Algorithmic Development for Informative and Intelligent Scientific Computation

Qiang Du @CM3

Department of Applied Physics and Applied Mathematics, and Data Science Institute

May 2021, ICSE

A word about CM3

Computational Mathematics and Multiscale Modeling (CM^3)

CM3 collaborates with various groups in Columbia: APAM/CS/CEEM/MECH, DSI, and Law School,...

Recent funded projects @CM3

- 2020-23 NSF-DMS: Numerical analysis of asymptotically compatible discretization of nonlocal models.
- 2020-21 AWS ML award: Multi-autonomous vehicle driving policy learning for efficient & safe traffic
- 2020-25 NSF-DMS: Research Training in Applied Mathematics at Columbia University
- 2019-23 NSF-DMREF: Complex nanofeatures in crystals: theory and experiment meet in the cloud
- 2017-20 NSF DMS: Numerical analysis of SPH type methods via nonlocal models
- 2017-20 NSF CCF (TRIPODS): From foundations to practice of data science and back
- 2015-19 NSF-DMREF: Deblurring our view of atomic arrangements in complex materials
- 2015-17 DARPA EQUiPS: Foundations of rigorous mathematics for UQ in large systems at extreme scale
- 2015-22 ARO-MURI: Fractional PDEs for conservation laws and beyond: theory, numerics & applications
- 2015-16 AFRL-STTR: Data infrastructure for materials genome with innovation and certification
- 2013-19 AFOSR-MURI center for Material Failure Prediction through Peridynamics

Recent funded projects @CM3

- 2020-23 NSF-DMS: Numerical analysis of asymptotically compatible discretization of nonlocal models.
- 2020-21 AWS ML award: Multi-autonomous vehicle driving policy learning for efficient & safe traffic
- 2020-25 NSF-DMS: Research Training in Applied Mathematics at Columbia University
- 2019-23 NSF-DMREF: Complex nanofeatures in crystals: theory and experiment meet in the cloud
- 2017-20 NSF DMS: Numerical analysis of SPH type methods via nonlocal models
- 2017-20 NSF CCF (TRIPODS): From foundations to practice of data science and back
- 2015-19 NSF-DMREF: Deblurring our view of atomic arrangements in complex materials
- 2015-17 DARPA EQUiPS: Foundations of rigorous mathematics for UQ in large systems at extreme scale **
- 2015-22 ARO-MURI: Fractional PDEs for conservation laws and beyond: theory, numerics & applications **
- 2015-16 AFRL-STTR: Data infrastructure for materials genome with innovation and certification
- 2013-19 AFOSR-MURI center for Material Failure Prediction through Peridynamics **

** Jointly with collaborators from other institutions and/or national labs.

Recent funded projects @CM3

- 2020-23 NSF-DMS: Numerical analysis of asymptotically compatible discretization of nonlocal models.
- 2020-21 AWS ML award: Multi-autonomous vehicle driving policy learning for efficient & safe traffic
- 2020-25 NSF-DMS: Research Training in Applied Mathematics at Columbia University
- 2019-23 NSF-DMREF: Complex nanofeatures in crystals: theory and experiment meet in the cloud
- 2017-20 NSF DMS: Numerical analysis of SPH type methods via nonlocal models
- 2017-20 NSF CCF (TRIPODS): From foundations to practice of data science and back
- 2015-19 NSF-DMREF: Deblurring our view of atomic arrangements in complex materials
- 2015-17 DARPA EQUiPS: Foundations of rigorous mathematics for UQ in large systems at extreme scale
- 2015-22 ARO-MURI: Fractional PDEs for conservation laws and beyond: theory, numerics & applications
- 2015-16 AFRL-STTR: Data infrastructure for materials genome with innovation and certification
- 2013-19 AFOSR-MURI center for Material Failure Prediction through Peridynamics

Reflecting the theme of **informative and intelligent scientific computing**.

Paradigm shift of scientific computing over the years:

- **O-O**: Operational and Optimized

  (executable, efficient, ...)

Paradigm shift of scientific computing over the years:

- **O-O**: Operational and Optimized

  (executable, efficient, ...)

- **V-V**: Validated and Verified

  (truthful, error bars, UQ, ...,)

Paradigm shift of scientific computing over the years:

- **O-O**: Operational and Optimized

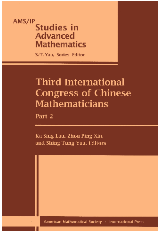
  (executable, efficient, ...)

