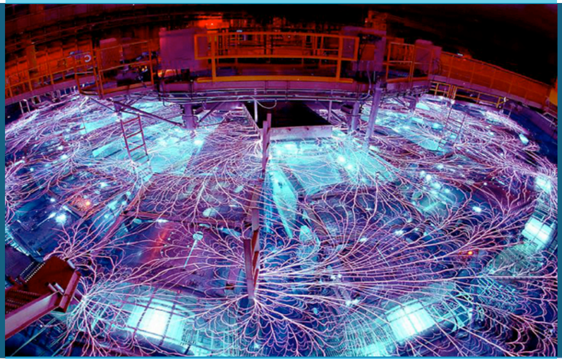
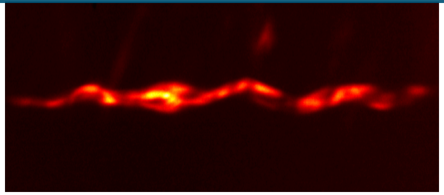
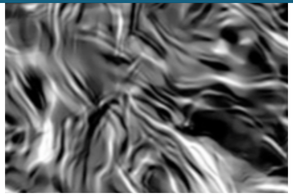
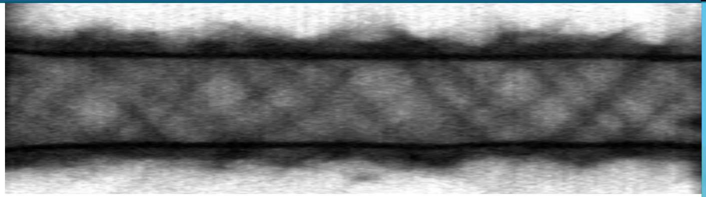
Data and Decision Science at Sandia National Labs









PRESENTED BY

Michael Glinsky

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2019-13763 PE

People involved:

- Michael Glinsky (0.5 FTE)
- Pat Knapp (0.3 FTE)
- Marc-Andre Schaeuble (0.25 FTE)
- Will Lewis (1 FTE)
- Evstati Evstatiev (0.25 FTE)
- Nikki Bennet (0.25 FTE)
- James Gunning (CSIRO, 0.15 FTE)
- Taisuke Nagayama (0.1 FTE)
- Chris Jennings (0.1 FTE)
- Brandon Klein (0.25 FTE)
- Amir Barati Farimani (ASC/CMU, 1 grad student, 1 undergrad student)
- Matt Martinez (0.25 FTE)
- Justin Brown (0.5 FTE)

Important characteristics of SNL approach

- Data Science feeds Decision Science
  - See MatrixDS: https://tinyurl.com/DAAG19-MatrixDS
  - Data Science provides statistical estimates of risks and uncertainties, inputs to Decision Science
  - Decision Science uses interview techniques based on "wisdom of crowds", essentially "bookmaker odds" for other risks and uncertainties
- Bayesian assimilation engine is at the core
  - Uses all experimental information, with optional simulation constraints
  - MLDL surrogates for physics of diagnostics
  - Estimates risks and uncertainties (covariance)
  - Estimates value of information (sensitivities of outputs to inputs, cross variance)
- **Focus on deficiency in model**
  - Largest uncertainty, probable bias, and significant distortion of PDF
  - Monitor diagnostics
  - Use of Mallat Scattering Transformation to keep "on manifold", topological curvature
  - Research on causal statistics (CMU)
- Python based, leveraging expertise of petroleum industry
- Researching fast surrogates for rad-MHD simulations (ASC funding of CMU)
  - cGAN and MST (state and transition kernel)
- Recognize need for "data lake" in the cloud

