

## UT-LANL DOE/ASCR ISICLES Project

### *PARADICE: Uncertainty Quantification for Large-Scale Ice Sheet Modeling and Simulation*

Omar Ghattas\*<sup>†‡</sup>

Representing the rest of the UT-Austin/LANL team:

Don Blankenship<sup>†</sup> Tan Bui-Thanh\* Carsten Burstedde\*  
Jim Gattiker<sup>¶</sup> Dave Higdon<sup>¶</sup> Toby Isaac\* Charles Jackson<sup>†</sup>  
James Martin\* Stephen Price<sup>§</sup> Georg Stadler\*  
Lucas Wilcox\* (Bill Lipscomb<sup>§</sup>)

\*Institute for Computational Engineering and Sciences, UT-Austin

<sup>†</sup>Jackson School of Geosciences, UT-Austin

<sup>‡</sup>Dept. of Mechanical Engineering, UT-Austin

<sup>§</sup>COSIM/Fluid Dynamics Group, Los Alamos National Laboratory

<sup>¶</sup>Statistical Sciences, Los Alamos National Laboratory

## Overall project goals

- ▶ Create scalable, parallel, adaptive full Stokes ice sheet simulator equipped with deterministic and statistical inversion capabilities
- ▶ Integrate as an alternative dynamics model with CISM
- ▶ Quantify uncertainties in predictions of ice sheet dynamics in West Antarctica through assimilation of observational data

## Research challenges

- ▶ Scalable linear solvers for variable-coefficient Stokes
- ▶ Scalable nonlinear solvers for nonlinear rheology
- ▶ Adaptivity to resolve flow transitions and high velocity gradients
- ▶ Inverse methods to identify unknown constitutive parameters, basal boundary conditions, and geothermal heat flux
- ▶ Statistical inference methods to estimate uncertainty in unknowns
- ▶ Assimilation of observational data into models to quantify uncertainties in ice sheet dynamics predictions

# Overview of current parallel AMR framework for creeping non-Newtonian Stokesian flow and transport

$$\nabla \cdot \mathbf{u} = 0$$

$$-\nabla \cdot \left[ \eta(T, \mathbf{u}) \left( \nabla \mathbf{u} + \nabla \mathbf{u}^\top \right) - p \mathbf{I} \right] = \rho_0 [1 - \alpha (T - T_0)] \mathbf{g}$$

$$\rho_0 c \left( \frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) - \nabla \cdot (k \nabla T) = H(\mathbf{u})$$

$$\mathbf{n} \times \mathbf{n} \times \left[ \eta \left( \nabla \mathbf{u} + \nabla \mathbf{u}^\top \right) - p \mathbf{I} \right] \mathbf{n} = \mathbf{0} \text{ on } \partial \Omega$$

$$\mathbf{u} \cdot \mathbf{n} = 0 \text{ and } T = T_{bc} \text{ on } \partial \Omega \text{ and } T = T_{inc} \text{ at } t = 0$$

Variables:

- ▶  $T(\mathbf{x}, t)$ ,  $\mathbf{u}(\mathbf{x}, t)$ ,  $p(\mathbf{x}, t)$  — temperature, velocity, pressure

Parameters:

