Model-Free Data-Driven Science: Cutting out the Middleman

Michael Ortiz
California Institute of Technology

Collaborators:
T. Kirchdoerfer, K. Karapiperis, J. Andrade (Caltech)
L. Stainier, J. Réthoré, A. Leygue (Central Nantes)
S. Conti, S. Müller (Bonn University)
R. Eggersmann, S. Reese (RWTH Aachen)

Mechanics Colloquium – February 19, 2021
Department of Engineering, Cambridge University

Data Science, Big Data, AI…
What's in it for us?

http://olap.com/forget-big-data-lets-talk-about-all-data/

Michael Ortiz
CAMBRIDGE 2021

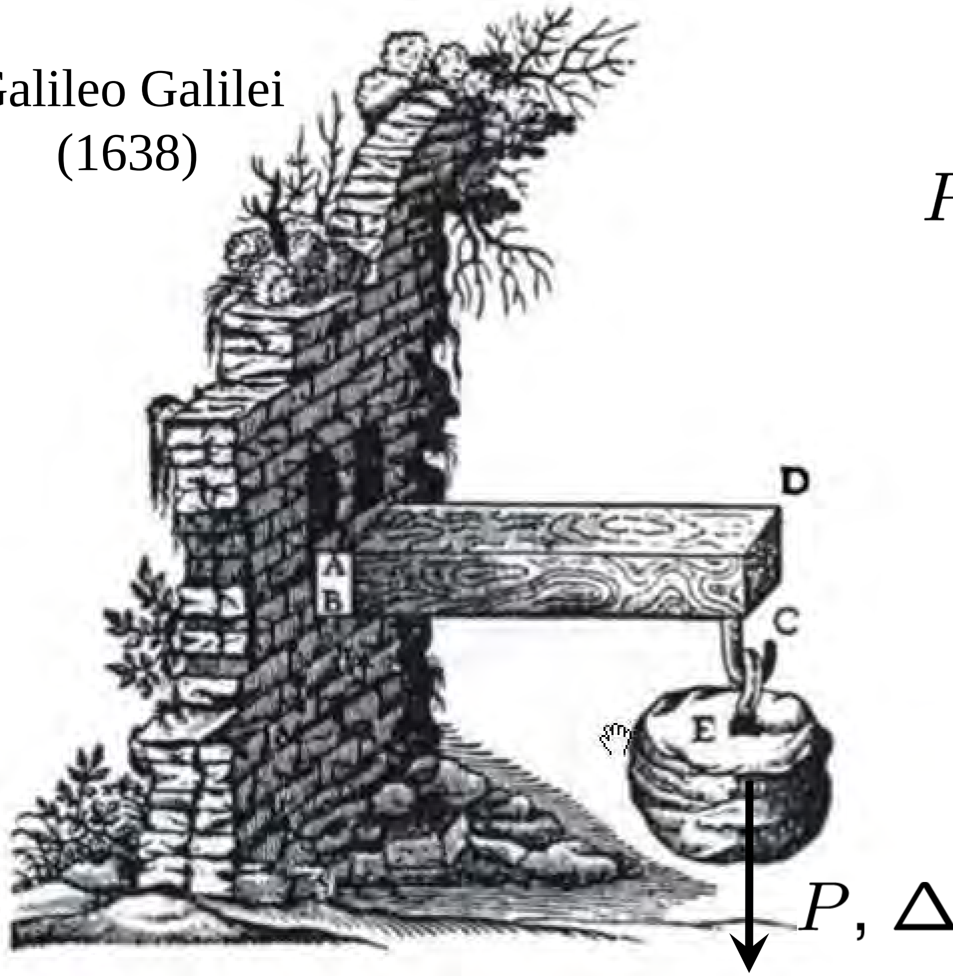
Data-Driven Science: Talking points

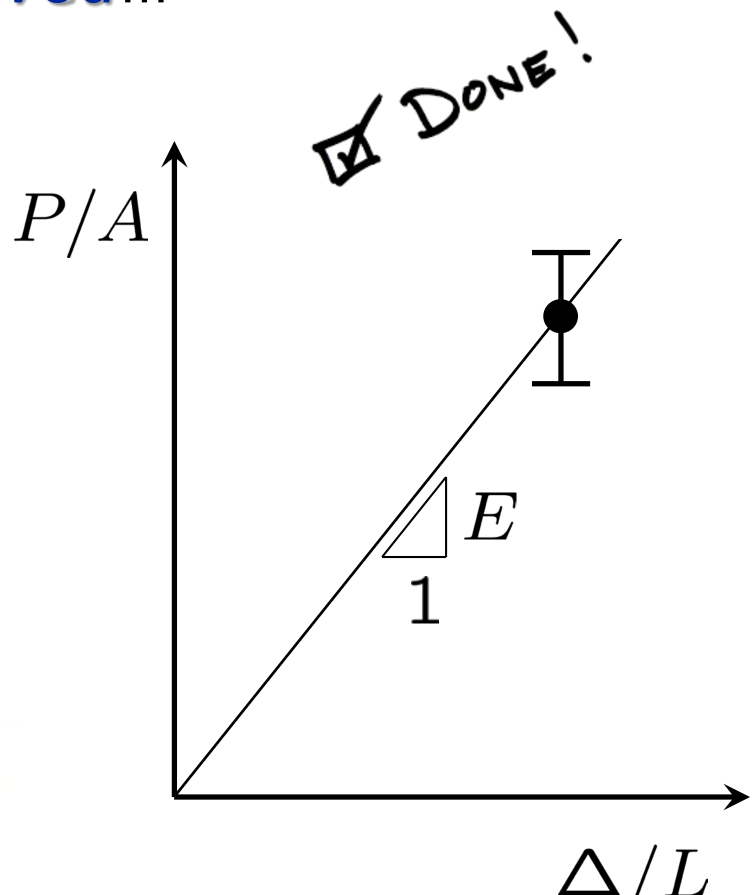
- New emerging paradigm: *Data-Driven Science*
- Why now? *What has changed*?
- How does Data Science intersect with the *physical sciences*? With *experimental science*?
- Is there *intellectual depth* to Data-Driven Science? Are there *opportunities* for fundamental, enduring, contributions?
- Data-Driven Science: *Theory vs. practice*
- Is Data-Driven Science likely to change *engineering practice*? Industry?
- Data-Driven ecosystem, *infrastructure*...

Materials data through the ages…

Traditionally, mechanics of materials has been *data starved*...

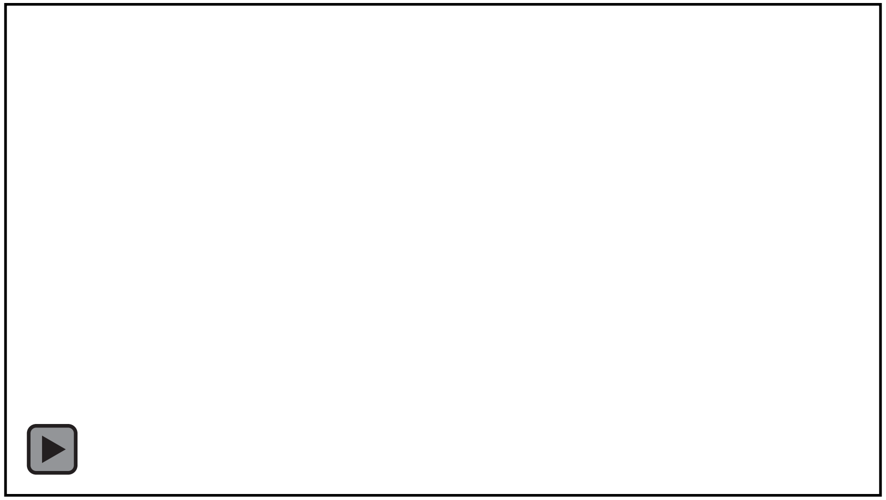
Galileo Galilei
(1638)



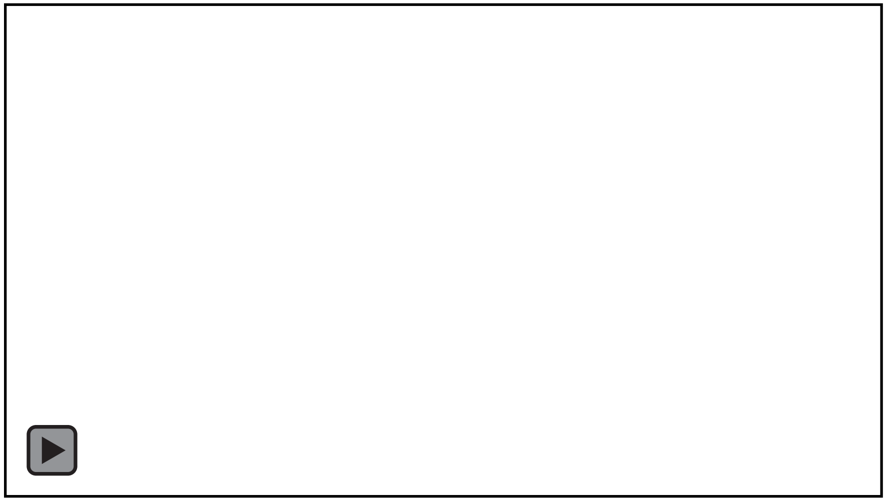


The new data-rich world…

- Material *data is currently plentiful* due to dramatic advances in experimental science (DIC, EBSD, microscopy, tomography…) and multiscale computing (DFT $\rightarrow$ MD $\rightarrow$ DDD $\rightarrow$ SM $\rightarrow$ Hom)



3D tomographic reconstruction of particles in battery electrode

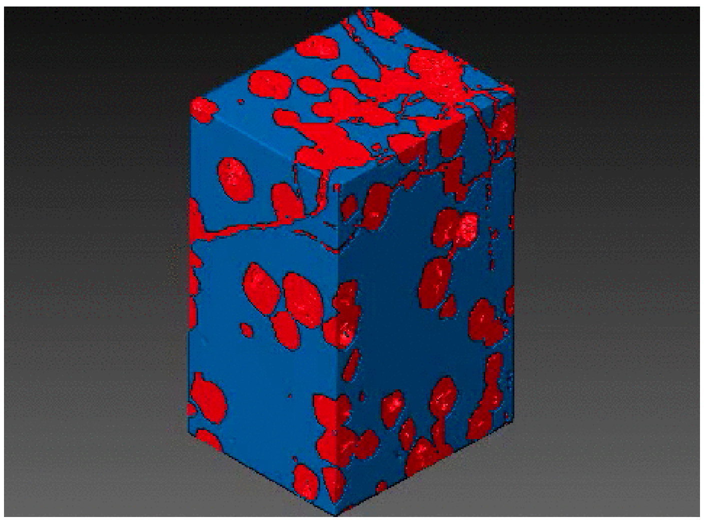


3D DIC-measured internal-strain full-field compressed PDMS sample

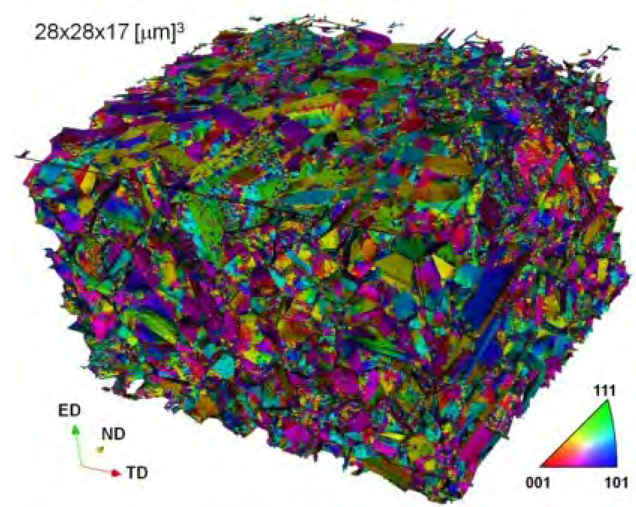
John Lambros, UIUC, https://lambros.ae.illinois.edu/moviesimages/

The new data-rich world…

- Material *data is currently plentiful* due to dramatic advances in experimental science (DIC, EBSD, microscopy, tomography…) and multiscale computing (DFT → MD → DDD → SM → Hom)



Two-phase μCT analysis of Ti2AlC/Al composite[1]



3D EBSD microstructure in Cu-0.17wt%Zr after ECAP[2]

[1]Hanaor *etal*, *Mater Sci Eng A*, **672** (2019) 247.
[2]Khorashadizadeh, *Adv Eng Mater*, **13** (2011) 237.

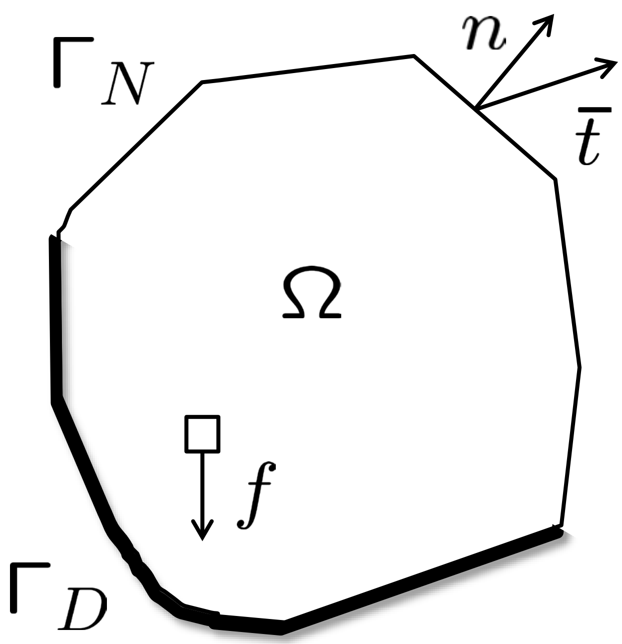
Data Science, Big Data…

- *Data Science* is the extraction of '*knowledge*' from large volumes of unstructured data[1]
- Data science requires sorting through *big-data* sets and extracting '*insights*' from these data
- Data science uses data management, statistics and machine learning to derive *mathematical models* for subsequent use in decision making
- Data science influences (*non-STEM*) fields such as marketing, advertising, finance, social sciences, security, policy, medical informatics...
- *But... The field theories of science have distinct math structure which must be accounted for!*

[1]Dhar, V., *Communications of the ACM*, **56**(12) (2013) p. 64.

Differential structure of field equations

- Anatomy of a ***field-theoretical*** STEM problem:



i) Kinematics + Dirichlet BC:

$$\left.\begin{aligned} &\epsilon(u) = 1/2(\nabla u + \nabla u^T) \\ &u = \bar{u}, \quad \text{on } \Gamma_D \end{aligned}\right\}$$

ii) Equilibrium + Neumann BC:

$$\left.\begin{aligned} &\operatorname{div} \sigma + f = 0 \\ &\sigma\, n = \bar{t}, \quad \text{on } \Gamma_N \end{aligned}\right\}$$

iii) ***Material law:*** $\sigma(x) = \sigma(\epsilon(x))$

Differential structure of field equations

- Anatomy of a *field-theoretical* STEM problem:

Universal laws!
Material independent!
Exactly known!
Uncertainty-free!
Similarly:
Maxwell, Einstein, Schrödinger
Mass transport...

i) Kinematics + Dirichlet BC:

$$\left.\begin{aligned} &\epsilon(u) = 1/2(\nabla u + \nabla u^T) \\ &u = \bar{u}, \quad \text{on } \Gamma_D \end{aligned}\right\} \delta!$$

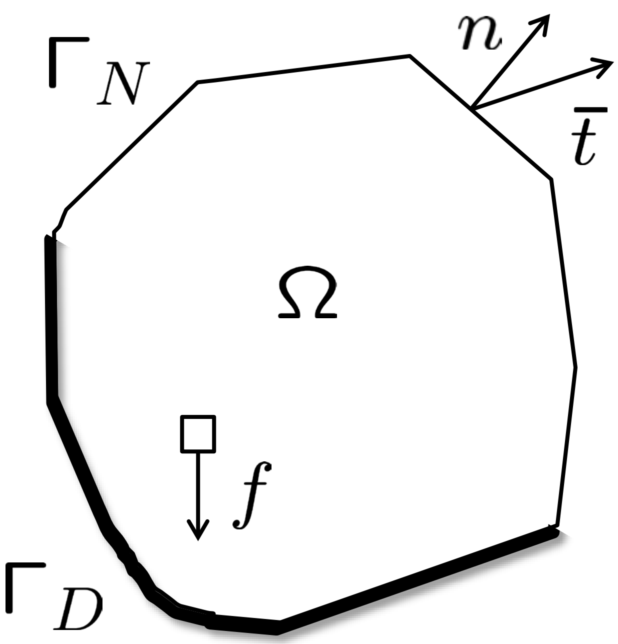
ii) Equilibrium + Neumann BC:

$$\left.\begin{aligned} &\operatorname{div}\sigma + f = 0 \\ &\sigma\, n = \bar{t}, \quad \text{on } \Gamma_N \end{aligned}\right\} \partial!$$

iii) *Material law:* $\sigma(x) = \sigma(\epsilon(x))$

Differential structure of field equations

- Anatomy of a *field-theoretical* STEM problem:



i) Kinematics + Dirichlet BC:

$$\left.\begin{array}{l} \epsilon(u) = 1/2(\nabla u + \nabla u^T) \\ u = \bar{u}, \quad \text{on } \Gamma_D \end{array}\right\}$$

ii) Equilibrium + Neumann BC:

$$\left.\begin{array}{l} \operatorname{div}\sigma + f = 0 \\ \sigma n = \bar{t}, \quad \text{on } \Gamma_N \end{array}\right\}$$

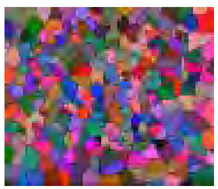
$$\underbrace{\text{iii) \textit{Material law:}} \quad \sigma(x) = \sigma(\epsilon(x))}_{\text{Unknown! Epistemic uncertainty!}}$$

Michael Ortiz
CAMBRIDGE 2021

Adapting to a new data-rich world...

Classical Model-Based Computational Science...

Modern microscopy generates massive data sets



Material data

Modeling funnel

$$\boxed{\sigma = \mathbb{C}\epsilon}$$

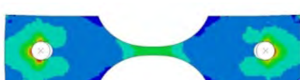
Material model

funnel

Simulation

Manufactured data

Modelling entails massive loss of information!



Adapting to a new data-rich world...

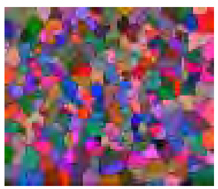
- Modeling entails *massive loss of information* from material data sets, inserts *uncertainty*
- Modeling is *ad hoc, open ended, ill-posed*
- There is *no theory* that determines models from first principles to a desired level of accuracy
- Modeling requires *heuristics and intuition*: Models are *only as good as the modeler*'s physical intuition, empirical knowledge
- *Machine learning* requires massive amounts of *ad hoc* modeling, regression, in practice
- Can a more *direct connection between data and prediction* be forged? (*the data, all the data, nothing but the data*) How?

