Model-Based Rigorous Uncertainty Quantification in Complex Systems

Michael Ortiz
California Institute of Technology

MECOM 2010

Buenos Aires, November 17, 2010

Work done in collaboration with: Marc Adams, Addis Kadani, Bo Li, Mike McKerns, Ali Lashgari, Jon Mihaly, Houman Owhadi, G. Ravichandran, Ares J. Rosakis, Tim Sullivan

ASC/PSAAP Centers

CALTECH PSAAP

STANFORD UNIVERSITY PSAAP

PRISM NNSA Center for Prediction of Reliability, Integrity and Survivability of Microsystems

PURDUE UNIVERSITY

CALIFORNIA INSTITUTE OF TECHNOLOGY 1891

CALTECH PSAAP

ICES The Institute for Computational Engineering and Sciences

CENTER FOR RADIATIVE SHOCK HYDRODYNAMICS

MICHIGAN

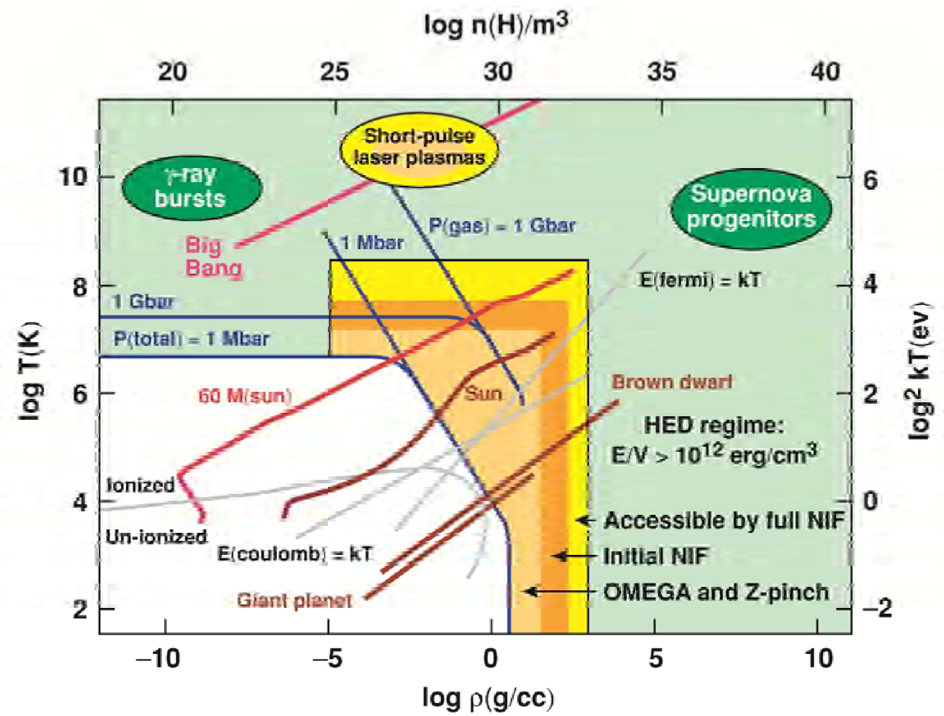
THE UNIVERSITY OF TEXAS AT AUSTIN

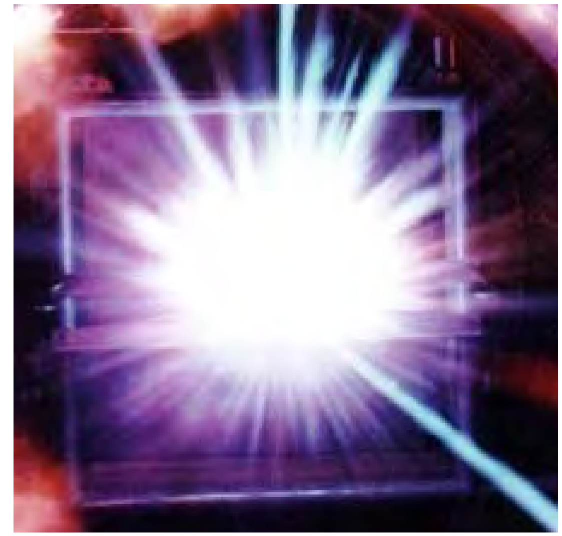
PSAAP: Predictive Science Academic Alliance Program

Hypervelocity impact as an example of a complex system

CALTECH PSAAP

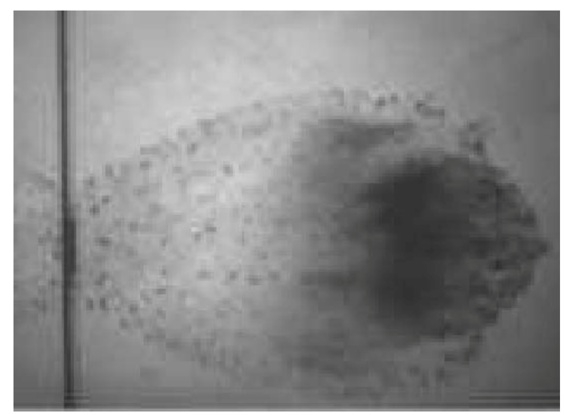
Challenge: Predict *hypervelocity impact* phenomena (10Km/s) with ***quantified margins and uncertainties***





NASA Ames Research Center Energy flash from hypervelocity test at 7.9 Km/s

Hypervelocity impact test bumper shield (Ernst-Mach Institut, Freiburg Germany)



Michael Ortiz

Quantification of margins and uncertainties (QMU)

CALTECH PSAAP

- Aim: *Predict mean performance and* ***uncertainty*** *in the behavior of complex physical/engineered systems*
- Example: Short-term weather prediction,
  - Old: Prediction that tomorrow will rain in Buenos Aires…
  - New: ***Guarantee*** same with 99% confidence…
- ***QMU*** is important for achieving confidence in high-consequence decisions, designs
- ***Paradigm shift*** in experimental science, modeling and simulation, scientific computing (***predictive science***):
  - Deterministic → Non-deterministic systems
  - Mean performance → Mean performance + uncertainties
  - Tight integration of experiments, theory and simulation
  - Robust design: Design systems to minimize uncertainty
  - Resource allocation: Eliminate main uncertainty sources

Certification view of QMU

CALTECH PSAAP

$$\overbrace{(X_1, \ldots, X_M)}^{\text{system inputs}} \xrightarrow{\underbrace{G}_{\text{response function}}} \overbrace{(Y_1, \ldots, Y_N)}^{\text{performance measures}}$$

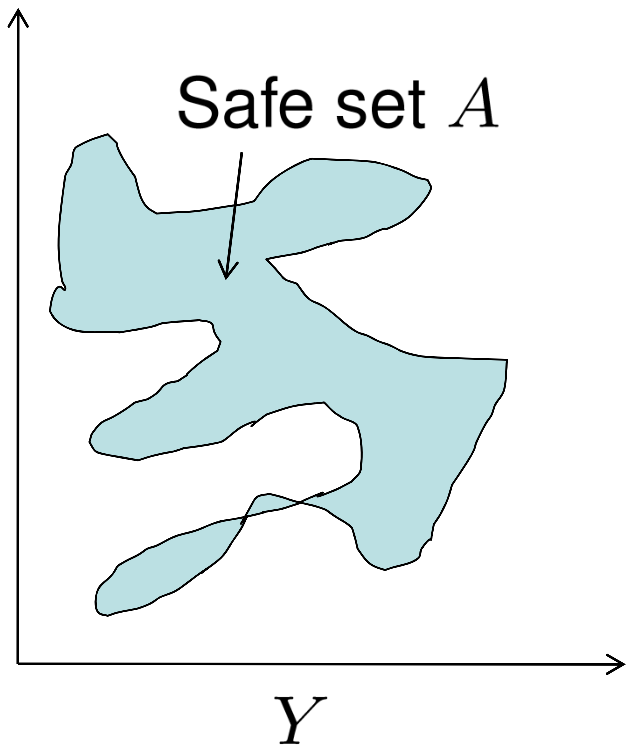
- Random variables
- Known or unknown pdfs
- Controllable, uncontrollable, unknown-unknowns

System as black box

- Observables
- Subject to performance specs
- Random due to randomness of inputs or of system

Certification view of QMU

- Certification = Rigorous guarantee that complex system will perform safely and according to specifications



- Certification criterion: Probability of failure must be below tolerance,

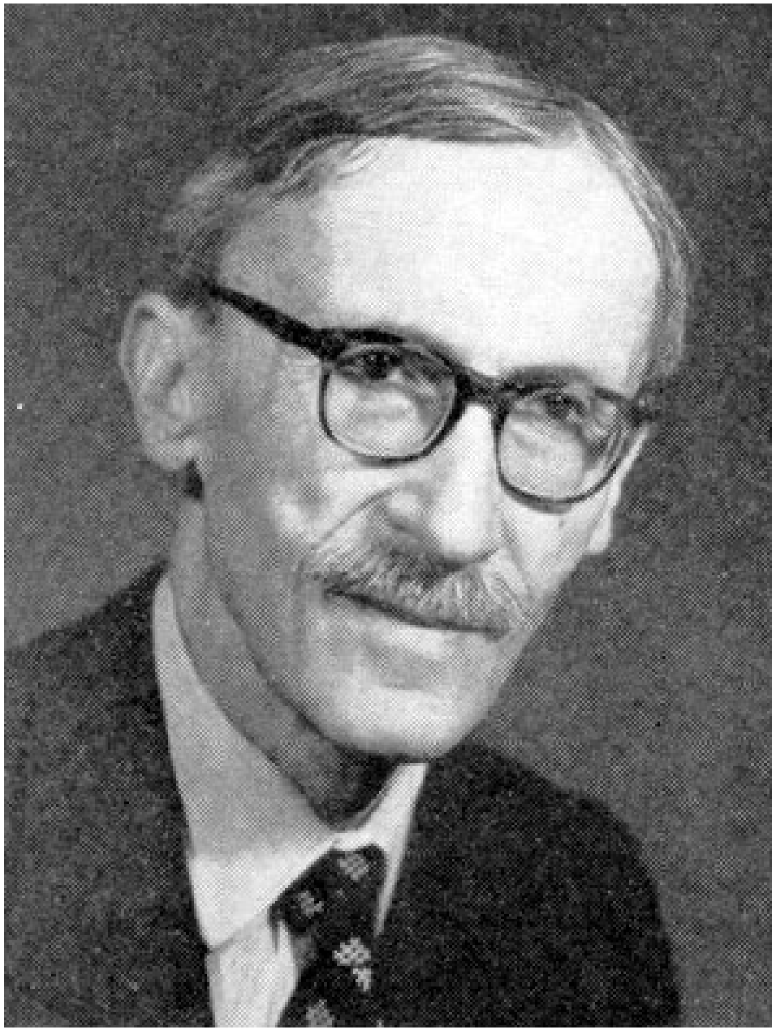
$$\mathbb{P}[Y \in A^c] \le \epsilon$$

- Alternative (conservative) certification criterion: Rigorous *upper bound* of probability of failure must be below tolerance,

$$\mathbb{P}[Y \in A^c] \le \boxed{\text{upper bound} \le \epsilon}$$

- Challenge: Rigorous, measurable/computable upper bounds on the probability of failure of systems