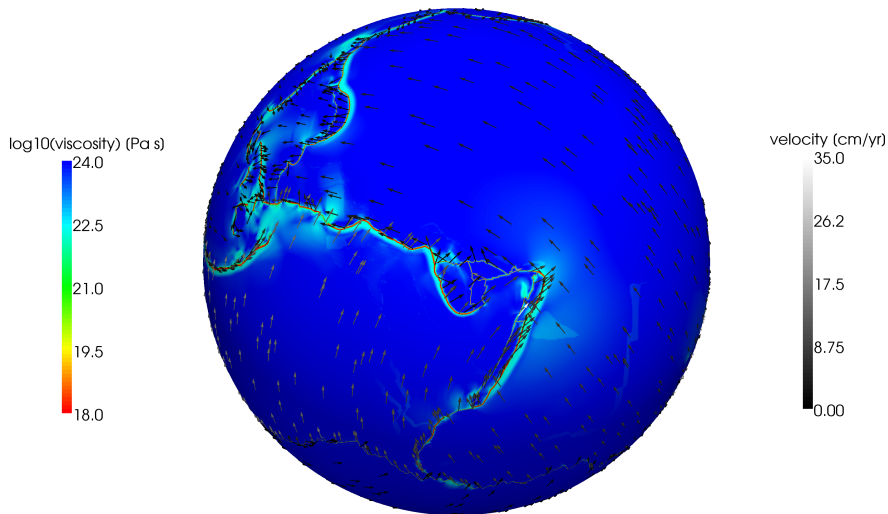
- ▶ Ra — Rayleigh number  $\sim 10^6$ – $10^9$
- ▶  $H(\mathbf{u})$ ,  $\eta(T, \mathbf{u})$  — heat production rate, viscosity
- ▶  $\rho_0$ ,  $T_0$  — reference density and temperature
- ▶  $k$ ,  $c$  — thermal conductivity, specific heat
- ▶  $\mathbf{g}$ ,  $\alpha$  — gravitational acceleration, coefficient of thermal expansion

# Mantle rheology

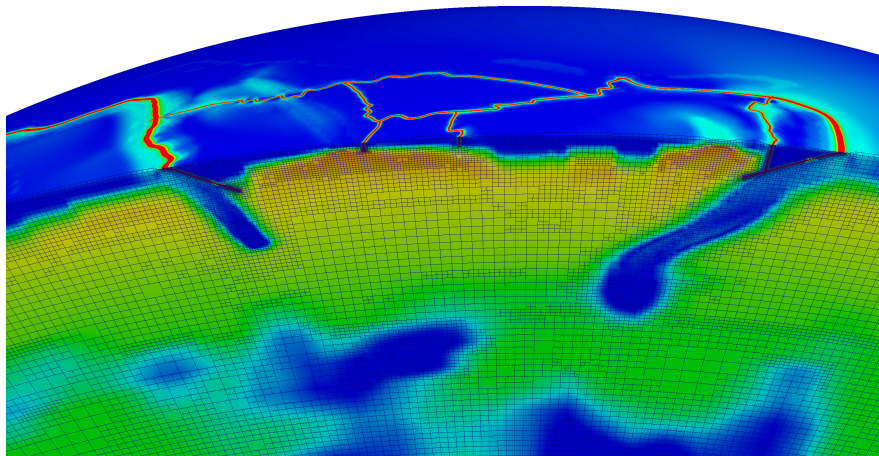
$$\eta(T, \mathbf{u}) = \left( \frac{d^q}{A C_{OH}^r} \right)^{\frac{1}{n}} \dot{\epsilon}_{II}^{\frac{1-n}{2n}} \exp\left( \frac{E_a + pV_a}{nRT} \right)$$

- ▶  $\eta(T, \mathbf{u})$  — effective viscosity
- ▶  $d$  — grain size
- ▶  $C_{OH}$  — water content in parts per million of silicon
- ▶  $\dot{\epsilon}_{II}$  — second invariant of strain rate tensor
- ▶  $E_a$  — the activation energy
- ▶  $p$  — pressure
- ▶  $V_a$  — activation volume
- ▶  $R$  — universal gas constant
- ▶  $T$  — temperature
- ▶  $A, n, r, q$  — parameters

