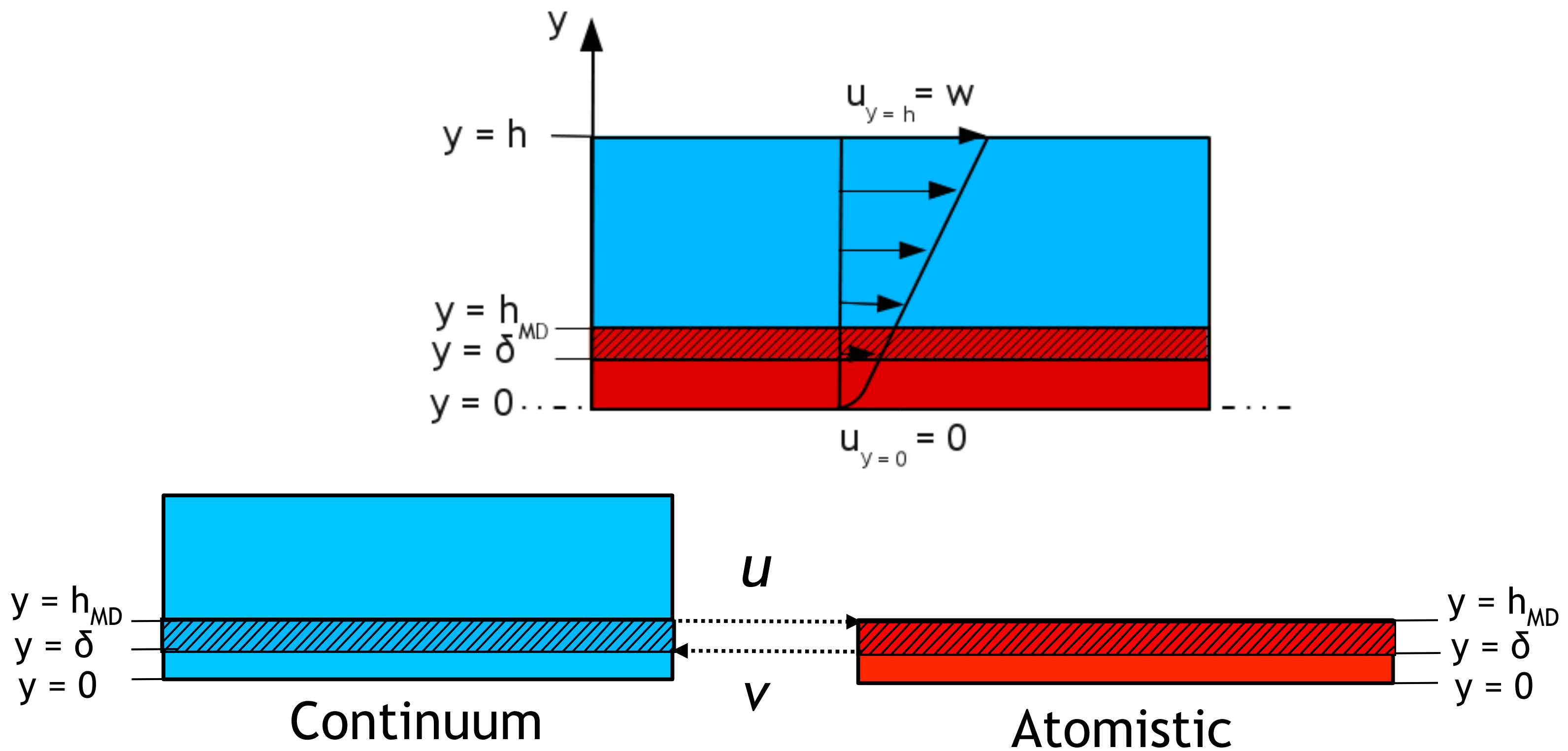
- Bayesian methods are used to quantify sampling noise in macroscale observables extracted from atomistic simulations
- Stochastic multiscale coupling approach accounts for sampling noise and parametric uncertainty
- Response surfaces for atomistic simulations allow coupling on the macroscale level
- Simple model problem here often allows for analytical solutions, but formulation is generally applicable
  - Application to nanopore ionic fluxes in progress
- More details
  - Salloum et al., SIAM MMS, submitted
  - Rizzi et al., J. Comp. Phys, submitted (Part I and II)

Papers

- Rizzi, F., Najm, H.N., Debusschere, B.J., Sargsyan, K., Salloum, M., Adalsteinsson, H., Knio, O.M., “Uncertainty Quantification in MD Simulations. Part I: Forward propagation”, *J. Comp. Phys.*, submitted, 2011
- Salloum, M., Sargsyan, K., Najm, H.N., Debusschere, B., Jones, R., Adalsteinsson, H. “A Stochastic Multiscale Coupling Scheme to account for Sampling Noise in Atomistic-to-Continuum Simulations” *SIAM Multiscale Modeling and Simulation*, submitted, 2011
- Rizzi, F., Najm, H.N., Debusschere, B.J., Sargsyan, K., Salloum, M., Adalsteinsson, H., Knio, O.M., “Uncertainty Quantification in MD Simulations. Part II: Inference of Force Field Parameters”, J. Comp. Phys., submitted, 2011

Extra material

Variables are exchanged across scale interfaces



Bayesian Inference

Let *m* be a hypothesis and *D* observed data.

Posterior → Likelihood → Prior

$$\mathcal{P}(m|D) = \frac{\mathcal{P}(D|m)\mathcal{P}(m)}{\int \mathcal{P}(D|m)\mathcal{P}(m)dm}$$

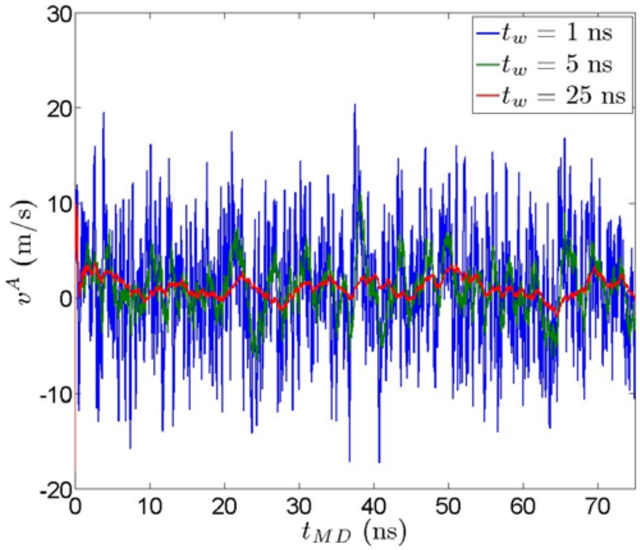
- The prior expresses the initial knowledge about the hypothesis *m* (*e.g.* uniform distribution, expert's knowledge...)