The layers of the paradigm

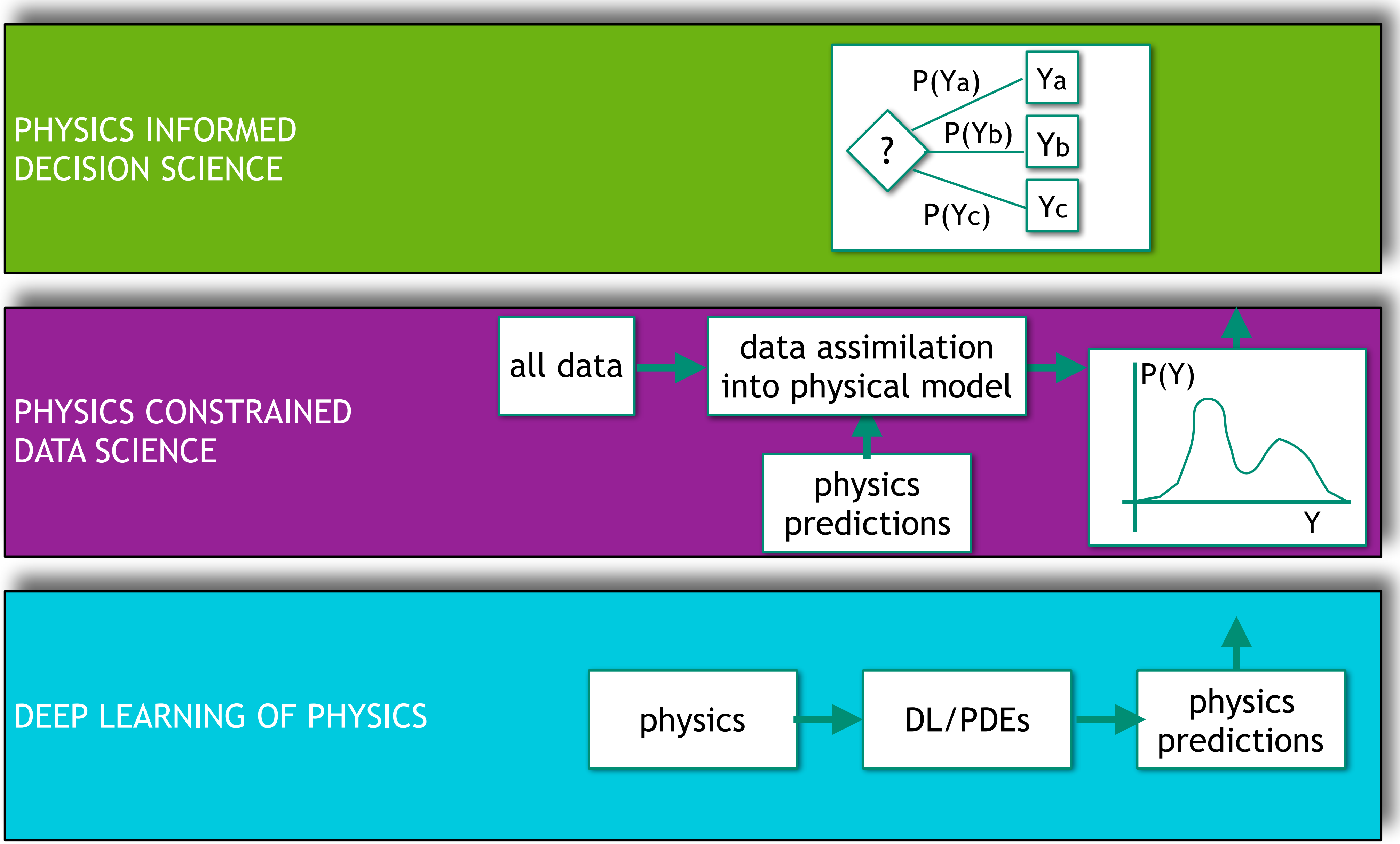
decision

=

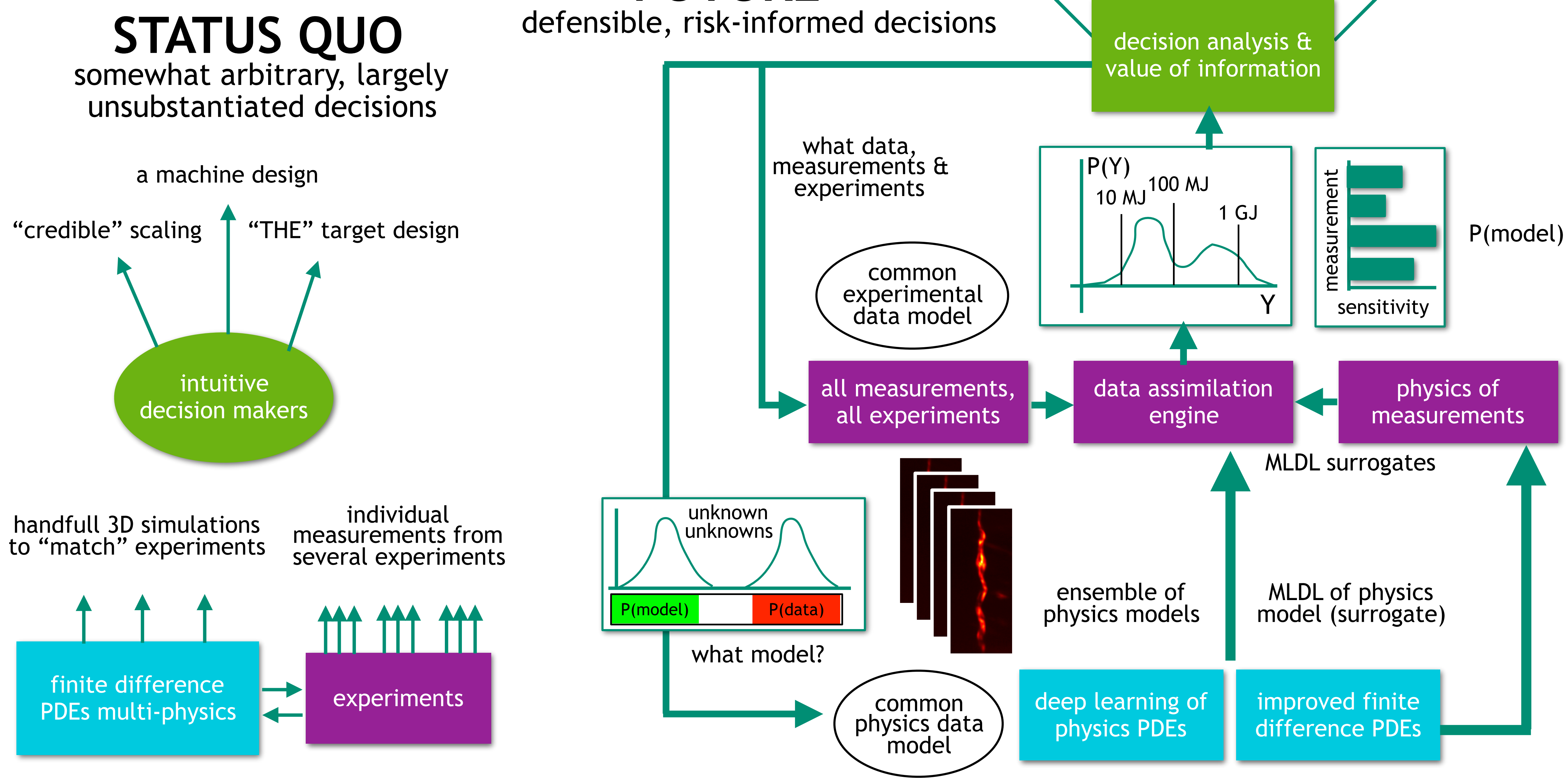
experimental data

+

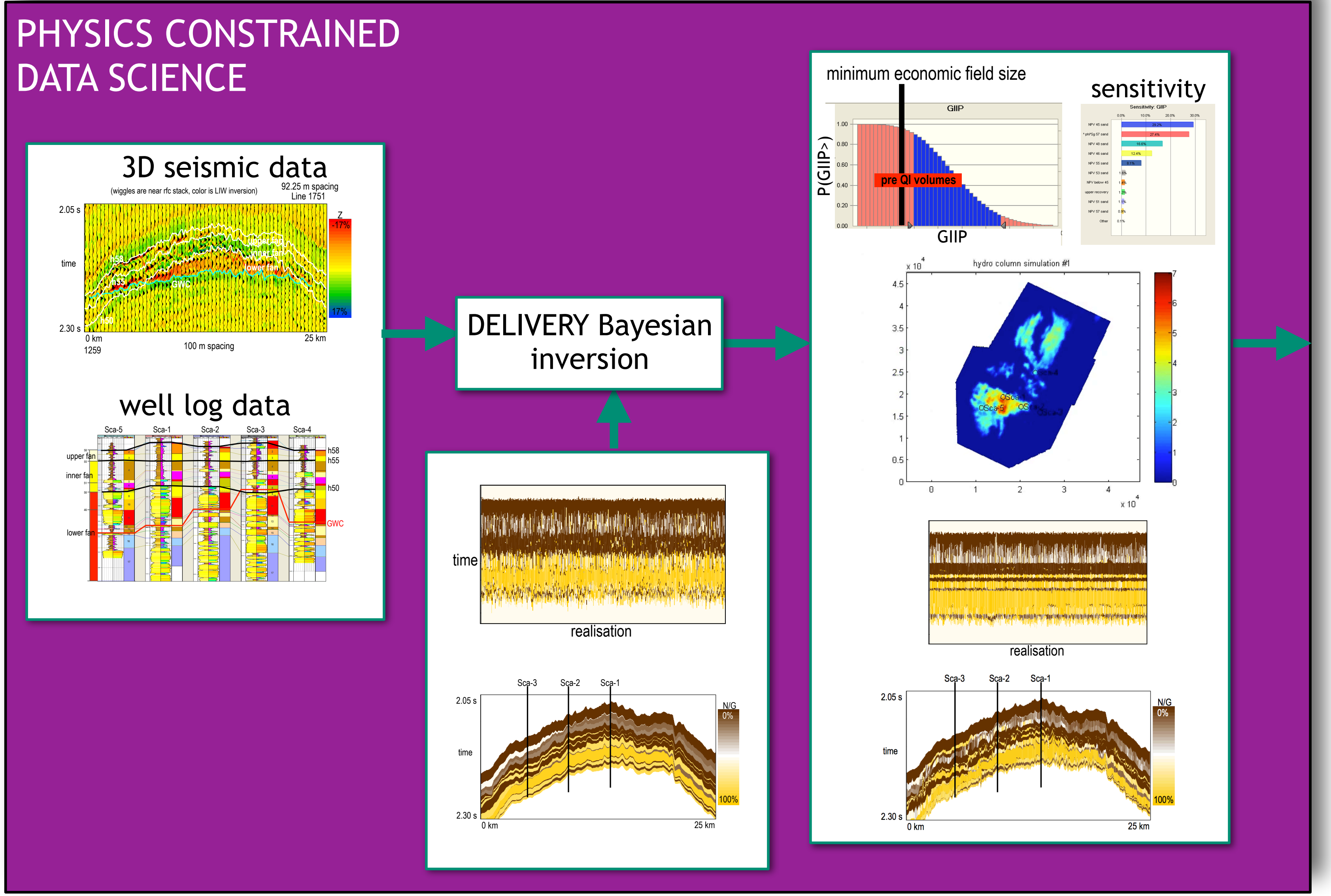