Adapting to a new data-rich world...

Classical Model-Based Computational Science...

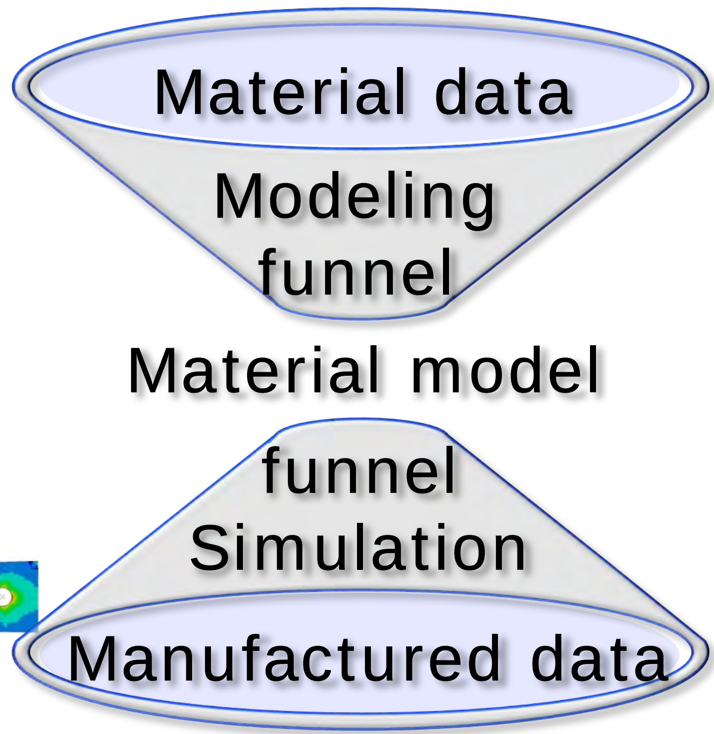
Alternative: Model-Free Data-Driven Computing!

Modern microscopy generates massive data sets



$$\sigma = \mathbb{C}\epsilon$$

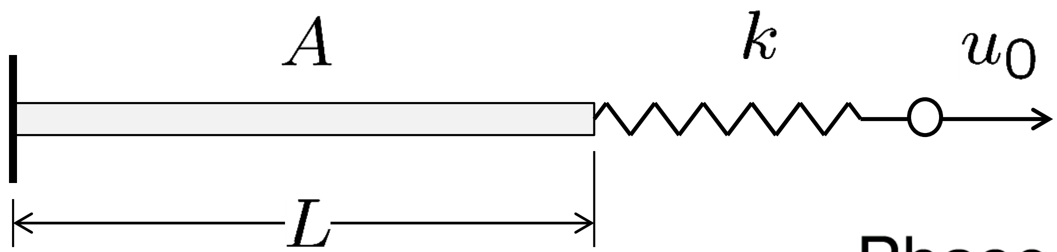
Modelling entails massive loss of information!

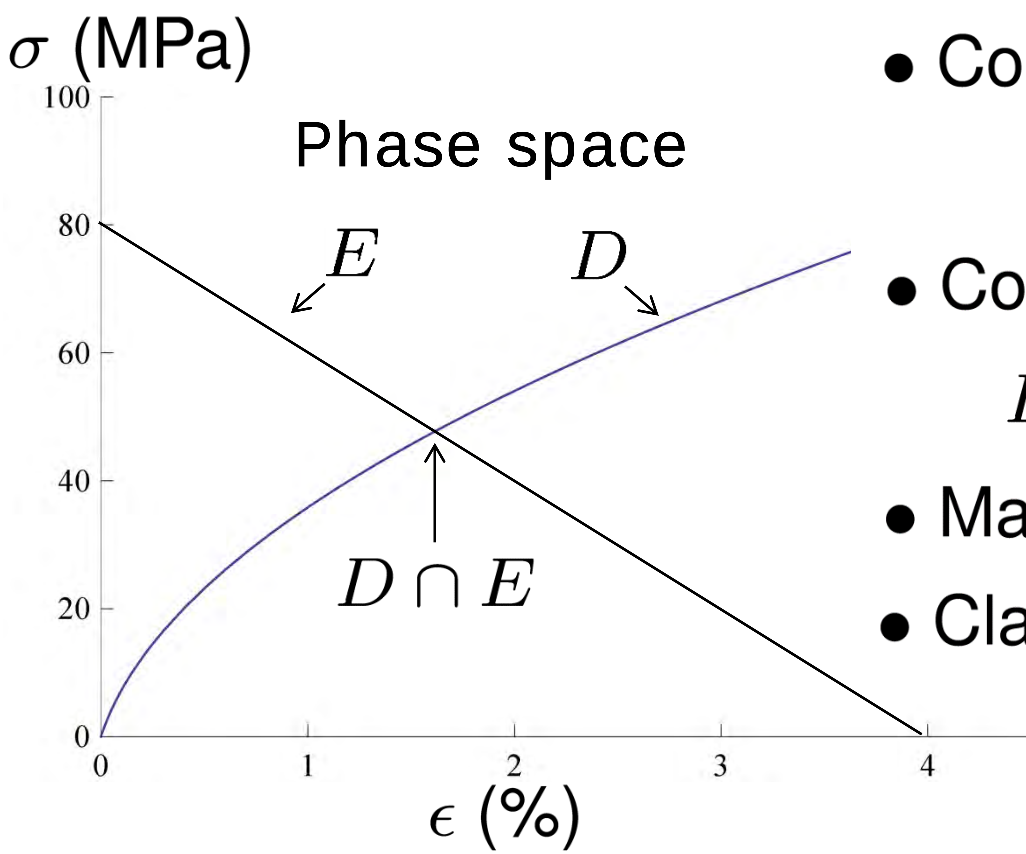


Set/solve problems directly from data!

Eliminate modeling bottleneck!

Elementary example: Bar and spring





- Phase space: $\{(\epsilon, \sigma)\} \equiv Z$
- Compatibility + equilibrium:

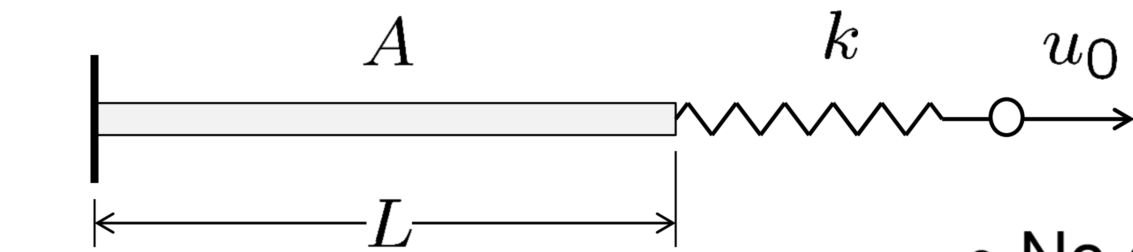
$$\sigma A = k(u_0 - \epsilon L)$$

- Constraint set:

$$E = \{\sigma A = k(u_0 - \epsilon L)\}$$

- Material data set: $D \subset Z$
- Classical solution set: $D \cap E$

Elementary example: Bar and spring

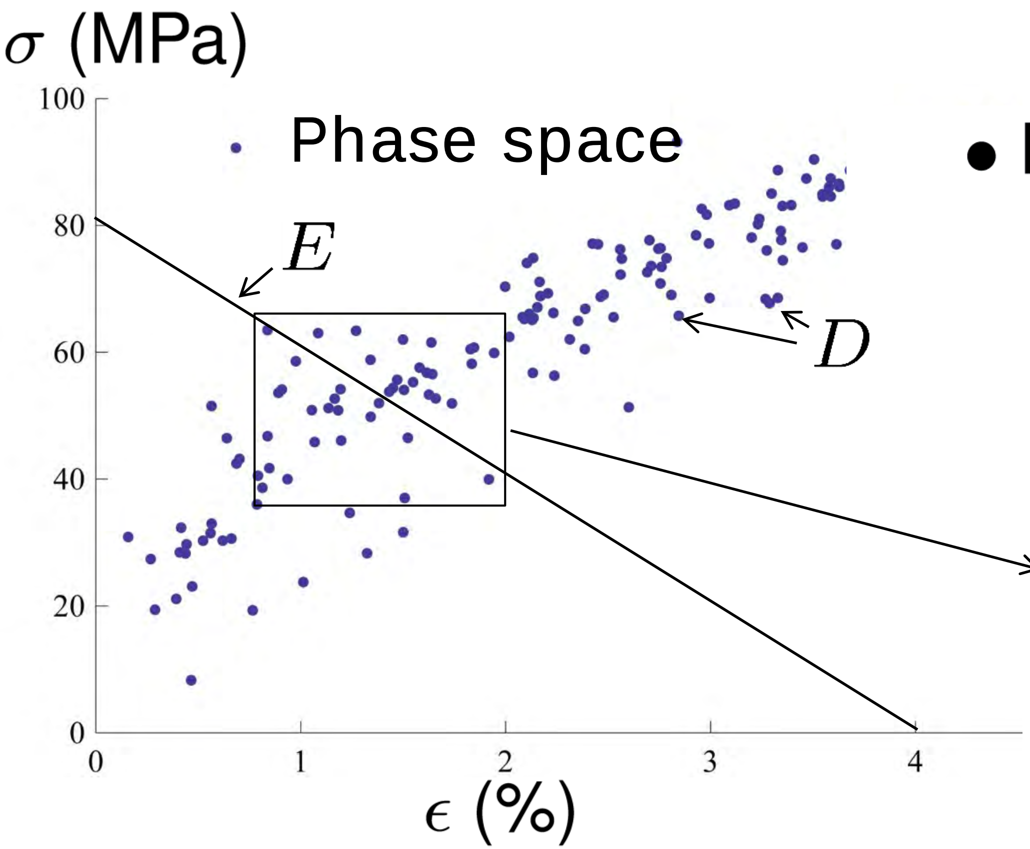


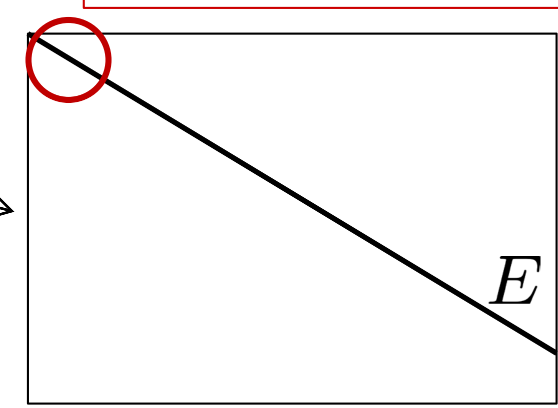
- No classical solutions!

$$D \cap E = \emptyset$$

- Data-driven solution:

$$\boxed{\min_{z \in E} \operatorname{dist}(z, D)}$$





The general Data-Driven (DD) problem

- The *Data-Driven paradigm*[1]: Given,
  - D = {*fundamental material data*},
  - E = {*compatibility + equilibrium*},

  Find: $\text{argmin}\{d(z, D),\ z \in E\}$
- *The aim of Data-Driven analysis is to find the admissible state (compatibility and equilibrium) closest to the material data set*
- Raw *fundamental material data* (stress & strain) is used (unprocessed) in calculations
- *No material modeling, no loss of information (the data, all the data, nothing but the data)*

[1]T. Kirchdoerfer and M. Ortiz (2015) arXiv:1510.04232.
[1]T. Kirchdoerfer and M. Ortiz, *CMAME*, **304** (2016) 81–101

Data-Driven elasticity – Well-posedness

Definition (Constraint set)

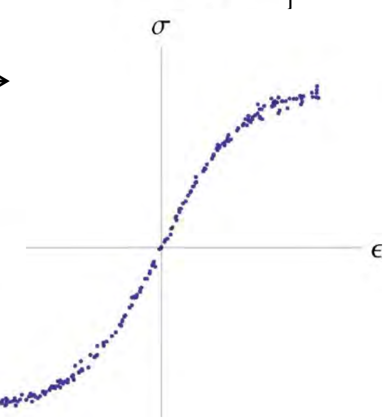
ii) Equilibrium,

$$\left.\begin{aligned} &\mathrm{div}\sigma + f = 0, \\ &\sigma\nu = h, \quad \text{on } \Gamma_N. \end{aligned}\right\} E$$

Definition (Material data set)

Hooke's law (linear) $D = \{\sigma = \mathbb{C}\epsilon\}$.

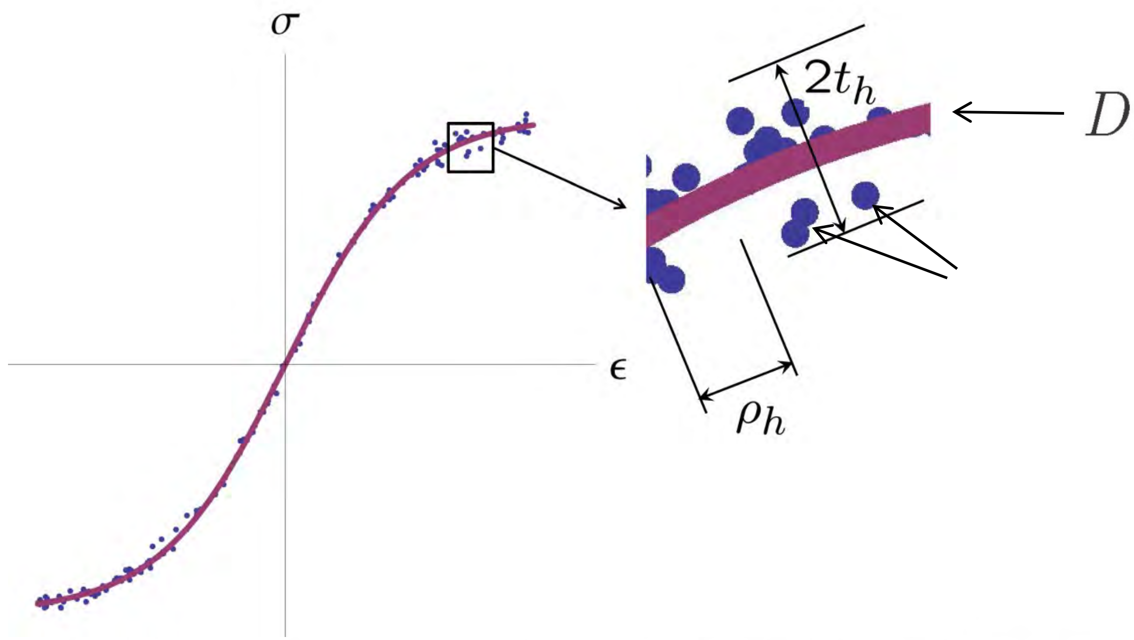
Hooke's law (monotone) $D = \{\sigma = \sigma(\epsilon)\}$.



$$\min\{d(z, D),\ z \in E\}$$

Conti, S., Müller, S. & Ortiz, M., *ARMA*, **229** (2018) 79-123.

Data-Driven elasticity – Δ-convergence



Theorem

Suppose D *monotone graph,* $\rho_h \downarrow 0$ *and* $t_h \downarrow 0$ *such that:*

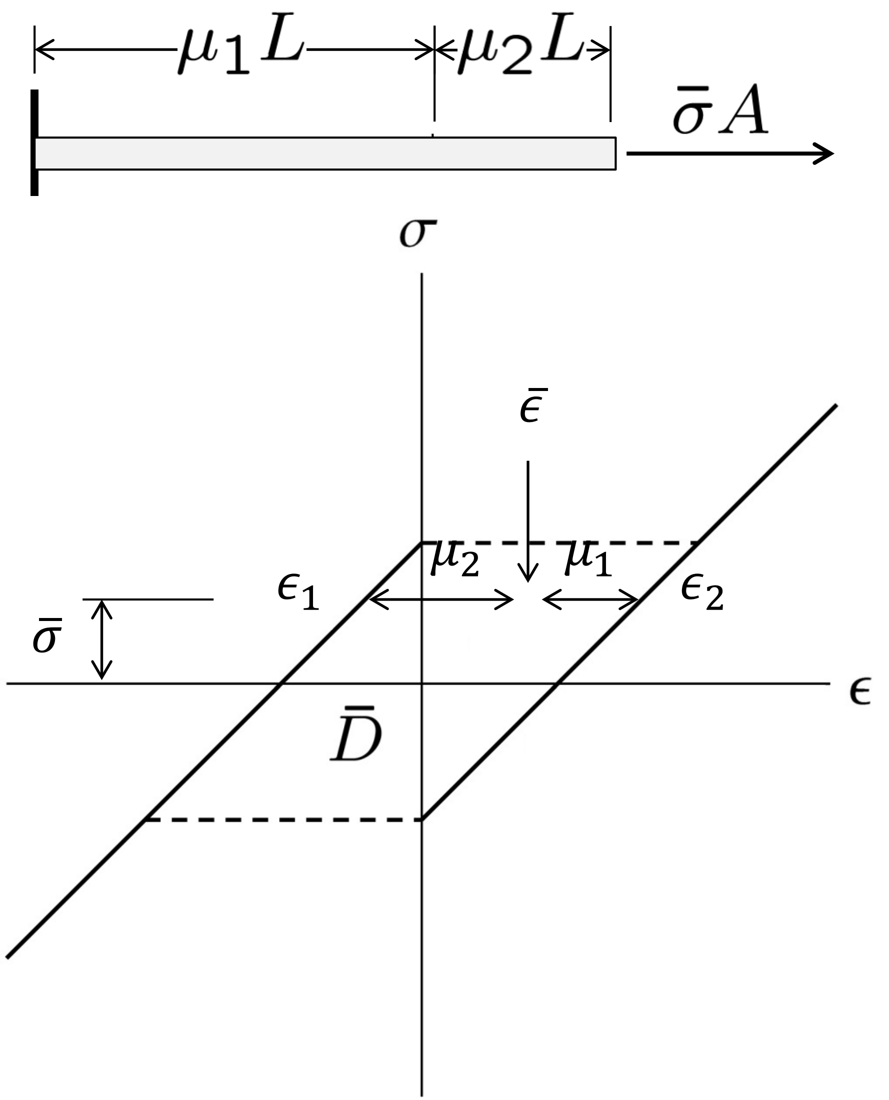
i) *Fine approximation:* $d(\xi, D_h) \leq \rho_h, \forall \xi \in D$.