- The likelihood is the probability of observing the data *D* given the hypothesis *m*. It encompasses the forward model of *m*.

- The denominator is a normalization constant.

- The posterior is the probability of the hypothesis *m* given the data *D* : offers an enhanced knowledge of *m*.

Quantification of sampling noise in atomistic simulations



$$v_j^A = v^C + s\eta_j \qquad d = \{v_j^A\} \qquad j = 1,\ldots,N$$

$$P(v^C, s^2 | d) \propto P(d | v^C, s^2) P(v^C, s^2)$$

$$\mathcal{P}(\boldsymbol{d} | v^C, s^2) = (2\pi)^{-N/2} (s^2)^{-N/2} \exp\left(-\frac{\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}}{2s^2}\right)$$

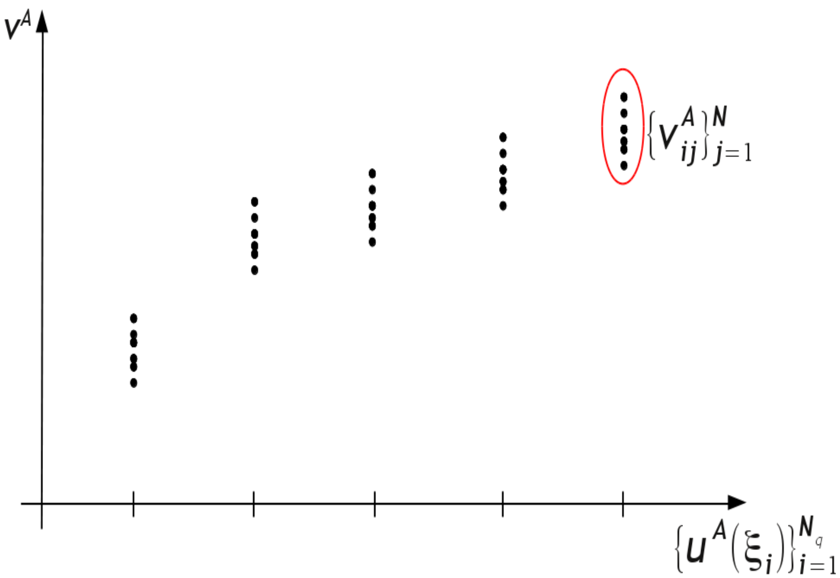
$$\epsilon_j = s\eta_j = v_j^A - v^C$$

$$\zeta \sim \mathcal{S}(0, 1, \gamma)$$

$$v^C = \mu + \Sigma\zeta$$

- Infer v^C from N short-term averaged MD velocity samples v_j^A
  - Gaussian model for data noise due to Central Limit Theorem (CLT)
  - Analytical solution gives v^C as student-t random variable
    - $P(v^C)$ marginalized over s is student-t distributed

Propagating parametric uncertainty and sampling noise through atomistic simulations



$$v_{ij}^A = v^C(\xi) + s\,\eta_{ij} \qquad v^C = \sum_{k=0}^{P} v_k^C \psi_k(\xi)$$

$$\begin{aligned} \tilde{v}^C &= \sum_{k=0}^{P} \tilde{v}_k^C \Psi_k(\xi) \\ &= \Psi(\xi)^T(\boldsymbol{\lambda} + \Lambda\boldsymbol{\zeta}) \end{aligned}$$

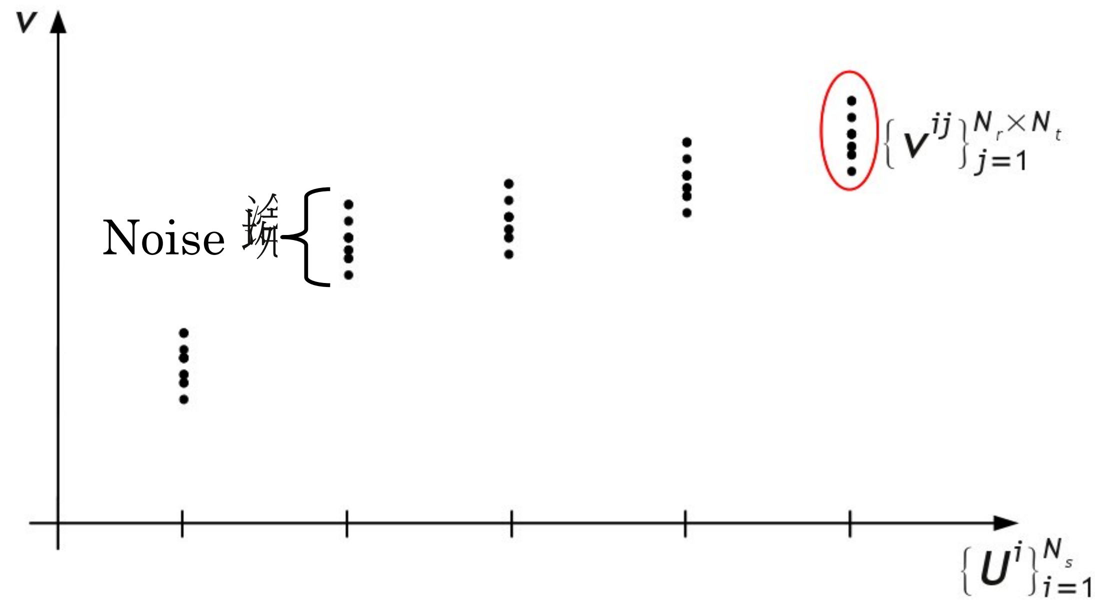
$$\zeta \sim \mathcal{S}(0, 1, \gamma)$$

- Infer v^C from N short-term averaged MD velocity samples v_{ij}^A
  - Sampled over a range of input velocities u_i^A
  - Gaussian model for data noise due to Central Limit Theorem (CLT)
  - Analytical solution gives v^C as student-t process over input velocity uncertainty

Inferring the Output Variable

$$\mathcal{P}(\tilde{\boldsymbol{u}}, \sigma^2 | D_v) \propto \mathcal{P}(D_v | \tilde{\boldsymbol{u}}, \sigma^2)\mathcal{P}(\tilde{\boldsymbol{u}}, \sigma^2)$$

$$v = \sum_{k=0}^{P} \tilde{u}_k \Psi_k(\xi) + \underbrace{\sigma\hat{\xi}}_{\text{Noise term}}$$



We draw samples from the posterior using Markov Chain Monte Carlo (MCMC) sampling.