Concentration of measure (CoM)



Paul Pierre Levy (1886-1971)

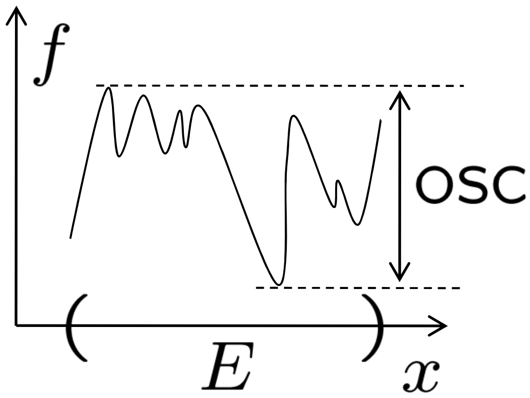
- CoM phenomenon (Levy, 1951): Functions over high-dimensional spaces with small local oscillations in each variable are almost constant
- CoM gives rise to a class of probability-of-failure inequalities that can be used for rigorous certification of complex systems

The *diameter* of a function

CALTECH PSAAP

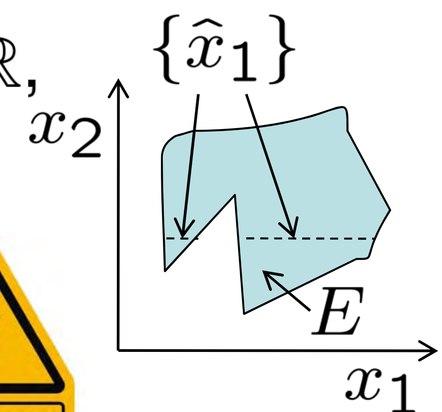
- Oscillation of a function of one variable:

$$\mathrm{osc}(f,E) = \sup_{x\in E} f(x) - \inf_{x\in E} f(x)$$
$$= \sup_{x,x'\in E} |f(x) - f(x')|$$



- Function subdiameters: $f : E \subset \mathbb{R}^N \to \mathbb{R}$,

$$D_i(f,E) = \sup_{\widehat{x}_i \in \mathbb{R}^{N-1}} \mathrm{osc}(f, E \cap \{\widehat{x}_i\}),$$
$$\widehat{x}_i = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N\}$$



- Function diameter:

$$D(f,E) = \sqrt{\sum_{i=1}^{N} D_i^2(f,E)}$$

evaluation requires global optimization!

McDiarmid's inequality

CALTECH PSAAP

ON THE METHOD OF BOUNDED DIFFERENCES

Colin McDiarmid

(1.2) Lemma: Let $X_1,\dots,X_n$ be independent random variables, with X_k taking values in a set A_k for each k. Suppose that the (measurable) function $f\colon \Pi A_k \to \mathbb{R}$ satisfies

(1.3)
$$|f(\underline{x}) - f(\underline{x}')| \le c_k$$

whenever the vectors $\underline{x}$ and $\underline{x}'$ differ only in the kth co–ordinate. Let Y be the random variable $f\left[X_1,\dots,X_n\right]$. Then for any $t > 0$,

$$P(|Y - E(Y)| \ge t) \le 2\exp\left[-2t^2/\Sigma c_k^2\right].$$

McDiarmid, C. (1989) "On the method of bounded differences". In J. Simmons (ed.), Surveys in Combinatorics*: London Math. Soc. Lecture Note Series* **141**. Cambridge University Press.

McDiarmid's inequality

Theorem [McDiarmid] *Suppose that:*

i) $\{x_1, \dots, x_N\}$ *are independent random variables,*

ii) $f : E \subset \mathbb{R}^N \to \mathbb{R}$ *is integrable.*

Then, for every $r \geq 0$

$$\mathbb{P}[|\, f - \mathbb{E}[f] \,| \geq r] \leq \exp\left(-2\frac{r^2}{D^2(f,E)}\right),$$

where $D(f,E)$ *is the diameter of* f *over* E*.*

- Bound does not require distribution of inputs
- Bound depends on two numbers only: Function ***mean*** and function ***diameter****!*

McDiarmid's inequality and QMU

Corollary *A conservative certification criterion is:*

$$\underbrace{\mathbb{P}[G \leq a]}_{\text{Probability of failure}} \leq \underbrace{\exp\left(-2\frac{(\mathbb{E}[G]-a)_+^2}{D_G^2}\right)}_{\text{Upper bound}} \leq \underbrace{\epsilon}_{\text{Failure tolerance}},$$

- Equivalent statement (confidence factor CF):

$$\text{CF} \equiv \frac{M}{U} \equiv \frac{(\mathbb{E}[G]-a)_+}{D_G} \geq \sqrt{\log\sqrt{\frac{1}{\epsilon}}} \Rightarrow \text{certification!}$$

- Rigorous definition of margin (*M*)
- Rigorous definition of uncertainty (*U*)

McDiarmid's inequality and QMU

CALTECH PSAAP

- CoM Uncertainty Quantification (UQ) 'does the job':
  - Rigorous upper bounds on PoFs for complex systems
  - Rigorous definitions of 'uncertainty' and 'margin'
  - Does not require knowledge of input parameters pdfs
  - Reduces UQ to determination of:
    - Mean performance $E[G]$
    - System diameter D_G
- But determination of response diameter is a global optimization problem over parameter space: Solution requires exceedingly many function evaluations
- Strictly experimental implementation is often impractical
- Alternative: ***Model-Based Uncertainty Quantification!***

Model-Based QMU

CALTECH PSAAP

$$\overbrace{(X_1,\ldots,X_M)}^{\text{system inputs}} \xrightarrow{\overbrace{\quad F \quad}^{\text{response function}}} \overbrace{(Y_1,\ldots,Y_N)}^{\text{performance measures}}$$

- Random variables
- Known or unknown pdfs
- Controllable, uncontrollable, unknown-unknowns

- Observables
- Subject to performance specs
- Random due to randomness of inputs or of system

System model

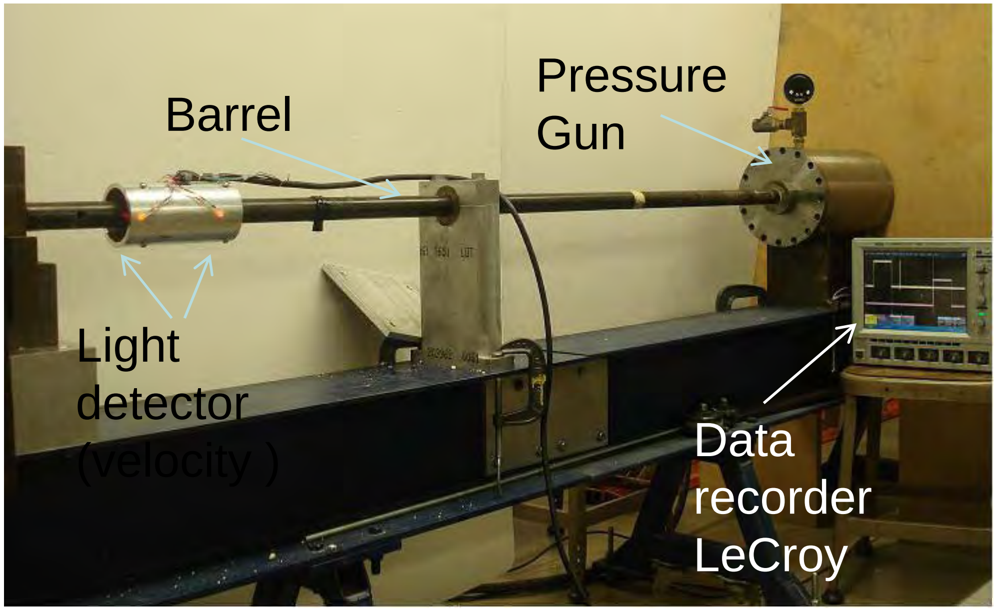
Model-based QMU – Perfect model

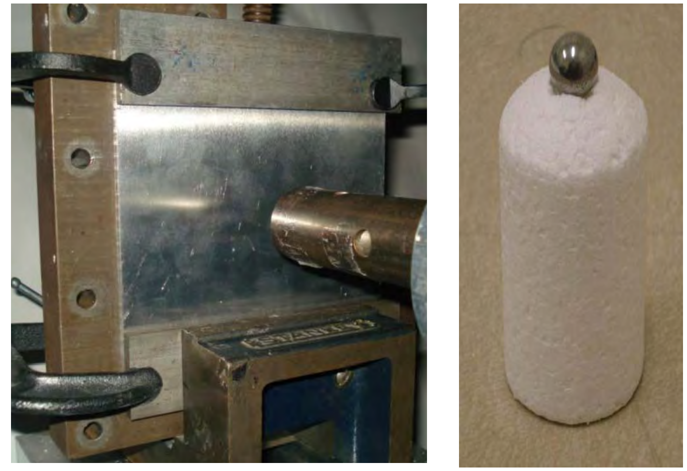
System model

- Assume ***deterministic system*** (no scatter)
- Assume ***model is perfect*** (F=G)
- Assume that *mean performance* and *system diameter* can be ***computed exactly***
- Then UQ can be carried out entirely in cyber-space, ***no experiments are required!***