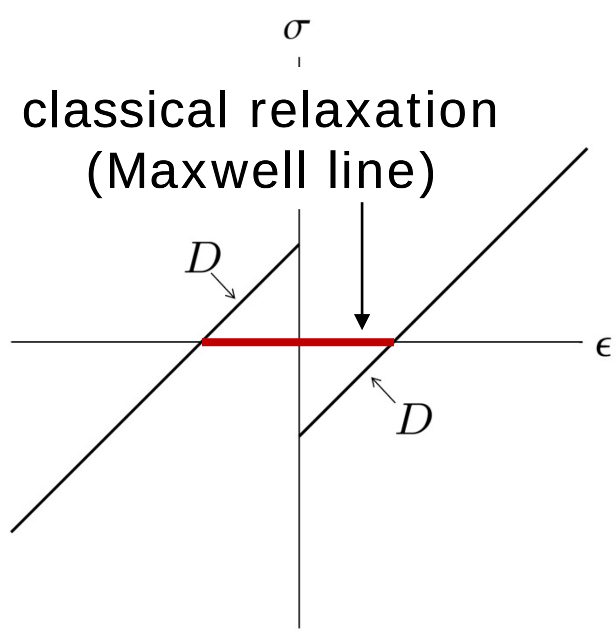
ii) *Uniform approximation:* $d(\xi, D) \leq t_h, \forall \xi \in D_h$.

Then, $(\epsilon_h, \sigma_h) \to (\epsilon, \sigma)$.

Conti, S., Müller, S. & Ortiz, M., *ARMA*, **229** (2018) 79-123.

Data-Driven elasticity - Relaxation

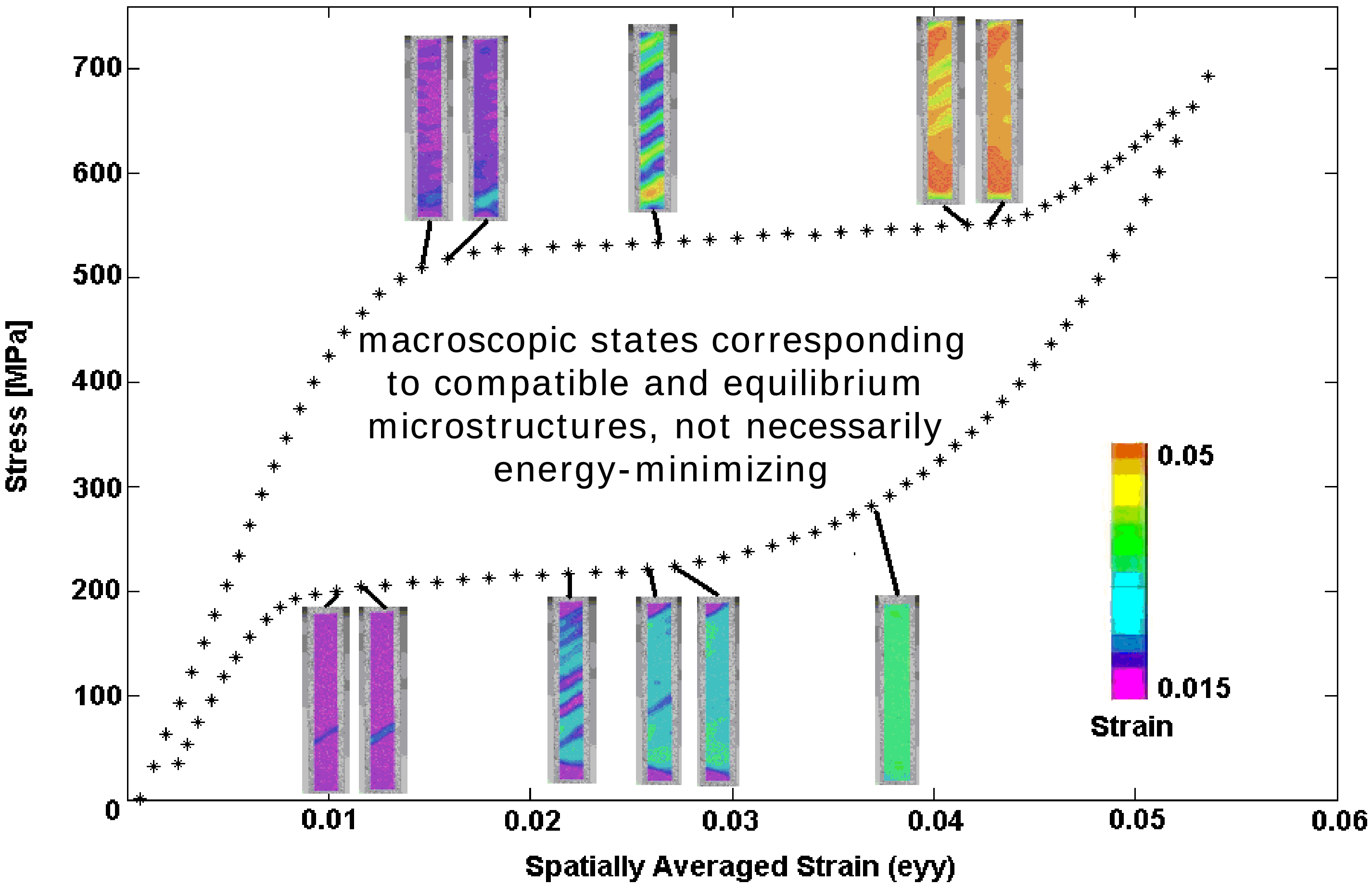


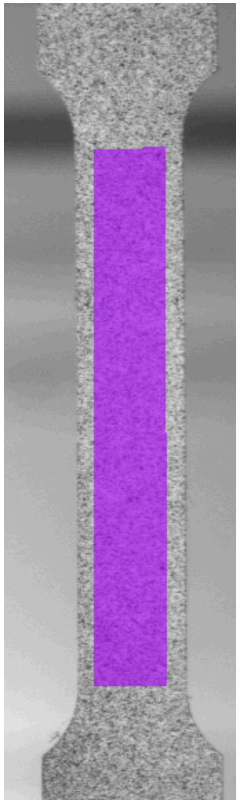


Michael Ortiz
CAMBRIDGE 2021

Conti, S., Müller, S. & Ortiz, M., *ARMA*, **229** (2018) 79-123.

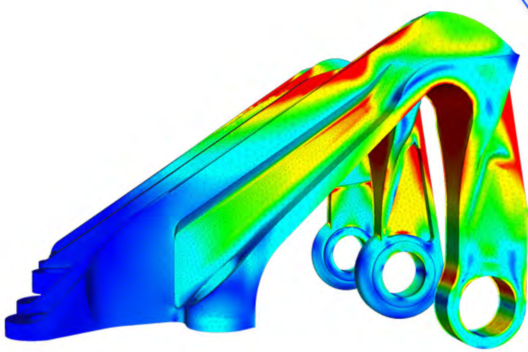
Phase transformation in nitinol (2D DIC)