ξ relates to the spread in U^i

Folding the input uncertainty and the sampling noise into one uncertain output

After marginalizing over σ^2, we obtain a joint posterior on the $\{\tilde{u}_k\}_{k=0}^{P}$:

$$\tilde{u} = \sum_{k=0}^{P} \tilde{u}_k \Psi_k(\xi)$$

We approximate $\{\tilde{u}_k\}_{k=0}^{P}$ as a Multivariate Normal Distribution (MVN) as follows:

$$\{\tilde{u}_k\}_{k=0}^{P} \sim \mathcal{MVN}(\boldsymbol{\mu}, \Sigma) = \boldsymbol{\mu} + L\zeta \quad \text{where} \quad L^T L = \Sigma$$

We obtain:

$$\tilde{u} = \boldsymbol{\Psi}(\xi)^T \cdot \boldsymbol{\mu} + \zeta \sqrt{\boldsymbol{\Psi}(\xi)^T \cdot \Sigma \cdot \boldsymbol{\Psi}(\xi)}$$

Folding the input uncertainty and the sampling noise into one uncertain output

$$\tilde{u} = \boldsymbol{\Psi}(\xi)^T \cdot \boldsymbol{\mu} + \zeta \sqrt{\boldsymbol{\Psi}(\xi)^T \cdot \Sigma \cdot \boldsymbol{\Psi}(\xi)}$$

This expression of $\tilde{u}$ is “cheap” for sampling in ζ and ξ !

Inverse Cumulative Distribution Function (CDF) transform

$F(.)$ is the CDF of $\tilde{u}$

$$u_k = \frac{\langle F^{-1}(\Phi(\xi))\Psi_k(\xi)\rangle}{\langle \Psi_k(\xi)^2 \rangle}$$