- **V-V**: Validated and Verified

  (truthful, error bars, UQ, ...,)

- **I - I**: Informative and Intelligient

  (smart algorithms, knowledge extraction/discovery, ...)



Informative and intelligent scientific computing
(Du: ICCM2004, published 2008, 731-748)

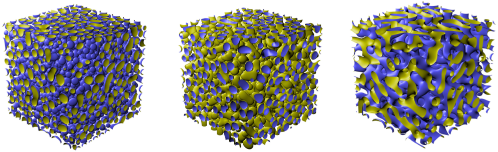
Algorithmic development for traditional scientific computing

- Operational & Optimized, physics/model-based simulations, from Tera, Peta, to Exa-scale, ...

[1] J.Zhang-C.Zhou-Y.Wang-L.Ju-Q.Du-X.Chi-D.Xu-D.Chen-Y.Liu-Z.Liu, SC16, IEEE 2016

Algorithmic development for traditional scientific computing

- Operational & Optimized, physics/model-based simulations, from Tera, Peta, to Exa-scale, ...
- Eg. *Record-breaking* phase field simulations of microstructure evolution, from nucleation to coarsening, at extreme-scale for Gordon-Bell competition: with peak over 50 PFlops, 21.5 trillion Grid-Points, 10 million Processing-Elements.



(3D microstructure evolution involving complex topological changes[1])

- Algorithmic development: high order discretization, localization, domain decomposition ⇒ **scalable**

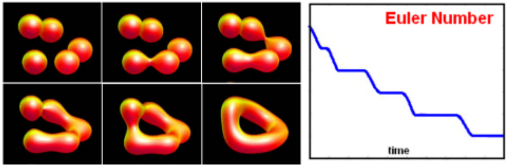
[1] J.Zhang-C.Zhou-Y.Wang-L.Ju-Q.Du-X.Chi-D.Xu-D.Chen-Y.Liu-Z.Liu, SC16, IEEE 2016

Algorithmic development for informative and intelligent scientific computing

Keys for I^{2}SC: extracting hidden information, learning from simulation/data, smart computing.

Illustrative examples (from the past):

- Extracting hidden topology of simulated microstructure with implicitly represented interfaces



Reverse thinking:

algorithms for computing Euler number to learn/detect topology change (Du et al 2005)

- CAMLET (Combined Ab-initio and Manifold LEarning Toolbox): NSF-CCF (CS/Math/Phys/Stat) 2004-07; smart computing with surrogate models learned from database of ab-initio simulations.

- Shrinking dimer dynamics (SDD) + KMC for transition state search and nucleation/coarsening.

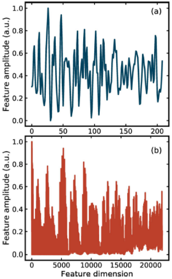
Algorithmic development for informative and intelligent scientific computing

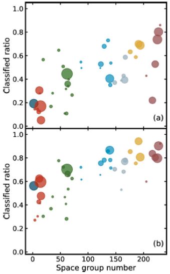
Keys for I^{2}SC: extracting hidden information, learning from simulation/data, smart computing.

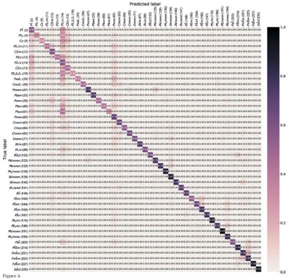
Illustrative examples (more recent ones):

- Nanostructure inverse problem (NSF-DMREF, with Billinge group, 2015-).

  Extracting materials structure information from x-ray data (PDF) with optimization and deep learning. E.g. group symmetry classification using CNN. (Liu et. al. 2019, Acta Crys. A)





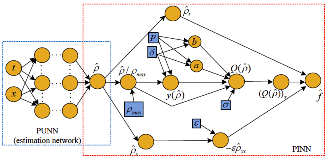


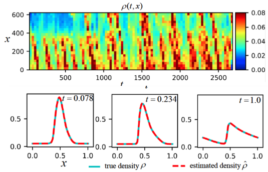
Algorithmic development for informative and intelligent scientific computing

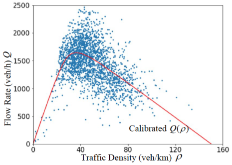
Keys for I^{2}SC: extracting hidden information, learning from simulation/data, smart computing.