theoretical physics

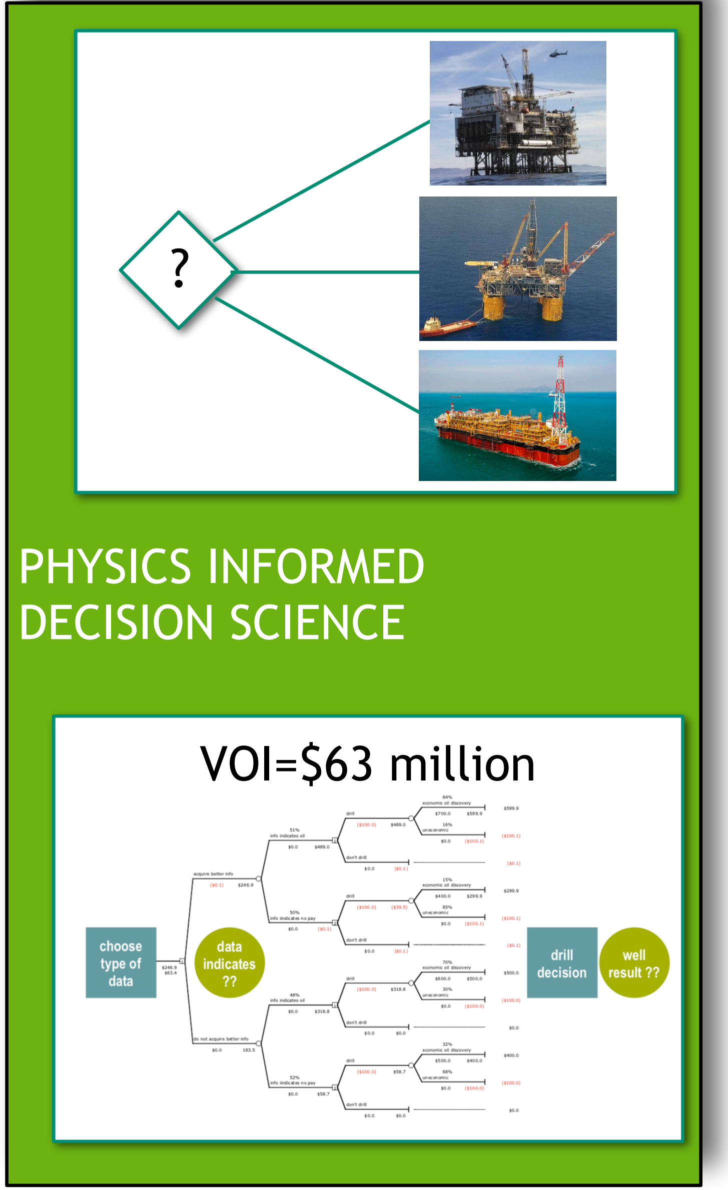


Status Quo and Future



Realized petroleum and mining technology



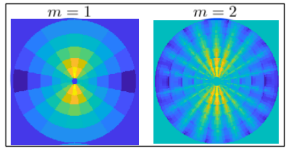


Early, proof-of-concept for NGPPF