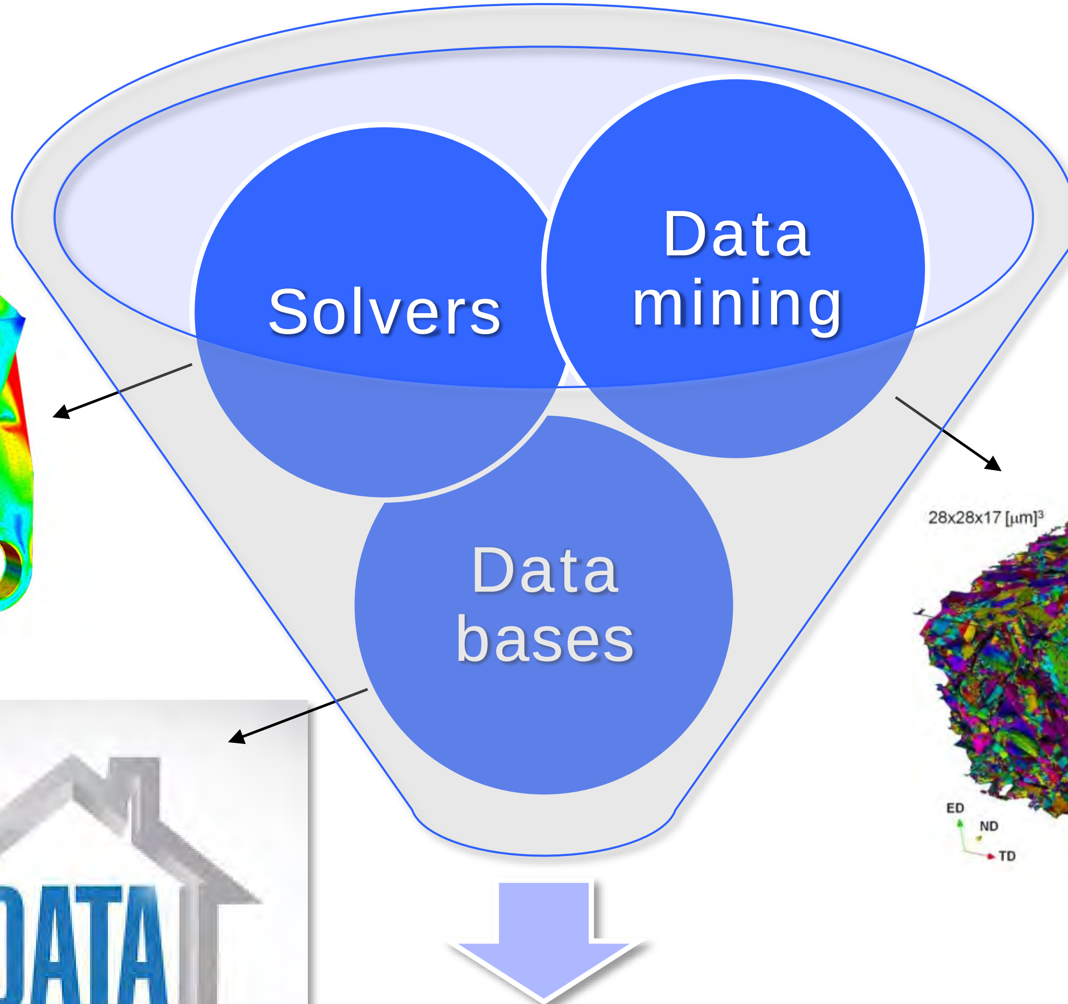


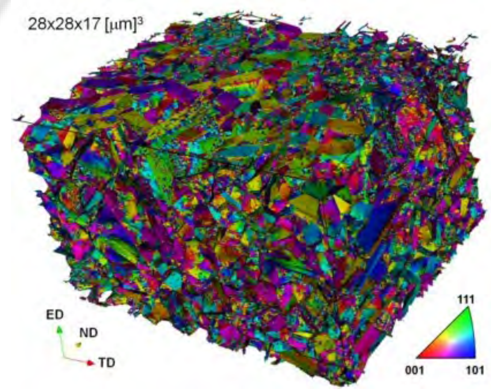


S. Daly, G. Ravichandran and K. Bhattacharya, *Acta Mater.*, **55** (2007) 3593.

The (model-free) DD ecosystem…

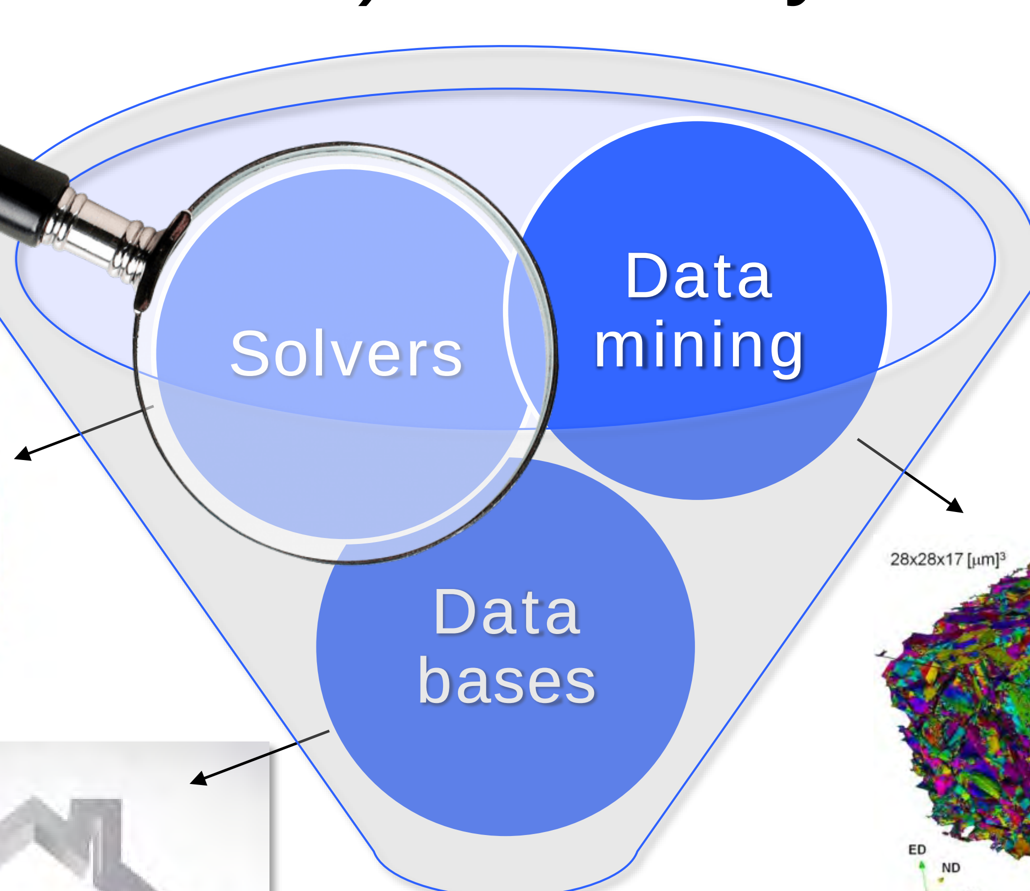


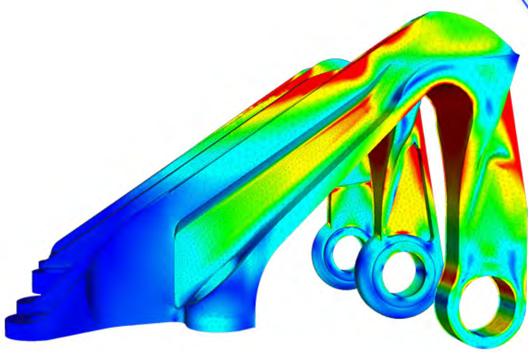


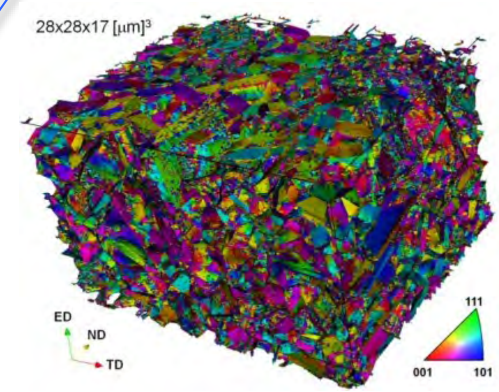


DD

The (model-free) DD ecosystem…



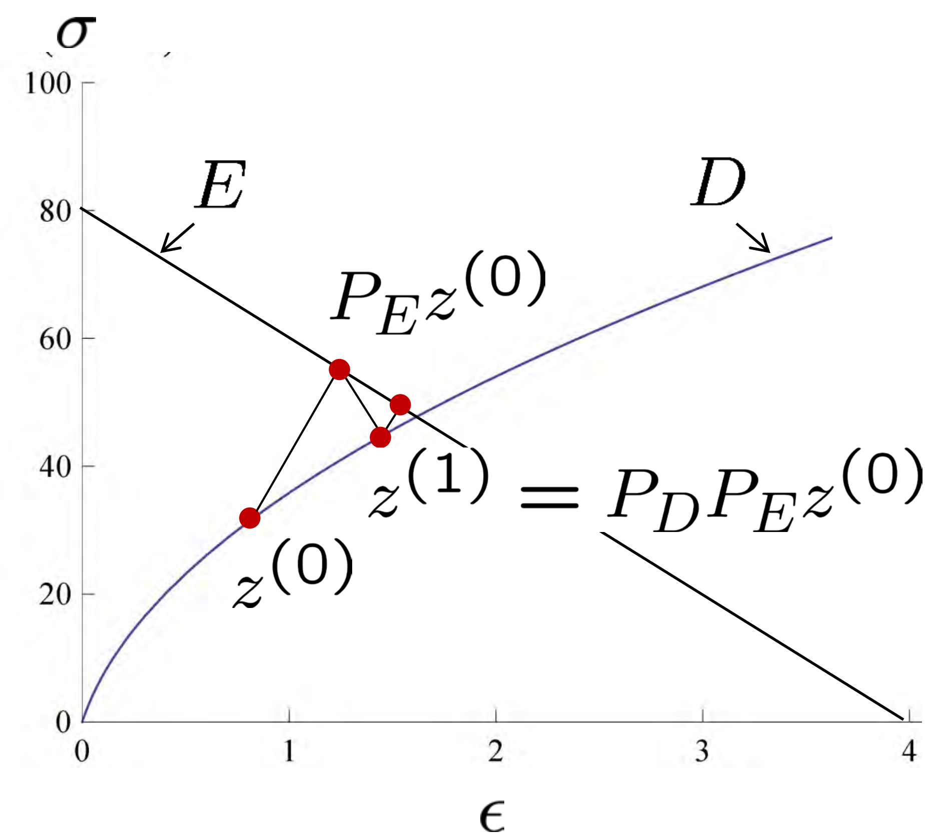


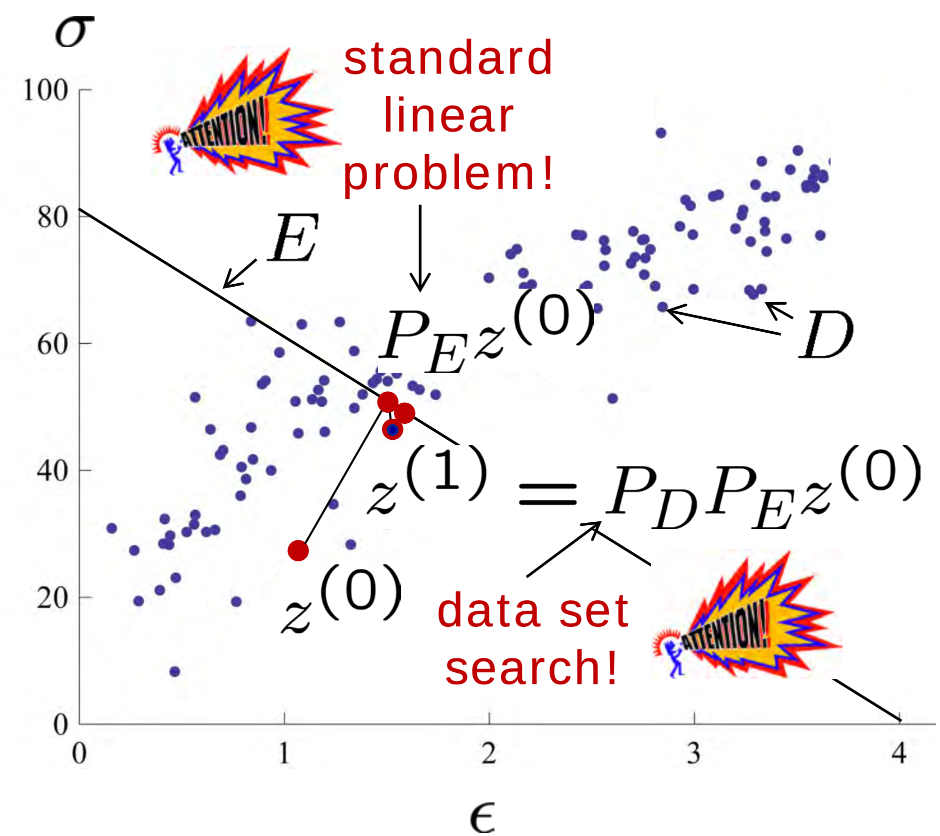


DD

DD solvers: Fixed-point iteration

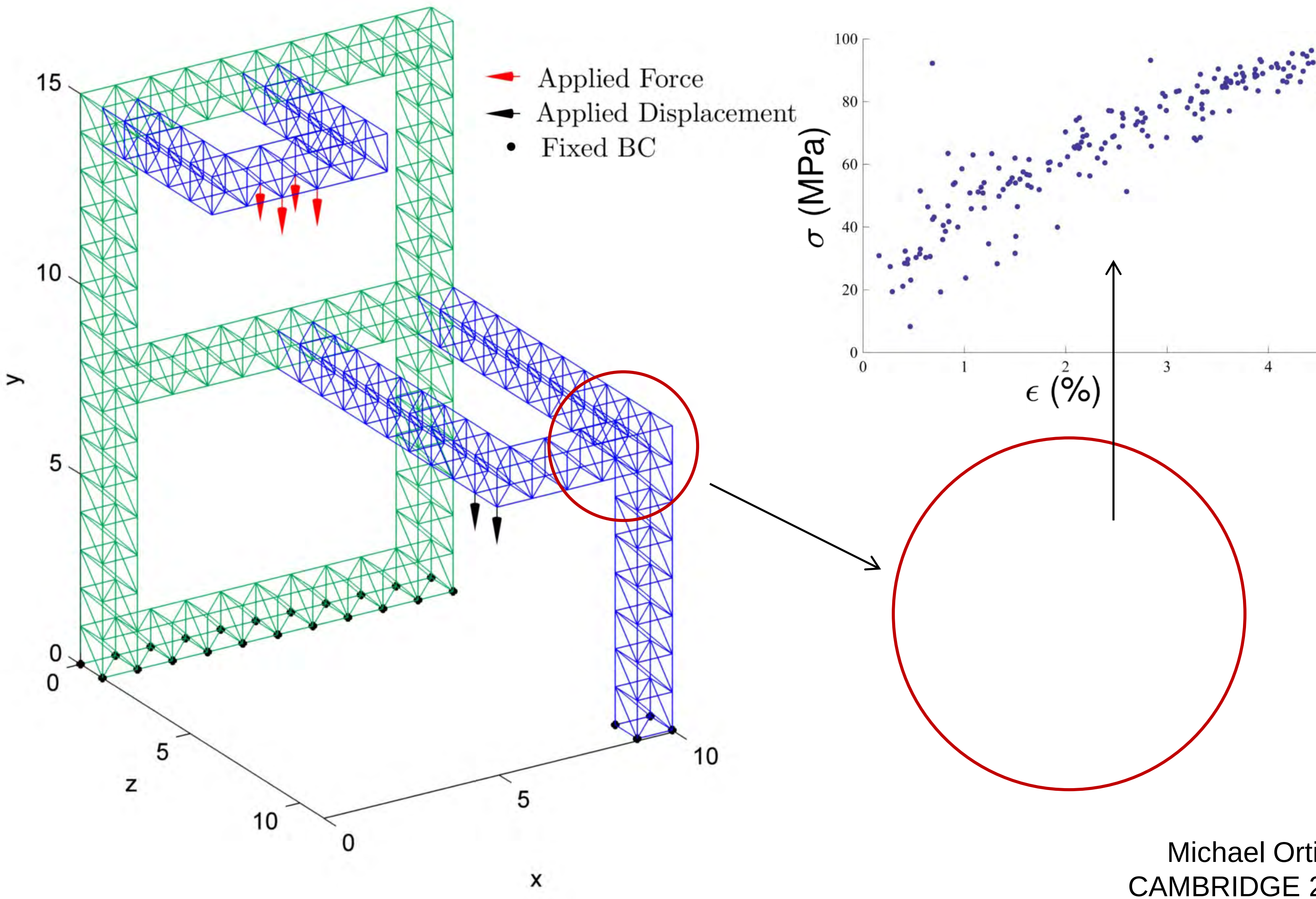
- Find: $\operatorname{argmin}\{d(z, D),\ z \in E\}$
- Fixed-point iteration[1]: $z^{(k+1)} = P_D P_E z^{(k)}$





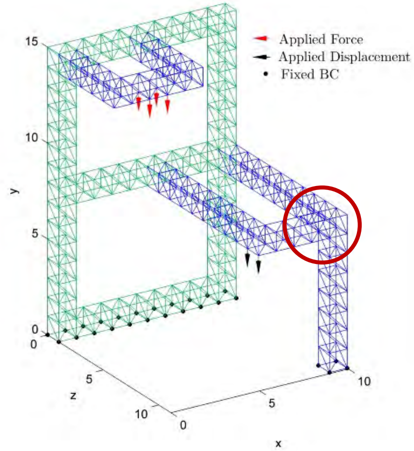
[1]T. Kirchdoerfer and M. Ortiz (2015) arXiv:1510.04232.
[1]T. Kirchdoerfer and M. Ortiz, *CMAME*, **304** (2016) 81–101

Example: 3D Truss



Michael Ortiz
CAMBRIDGE 2021

Example: 3D Truss

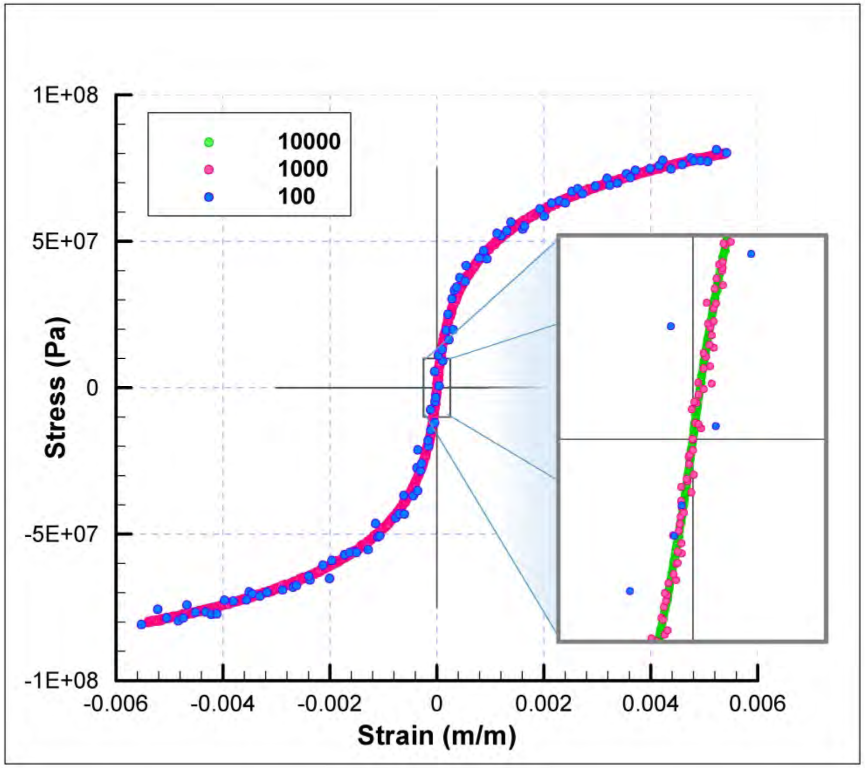


- Degrees of freedom: $(u_i)_{i=1}^{n}$
- Phase space: $Z = \{(\epsilon_e, \sigma_e)_{e=1}^{m}\}$
- Norm: $\|(\epsilon, \sigma)\|^2 = \sum_{e=1}^{m} w_e \left(\mathbb{C}\epsilon_e^2 + \mathbb{C}^{-1}\sigma_e^2\right)$
- Constraint set: $E = \{\epsilon = Bu, \ \ B^T\sigma = f\}$
- DD problem: $\min\limits_{(\epsilon', \sigma') \in D} \left(\min\limits_{(\epsilon, \sigma) \in E} \|(\epsilon - \epsilon', \sigma - \sigma')\|^2 \right)$
- DD iteration: $\epsilon^{(k)} = B(B^T\mathbb{C}B)^{-1}B^T\mathbb{C}\epsilon^{*(k)}$

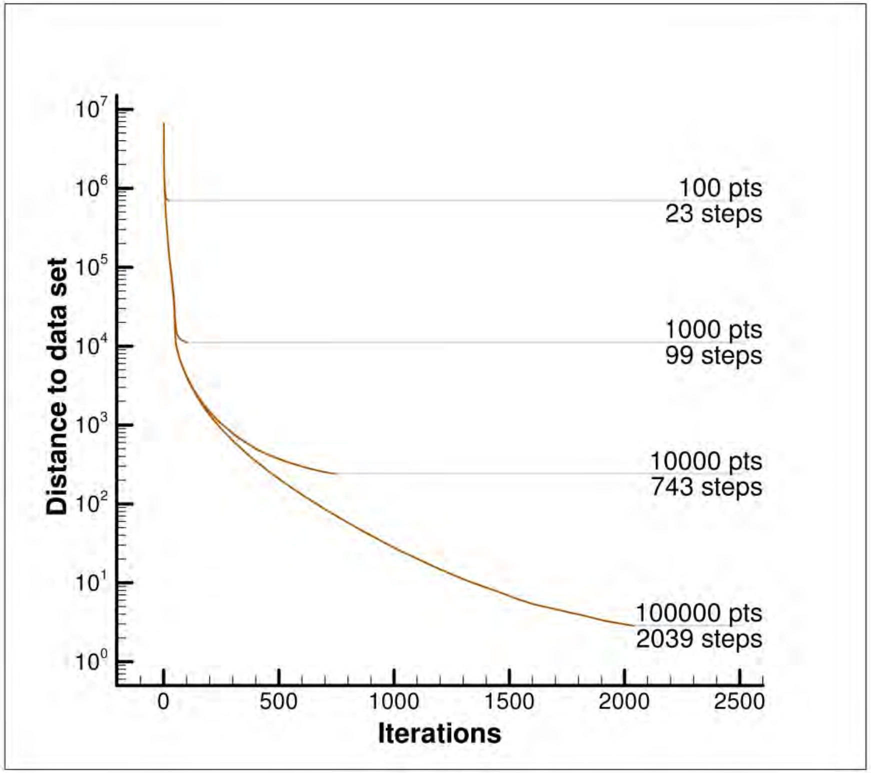
$$\sigma^{(k)} = \sigma^{*(k)} + \mathbb{C}B\underbrace{(B^T\mathbb{C}B)^{-1}}(f - B^T\sigma^{*(k)})$$

standard stiffness matrix!

Convergence of fixed-point solver



Radomized material-data sets of increasing size

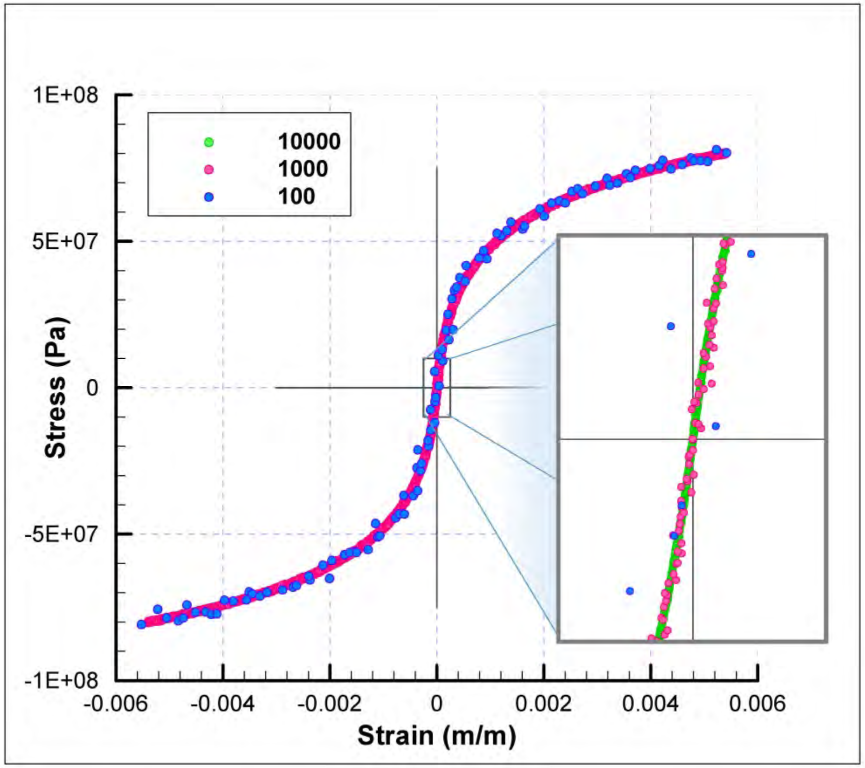


Convergence of fixed-point iteration

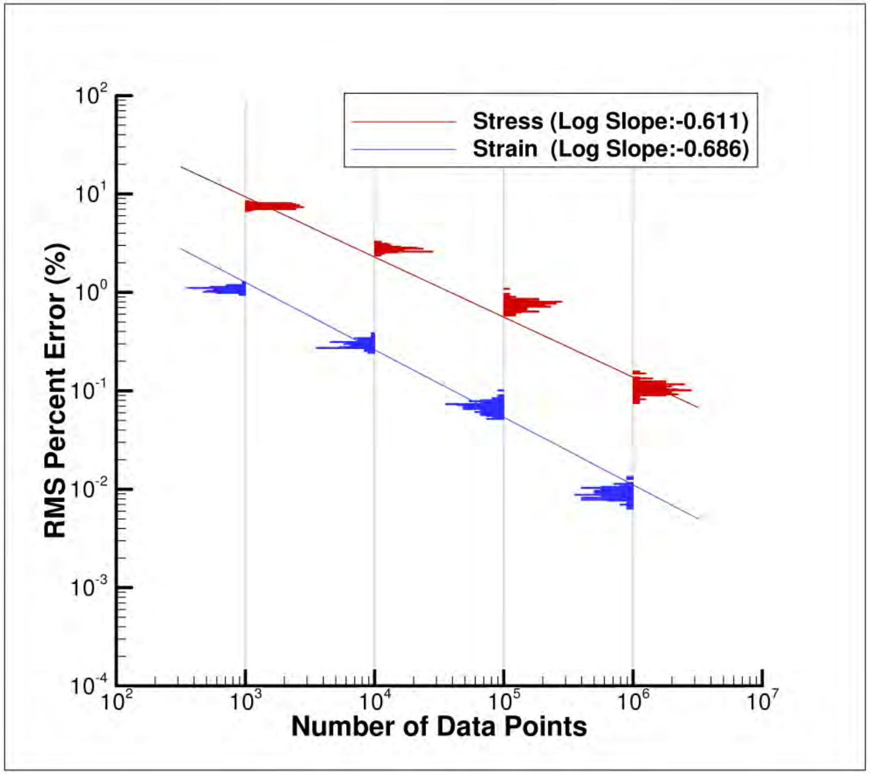
T. Kirchdoerfer and M. Ortiz, *CMAME*, **304** (2016) 81–101.