Illustrative examples (more recent ones):

- Physics-informed deep learning of traffic state estimation (with X. Di of CEEM Shi et al, 2020)







- Autonomous experiment design with meta-modeling and machine learning:
  - exploring structure-synthesis-property relation for materials discovery,
  - cooperative and non-cooperative games and reinforcement learning for constitutive modeling, (e.g., work with S. Sun group of CEEM, Wang et. al. 2020).

Physics-based and data-driven modeling and simulations

Advent of computing technology

$+$

Access to large amount data

$+$

Improved machine learning

$\Downarrow$

Physics-based/data-driven modeling

$\Rightarrow$ Capturing the essence of I^2SC:

Iteratively solve forward+inverse problems.

Physics-based and data-driven modeling and simulations

Advent of computing technology

$+$

Access to large amount data

$+$

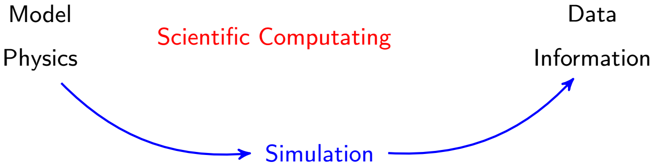
Improved machine learning

$\Downarrow$

Physics-based/data-driven modeling

$\Rightarrow$ Capturing the essence of I^{2}SC:

Iteratively solve forward+inverse problems.



Physics-based and data-driven modeling and simulations

Advent of computing technology

$+$

Access to large amount data

$+$

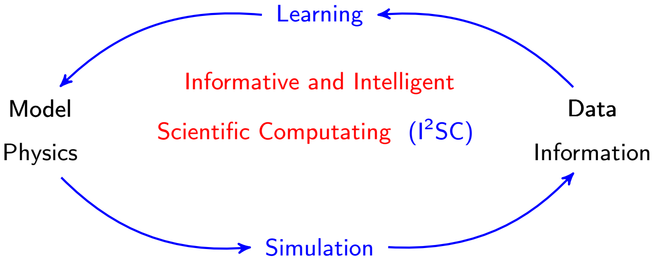
Improved machine learning

$\Downarrow$

Physics-based/data-driven modeling

$\Rightarrow$ Capturing the essence of I²SC:

Iteratively solve forward+inverse problems.



Physics-based and data-driven modeling and simulations

Advent of computing technology

$+$

Access to large amount data

$+$

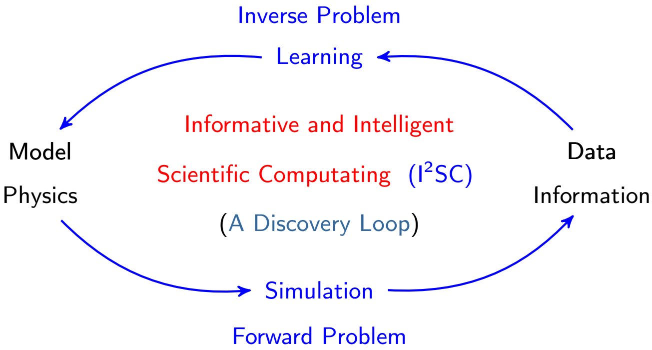
Improved machine learning

$\Downarrow$

Physics-based/data-driven modeling

$\Rightarrow$ Capturing the essence of I^{2}SC:

Iteratively solve forward+inverse problems.

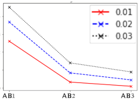


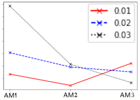
Informative and Intelligent Scientific Computing (I^2SC): Forward + Inverse

Eg: Integrating given dynamics (forward) and learning unknown dynamics (inverse).

⋆ Learning/predicting dynamics from video sequences (jointly with Hod Lipson group)

⋆ Numerical integrator: good (stable) for given dynamics $\neq$ good for dynamics discovery.





New stability/consistent theory for learning with LMMs (2020)

I^2SC $\Leftrightarrow$ computing and learning $\Leftrightarrow$ knowledge discovery (robust/predictive/trustworthy/automated)

Learning$\neq$black box $\Rightarrow$ new challenges to algorithmic development: adaptive/multiscale computation, error/uncertainty quantification, dimension reduction (thus modeling nonlocal interactions), ...

Thank you!