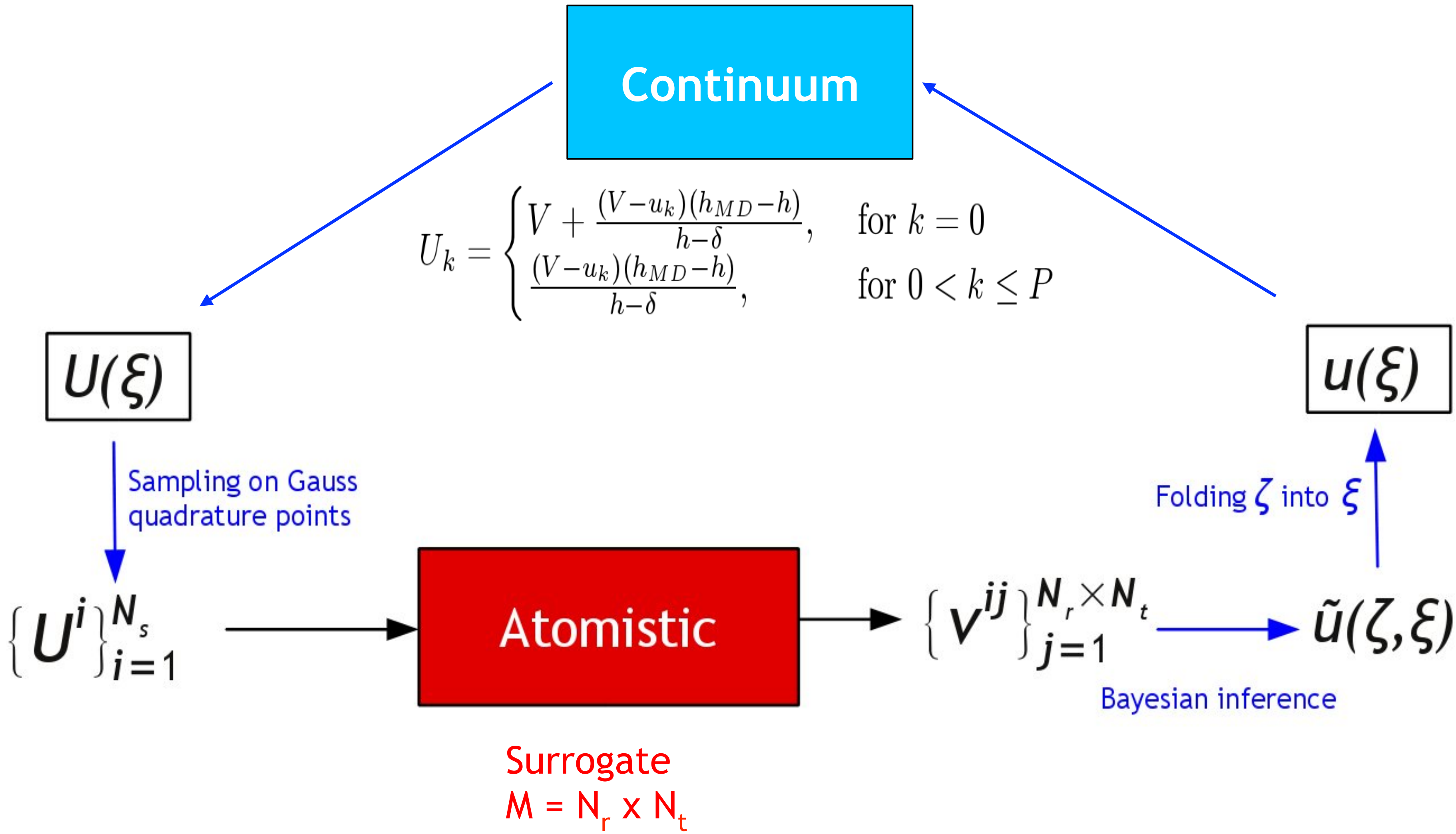
$$u = \sum_{k=0}^{P} u_k \Psi_k(\xi)$$

ξ is the degree of freedom associated with the sampling noise

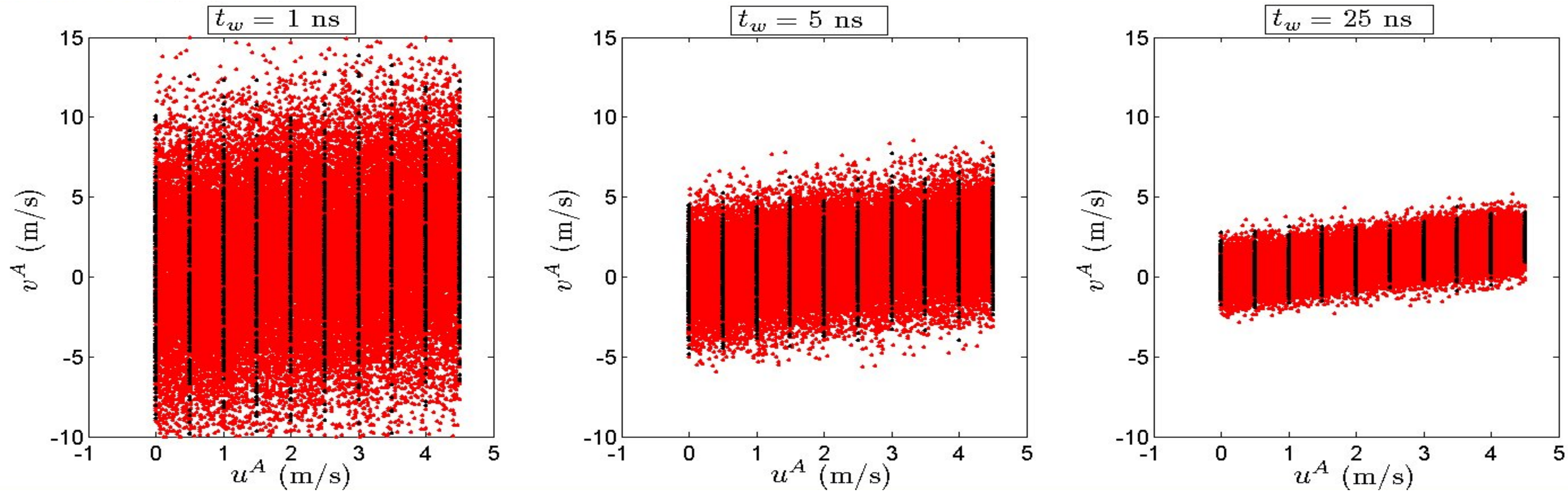
Summary of the Different Steps for Coupling

Laminar Newtonian Couette flow

The analytical solution is available



The surrogate reflects the original short-time averaged data



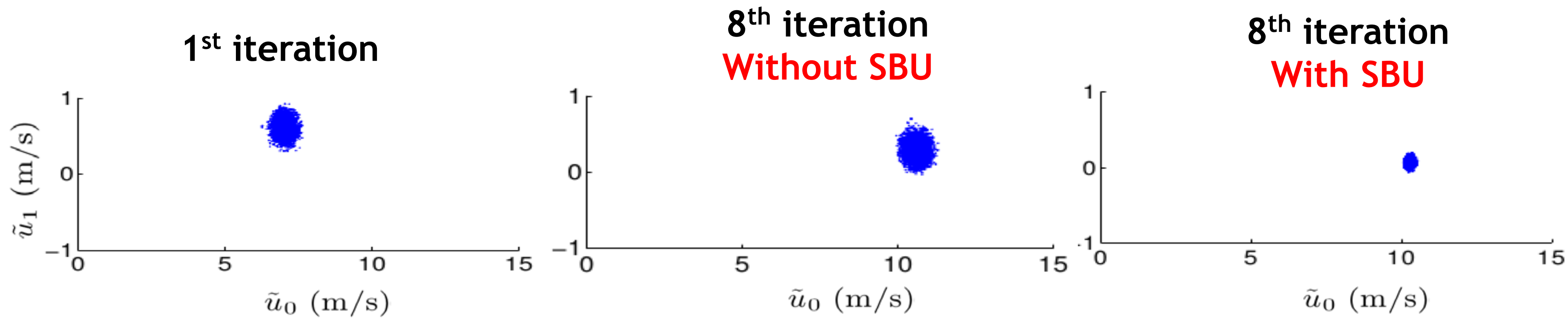
t_w = Time averaging window width

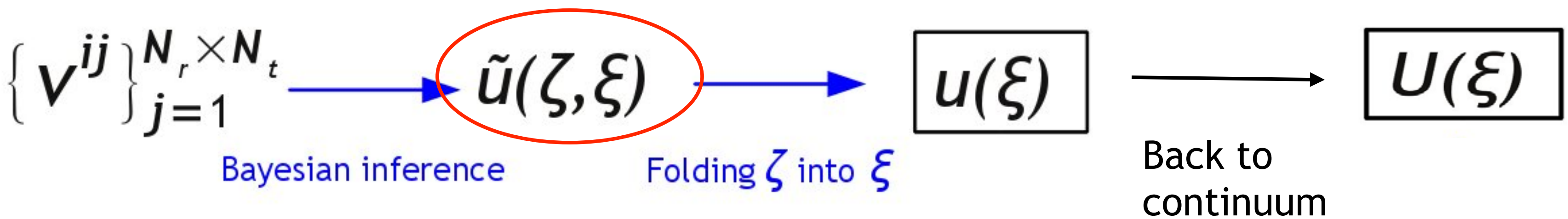
$$v^A = \alpha_0 + \alpha_1 u^A + s\eta$$

→ Spread in the short-time averaged MD data

- For a given t_w, the properties of the joint posterior on $\{a_0, a_1, s\}$ are tabulated.
- This joint posterior is a surrogate to the original atomistic scale short-time averaged MD data.