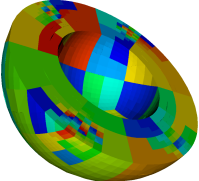
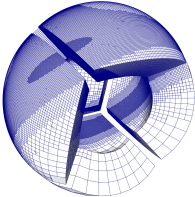
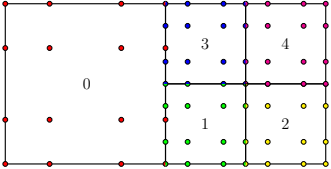
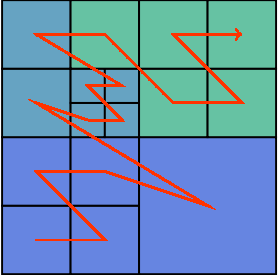
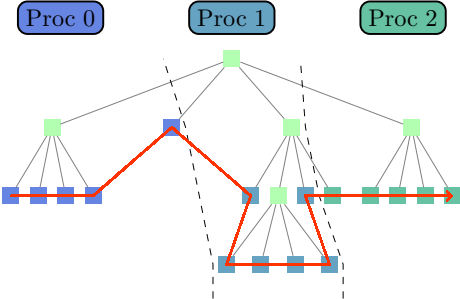
# Global mantle flow AMR simulations with realistic rheology and 1km local resolution at plate boundaries using *Rhea*



Global mantle flow AMR simulations with realistic rheology and 1km local resolution at plate boundaries using *Rhea*



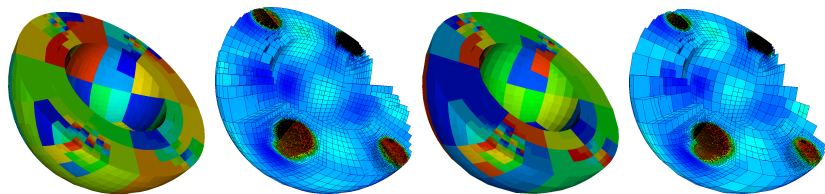
*p4est*: Adaptive parallel forests of octrees library





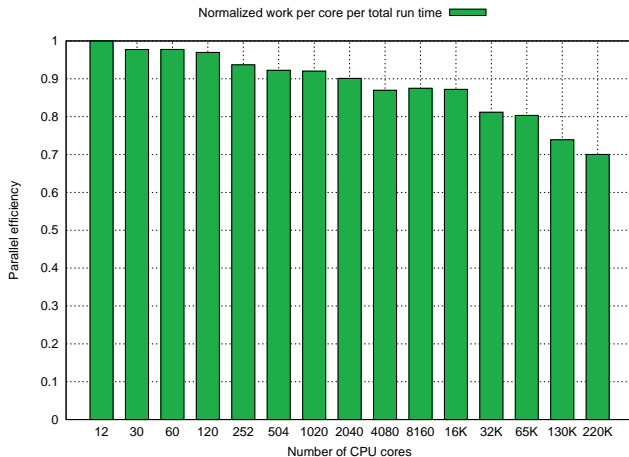
## *mangll*: High-order spectral element discretization library

- ▶ tailored to forest-of-octrees mesh library (*p4est*)
- ▶ Hexahedral elements with 2:1 nonconforming adapted faces
- ▶ Arbitrary order spectral element basis functions (nodal, GLL)
- ▶ Continuous and discontinuous Galerkin approximations
- ▶ Isoparametric or piecewise diffeomorphic geometry mapping



# Weak scalability of AMR for advection on spherical shell

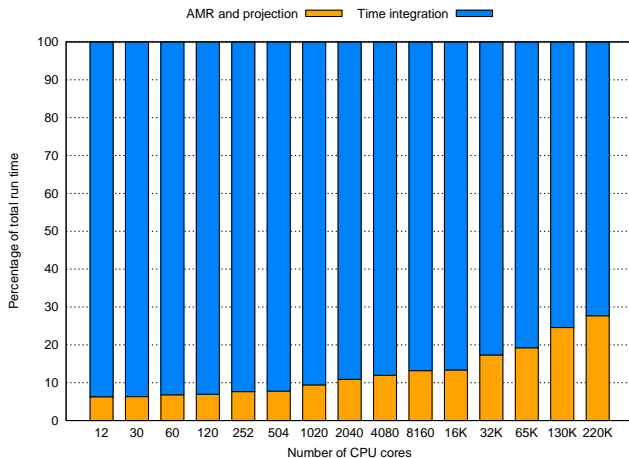
Excellent scalability from 12 to 220,000 cores