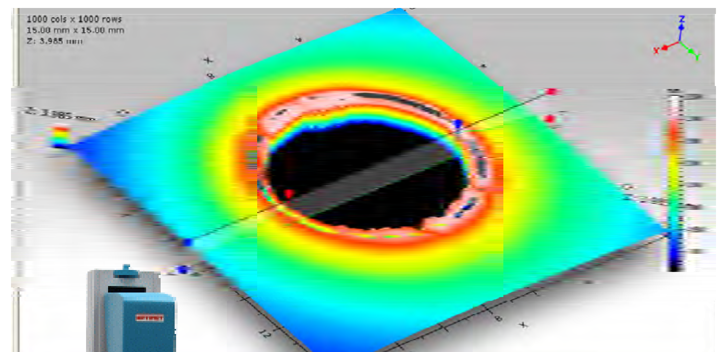
Case Study – Steel/Al ballistics

CALTECH PSAAP





Target and projectile

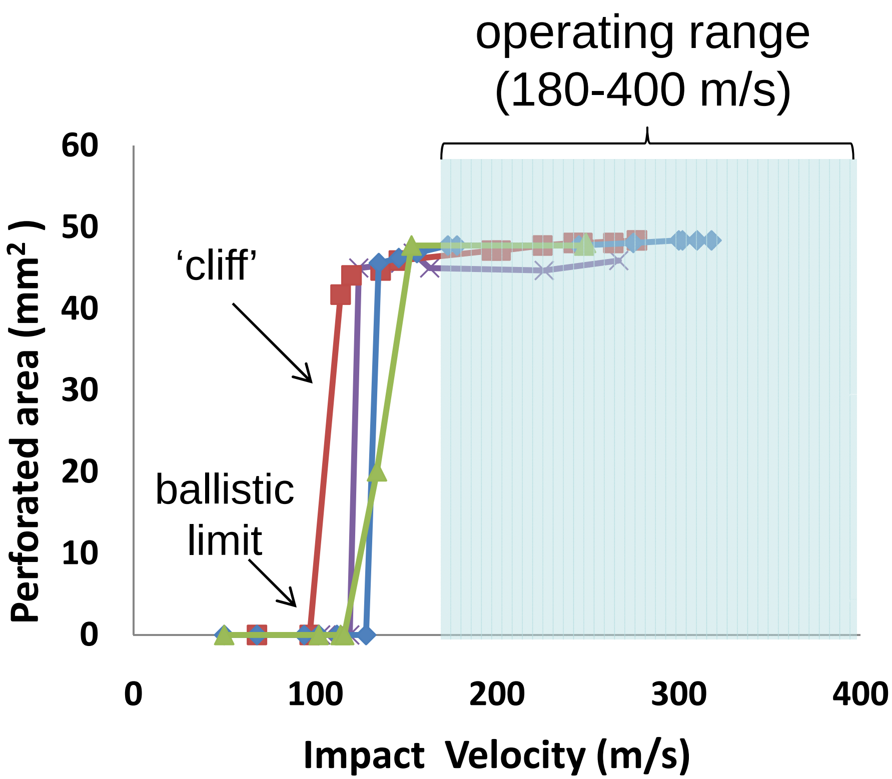


Optimet MiniConoscan 3000

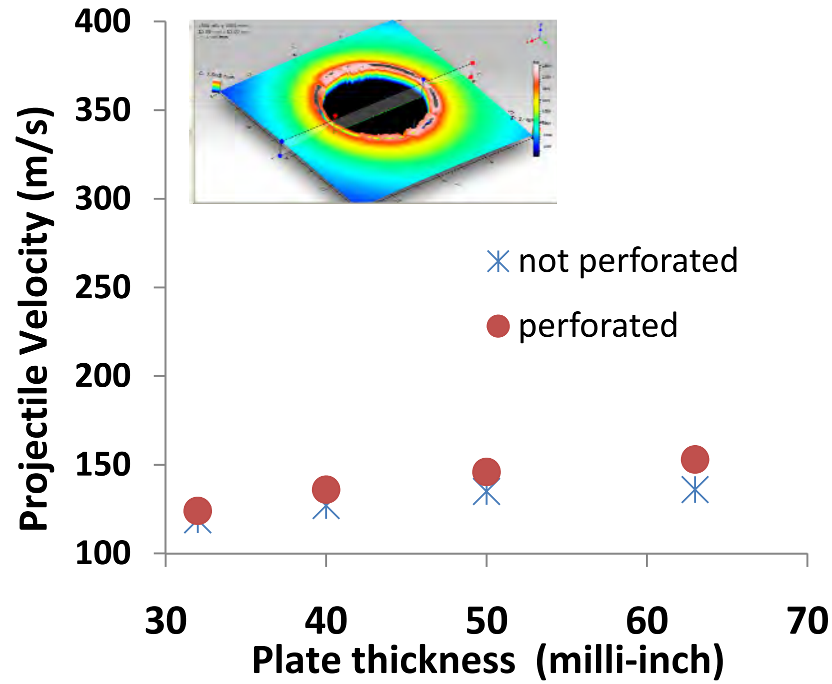
- Target/projectile materials:
  - Target: Al 6061-T6 plates (6"x 6")
  - Projectile: S2 Tool steel balls (5/16")
- Model input parameters (*X*):
  - Plate thickness (0.032''-0.063'')
  - Impact velocity (200-400 m/s)

Case Study – Steel/Al ballistics

CALTECH PSAAP



Perforation area *vs*. impact velocity (note small data scatter!)



Perforation/non-perforation boundary

- System output (*Y*): ***Perforation area!***
- Certification criterion: Y>0 (lethality)

Michael Ortiz

Optimal-Transportation Meshfree (OTM) model of terminal ballistics

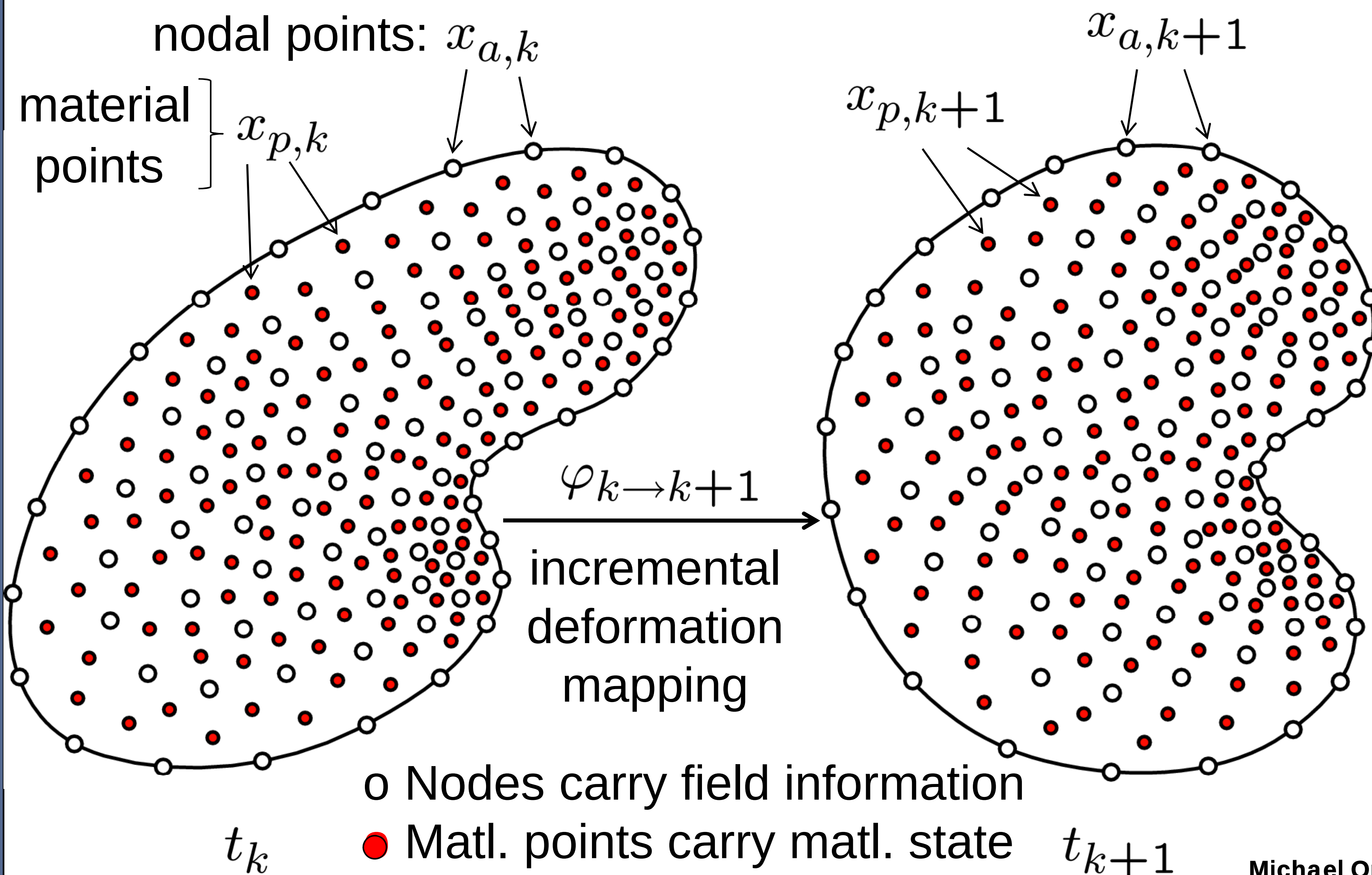
CALTECH PSAAP

- Optimal transportation theory is a useful tool for generating geometrically-exact discrete Lagrangians for flow problems
- Inertial part of discrete Lagrangian measures distance between consecutive mass densities (in sense of Wasserstein)
- Discrete Hamilton principle of stationary action: Variational time integration scheme:
  - Symplectic, time reversible, exact conservation
  - Variational convergence (Γ-convergence, B. Schmidt)
- Extension to inelasticity: Variational constitutive updates