Joint Posterior of $\{\tilde{u}_0, \tilde{u}_1\}$

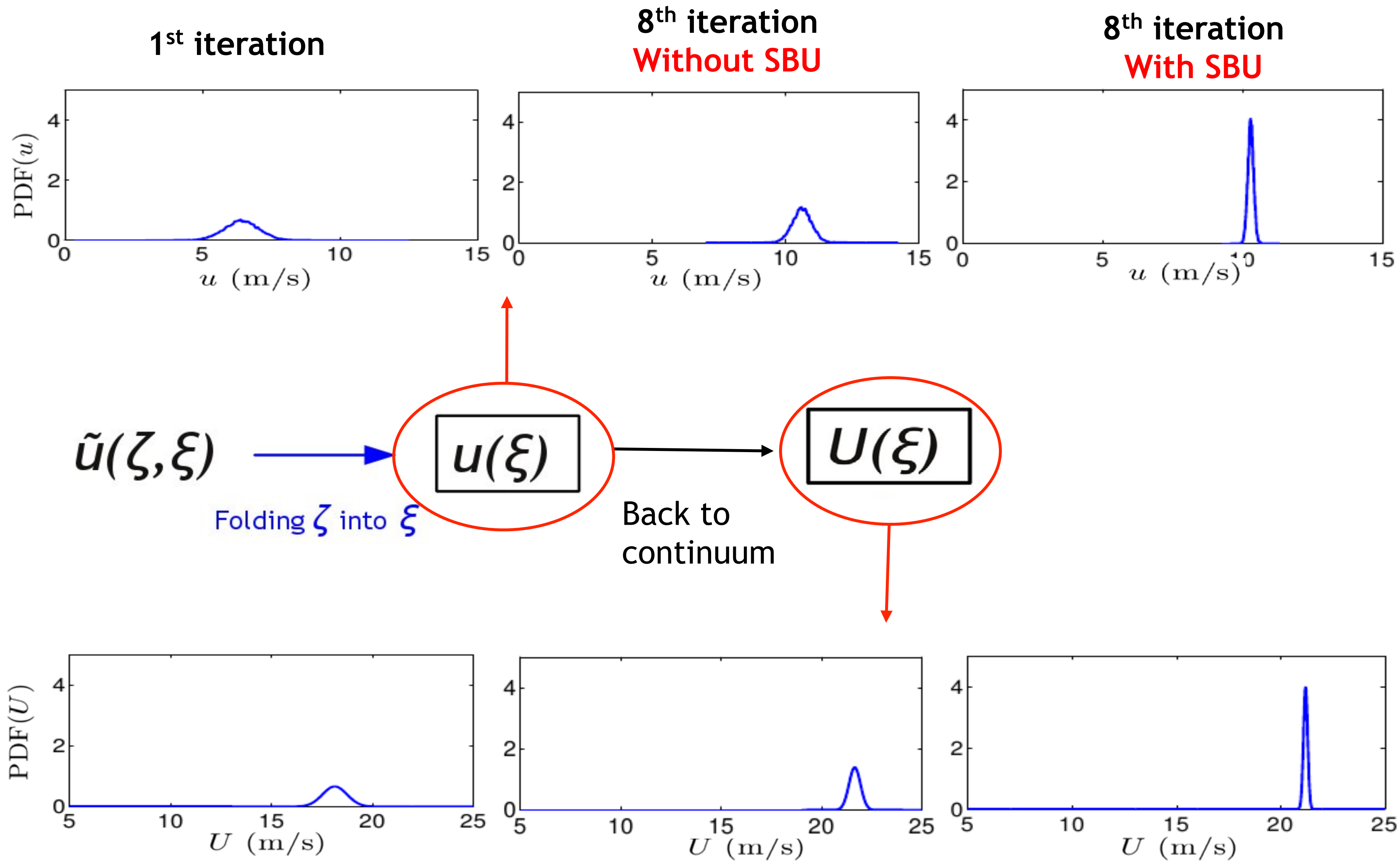




Sequential Bayesian Updating (SBU)

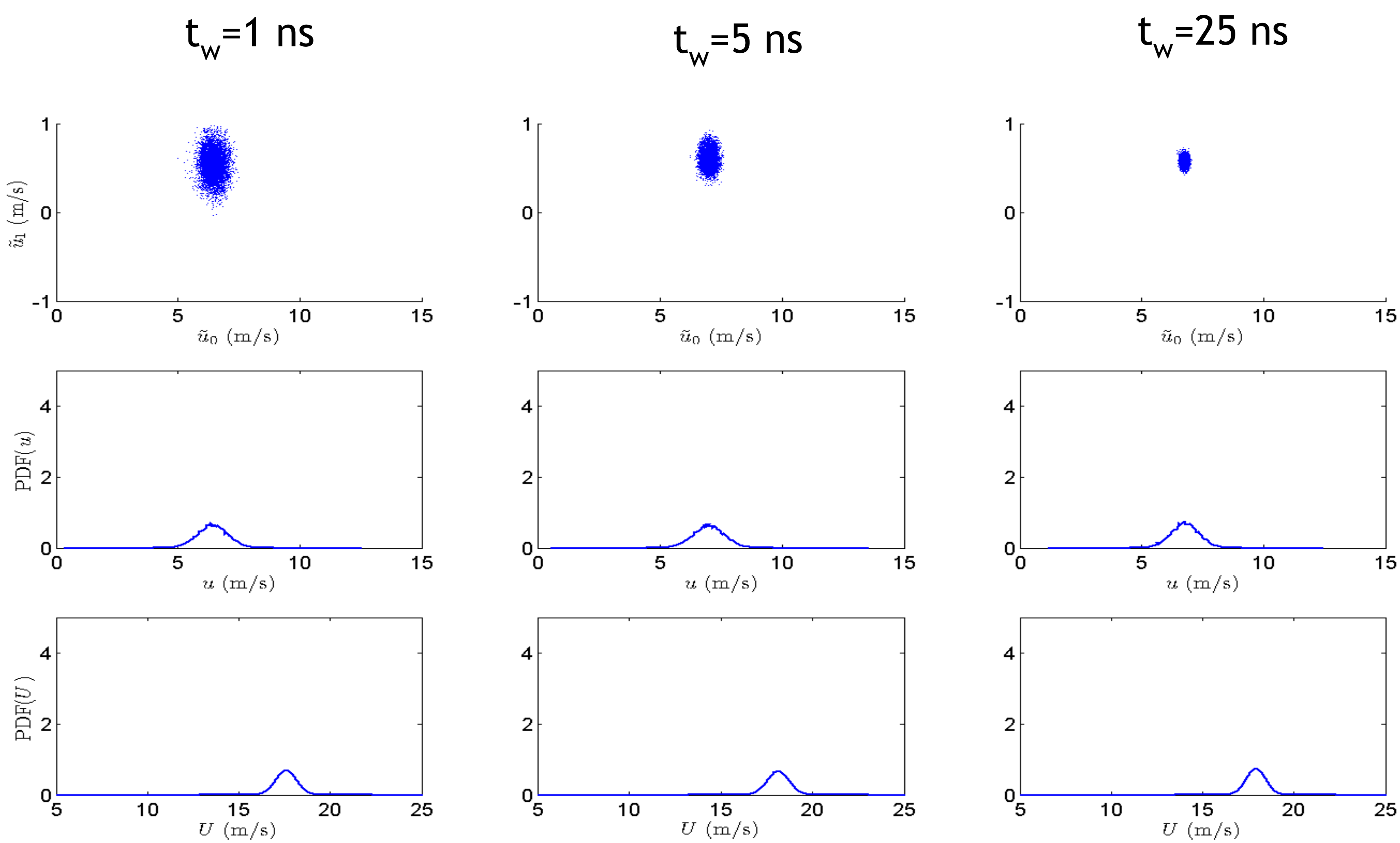
The posterior of the previous iteration is used as the prior in the current iteration.