**Figure:** mesh is adapted every 32 time steps; explicit time stepping; discontinuous Galerkin, 3rd order spectral elements, 3200 elements/core (45B DOF on 220K cores); ORNL Jaguar system

# Weak scalability of AMR for advection on spherical shell

AMR consumes < 30% of overall time (unoptimized implementation)



**Figure:** mesh is adapted every 32 time steps; explicit time stepping; discontinuous Galerkin, 3rd order spectral elements, 3200 elements/core (45B DOF on 220K cores); ORNL Jaguar system

# Rhea: adaptive mantle convection code

## Summary of discretization and solution

- ▶ Trilinear FEM for temperature, velocity, and pressure
  - ▶ Conforming approximation by algebraic elimination of “hanging nodes”
- ▶ FEM stabilization:
  - ▶ Streamline Upwind/Petrov Galerkin (SUPG) for advection-diffusion system
  - ▶ Polynomial pressure projection for stabilization of Stokes equation (Dohrmann and Bochev)
- ▶ Explicit integration of energy equation decouples temperature update from nonlinear Stokes solve
- ▶ Nonlinear Stokes solver: lagged-viscosity (Picard) iteration
- ▶ Linear Stokes solver: MINRES iteration with block preconditioner based on AMG V-cycle and inverse viscosity mass matrix approximation of Schur complement

## MINRES preconditioner for linear Stokes solve

- ▶ Block factorization of Stokes system:

$$\begin{pmatrix} A & B^\top \\ B & -C \end{pmatrix} = \begin{pmatrix} I & 0 \\ BA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & -(BA^{-1}B^\top + C) \end{pmatrix} \begin{pmatrix} I & A^{-1}B^\top \\ 0 & I \end{pmatrix}$$

where  $A$  is discrete viscous operator,  $B$  is discrete divergence, and  $C$  is pressure stabilization

- ▶ Suggests preconditioner of form:

$$P = \begin{pmatrix} A & 0 \\ 0 & S \end{pmatrix} \quad \text{with} \quad S = BA^{-1}B^\top + C$$

- ▶ Approximate inverse of  $A$  with one V-cycle of Algebraic Multigrid (e.g. *ML* or *BoomerAMG*)
- ▶ Approximate  $S$  with diagonal inverse-viscosity lumped-mass matrix  $\tilde{M}$  (spectrally equivalent in isoviscous case)

# Independence of solver w.r.t. viscosity variation

Using BoomerAMG on Cartesian geometry

$\mu_{\min}$	$\mu_{\max}$	# MINRES iterations	AMG setup time (s)	solve time per iteration (s)
1.00e-0	1.00	86	25.29	5.82
4.98e-2	1.00	80	28.02	5.80
5.53e-4	1.00	75	25.26	5.62
5.53e-4	1.00	90	28.44	5.75
5.53e-4	1.00	91	26.97	5.35
6.14e-6	1.00	95	28.42	5.70
3.06e-7	1.00	93	31.35	6.46

Figure: 216M unknowns, 512 cores

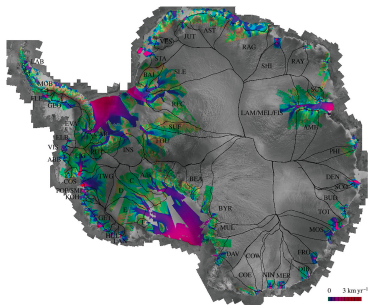
## Weak scalability of Stokes iterative solver in *Rhea*