Li, B., Habbal, F. and Ortiz, M., *IJNME*, **83** (2010) 1541-1579

OTM – Spatial discretization

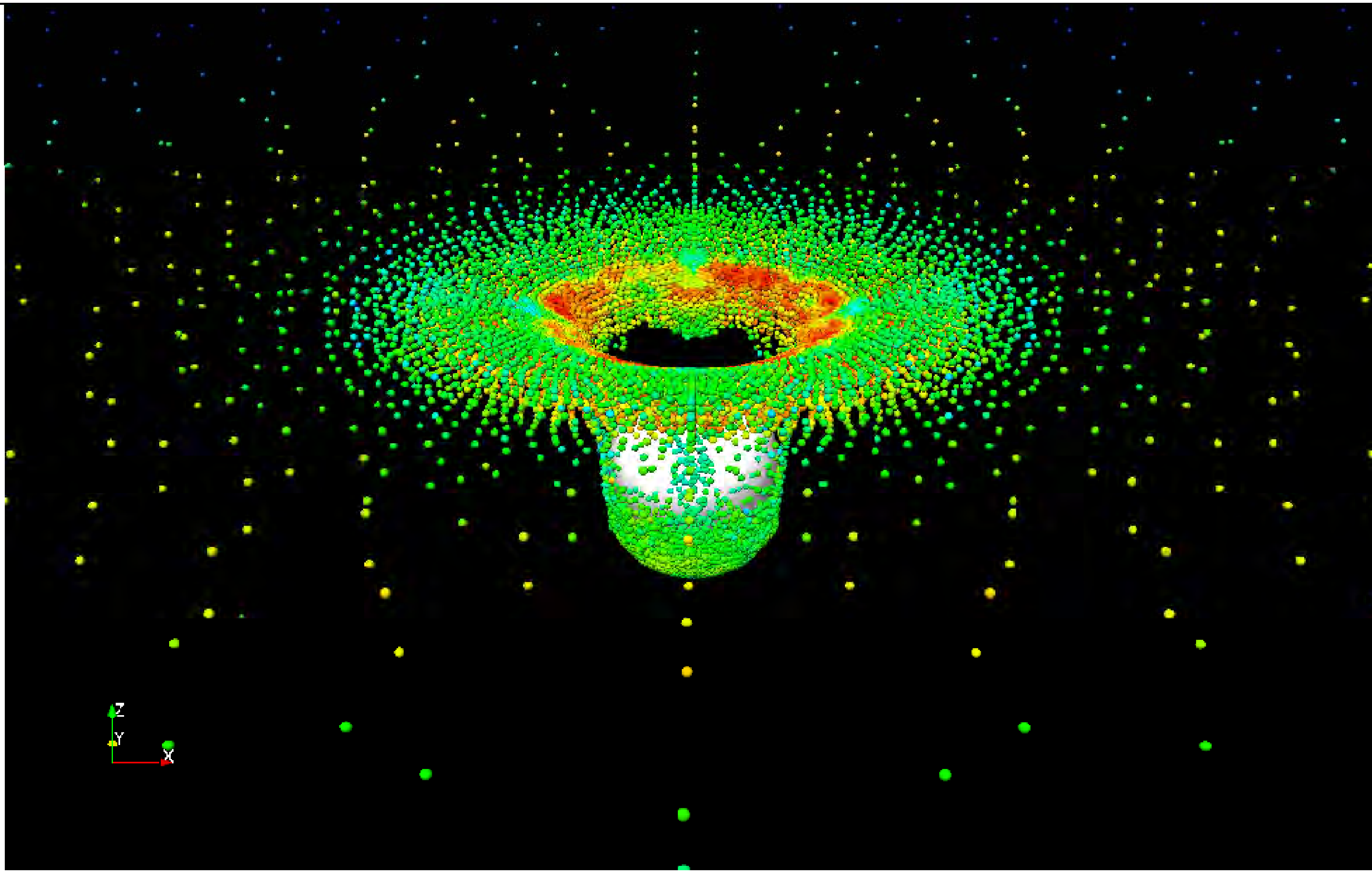
CALTECH PSAAP



o Nodes carry field information

● Matl. points carry matl. state

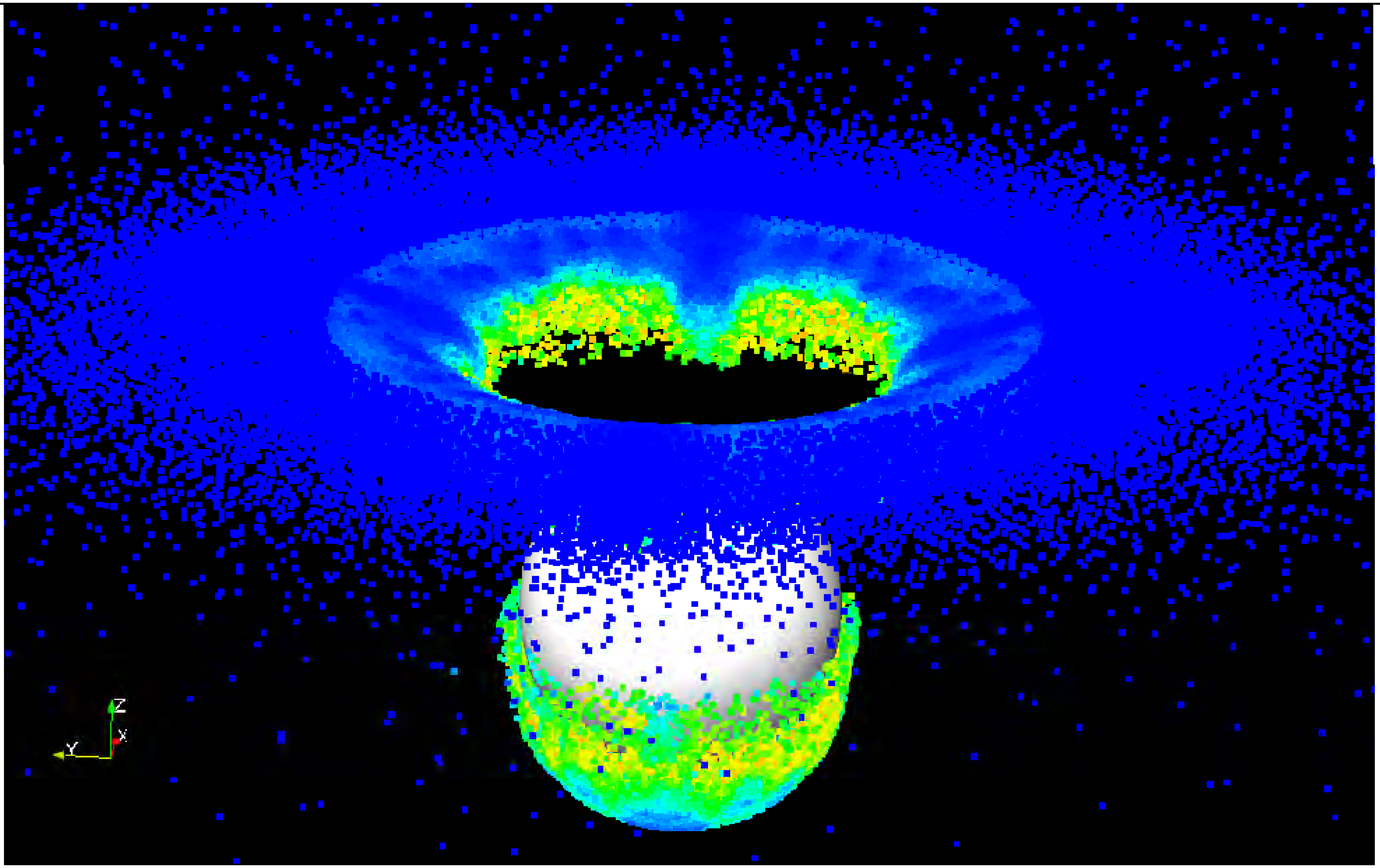
OTM – Nodal point set



Steel projectile/aluminum plate: Nodal set

OTM – Material point set

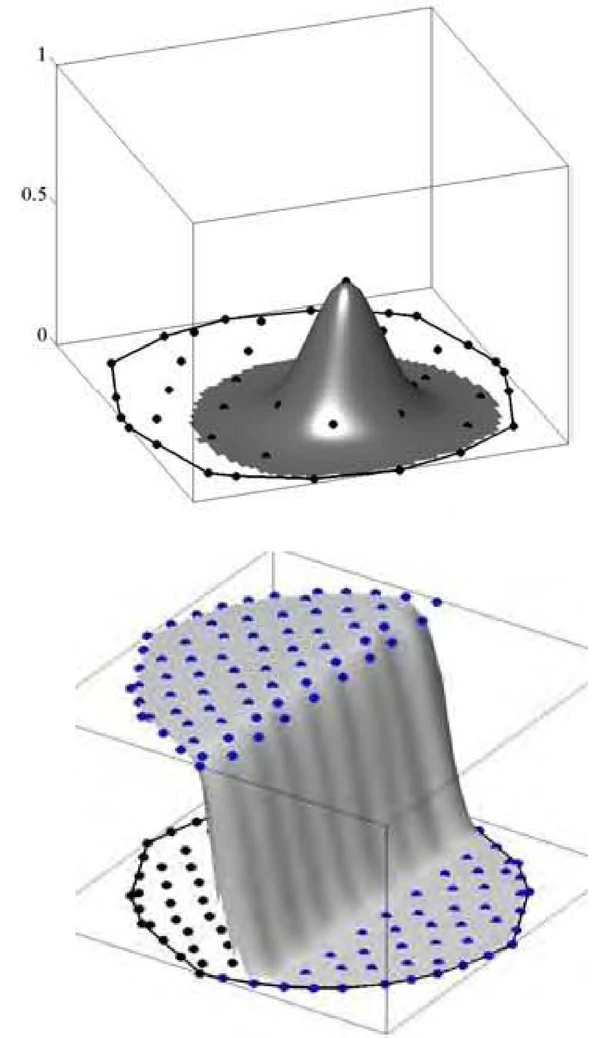
CALTECH PSAAP



Steel projectile/aluminum plate: Material point set

OTM — Max-ent interpolation

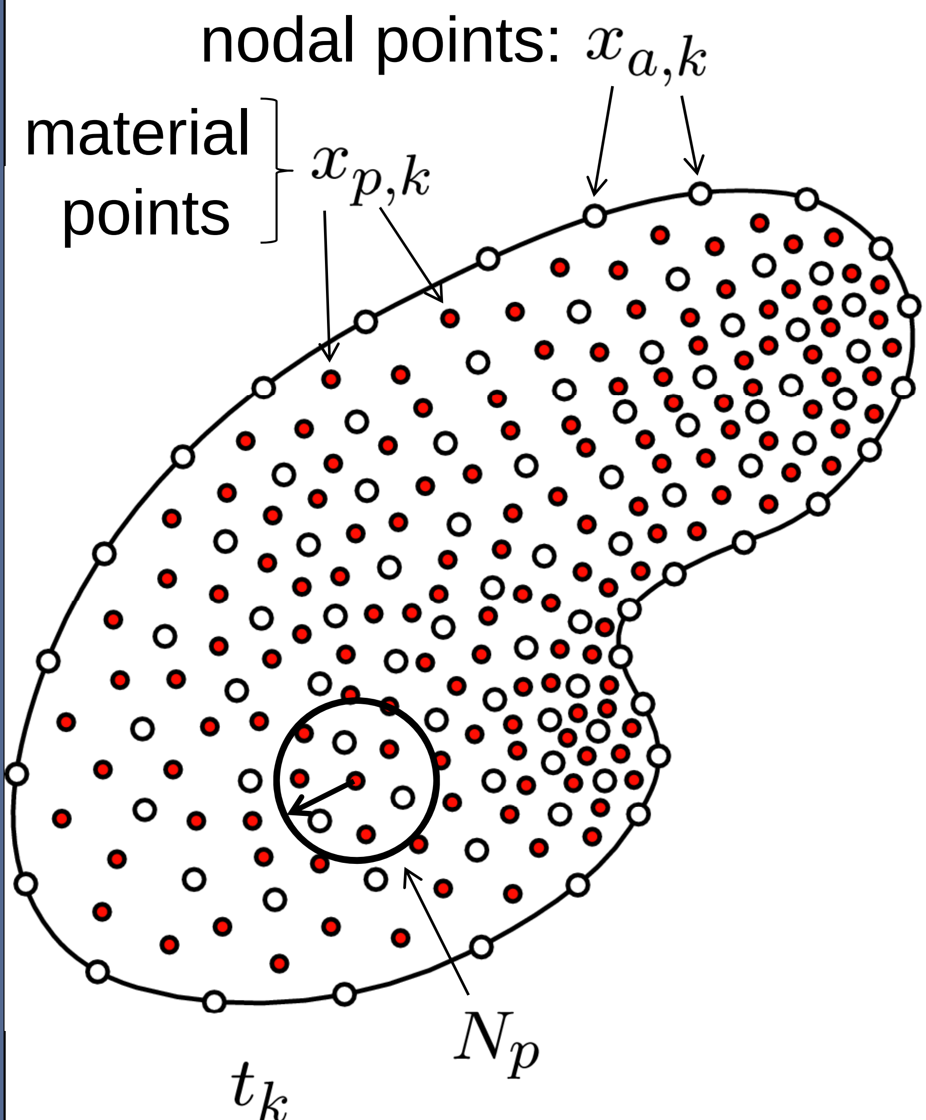
- Max-ent interpolation is smooth, meshfree
- Finite-element interpolation is recovered as a limit
- Rapid decay, short range
- Monotonicity, maximum principle
- Good mass lumping properties
- Kronecker-delta property at the boundary:
  - Displacement boundary conditions
  - Compatibility with finite elements



Arroyo, M. and Ortiz, M., *IJNME*, **65** (2006) 2167-2202

OTM — Spatial discretization

CALTECH PSAAP



- Max-ent interpolation at material point *p* determined by nodes in its local environment *Np*
- Local environments determined 'on-the-fly' by range searches
- Local environments evolve continuously during flow (dynamic reconnection)
- Dynamic reconnection requires no remapping of history variables!