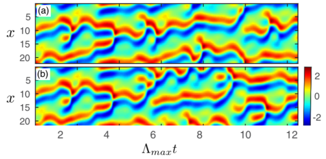
DEEP LEARNING OF PHYSICS

MST transformation that conserves physical diffeomorphism & group symmetries

$\varphi_0(\lambda)$ $S_2^{-1}(\lambda, \lambda')$



finite difference solution

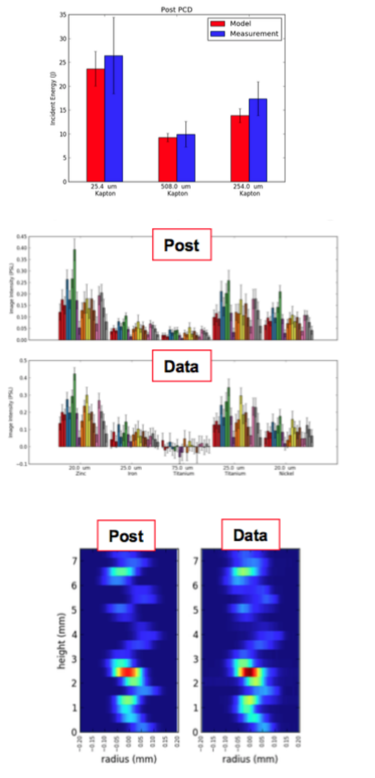


machine learning solution

Kuramoto-Sivashinsky equation
Pathak et al., PRL **120**, 024102 (2018)