#cores	#elem/ core	#elem	#dofs	#iter	setup time [s]	matvecs & inner prod. [s]	Vcycle time [s]
120	5800	700K	2.68M	24	1.39	2.75	2.88
960	4920	4.72M	18.7M	22	2.30	3.94	2.89
7680	4805	36.9M	146M	23	4.07	3.99	5.72
61440	5145	316M	1.26B	21	34.2	4.60	9.03
122880	5135	631M	2.51B	26	112.48	6.29	8.39

- ▶ Mid-ocean ridge benchmark problem
- ▶ Weak scaling with  $\sim 5000$  elements per core on ORNL Jaguar system
- ▶ Mesh contains elements over a range of three refinements
- ▶ Viscosity varies over an order of magnitude
- ▶ Number of MINRES iterations for decrease of residual by factor of  $10^4$
- ▶ AMG setup and V-cycle time based on *ML* from *Trilinos* with RCB/Zoltan repartitioning within multigrid hierarchy
- ▶ Matvec/inner product time includes all MINRES time other than preconditioner
- ▶ Note that in practice, AMG setup is amortized over numerous linear solves

# Inverse ice sheet modeling

- ▶ Observations: internal layering structure (ice-penetrating radar), surface flow velocity (InSAR), age (from ice cores), surface elevation (altimetry)
- ▶ Inversion variables: constitutive parameters, basal boundary conditions, geothermal heat flux, initial temperature



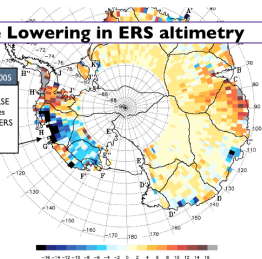


# Inverse ice sheet modeling, cont.

## Surface Lowering in ERS altimetry

Davis et al., 2005

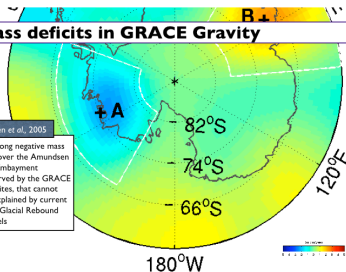
Ongoing surface lowering in the ASE during the nineties observed by the ERS satellite's radar altimeters



## Mass deficits in GRACE Gravity

Chen et al., 2005

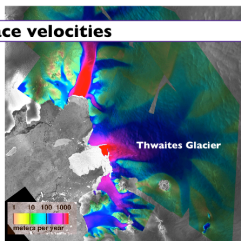
A strong negative mass loss over the Amundsen Sea Embayment observed by the GRACE satellites, that cannot be explained by current Post Glacial Rebound models



## InSAR surface velocities

Rignot et al., 2005

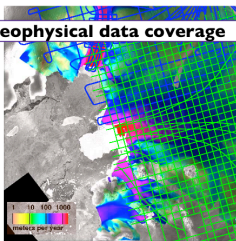
Significant accelerations of surface ice flow velocities seen between 1992 and 2005



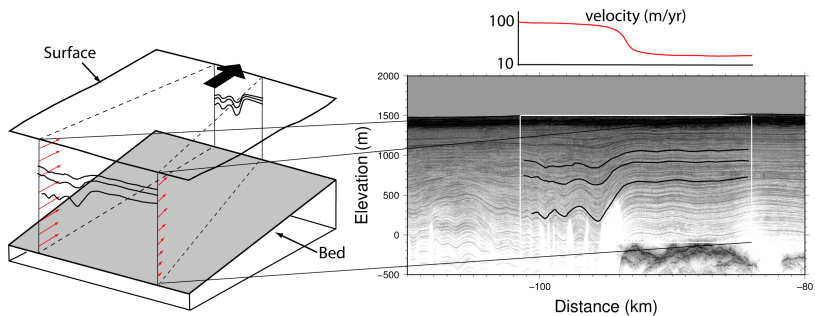
## UTIG aerogeophysical data coverage

Holt et al., 2006

Regional airborne survey probes underlying bedrock and demonstrates potential longer term vulnerability of the Thwaites Glacier Catchment.