Michael Ortiz

OTM – Flow chart

(i) Explicit nodal coordinate update:

$$x_{k+1} = x_k + (t_{k+1} - t_k)(v_k + \frac{t_{k+1} - t_{k-1}}{2} M_k^{-1} f_k)$$

(ii) Material point update:

position: $x_{p,k+1} = \varphi_{k\to k+1}(x_{p,k})$

deformation: $F_{p,k+1} = \nabla\varphi_{k\to k+1}(x_{p,k}) F_{p,k}$

volume: $V_{p,k+1} = \det \nabla\varphi_{k\to k+1}(x_{p,k}) V_{p,k}$

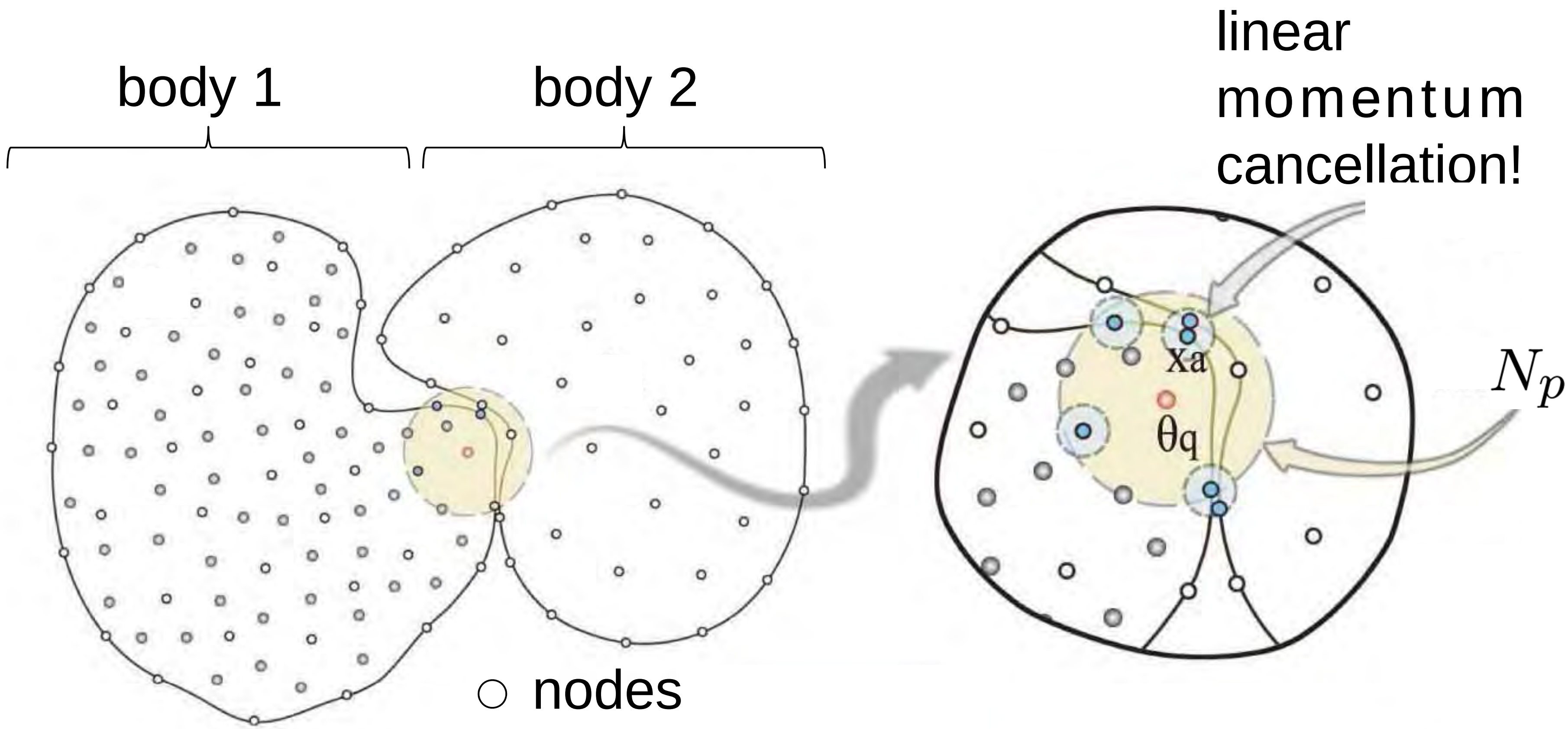
density: $\rho_{p,k+1} = m_p / V_{p,k+1}$

(iii) Constitutive update at material points

(iv) Reconnect nodal and material points (range searches), recompute max-ent shape functions

Michael Ortiz

OTM – Seizing contact

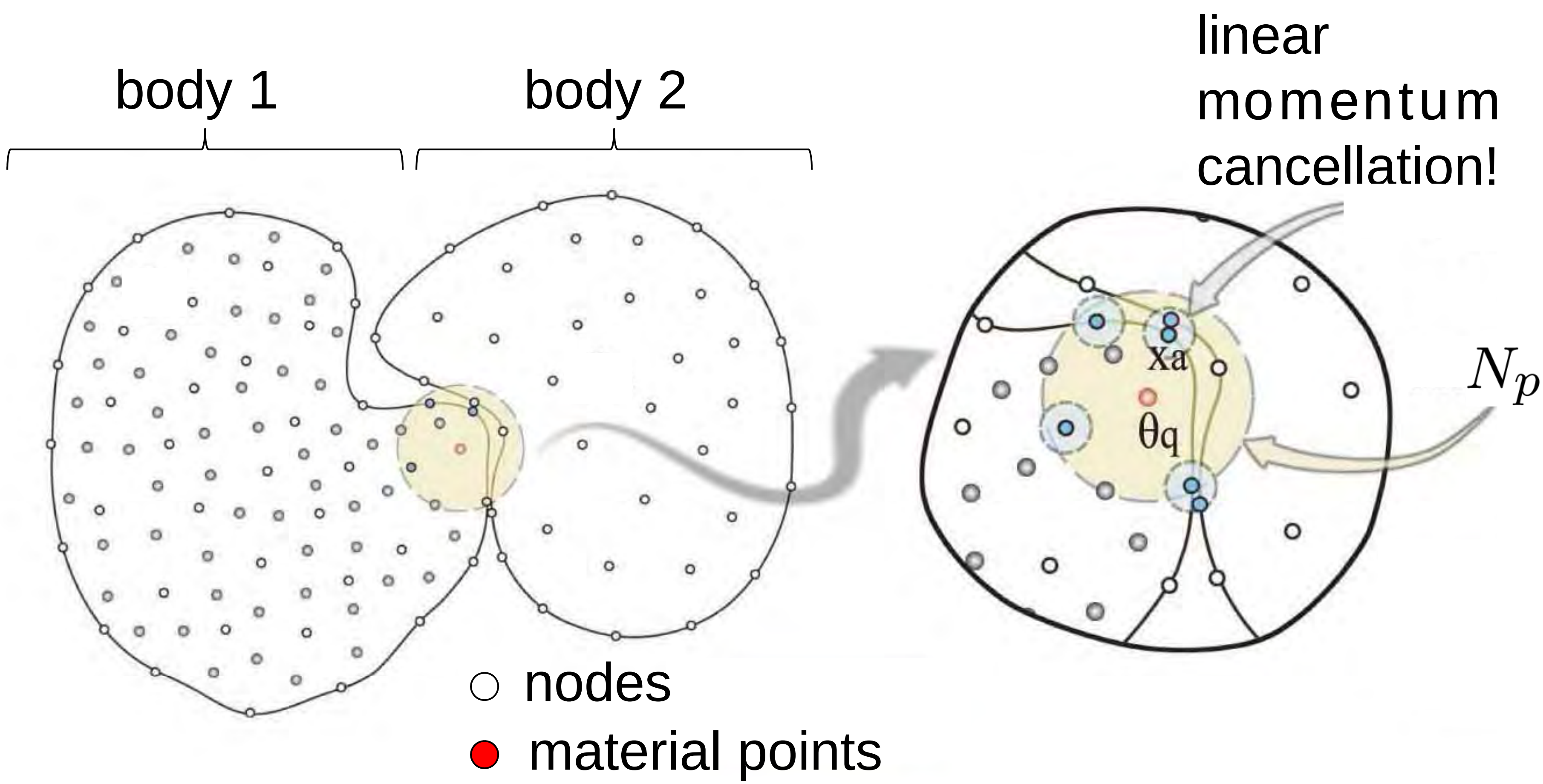


○ nodes

● material points

Seizing contact (infinite friction) is obtained for free in OTM! (as in other material point methods)

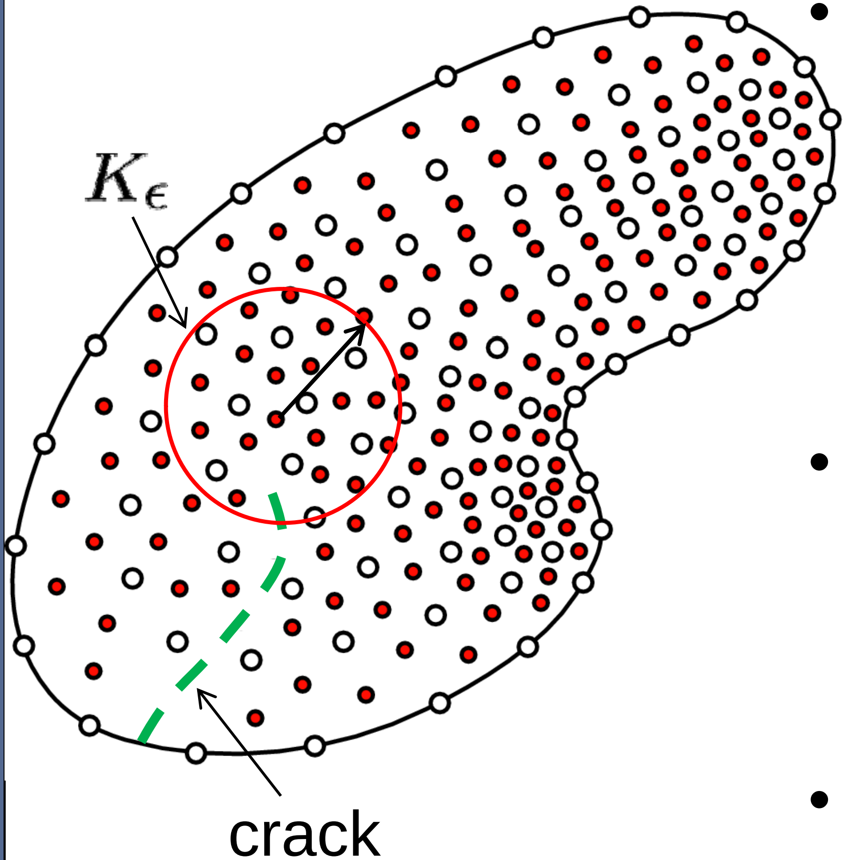
OTM — Seizing contact



Seizing contact (infinite friction) is obtained for free in OTM! (as in other material point methods)

Michael Ortiz

OTM - Fracture & fragmentation



$$G_\epsilon \sim \frac{h^2}{|K_\epsilon|} \int_{K_\epsilon} W(\nabla u)\, dx$$

- Proof of convergence of variational element erosion to Griffith fracture:
  - Schmidt, B., Fraternali, F. and Ortiz, M., SIAM J. Multiscale Model. Simul., **7**(3) (2009) 1237-1366.
- OTM implentation: Variational erosion of material points (by ε-neighborhood construction),