Michael Ortiz
CAMBRIDGE 2021

Convergence of fixed-point solver



Randomized material-data sets of increasing size and decreasing scatter



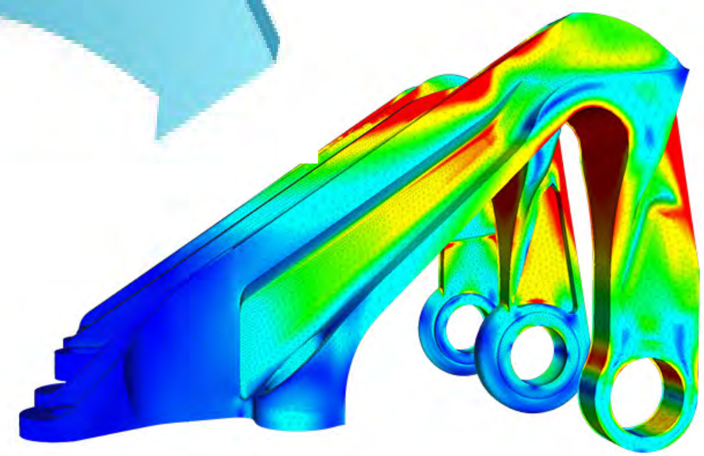
Convergence with respect to data set

T. Kirchdoerfer and M. Ortiz, *CMAME*, **304** (2016) 81–101.

The DD information flow

Material data assignment

DD

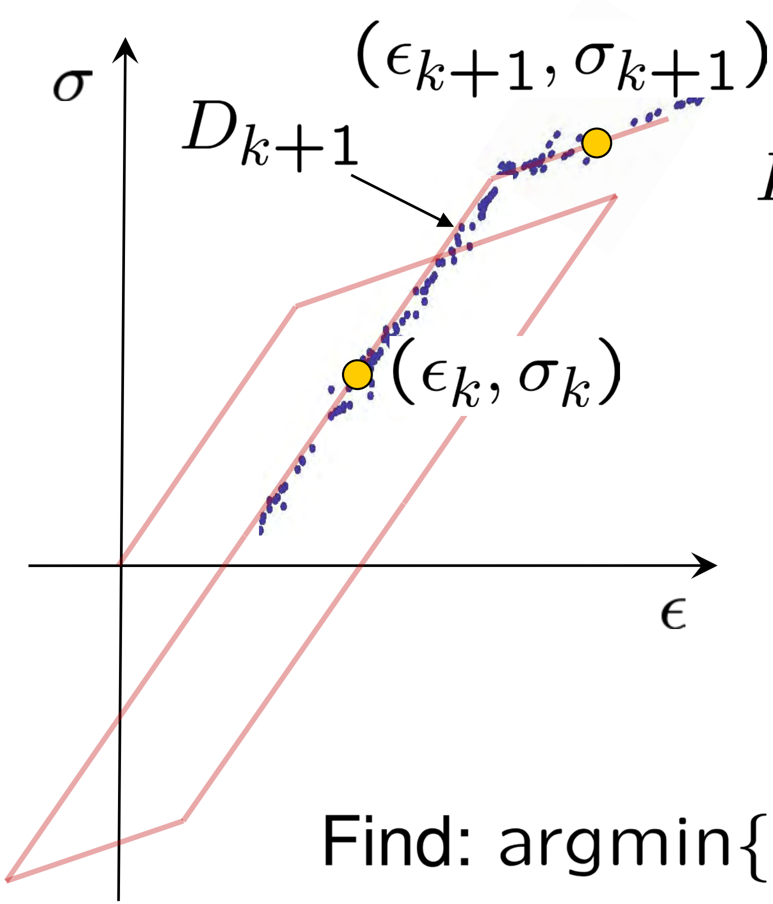


Material data lookup

Gauss point material states

Material-independent linear problem

Data-Driven inelasticity



- Material set representation:

$$D_{k+1} = \left\{ (\epsilon_{k+1}, \sigma_{k+1}) \ : \ \text{history} \right\}$$

- Need ***material history data***! (from material testing along selected loading paths...)
- History data must provide adequate ***path coverage***...
- Data-driven problem:

$$\text{Find: argmin}\{d(z, D_{k+1}),\ z \in E_{k+1}\}$$

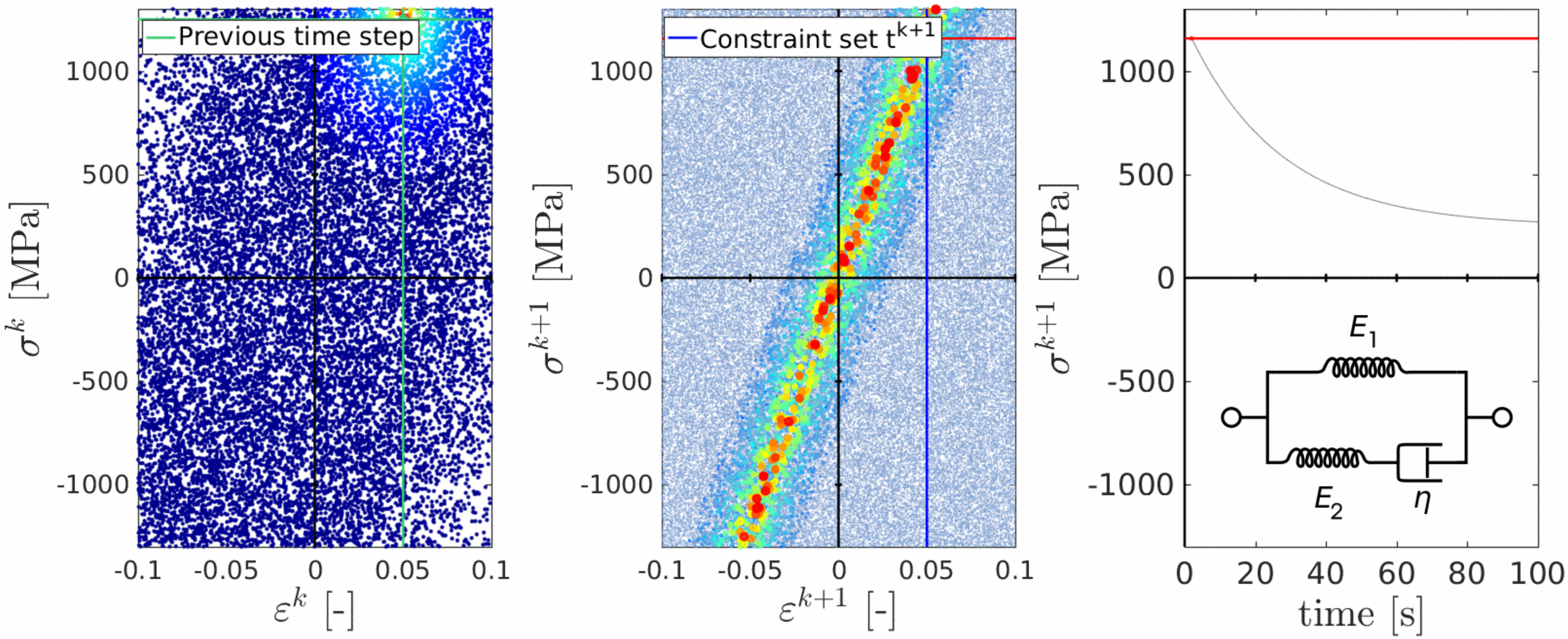
time dependent!

R. Eggersmann, T. Kirchdoerfer, L. Stainier, S. Reese and M. Ortiz, *CMAME*, **350** (2019) 81-99.

Michael Ortiz
CAMBRIDGE 2021

Data-Driven Inelasticity

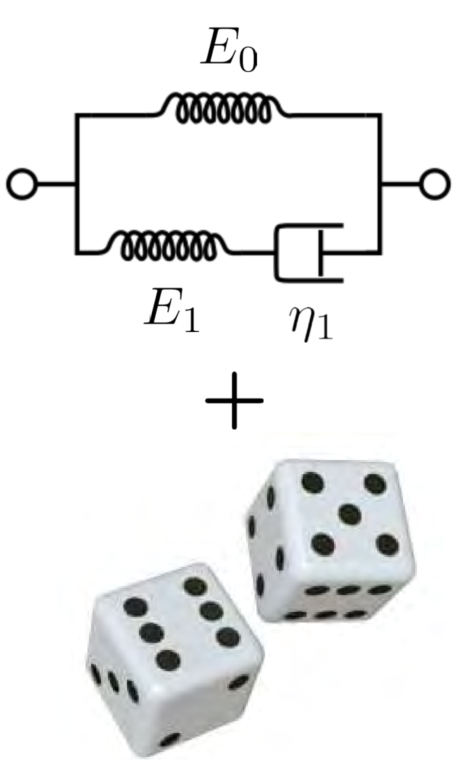
- Example: Data-Driven *linear viscoelasticity*
  - *(Randomized) Standard Linear Solid*
  - *Relaxation test (constant strain)*

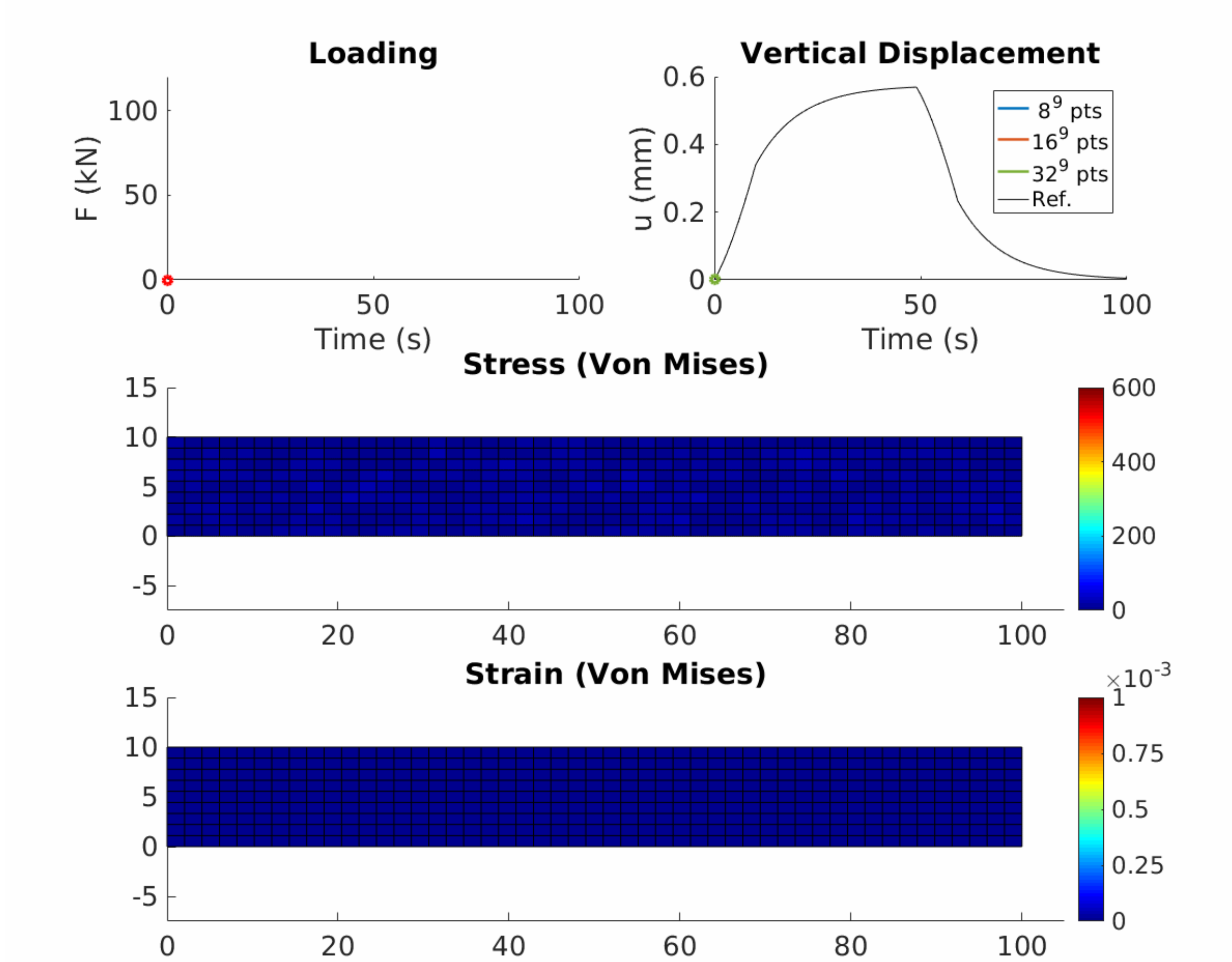


R. Eggersmann, T. Kirchdoerfer, L. Stainier, S. Reese and M. Ortiz, *CMAME*, **350** (2019) 81-99.

Data-Driven viscoelasticity

- Material set: $D_{k+1} = \left\{ (\epsilon_{k+1}, \sigma_{k+1}) \,:\, (\epsilon_k, \sigma_k) \right\}$





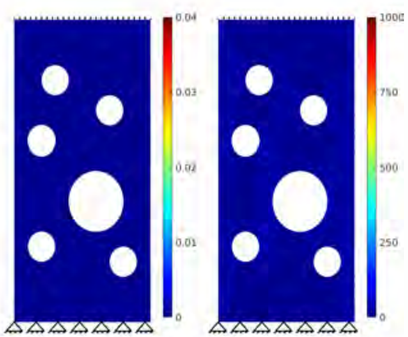
R. Eggersmann *et al.*, *CMAME*, **350** (2019) 81-99.

Data-Driven plasticity

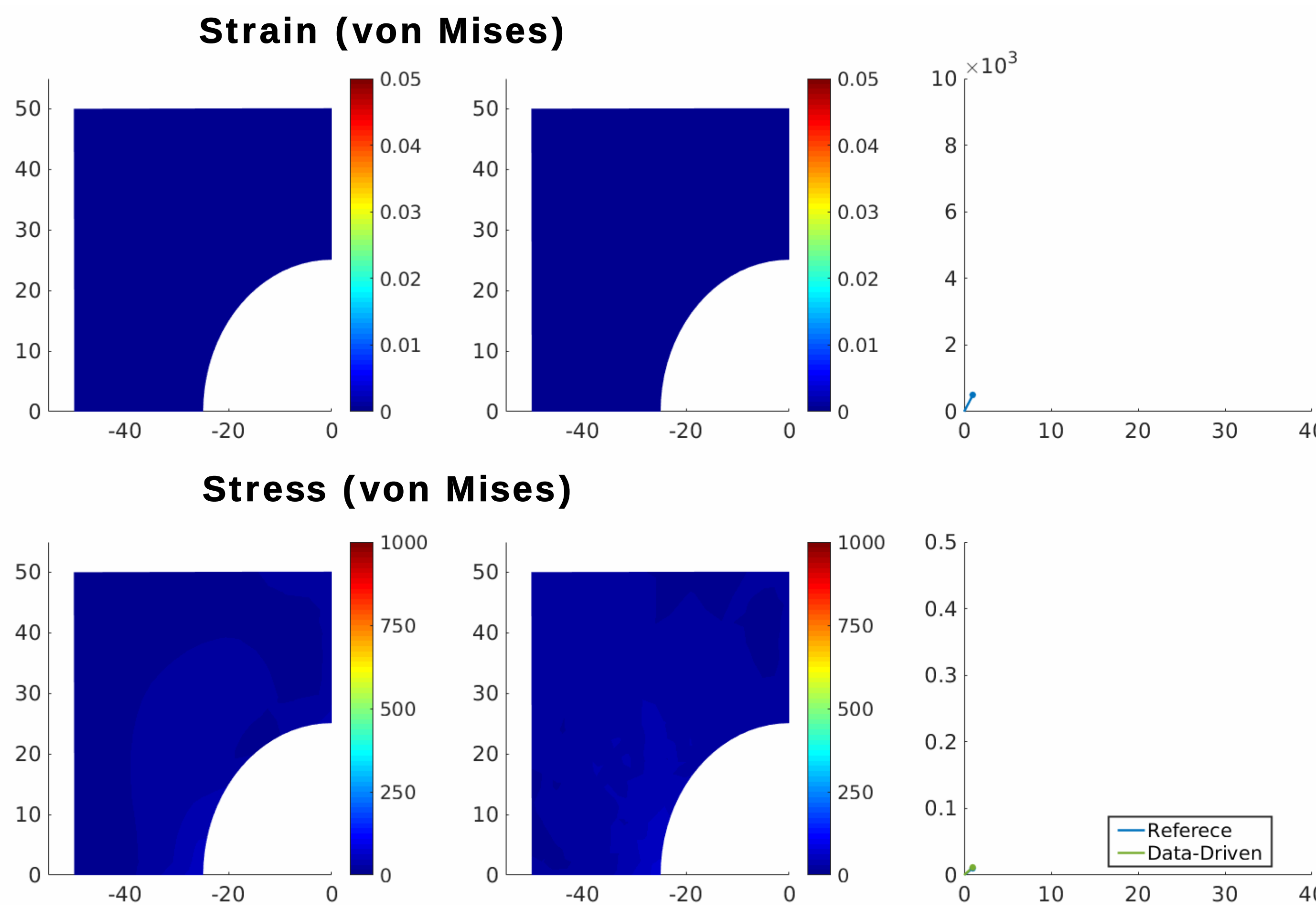
- Material set: $D_{k+1} = \left\{ (\epsilon_{k+1}, \sigma_{k+1}) \,:\, (\epsilon_k, \sigma_k) \right\}$

J2 plasticity kinematic hardening

+



Virtual test



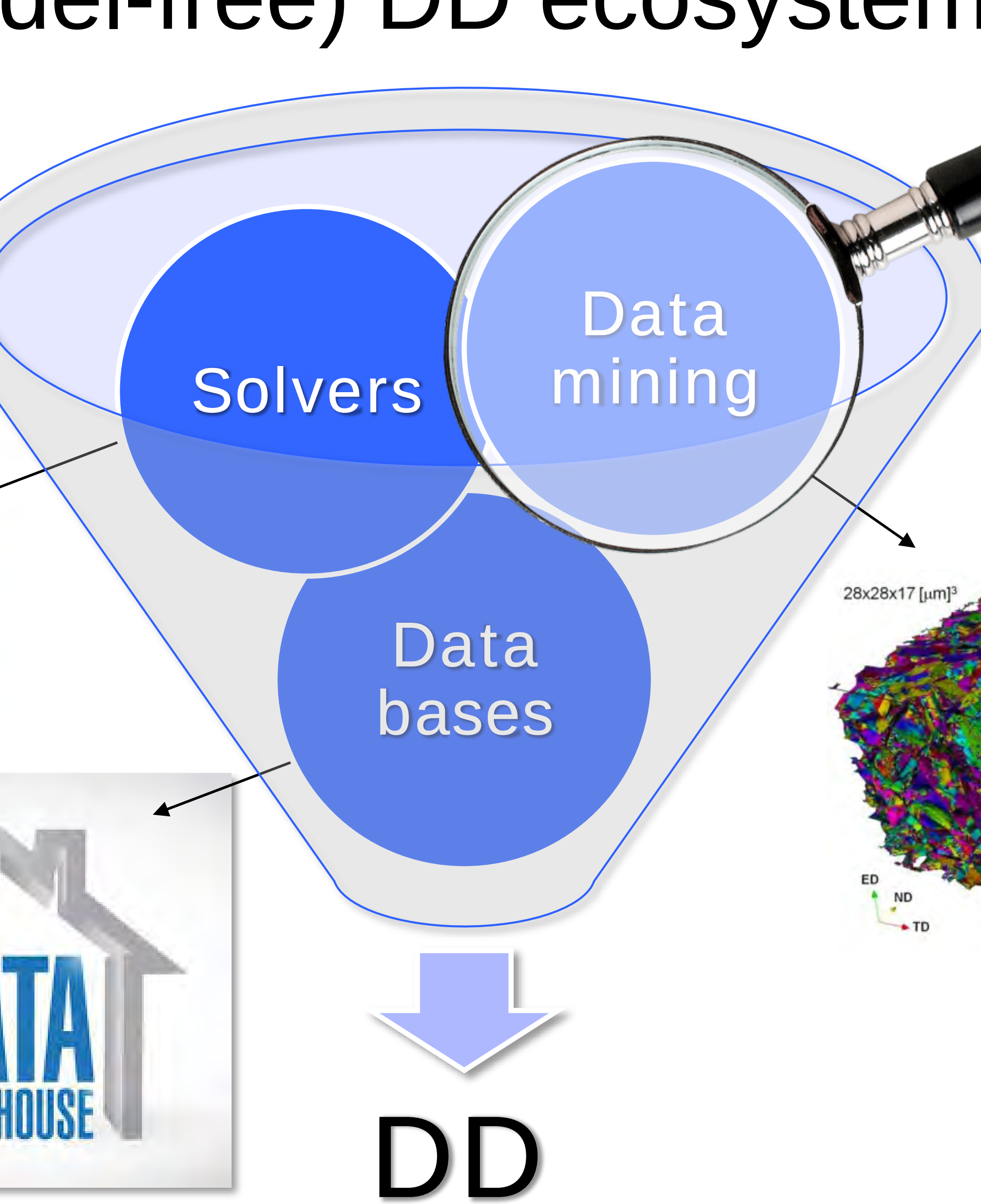
(R. Eggersmann & S. Reese, 2019)

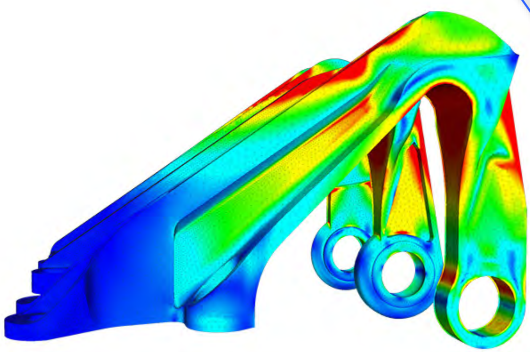
(Model-Free) Data-Driven solvers

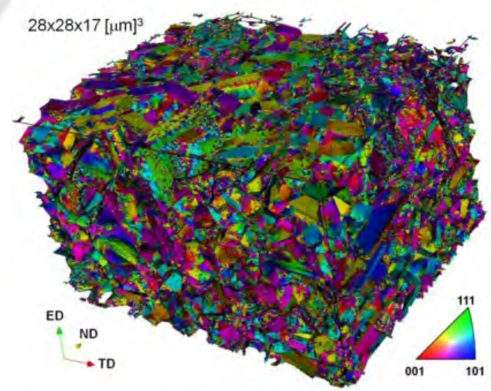
- Model-free DD computing is *achieving maturity*:
  - *Analysis, existence of solutions, well-posedness*[1]
  - *Finite elasticity, finite kinematics*[2,3]
  - *Dynamics, time discretization*[4]
  - *Noisy data, clustering (k-means), outliers*[5]
  - *Inelasticity, viscoelasticity, plasticity*[6]*, fracture*[7]
  - *Multiscale data mining, model-free upscaling*[8]
  - *Probabilistic extension, model-free, prior-free*
  - *Fast Data searching, k-d trees, scalable ANN*

[1]Conti, S., Müller, S. & Ortiz, M., *ARMA*, **229** (2018) 79-123.
[2]L.T.K. Nguyen and M.A. Keip, *Comput. Struct.*, **194** (2018) 97–115
[3]S. Conti, S. Müller and M. Ortiz, *ARMA*, **237** (2020) 1–33.
[4]T. Kirchdoerfer and M. Ortiz, *IJNME*, **113**(11) (2018) 1697-1710.
[5]T. Kirchdoerfer and M. Ortiz, *CMAME*, **326** (2017) 622-41.
[6]R. Eggersmann, S. Reese, L. Stainier, M. Ortiz, *CMAME*, **350** (2019) 81-99.
[7]P. Carrara, L. Lorenzis, L. Stainier, M. Ortiz, *CMAME*, **372** (2020) 113390.
[8]K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, *JMPS*, (2020) 104239.