# Inverse ice sheet modeling, cont.



UTIG ice penetrating radar

# A general ice sheet inverse problem

minimize  $\mathcal{F}(n, A_0, Q, q_B, \beta_B) := \| \mathbf{u} - \mathbf{u}^{\text{obs}} \|_{\Gamma_{FS}} + \| z - z^{\text{obs}} \|_{\Gamma_{FS}} + \| \phi - \phi^{\text{obs}} \|_{\Omega}$   
 $+ \| n - n^{\text{pr}} \|_{\Omega} + \| A_0 - A_0^{\text{pr}} \|_{\Omega} + \| Q - Q^{\text{pr}} \|_{\Omega} + \| q - q^{\text{pr}} \|_{\Gamma_B} + \| \beta - \beta^{\text{pr}} \|_{\Gamma_B}$   
 subject to:

$$\nabla \cdot \mathbf{u} = 0, \quad -\nabla \cdot \left[ \eta(T, \mathbf{u}) (\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{I}p \right] = \rho g$$

$$\eta(T, \mathbf{u}) = \frac{1}{2} A^{-\frac{1}{n}} \dot{\epsilon}_{\text{II}}^{\frac{1-n}{2n}}, \quad \dot{\epsilon}_{\text{II}} = \frac{1}{2} \text{tr}(\dot{\epsilon}^2), \quad \dot{\epsilon} = \frac{1}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T), \quad A = A_0 \exp \left( -\frac{Q}{RT} \right)$$

$$\frac{Dz}{Dt} |_{\Gamma_{FS}} = a, \quad \boldsymbol{\sigma} \mathbf{n} |_{\Gamma_{FS}} = \mathbf{0}, \quad \mathbf{u} \cdot \mathbf{n} |_{\Gamma_B} = 0, \quad \mathbf{n} \times \mathbf{n} \times \boldsymbol{\sigma} \mathbf{n} + \beta_B \mathbf{n} \times \mathbf{n} \times \mathbf{u} |_{\Gamma_B} = \mathbf{0}$$

$$\rho c \mathbf{u} \cdot \nabla T - \nabla \cdot (K \nabla T) = \eta \text{tr}(\dot{\epsilon}^2), \quad T |_{\Gamma_{FS}} = T_{FS}, \quad K \nabla T \cdot \mathbf{n} |_{\Gamma_B} = q_B$$

$$(\nabla \phi \cdot \nabla \phi)^{\frac{1}{2}} = \frac{1}{|\mathbf{u}|}, \quad \phi |_{\Gamma_{FS}} = 0$$

## Example: Inverse problem for parameter $n$

$$\text{minimize } \mathcal{F} := \| \mathbf{u} - \mathbf{u}^{\text{obs}} \| + \| n \|$$

subject to:

$$\nabla \cdot \mathbf{u} = 0$$

$$-\nabla \cdot [\eta(T, \mathbf{u}) (\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{I}p] = \rho \mathbf{g}$$

$$\rho c \left( \frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) - \nabla \cdot (K \nabla T) = \eta \text{tr}(\dot{\boldsymbol{\epsilon}}^2)$$

$$\eta(T, \mathbf{u}) = \frac{1}{2} A^{-\frac{1}{n}} \dot{\boldsymbol{\epsilon}}_{\text{II}}^{\frac{1-n}{2n}}$$

$$\dot{\boldsymbol{\epsilon}}_{\text{II}} = \frac{1}{2} \text{tr}(\dot{\boldsymbol{\epsilon}}^2)$$

$$\dot{\boldsymbol{\epsilon}} = \frac{1}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T)$$

$$A = A_0 \exp \left( -\frac{Q}{RT} \right)$$

# Optimality conditions

state equations:

