CESMIX: Center for the Exascale Simulation of Materials in Extreme Environments

Project Overview

Youssef Marzouk
MIT PSAAP-3 Team
18 August 2020
The CESMIX team

- Our team integrates expertise in quantum chemistry, atomistic simulation, materials science; hypersonic flow; validation & uncertainty quantification; numerical algorithms; parallel computing, programming languages, compilers, and software performance engineering
Project objectives

- **Exascale simulation of materials in extreme environments**
  - In particular: ultrahigh temperature ceramics in **hypersonic** flows
    - Complex materials, e.g., metal diborides
    - Extreme aerothermal and chemical loading
    - Predict materials *degradation* and damage (oxidation, melting, ablation), capturing the central role of *surfaces* and *interfaces*

- **New predictive simulation paradigms and new CS tools for the exascale**
Broad relevance

• Intense current interest in reentry vehicles and hypersonic flight
  – A national priority!
  – Materials technologies are a key limiting factor
• Material properties are of cross-cutting importance:
  – Oxidation rates
  – Thermo-mechanical properties: thermal expansion, creep, fracture
  – Melting and ablation
  – Void formation
• New systems being proposed and fabricated (e.g., metallic high-entropy alloys)
• May have relevance to materials aging

• Yet extreme environments are largely inaccessible in the laboratory
  – Predictive simulation is an essential path…
Demonstration problem: specifics

• Aerosurfaces of a hypersonic vehicle…
• Hafnium diboride (HfB$_2$) promises necessary temperature resistance
• Yet predictive simulation (with chemical accuracy) remains out of reach:
  – High temperatures, extreme and localized loading
  – Complex surface chemistry
  – Challenging electronic structures
  – Materials damage starts at surface and propagates into bulk; dominant role of surfaces and interfaces
  – *Need to understand oxidation resistance and mechanical properties!*
Physical processes...

Transition and shock-boundary layer interaction have dramatic impact on surface heat flux and pressure

Chemical species profiles near surface; reactive environment drives materials degradation

ZrB$_2$-SiC after oxidation at 1600 °C
“The accident investigation pointed to the existence of shock waves associated with peeling skin that were 100 times stronger than the vehicle was designed to withstand”
Eswar Josyula (Air Force)

Demonstration problem: specifics

- Our focus is on simulating **materials degradation**
- External flow calculations, in canonical configurations, will provide realistic aerothermal and chemical inputs
- In-house experimental effort, DARPA MACH data, and literature data for validation
Computing goals

- Targeted capabilities:
  - *In predictive simulation:*
    - An example of autonomous and validated bridging of first principles (quantum mechanical) modeling with the engineering scale
    - Comprehensive uncertainty quantification (UQ) for all predictions
    - Few efforts currently seek to *unify* across all these scales
  - *Enabled by exascale computing:*
    - Compiler technologies for portability and composable performance
    - High-level and domain-specific languages
    - Differentiable programming
    - Software toolchains for efficient UQ and inference

- These advances will have fundamental and broad applicability…
Simulation approach

Simulation "decision engine"

First principles

DFT and beyond

TeraChem

ab initio

MD

QM/MM

LAMMPS

CAMPS

Hypersonic flow simulation

Bayesian inference & uncertainty quantification

MUQ

Differentiable programming: CSI, Enzyme,

Validated predictions
Exascale computing

• Key challenges:
  – End of Moore’s Law
  – *Software performance engineering* and *new language and compiler technologies* are essential for future gains:
    • Heterogeneous computing environments
    • Combining legacy codes with new
    • Achieving composable performance
    • Ease of use, and abstractions for the long term

• Our approach: a unified software stack that enables *holistic* performance engineering across all components of an exascale scientific simulation
Proposed software stack

- Sparse algebra kernels
- Dense algebra kernels
- Novel HP scientific codes
- Existing scientific packages

- TACO DSL
- Tiramisu optimization framework
- Julia front-end
- C/C++ front-ends
- Fortran front-end

- Productivity tools
- CSI
- OpenCilk
- Enzyme
- Exascale back-end
- CPU
- GPU
- Accelerator
Simulation software

**Written using**
- taco
- julia
- C++
- MPI

**Compiled using**
- OpenCilk (CSI, Enzyme)

**Existing packages**
- CAMPS
- TeraChem

**Productivity tools**
- Fortran
- NVIDIA CUDA
- Aurora
- Frontier

**(based on Tapir/LLVM)**

**Compiled using**
- MUQ...
Uncertainty quantification

- Many uncertainties arise in traversing scales
  - Boundary effects in QM/MM and DFT
  - Disorder (configurational uncertainty) in real materials
  - MD with fixed potentials: parameter uncertainty and structural limitations
  - Different levels of DFT theory (specifically) and model fidelity (in general)

- Goal: choosing simulations and information sources
  - What kind of model to run and where? How to sample configuration space? When to switch from classical to ab initio MD? When to use or update reactive or fixed-charge potentials?
  - Take into account computational costs
  - Quantify uncertainty in the results!