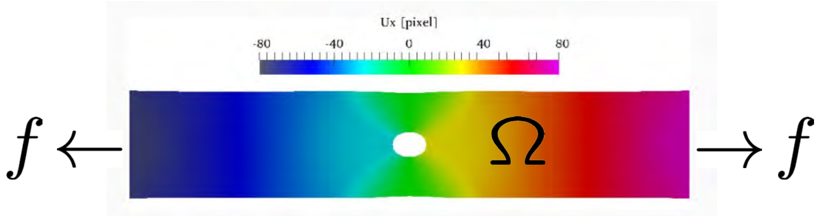
The (model-free) DD ecosystem…







DD material identification (DDMI)



- Full-field measurements (DIC): $(\epsilon'(x), f(x))$
- ***Stress field cannot be measured directly!***
- ***Density of states*** in phase space:

$$\rho(\epsilon,\sigma) = \frac{1}{|\Omega|}\int_\Omega \delta(\epsilon - \epsilon'(x), \sigma - \sigma'(x))\, dx$$

- Maximize ***information-theoretical entropy***: $S(\rho)$

  subject to ***equilibrium***: $\mathrm{div}\,\sigma'(x) + f(x) = 0$
- Solve (inverse) problem using ***DD k-means***

J. Rethore, HAL Id: hal-01454432, Feb. 2017.
J. Rethore and A. Leygue, HAL Id: hal-01452494, Feb. 2017.

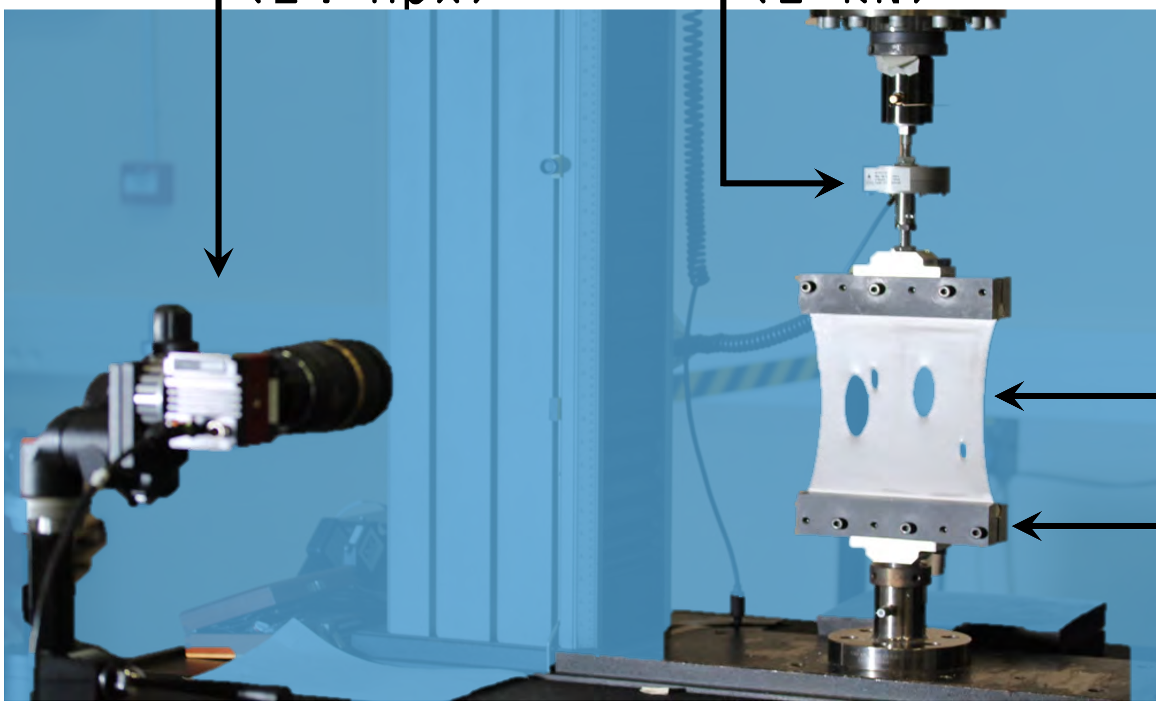
DDMI at Central Nantes

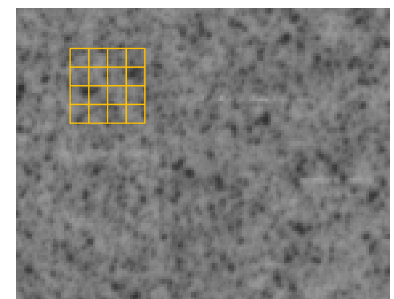
- Tests on perforated silicone sheets

Camera (29 Mpx)

Force cell (1 kN)

Membrane with speckles and square grid:





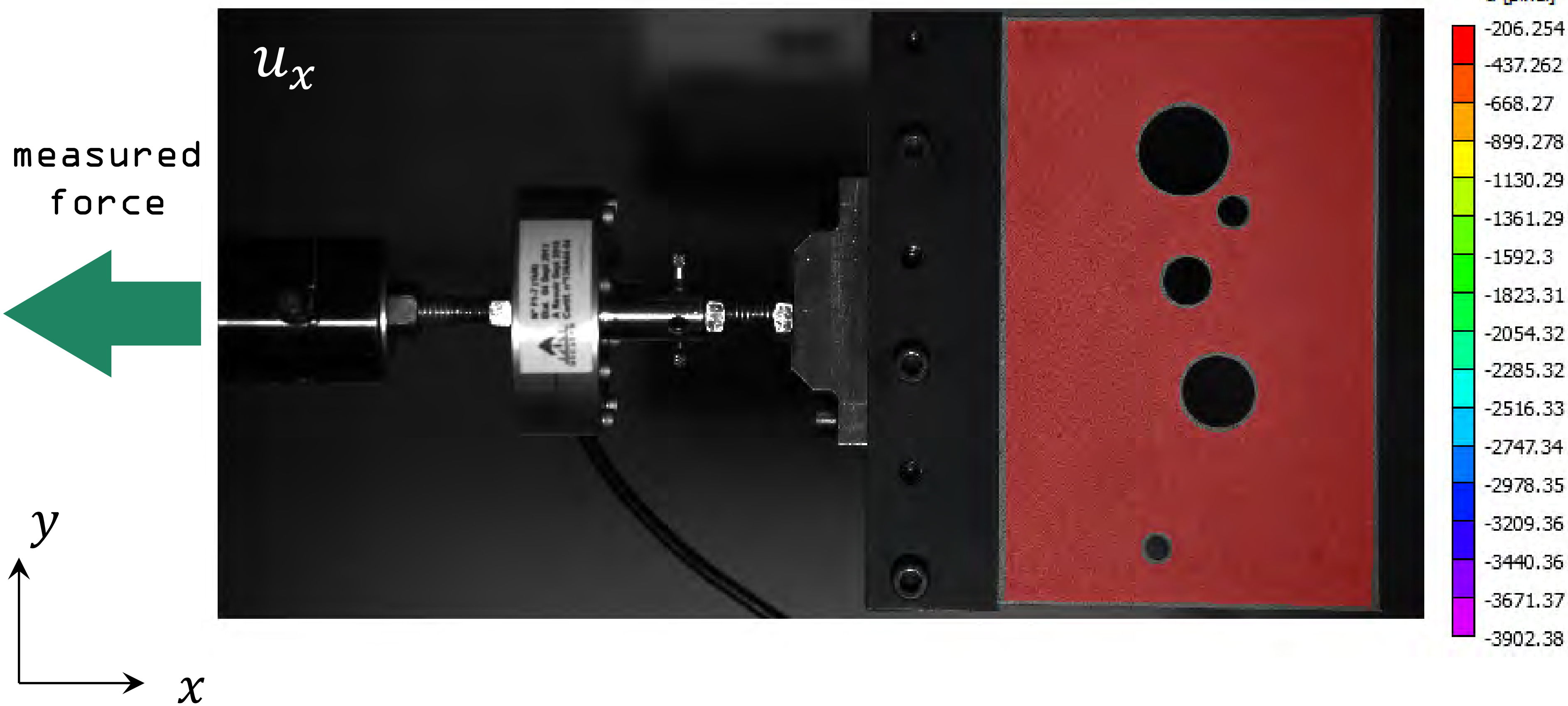
8,8 mm

Clamping system

M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

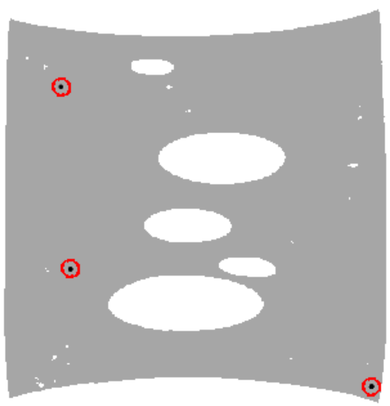
DDMI at Central Nantes

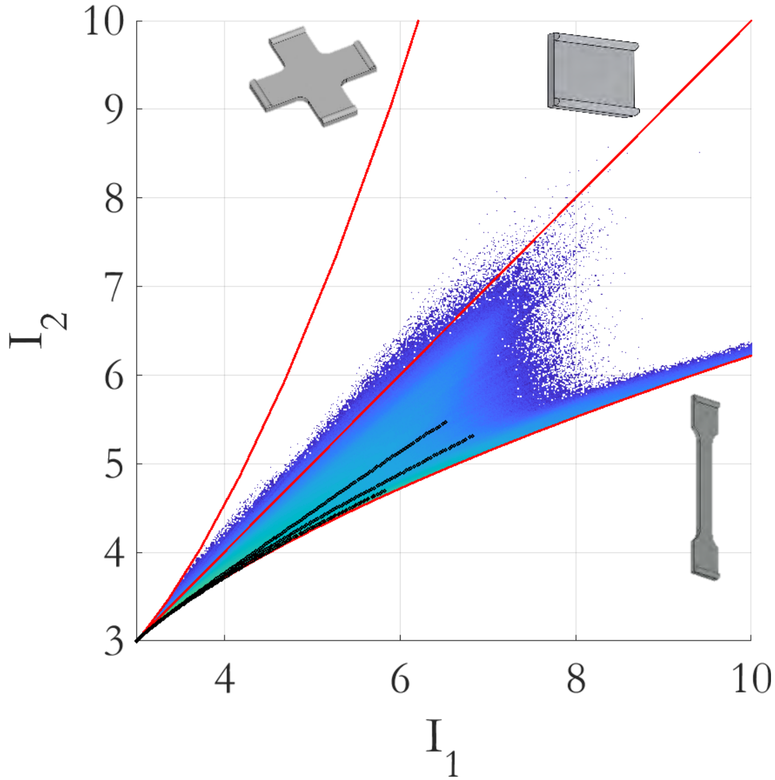
Displacement field (DIC)