$$\begin{aligned}\nabla \cdot \mathbf{u} &= 0 \\ -\nabla \cdot \left[ \eta(T, \mathbf{u}) (\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{I}p \right] &= \rho \mathbf{g} \\ \rho c \left( \frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) - \nabla \cdot (K \nabla T) &= \eta \operatorname{tr}(\dot{\boldsymbol{\epsilon}}^2)\end{aligned}$$

adjoint equations:

$$\begin{aligned}\nabla \cdot \mathbf{v} &= \mathcal{D}_p \mathcal{F} \\ -\nabla \cdot \left[ \eta (\nabla \mathbf{v} + \nabla \mathbf{v}^T) - \mathbf{I}q \right] + \frac{\mathcal{D}_u \eta}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T) : (\nabla \mathbf{v} + \nabla \mathbf{v}^T) &= \\ \rho c S \nabla T + S \mathcal{D}_u \eta \operatorname{tr}(\dot{\boldsymbol{\epsilon}}^2) - \nabla \cdot \left[ S \eta (\nabla \mathbf{u} + \nabla \mathbf{u}^T) \right] - \mathcal{D}_u \mathcal{F} & \\ -\rho c \left( \frac{\partial S}{\partial t} + \mathbf{u} \cdot \nabla S \right) - \nabla \cdot (K \nabla S) - S \mathcal{D}_T \eta \operatorname{tr}(\dot{\boldsymbol{\epsilon}}^2) &= \\ -\frac{\mathcal{D}_T \eta}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T) : (\nabla \mathbf{v} + \nabla \mathbf{v}^T) - \mathcal{D}_T \mathcal{F} &\end{aligned}$$

control equation (e.g. for parameter  $n$ ):

$$\int_{t_0}^{t_1} \left[ \frac{1}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T) : (\nabla \mathbf{v} + \nabla \mathbf{v}^T) - S \operatorname{tr}(\dot{\boldsymbol{\epsilon}}^2) \right] \mathcal{D}_n \eta - \mathcal{D}_n \mathcal{F} = 0$$

# Mathematical and computational issues encountered in deterministic inversion

Beyond generic inverse solver issues, the following must be addressed in the context of the inverse ice sheet problem:

- ▶ Gradient-consistent adjoint discretization schemes
- ▶ Inexact Newton-CG for the inverse problem
- ▶ Hessian preconditioning, multilevel methods
- ▶ Appropriate regularization

# Bayesian inference for inverse problem

- ▶ Bayesian framework for statistical inverse problem: when data and/or model have uncertainties, solution of inverse problem expressed as a posterior probability density function
- ▶ Central challenge: for inverse problems characterized by high-dimensional parameter spaces, method of choice is to sample the posterior density using Markov chain Monte Carlo (MCMC)
- ▶ For inverse problems characterized by expensive forward simulations, contemporary MCMC methods become prohibitive
- ▶ Intractability of MCMC methods for large-scale statistical inverse problems can be traced to their black-box treatment of the parameter-to-observable map (the forward code)
- ▶ Goal: develop methods that exploit the structure of the parameter-to-observation map (including its derivatives), as has been done successfully in deterministic PDE-constrained optimization
  - ▶ Hessian-informed Gaussian process response surface approximation
  - ▶ Hessian-preconditioned Langevin methods

# Bayesian formulation of statistical inverse problem

Given:

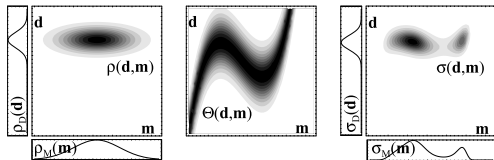
$\pi_{\text{pr}}(\mathbf{x})$  := prior p.d.f. of model parameters  $\mathbf{x}$

$\pi_{\text{obs}}(\mathbf{y})$  := prior p.d.f. of the observables  $\mathbf{y}$

$\pi_{\text{model}}(\mathbf{y}|\mathbf{x})$  := conditional p.d.f. relating  $\mathbf{y}$  and  $\mathbf{x}$