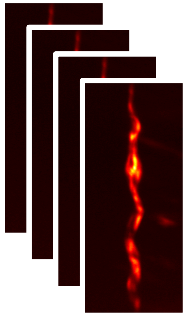
PHYSICS CONSTRAINED DATA SCIENCE

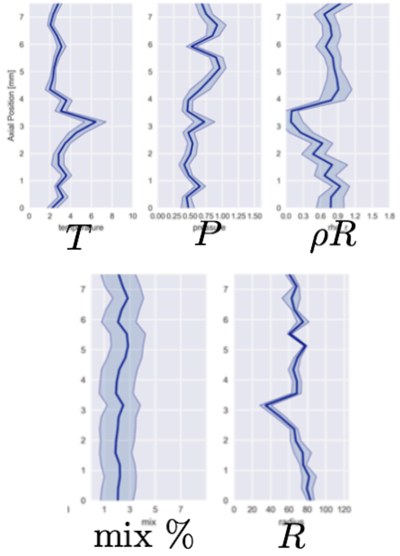
$Y_{DD}^{meas} = 3.2 \times 10^{12} \pm 20\%$
$Y_{DD}^{post} = 3.1 \times 10^{12} \pm 10\%$



Bayesian inversion

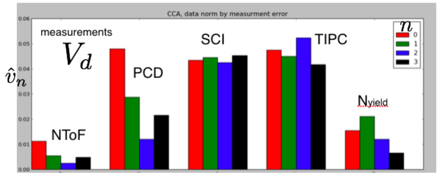
3D Gorgon



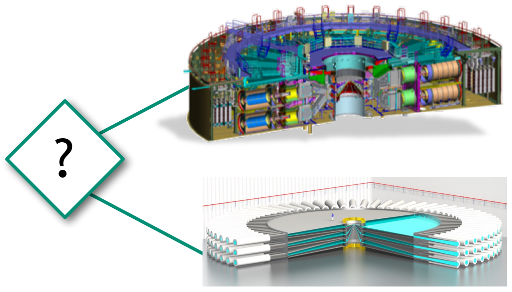


$E_{HS} = 10.8 \pm 1.1$ kJ
$\langle P \rangle = 0.63 \pm 0.17$ Gbar

experimental sensitivity



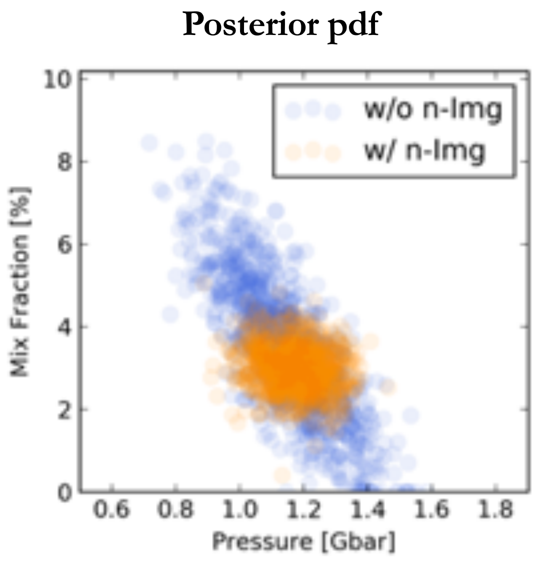
PHYSICS INFORMED DECISION SCIENCE



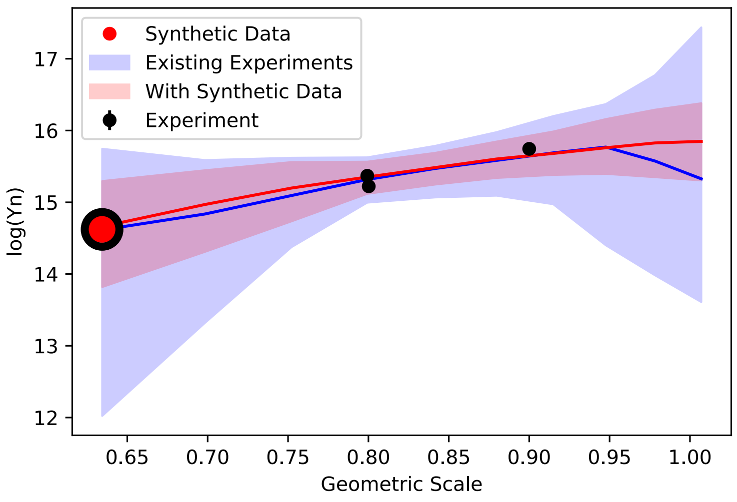
Uses of value of information

- decisions, metrics for making the decisions
  - diagnostics?