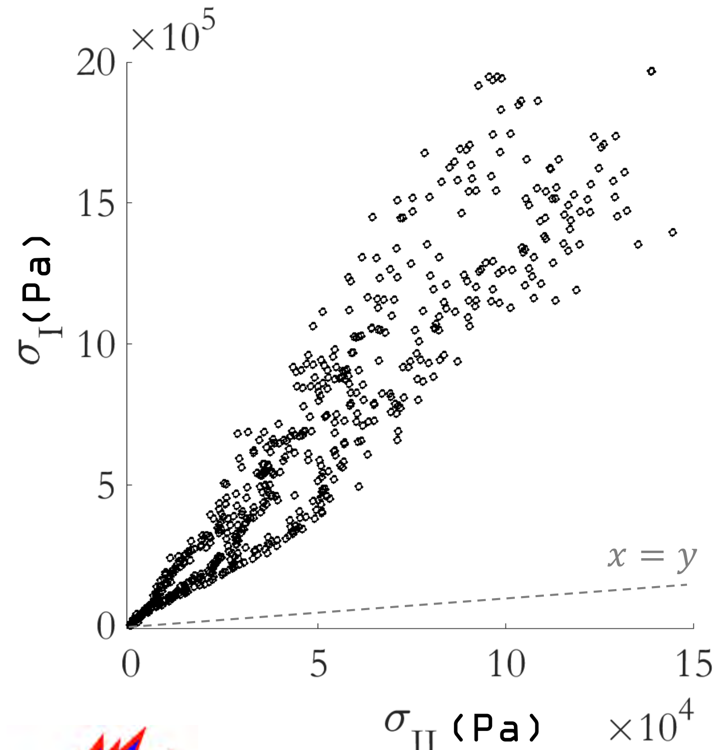


M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

DDMI at Central Nantes







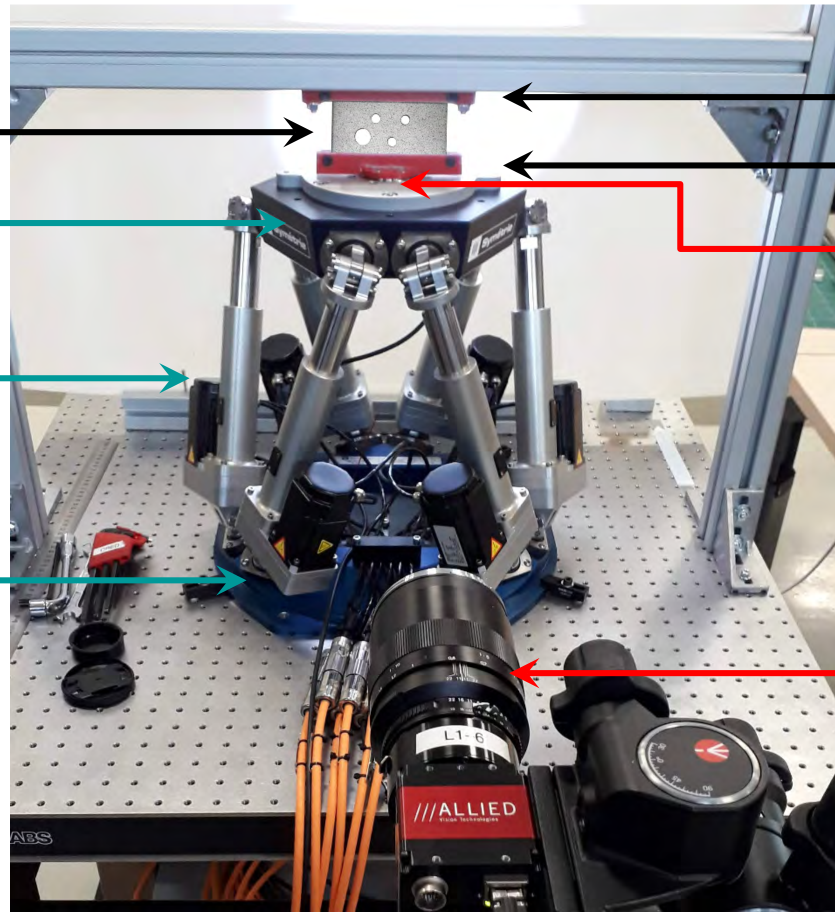
Strain data

Stress data

M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

DDMI at Central Nantes

- Hexapod testing machine



M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

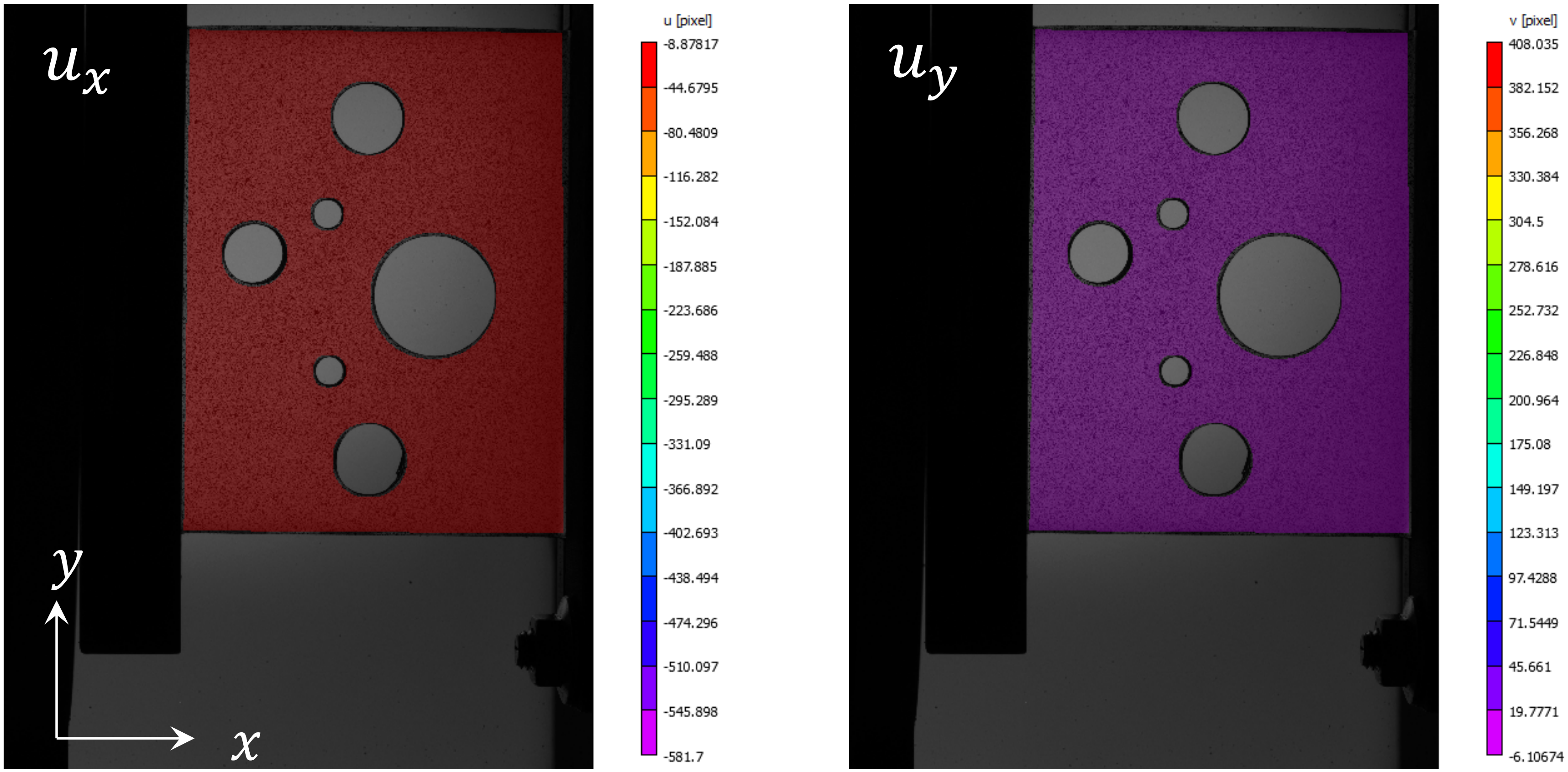
DDMI at Central Nantes

Hexapod in operation

M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

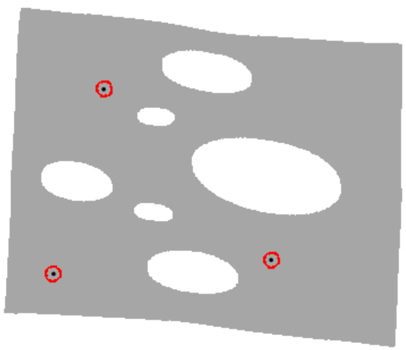
DDMI at Central Nantes

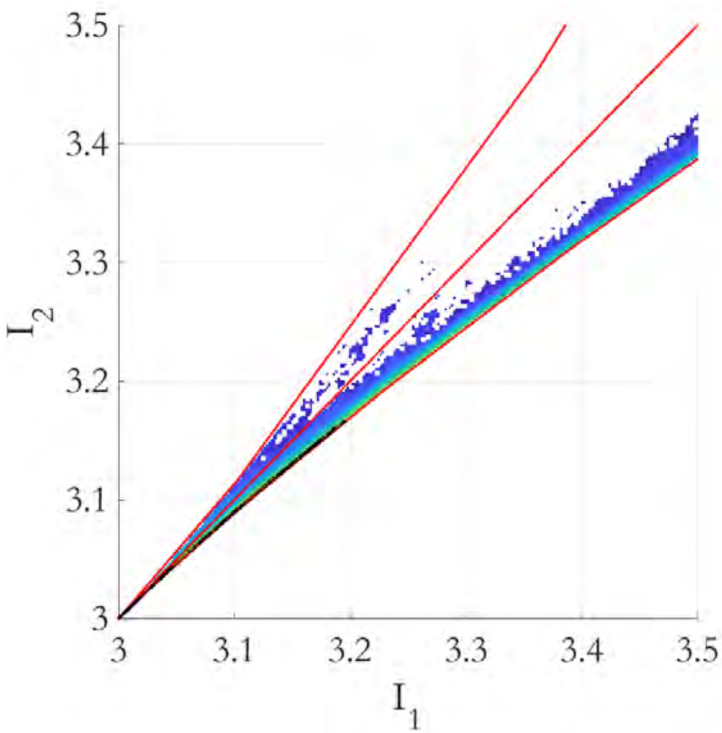
Displacement field (DIC)



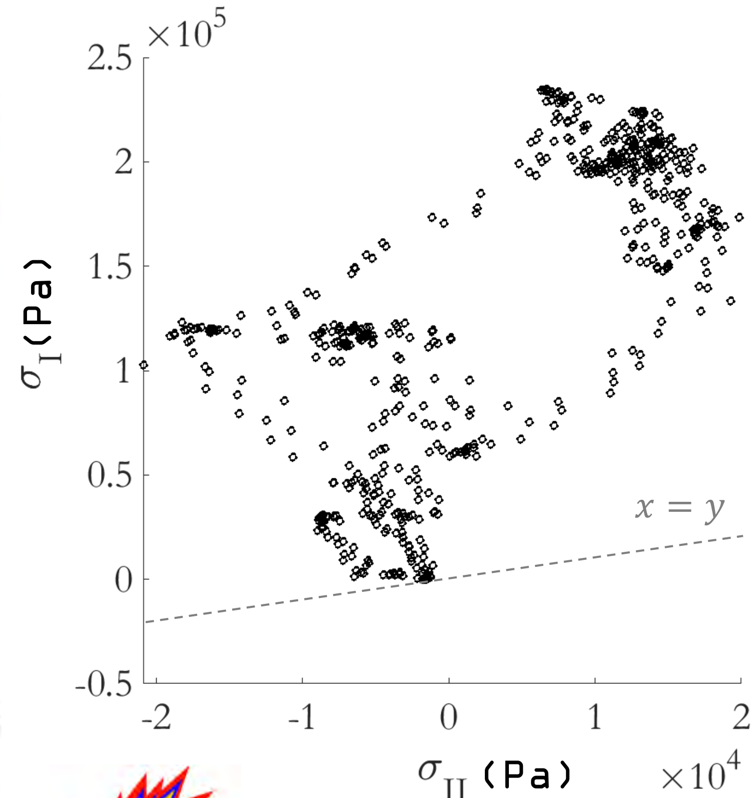
M. Dalémat, M. Coret, A. Leygue and E. Verron,
Mechanics of Materials, **136** (2019) 103087.

DDMI at Central Nantes





Strain data

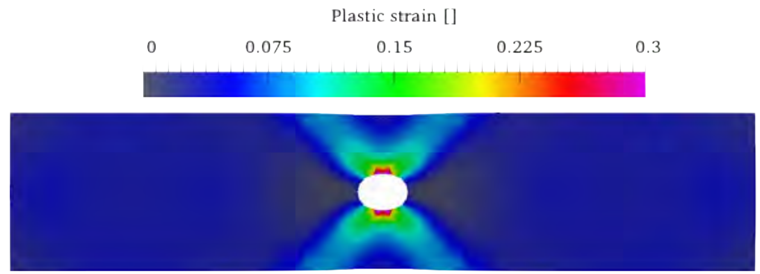


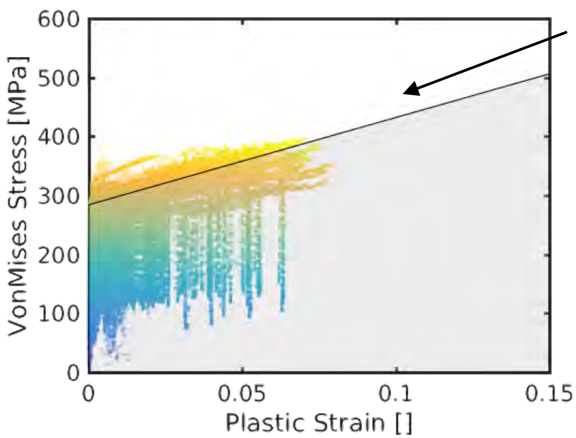
Stress data

M. Dalémat, M. Coret, A. Leygue and E. Verron, *Mechanics of Materials*, **136** (2019) 103087.

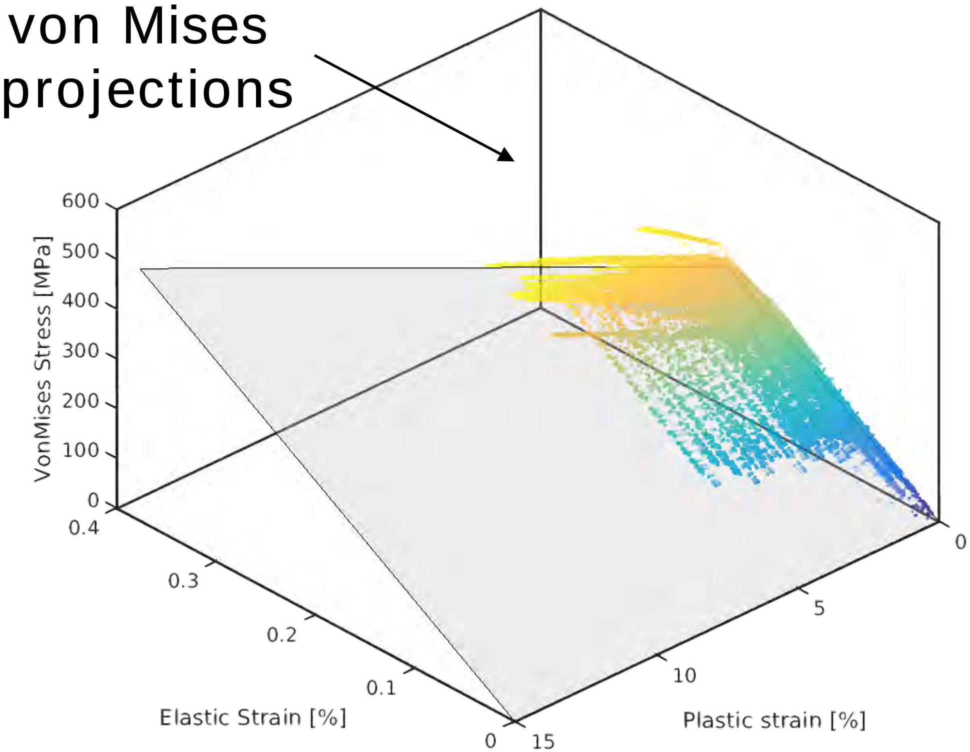
DD material identification (DDMI)

304L stainless steel perforated specimen





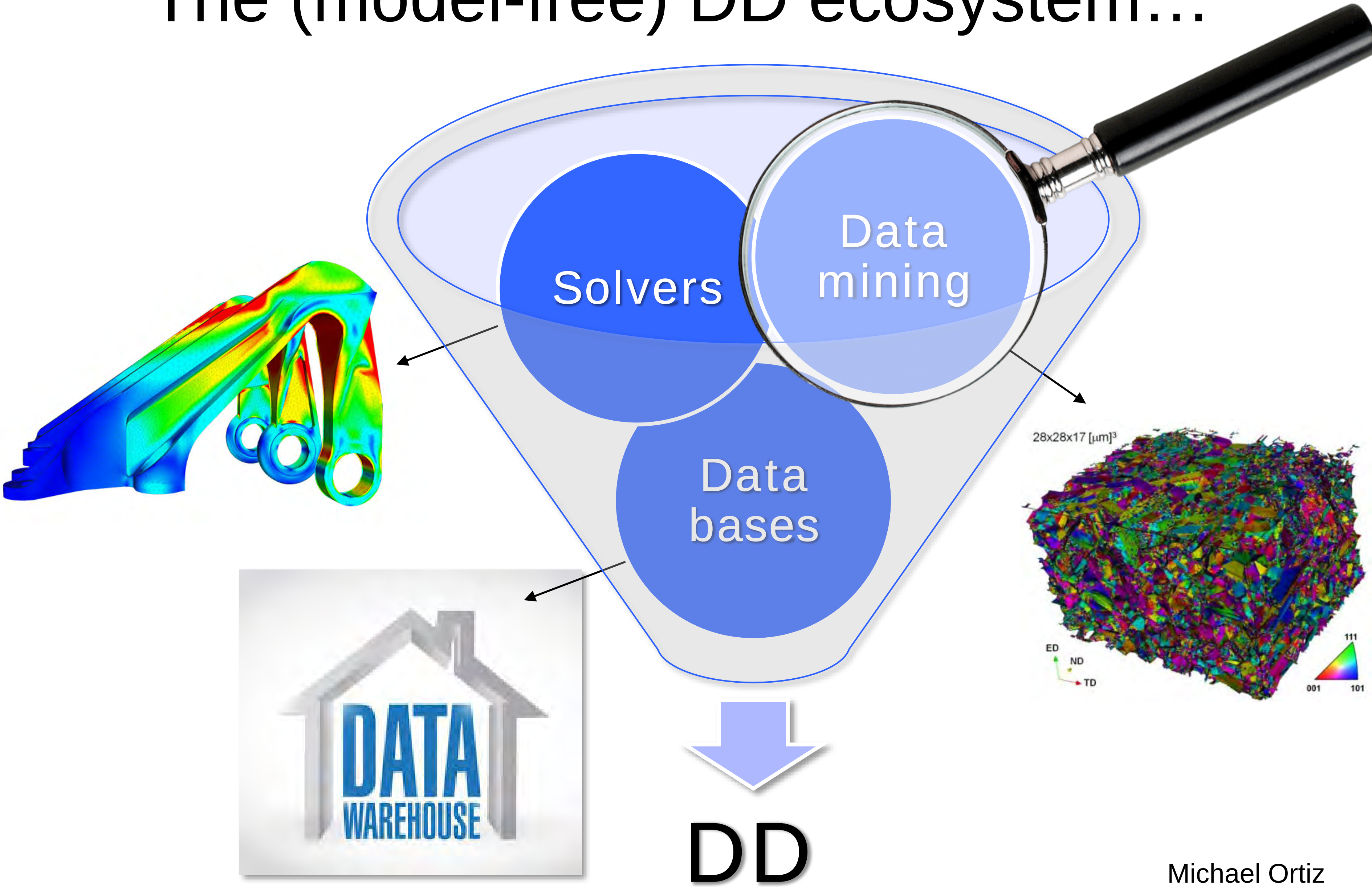
von Mises projections



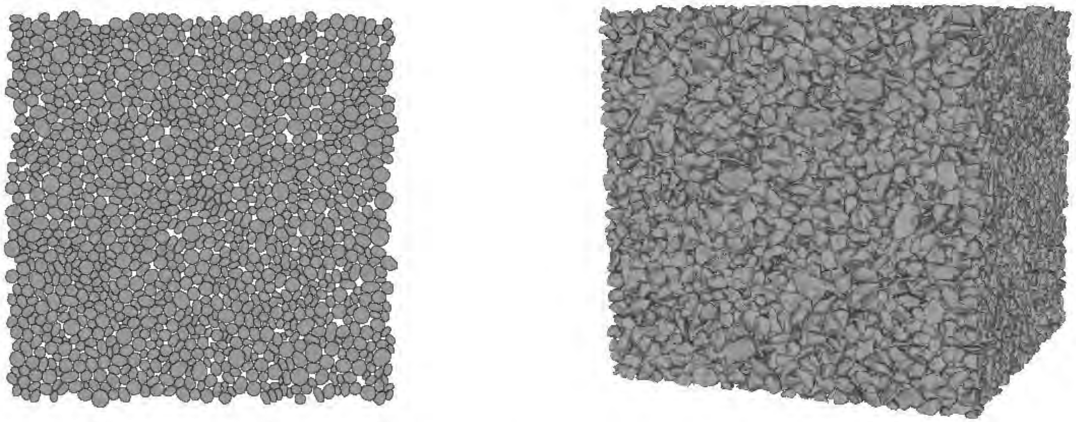
Elastic-plastic material data base, 600000 points in dimension 12

J. Rethore, HAL Id: hal-01454432, Feb. 2017.
J. Rethore and A. Leygue, HAL Id: hal-01452494, Feb. 2017.

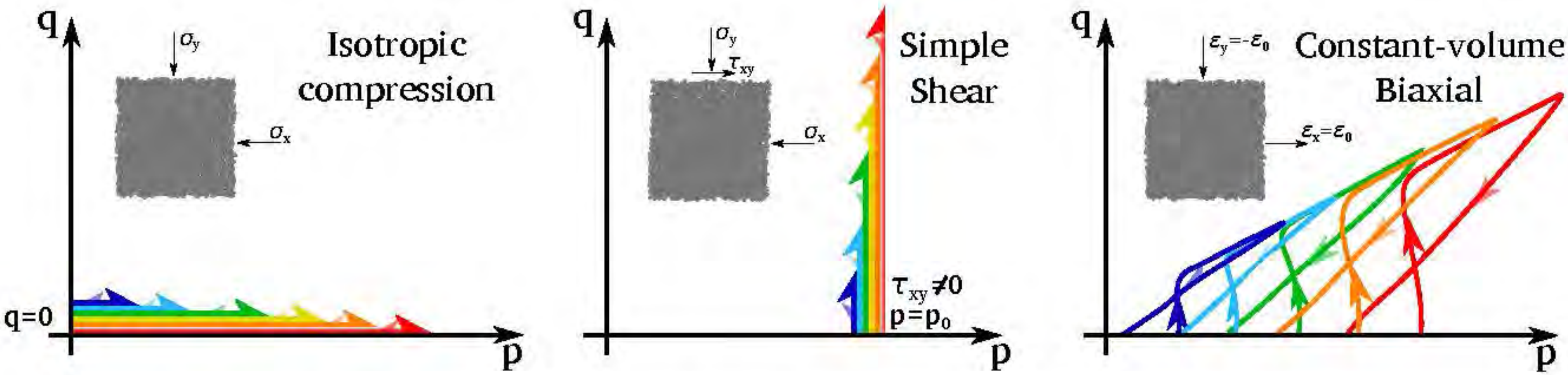
The (model-free) DD ecosystem…



Multiscale data mining − Sand



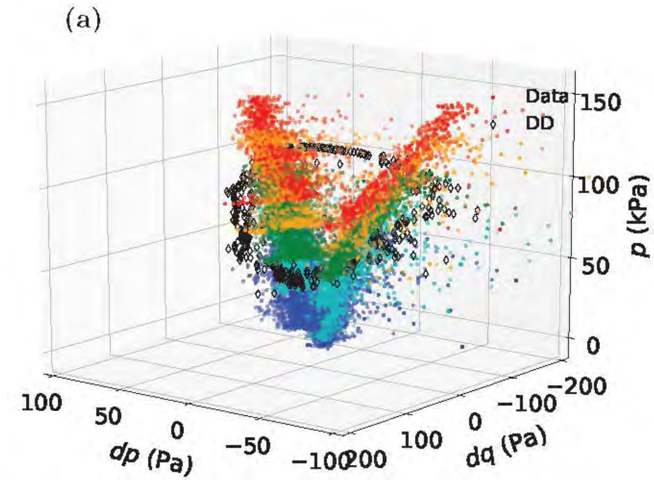
2D and 3D RVE and granular assemblies for LS-DEM analysis



Stress paths for material point simulations and data mining

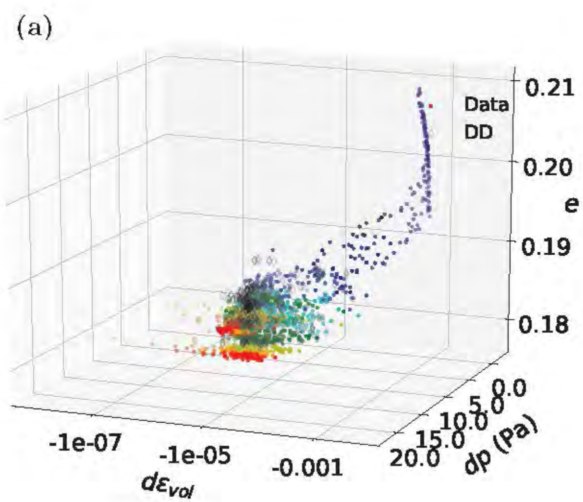
K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, *JMPS*, (2020) 104239.