PDFs of *u* and *U*



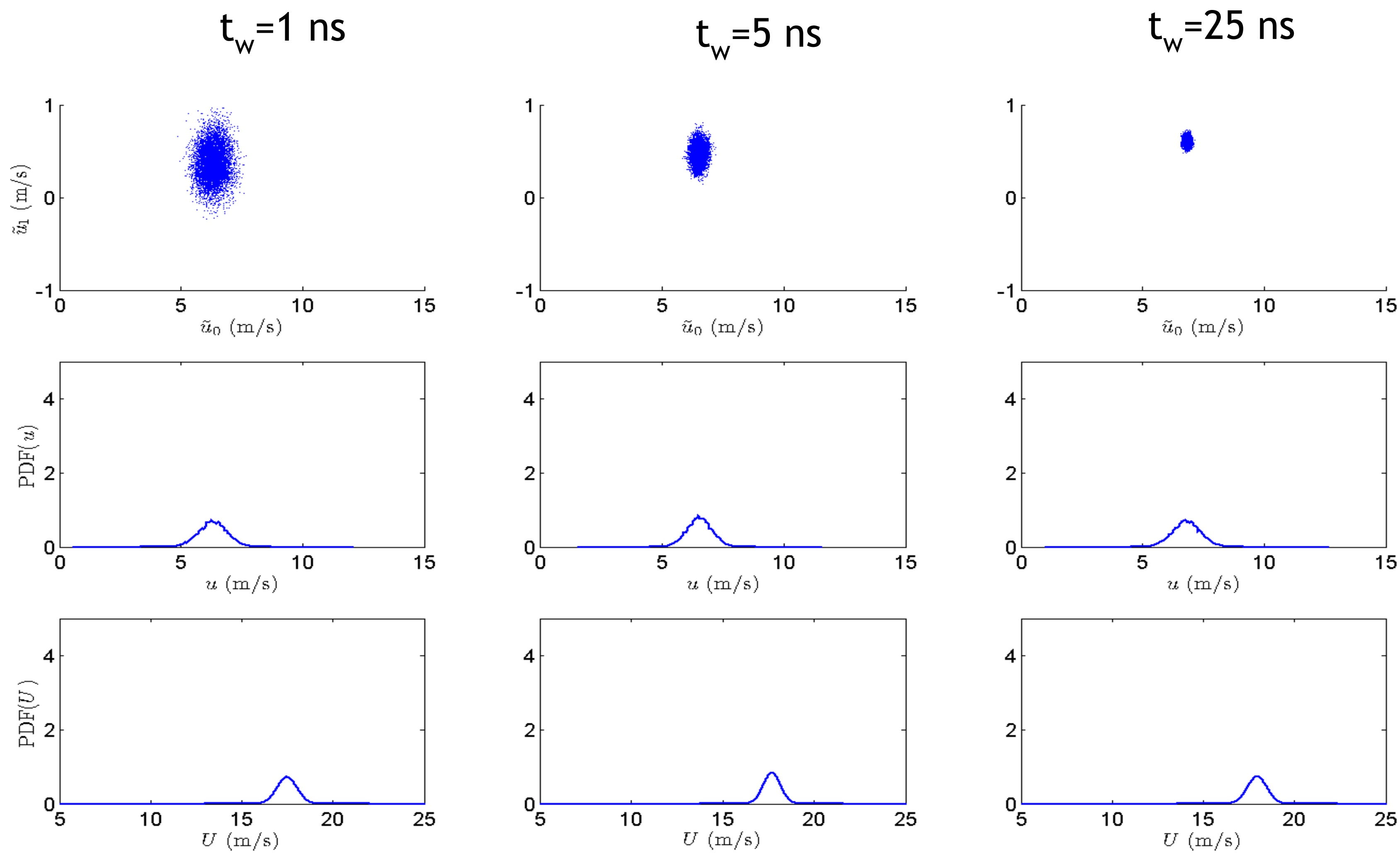
Stochastic Coupling Algorithm Convergence *Without* SBU