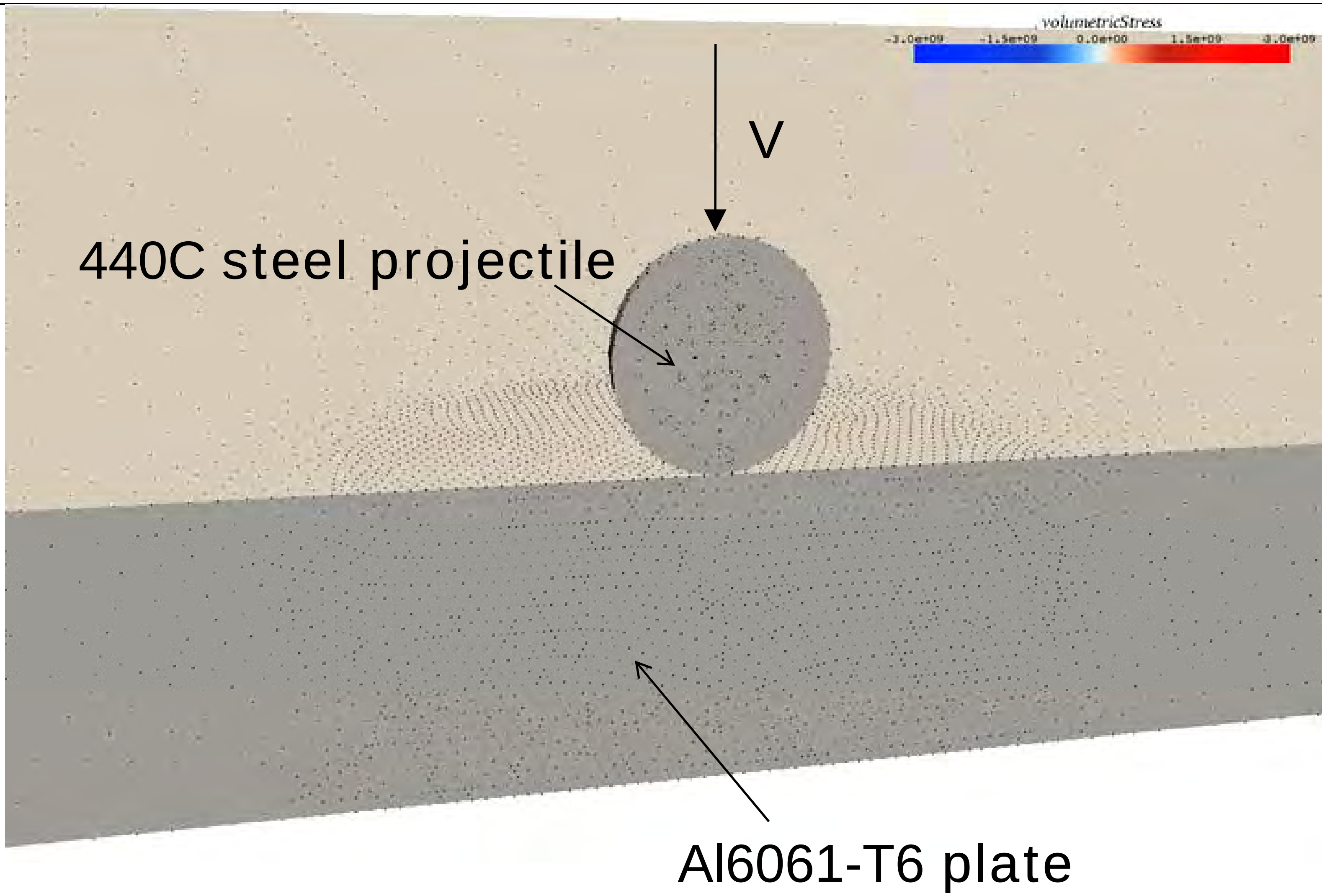
$$G_\epsilon \geq G_c$$

- Alternatively: Material point failure + comminution:
  - Pandolfi, A., Conti, S. and Ortiz, M., JMPS, **54** (2006) 1972-2003