- UQ and inference will be the “glue” across our multiscale simulation approaches
Validation experiments

- **Thermogravimetric analysis** of phase-pure ceramic material in oxidizing atmospheres, in novel high temperature devices (e.g., all-ceramic TGA, induction furnace)
- Pre- and post-characterizations: XRD for phase recognition, SEM for morphology, and EDX for elemental mapping

TGA: mass change observable during isothermal reduction at different temperatures
Numerous computational linkages

- OpenCilk (based on Tapir/LLVM) as a common middle-end
  - Integrate programs written in different languages and concurrency platforms
- Software performance engineering for molecular dynamics with complex potentials and for hypersonic flow sim (on GPUs and beyond)
- TACO DSL to revisit selected DFT kernels and improve portability to new architectures
- CSI for performance analysis and to augment floating-point computation in legacy codes
- Julia code for UQ and inference across scales, ML-driven error indicators, and simulation decisions
The need for exascale

• We seek an automated UQ-driven *multiscale* approach of unprecedented predictive fidelity:
  – System-level MD or QM/MM simulations will trigger numerous DFT calculations asynchronously
    • To update reactive or NN potentials on-the-fly as uncertainty thresholds are violated (e.g., in new regions of phase space)
    • To resolve dynamically chosen QM regions in QM/MM
  – Multiple concurrent runs of *ab initio* MD to discover reaction events and intermediates; driven by accelerated sampling
  – Within DFT: multiple fidelities (i.e., levels of theory)
  – Outer loops for forward UQ: impact of parametric and configurational uncertainties
  – Additional outer loops for coupling to continuum scale
  – Inner loops: many-query runs of MD models for Bayesian learning of potentials
Example: online updating of MD potentials

Bayesian learning of reactive or NN potentials

improved potentials

MD simulation with integrated UQ or error indicators

New DFT calculations

uncertainties too large?
The need for exascale

• We seek an automated UQ-driven *multiscale* approach of unprecedented predictive fidelity:
  – System-level MD or QM/MM simulations will trigger numerous DFT calculations asynchronously
    • To update reactive or NN potentials on-the-fly as uncertainty thresholds are violated (e.g., in new regions of phase space)
    • To resolve dynamically chosen QM regions in QM/MM
  – Multiple concurrent runs of *ab initio* MD to discover reaction events and intermediates; driven by accelerated sampling
  – Within DFT: multiple fidelities (i.e., levels of theory)
  – Outer loops for forward UQ: impact of parametric and configurational uncertainties
  – Additional outer loops for coupling to continuum scale
  – Inner loops: many-query runs of MD models for Bayesian learning of potentials
We are excited to dive in!

Questions?
Extra materials
CESMIX simulation version zero

- **DFT for HfB$_2$**
- **Interfaces* + glue code**
- **Potential fitting**
- **MD**
- **Hypersonic flow sim**
- **CAMPS**
- **LAMMPS**

- **OpenCilk**

*(based on Tapir/LLVM)*

*All interfaces package-independent*

**Not in version zero:**
- Non-equilibrium chemistry, melting, ablation
- Uncertainty quantification
- QM/MM, AIMD, and simulation decisions
- ML-driven surrogates and error indicators
Full system prediction goals

- **Year 1½:**
  - Oxidation profile in HfB$_2$, canonical flow geometry (e.g., double-ramp and cylindrical cowl leading edge)
    - Materials simulation using reactive potentials
    - One-way coupling from flow to materials simulation, and chemically reacting equilibrium gas models

- **Year 5:**
  - Oxidation profile + thermal properties + mechanical properties in HfB$_2$ with defects. Range of hypersonic flow conditions.
    - Materials simulation including QM/MM acceleration with *ab initio* MD, reactive/NN, and fixed-charge potentials deployed and updated dynamically, i.e., a complete UQ-driven multiscale simulation engine
    - Two-way coupling between flow and materials simulation
Uncertainty quantification

• **Key ingredients:**
  – Parametric forward UQ: sensitivity analysis, dimension reduction
  – Bayesian statistical inference for learning potentials
  – New approaches to inference and UQ under model misspecification
  – Tying error and uncertainty estimates to simulation decisions

• Enabled by **differentiable programming** and composable software
Simulation software

Existing packages

CAMPS  TeraChem  LAMMPS  QuantumEspresso  MUQ  ...

Written using

taco

julia

C++

MPI

NVIDIA

CUDA

Compiled using

Productivity tools

Tapir (CSI, Enzyme)

The LLVM compiler

Fortran

Aurora

Frontier