t_w = Time averaging window width



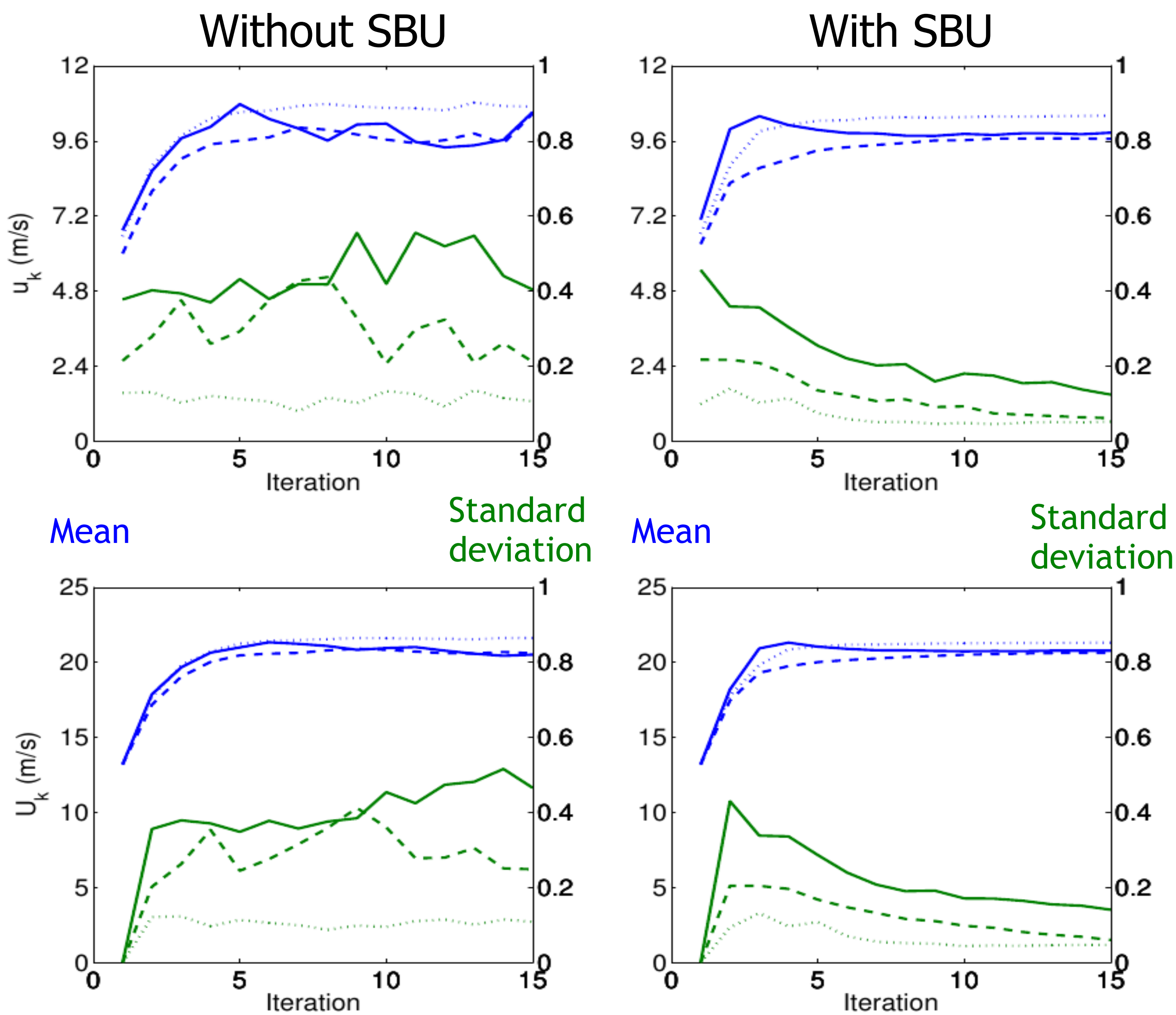
Stochastic Coupling Algorithm Convergence *With* SBU

t_w = Time averaging window width



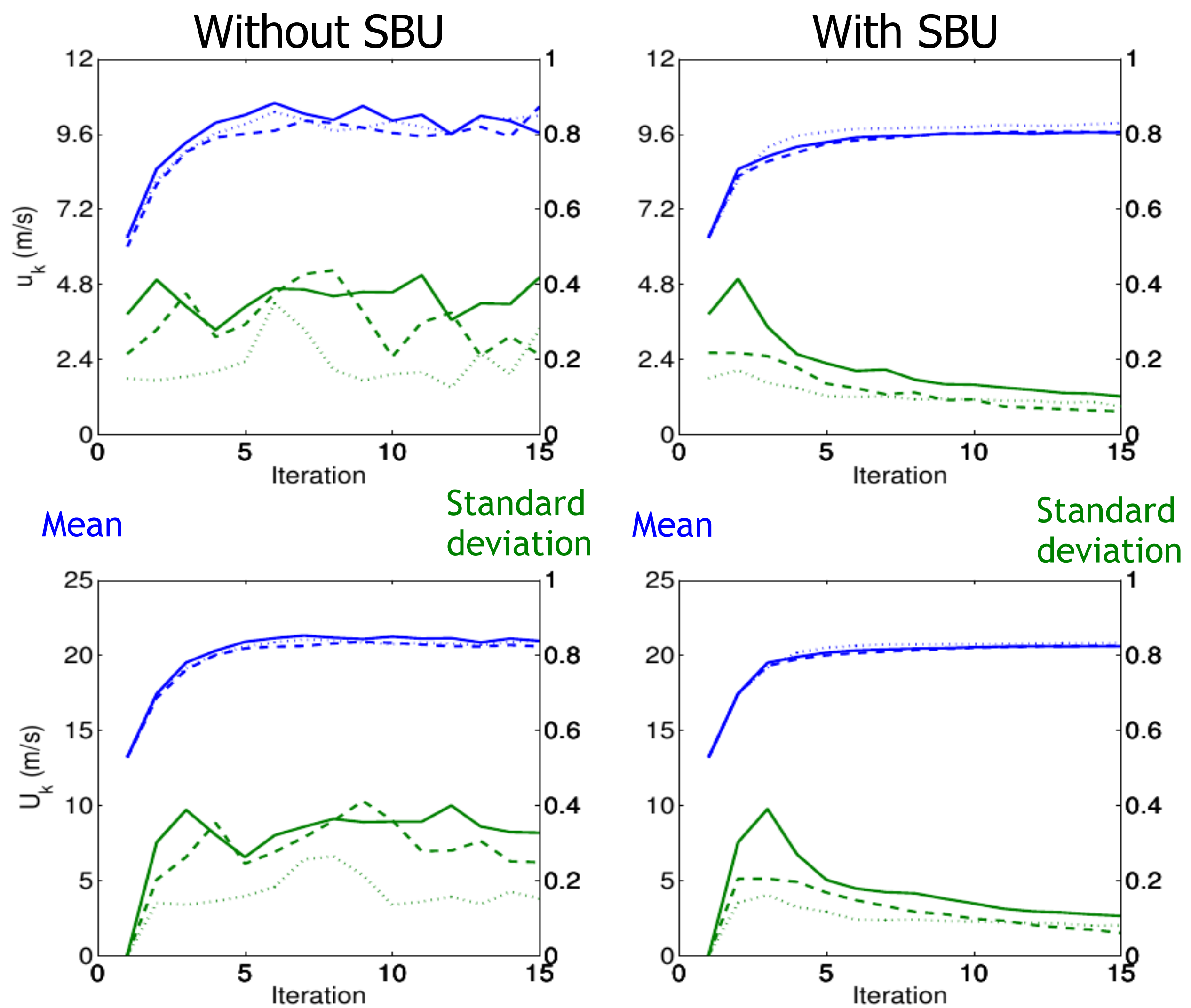
Effect of the time averaging window t_w on the convergence of the mean and standard deviation

Solid: t_w=1 ns Dashed: t_w=5 ns Dotted: t_w=25 ns



Effect of the number of samples *M* on the convergence of the mean and standard deviation