Michael Ortiz

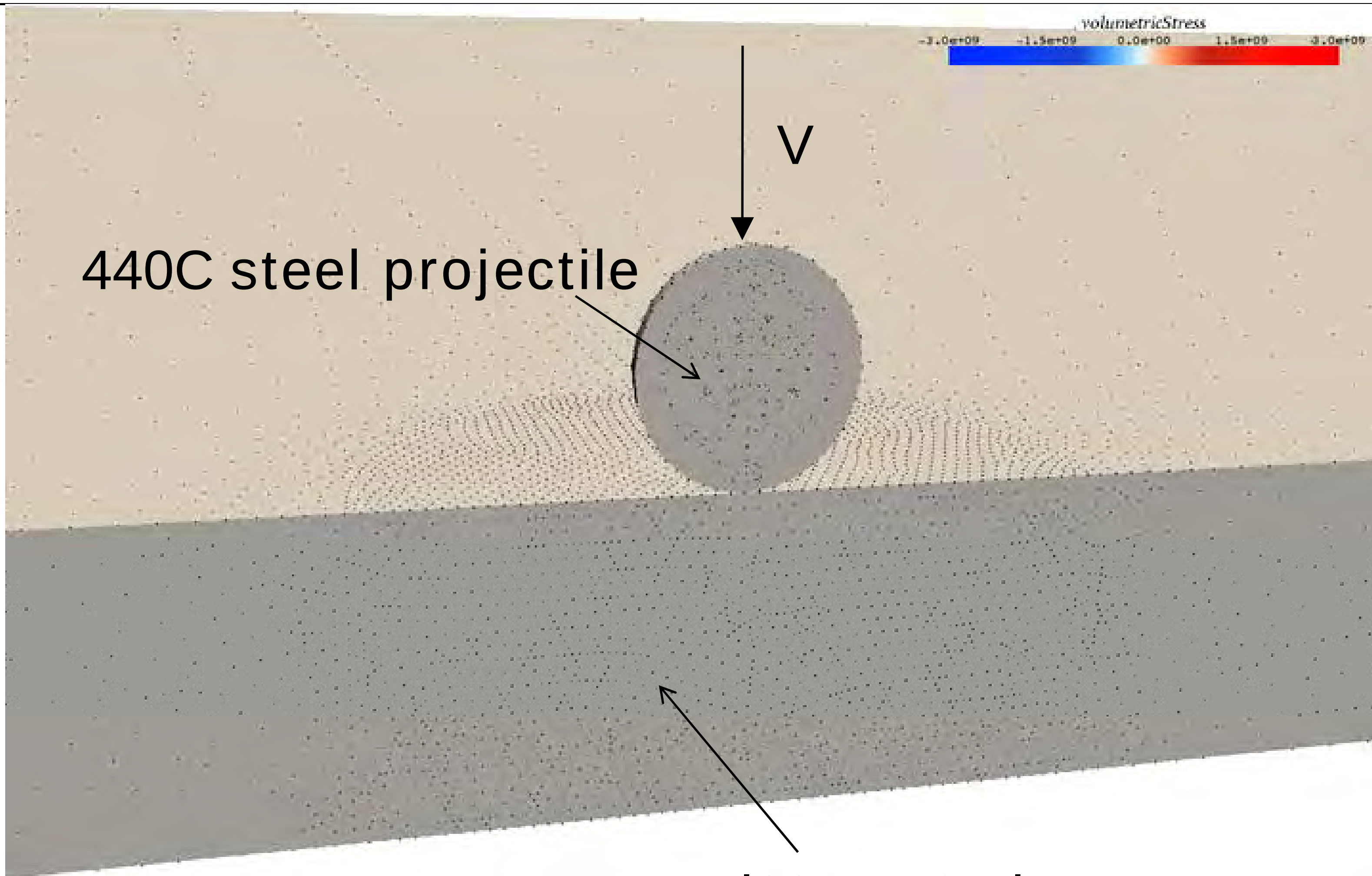
OTM – Terminal ballistics

CALTECH PSAAP



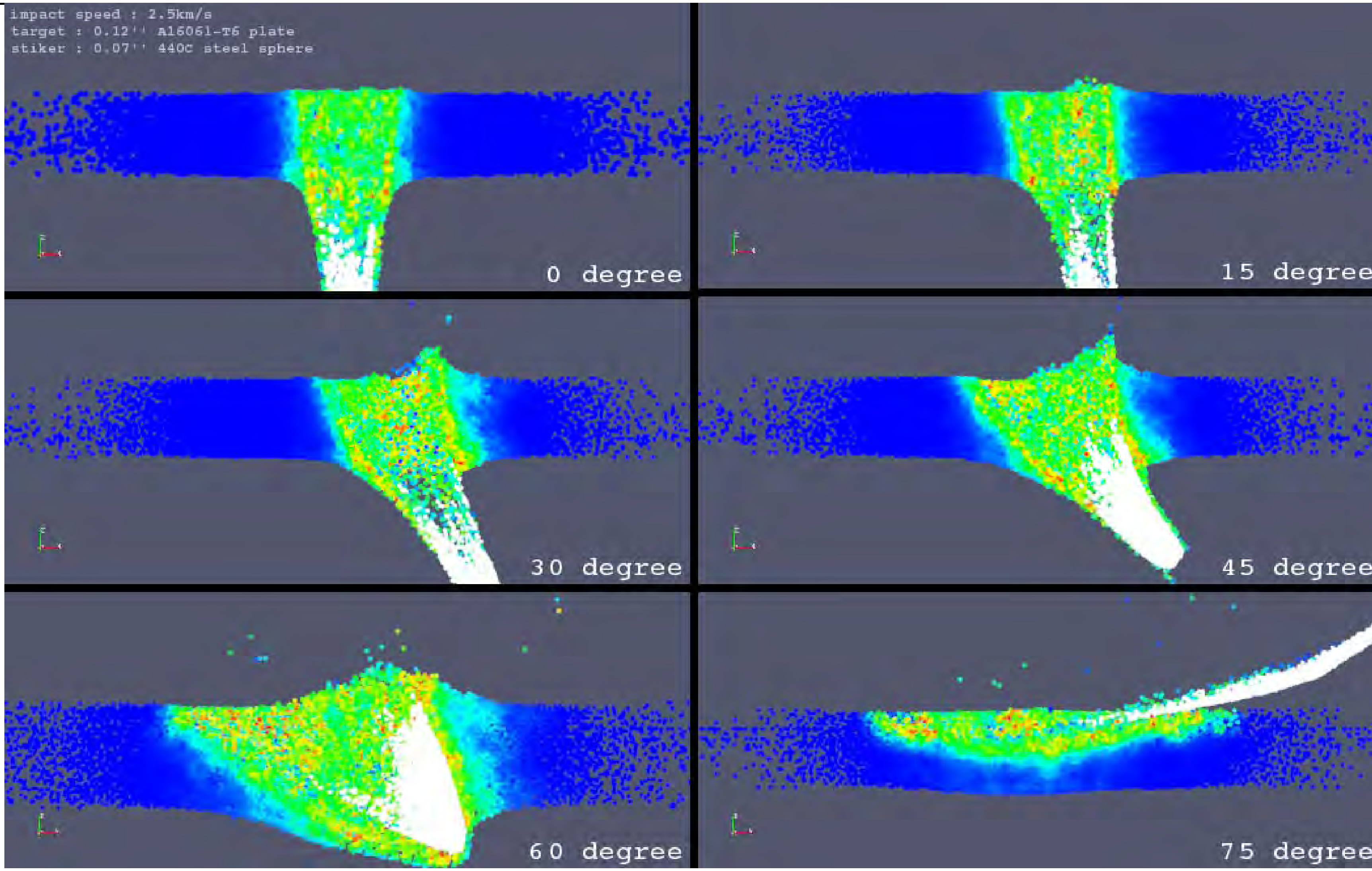
Michael Ortiz

OTM – Terminal ballistics



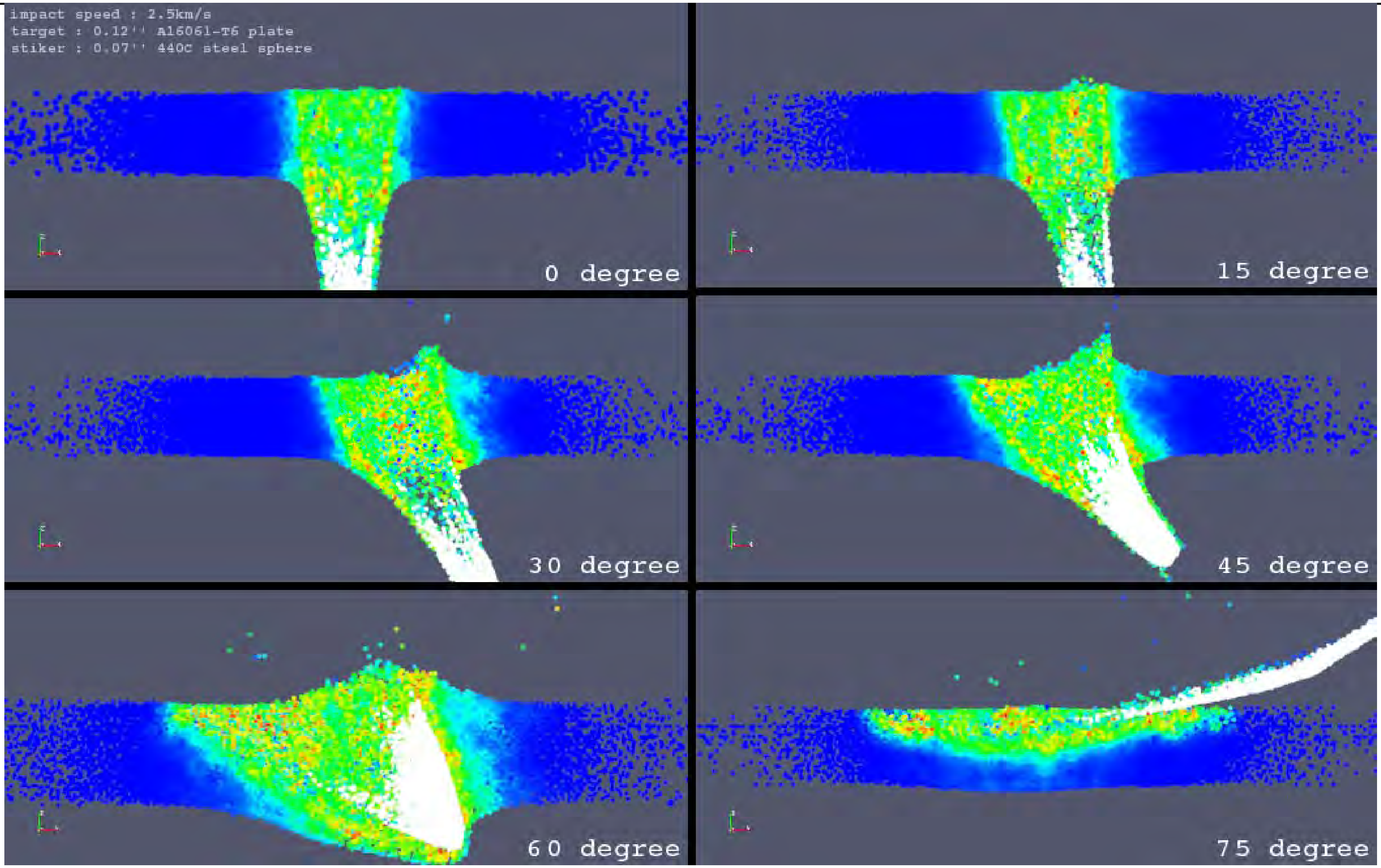
Michael Ortiz

OTM – Terminal ballistics



Michael Ortiz

OTM – Terminal ballistics



Case Study I – Steel/Al ballistics

CALTECH PSAAP

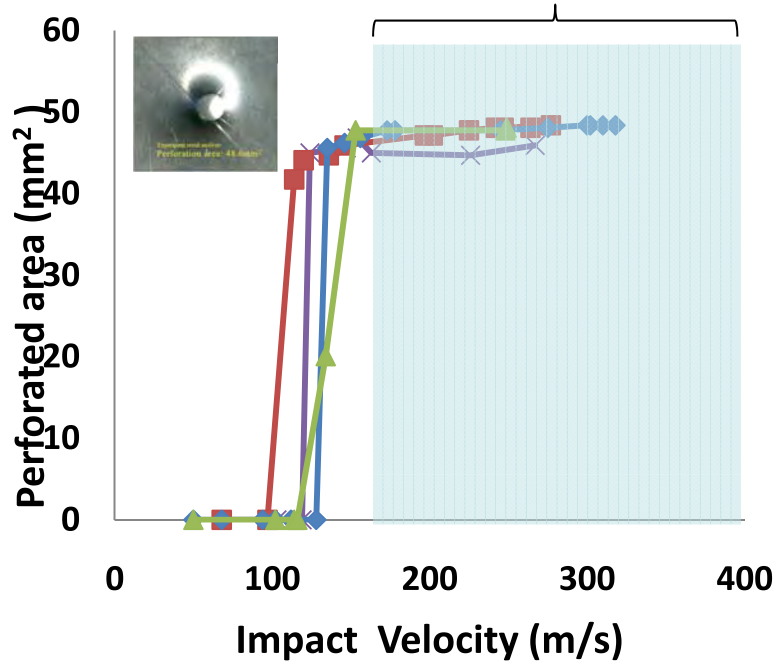
- McDiarmid inequality:

$$\mathrm{PoF} = \mathbb{P}[F \le a] \le \mathrm{e}^{-2\mathrm{CF}^2}$$

$$CF = \frac{M}{U} = \frac{(\mathbb{E}[F]-a)_+}{D_F}$$

Model diameter D_F	thickness	4.33 mm^2
	velocity	4.49 mm^2
	total	6.24 mm^2
Model mean *E[F]*		47.77 mm^2
Confidence factor *M/U*		7.66

operating range (180-400 m/s)



Steel/Al ballistics response function

- Lethality can be certified with ~ 10^{-51} confidence!
- Number of response function evaluations ~ 2,000

Uncertainty quantification ‘crimes’

- Models are inexact in general!
- How does lack of model fidelity contribute to uncertainty?
- Is rigorous model-based certification possible in the face of modeling error?
- Mean performance *E[F]* cannot be computed exactly for complex systems
- Instead, mean performance *E[F]* is approximated by empirical mean:

$$\mathbb{E}[G] \approx \frac{1}{m}\sum_{i=1}^{m} Y^i \equiv \langle Y \rangle$$

- What is the effect of the empirical mean approximation on uncertainty quantification?

UQ crime and punishment

CALTECH PSAAP

- Two functions that describe the system:
  - Experiment: *G(X)*
  - Model: *F(X)*

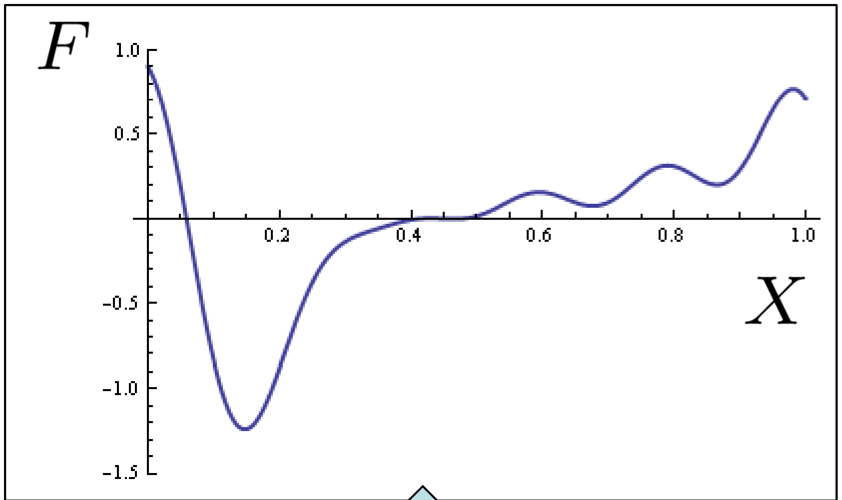
  F(X)-G(X) ≡ Modeling-error function
- McDiarmid bound monotonic in diameter
- Triangular inequality: $D_G \leq D_F + D_{F-G}$
- Conservative certification criterion:

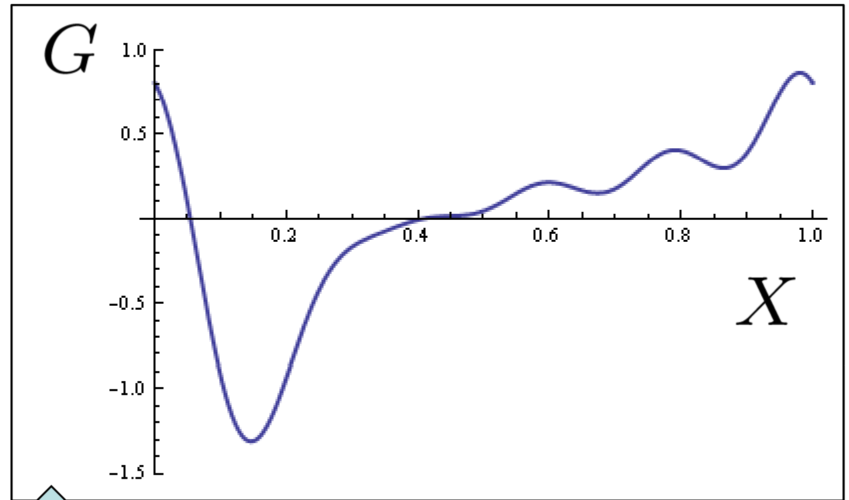
$$\mathbb{P}[G \leq a] \leq \exp\left(-2\frac{(\langle Y\rangle - a + \alpha)_+^2}{(D_F + D_{F-G})^2}\right) \leq \epsilon,$$