Multiscale data mining − Sand

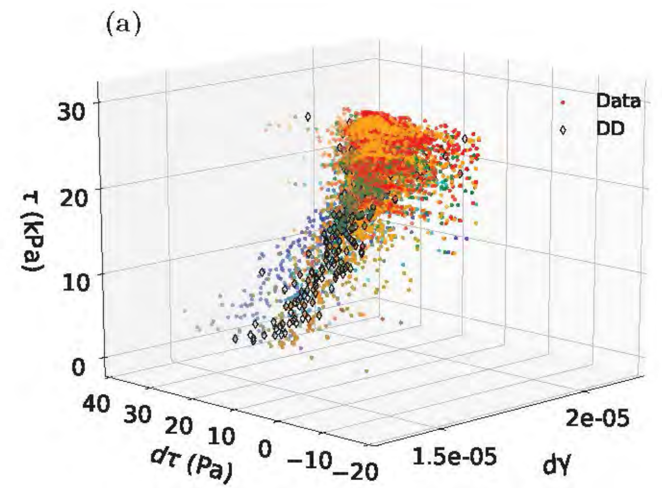


Isotropic compression

Simple shear



Biaxial compression

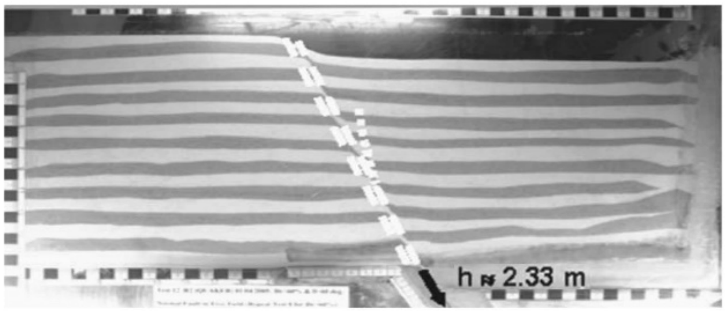


Data computed from RVE LS-DEM calculations

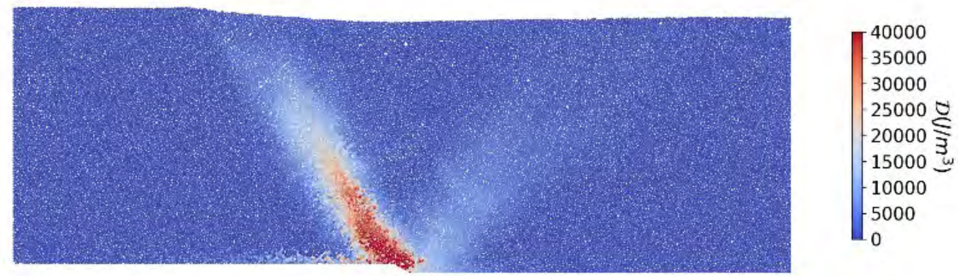
K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, *JMPS*, (2020) 104239.

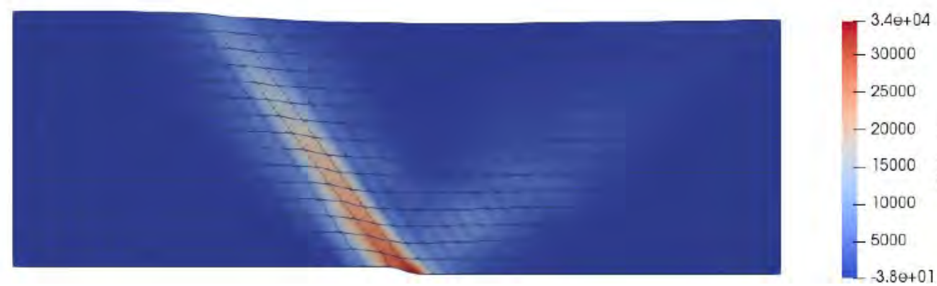
Michael Ortiz
CAMBRIDGE 2021

(Model-free) Data-Driven – Sand

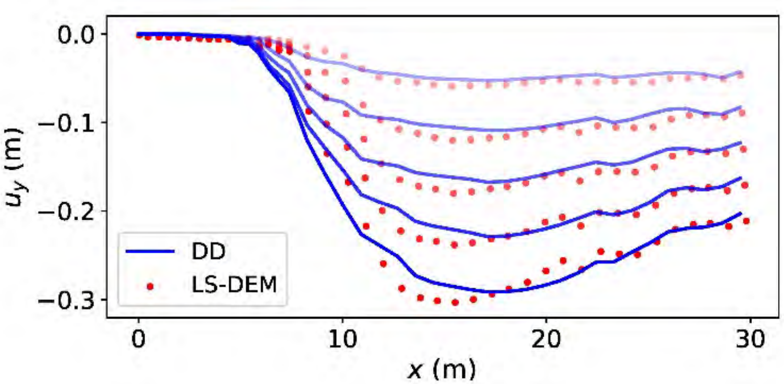


Experimental fault rupture experiment (Anastapoulos etal.. 2007)





Top: LS-DEM simulation
Bottom: DD simulation

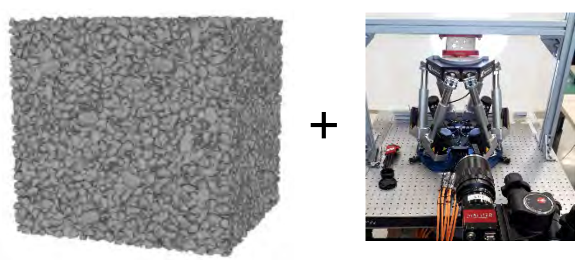


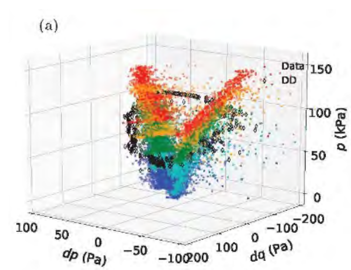
Evolution of surface settlement, LS-DEM vs. DD simulations

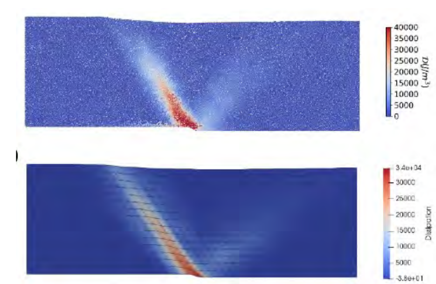
K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, *JMPS*, (2020) 104239.

Data mining, generation, upscaling

- DDMI can extract data from *full-field experimental microscopy* (TEM, SEM, DIC, EBSD...)
- DDMI generates *fundamental data without prior assumptions* on the form of constitutive response
- Data can be *mined* from lower-scale calculations, used in upper-scale calculations (*DD upscaling*)
- *DD sets forth new opportunities for synergism between experimental science and scientific computing, new multiscale analysis paradigms*

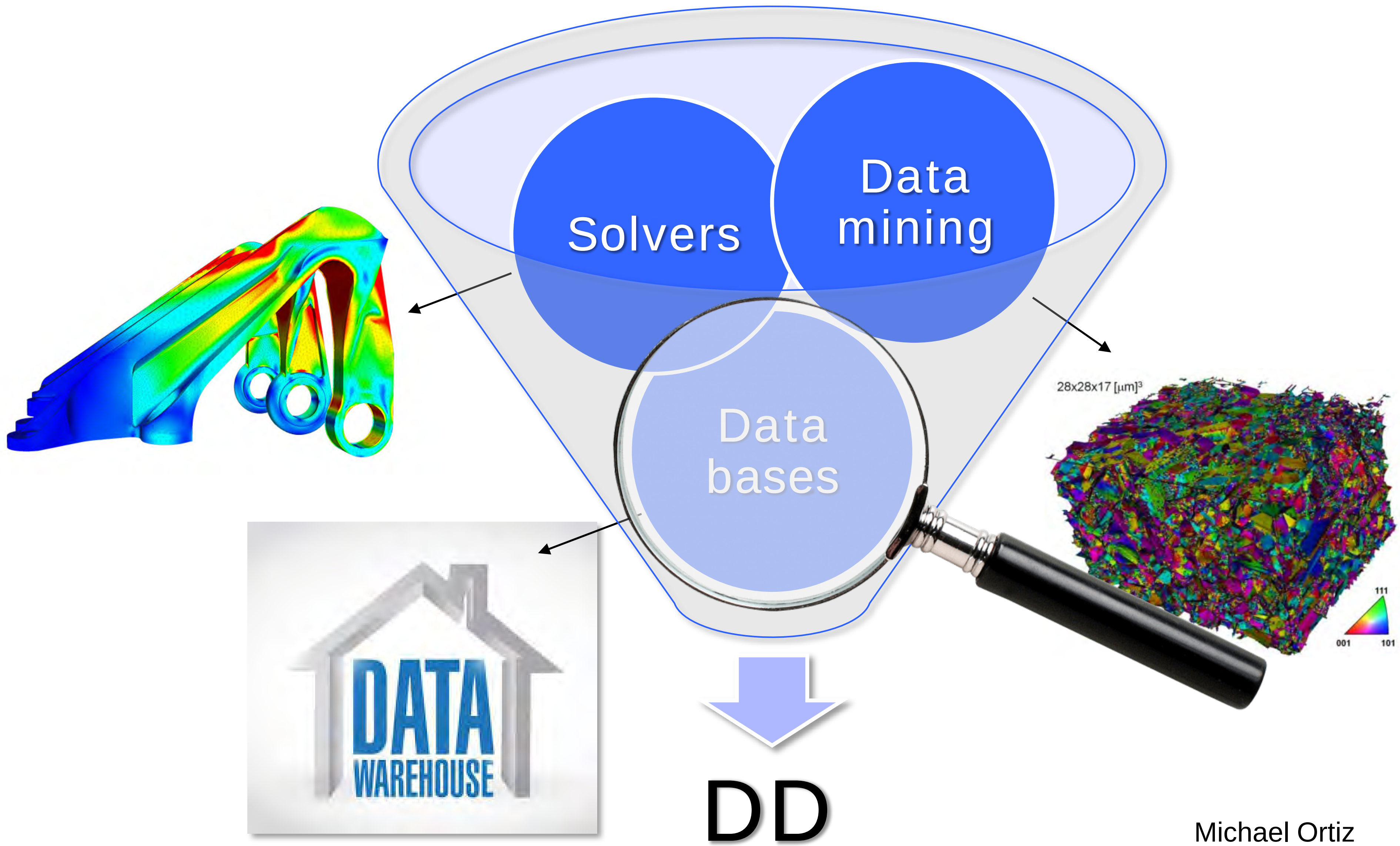




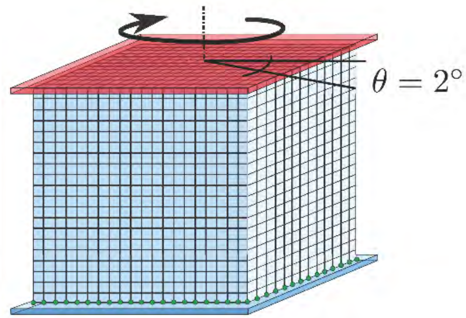


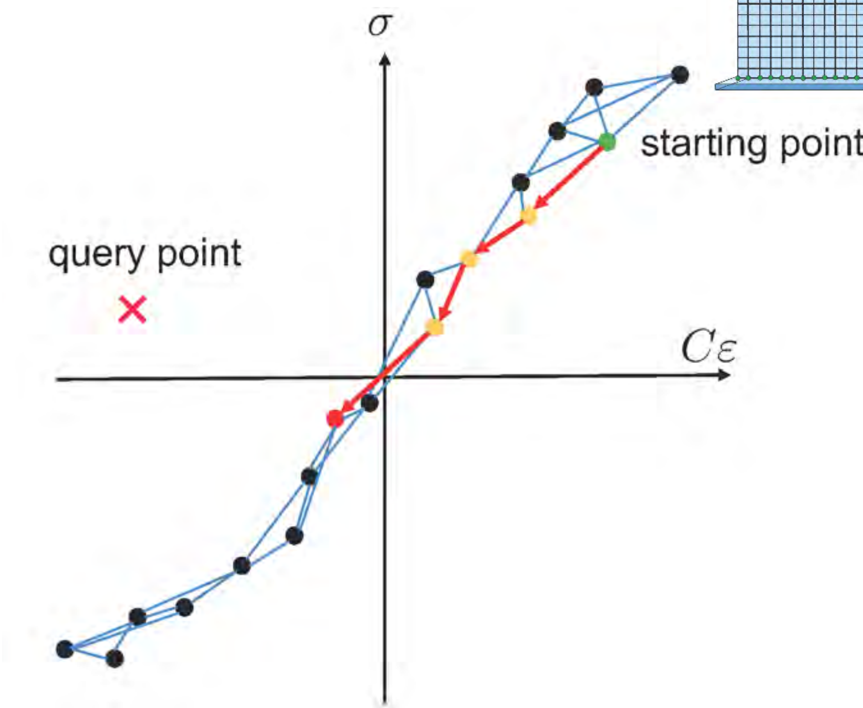
J. Rethore and A. Leygue, HAL Id: hal-01452494, Feb. 2017.
K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, *JMPS*, (2020) 104239.

The (model-free) DD ecosystem…

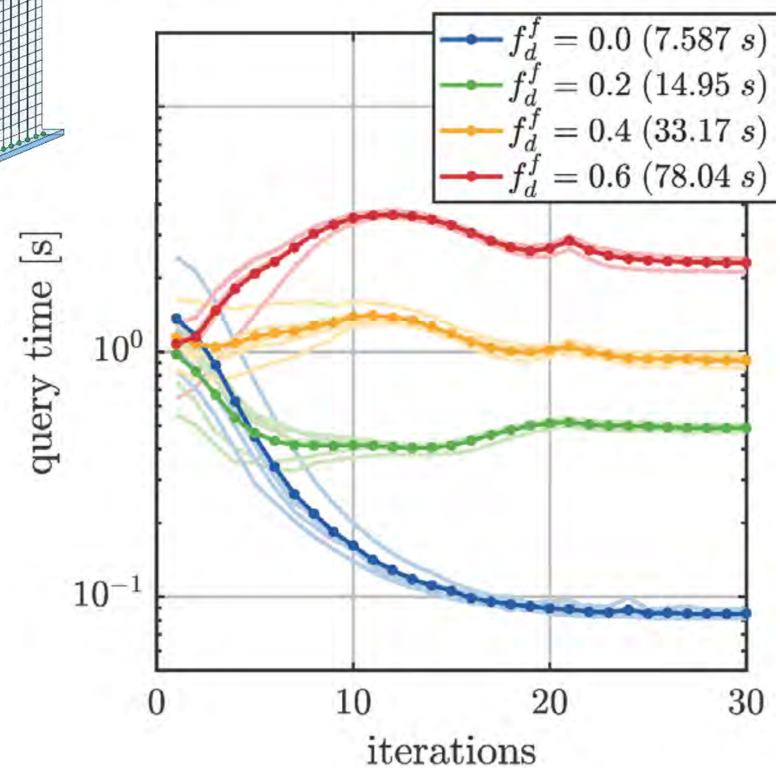


Data structures and searching algorithms





kNN data structure and searching algorithm



Performance of approximate ***k-means*** trees on ***billion-point data set***

Data structures and searching algorithms

- Searching large data sets requires specialized *data structures* (k-d trees, kNN trees, k-means)
- The data structures encode the structure of the data sets, clustering, NN relations...
- Structuring the data is a form of *set-oriented machine learning* (‘learning’ the data), where the result of machine learning is a *data structure*
- Unlike regression by ANNs, set-oriented machine learning is *unbiased* and *lossless* (*the data, all the data and nothing but the data*)

Publically-editable data repositories

- Reliance on *fundamental data* (stress and strain only, no model-dependent data) makes *material data fungible*, mergeable, interchangeable...
- *Publically editable repositories* have proven exceptional capacity for *organic growth*...
- *Publicly-editable material data repository! (Wikimat)*
  - *Fundamental, model-independent data:*
    - *Stress-strain*
    - *Temperature-entropy*
    - *Gradients, rates*
  - *Tools/Scripts for interfacing with commercial FE packages*

WIKIPEDIA
The Free Encyclopedia

Concluding remarks

- *The Model-Free Data-Driven ecosystem*:
  - Solver standardization (material independent)
  - Data identification from full-field microscopy
  - Multiscale analysis and upscaling
  - Data repositories (publically editable)
- Vision and development thereof falls squarely within the purview of the *Cambridge Centre for Data-Driven Discovery* and the *Alan Turing Institute*
- *Data-driven computing is likely to be a growth area in an increasingly data-rich world and to change the way in which data is mined, stored, exchanged, disseminated and utilized in science and in industry!*

Concluding remarks

Thank you!