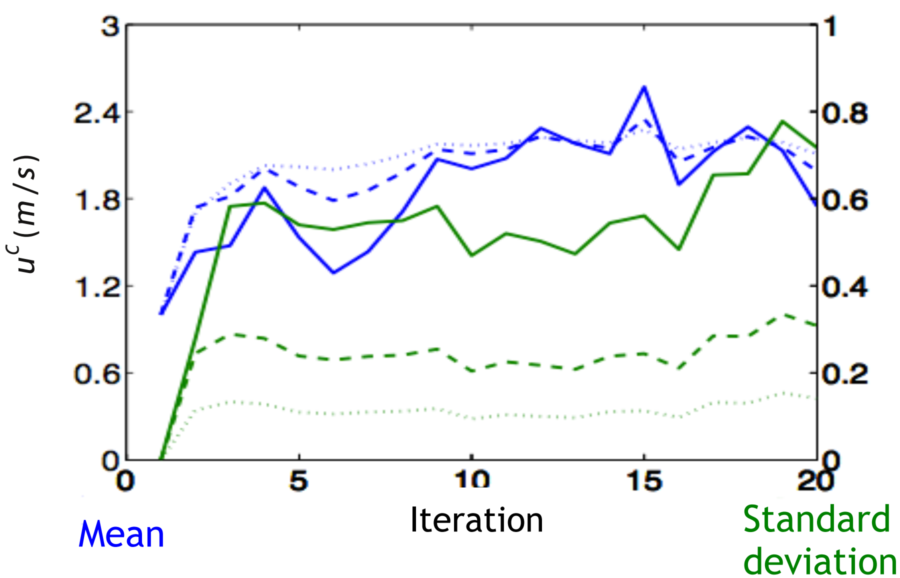
Solid: M=10 Dashed: M=20 Dotted: M=40



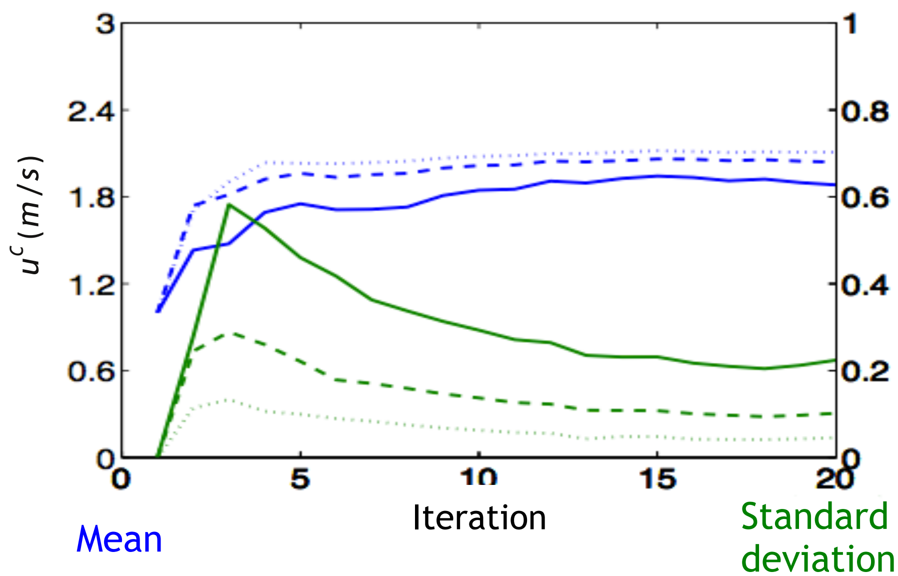
Effect of the time averaging window t_w on the convergence of the mean and standard deviation

Solid: t_w=1 ns   Dashed: t_w=5 ns   Dotted: t_w=25 ns

Without SBU



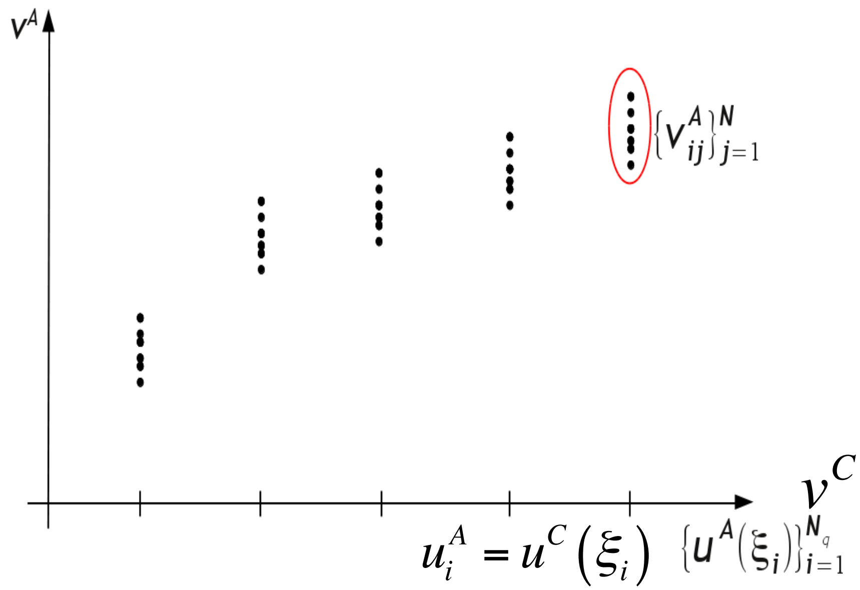
With SBU



Sequential Bayesian Updating (SBU)

- The posterior of the previous iteration is used as the prior in the current iteration.
- The accuracy in the converged variables by including additional data at each iteration.

Representation of response surface with student-t process



$$v^C = f^A\left(u^C, \zeta\right)$$

$$u^C = \sum_{k=0}^{P} u_k^C \psi_k(\xi) \qquad v^C = \sum_{k=0}^{P} v_k^C \psi_k(\xi)$$

$$v^C = \psi(\xi)^T (\lambda + \Lambda\zeta)$$

$$v^C = \psi(\xi)^T \overline{\mathbf{v}}^C + \zeta\sqrt{\psi(\xi)^T \Lambda^T \Lambda \psi(\xi)}$$

$$\zeta \sim S(0,1,\gamma)$$

- Sum of P+1 student-t RVs mapped into 1 student-t RV
- Well approximated with Gaussian Process if γ large enough
  - Satisfied if enough MD samples used

Comparison student-t and Gaussian