Then *posterior p.d.f. of model parameters* is given by:

$$\begin{aligned}\pi_{\text{post}}(\mathbf{x}) &\stackrel{\text{def}}{=} \pi_{\text{post}}(\mathbf{x}|\mathbf{y}_{\text{obs}}) \\ &\propto \pi_{\text{pr}}(\mathbf{x}) \int_{\mathcal{Y}} \frac{\pi_{\text{obs}}(\mathbf{y})\pi_{\text{model}}(\mathbf{y}|\mathbf{x})}{\mu(\mathbf{y})} d\mathbf{y} \\ &\propto \pi_{\text{pr}}(\mathbf{x}) \pi(\mathbf{y}_{\text{obs}}|\mathbf{x})\end{aligned}$$



From A. Tarantola, *Inverse Problem Theory*, SIAM, 2005



## Gaussian additive noise

Given the parameter-to-observable map  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ , a common noise model is Gaussian additive noise:

$$\mathbf{y}_{\text{obs}} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$$

If the prior is taken as Gaussian with mean  $\mathbf{x}_{\text{pr}}$  and covariance  $\boldsymbol{\Gamma}_{\text{pr}}$ , then the posterior can be written

$$\pi_{\text{post}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}_{\text{obs}}\|_{\boldsymbol{\Gamma}_{\text{noise}}^{-1}}^2 - \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{\text{pr}}\|_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}}^2\right)$$

Note that “most likely” point is given by

$$\begin{aligned} \mathbf{x}_{\text{MAP}} &\stackrel{\text{def}}{=} \arg \max_{\mathbf{x}} \pi_{\text{post}}(\mathbf{x}) \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}_{\text{obs}}\|_{\boldsymbol{\Gamma}_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{\text{pr}}\|_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}}^2 \end{aligned}$$

This is an (appropriately weighted) deterministic inverse problem!

## Gaussian additive noise, linear inverse problem

Suppose further the parameter-to-observable map is linear, i.e.

$$\mathbf{y} = \mathbf{F}\mathbf{x}$$

Then the posterior can be written

$$\pi_{\text{post}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \|\mathbf{F}\mathbf{x} - \mathbf{y}_{\text{obs}}\|_{\mathbf{\Gamma}_{\text{noise}}^{-1}}^2 - \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{\text{pr}}\|_{\mathbf{\Gamma}_{\text{pr}}^{-1}}^2\right)$$

The posterior is then Gaussian with

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}_{\text{MAP}}, \mathbf{\Gamma}_{\text{post}})$$

The covariance is the inverse Hessian of the negative log posterior:

$$\begin{aligned}\mathbf{\Gamma}_{\text{post}}^{-1} &= \mathbf{F}^T \mathbf{\Gamma}_{\text{noise}}^{-1} \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1} \\ &= \nabla_{\mathbf{x}}^2 (-\log \pi_{\text{post}})\end{aligned}$$

I.e., the covariance is given by the inverse Hessian of the regularized misfit function that is minimized by deterministic methods

# Summary of mathematical/computational research agenda

- ▶ Overall goal: Scalable, parallel, adaptive full Stokes ice sheet simulator equipped with deterministic and statistical inversion capabilities
- ▶ Base level enhancements to existing nonlinear Stokes code
  - ▶ free surface
  - ▶ basal boundary conditions
- ▶ Further forward solver enhancements (as needed)
  - ▶ high order spectral element DG/CG discretization
  - ▶ fully implicit time integration
  - ▶ full Newton solver
- ▶ Deterministic inversion
  - ▶ gradient-consistent/adjoint-appropriate discretization
  - ▶ globalization of inexact Newton-CG solver
  - ▶ multilevel Hessian preconditioner
  - ▶ regularization
- ▶ Statistical inversion
  - ▶ reduce-then-sample: Gaussian process response surface methods
  - ▶ sample-then-reduce: Langevin-based proposals for MCMC