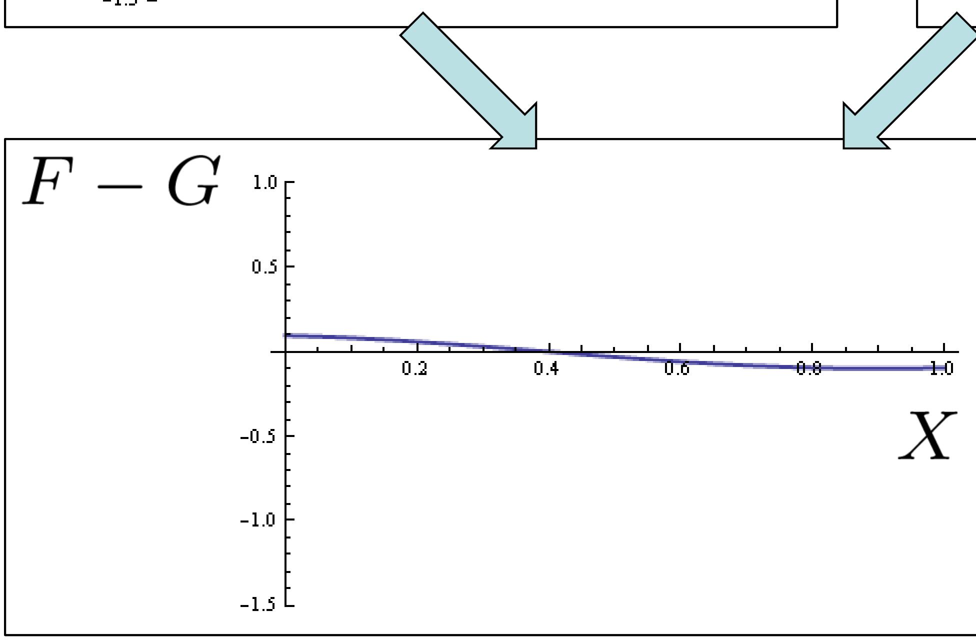
- $\alpha = U m^{-\frac{1}{2}}(-\log \epsilon')^{\frac{1}{2}}$ *: **Margin hit** (emp. mean)*
- D_F: ***Model diameter*** (*variability* of model)
- D_{F-G}: ***Modeling error*** (*badness* of model)

Model-based QMU – McDiarmid

CALTECH PSAAP







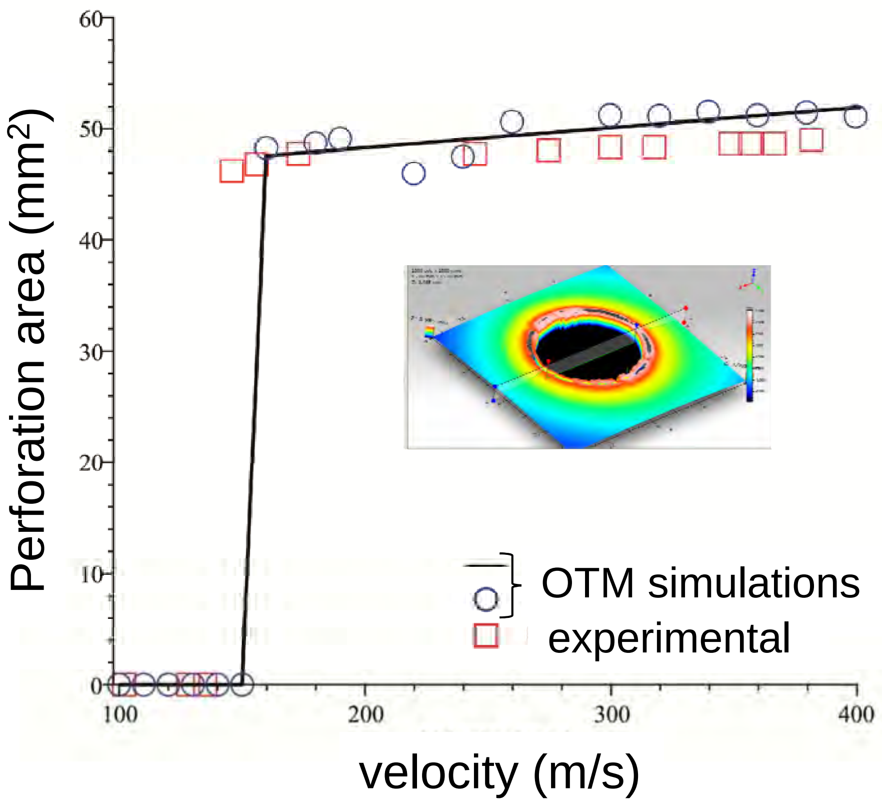
- Working assumptions:
  - *F-G* far more regular than *F* or *G* alone
  - Global optimization for $D_{F\text{-}G}$ converges fast (e.g. BFGS)
  - Evaluation of $D_{F\text{-}G}$ requires few experiments

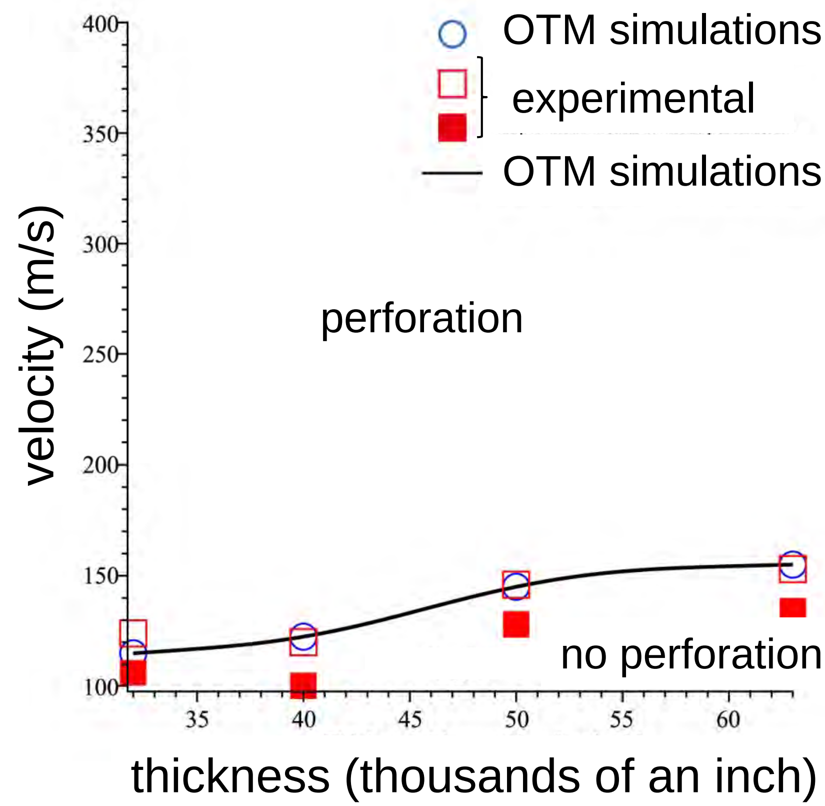
Model-based QMU – McDiarmid

- Calculation of D_F requires exercising model only
- Uncertainty Quantification burden mostly shifted to modeling and simulation!

- Evaluation of $D_{F\text{-}G}$ requires (few) experiments
- Rigorous certification not achievable by modeling and simulation alone!

Case Study – OTM modeling error



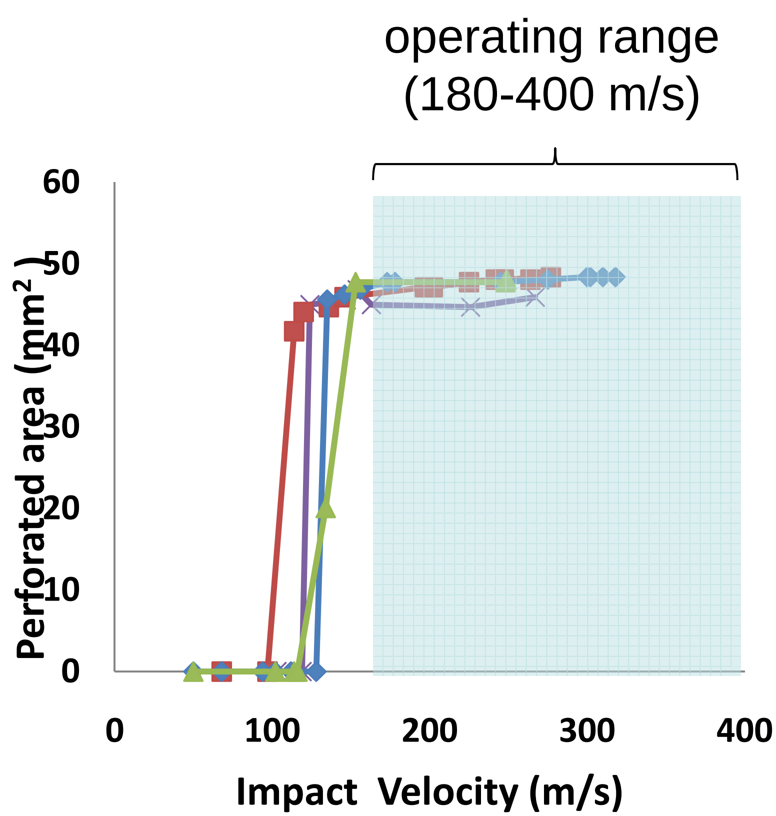


Measured *vs*. computed perforation area

Sample UQ Analysis – Ballistic range

CALTECH PSAAP

Model diameter D_F	thickness	4.33 mm^2
	velocity	4.49 mm^2
	total	6.24 mm^2
Modeling error $D_{F\text{-}G}$	thickness	4.96 mm^2
	velocity	2.16 mm^2
	total	5.41 mm^2
Uncertainty $D_F + D_{F\text{-}G}$		11.65 mm^2
Empirical mean $\langle Y \rangle$		47.77 mm^2
Margin hit α (ε'=0.1%)		4.17 mm^2
Confidence factor M/U		3.74



- Perforation can be certified with ~ 1-10^{-12} confidence!
- Total number of experiments ~ 50 → Approach feasible!

Michael Ortiz

Beyond McDiarmid - Extensions

- A number of extensions of McDiarmid may be required in practice:
  - Some input parameters cannot be controlled
  - There are unknown input parameters (unknown unknowns)
  - There is experimental scatter (*G* defined in probability)
  - McDiarmid may not be tight enough (convergence?)
  - Model itself may be uncertain (epistemic uncertainty)
  - Data may not be available 'on demand' (legacy data)
- Extensions of McDiarmid that address these challenges include:
  - Martingale inequalities (unknown unknowns, scatter…)
  - Partitioned McDiarmid inequality (convergent upper bounds)
  - Optimal Uncertainty Quantification (OUQ)
  - Optimal models (least epistemic uncertainty)

Concluding remarks…

CALTECH PSAAP

- QMU represents a paradigm shift in predictive science:
  - Emphasis on predictions with *quantified uncertainties*
  - Unprecedented integration between simulation and experiment
- QMU supplies a powerful organizational principle in predictive science: *Theorems run entire centers*!
- QMU raises theoretical and practical challenges:
  - Tight and measureable/computable probability-of-failure upper bounds (need theorems!)
  - Efficient global optimization methods for highly non-convex, high-dimensionality, noisy functions
  - Effective use of massively parallel computational platforms, heterogeneous and exascale computing
  - High-fidelity models (multiscale, effective behavior…)
  - Experimental science for UQ (diagnostics, rapid-fire testing…)…

Concluding remarks…