    - what instruments?
    - design of instruments?
      - lines of sight?
      - spectral ranges?
      - what needs to be improved?
    - what calibrations?
  - what is relevant physics?
  - are we neglecting something?
  - what experiments should we do?
  - how does physics extrapolate?
  - what models should be used in the analysis?

value of additional diagnostic

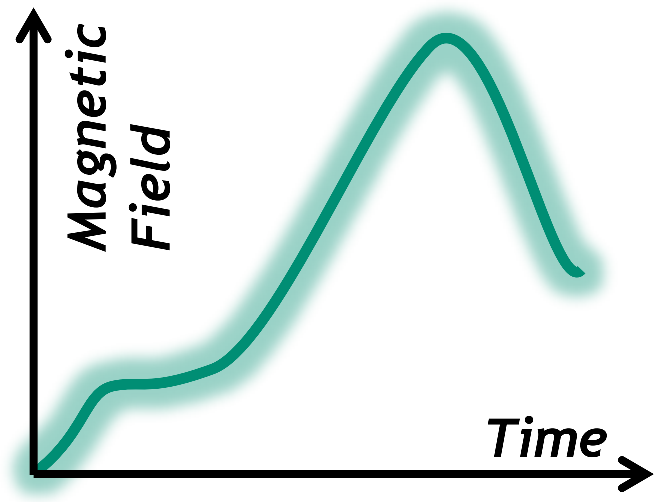


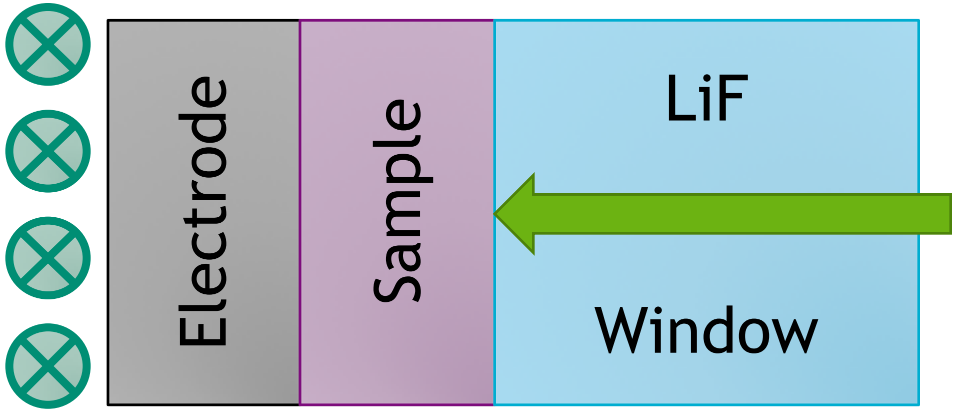
value of experimental point



A Bayesian calibration framework has been developed for interpreting DMP experiments

Z experiments are designed to be accurately modeled through 1D MHD simulations



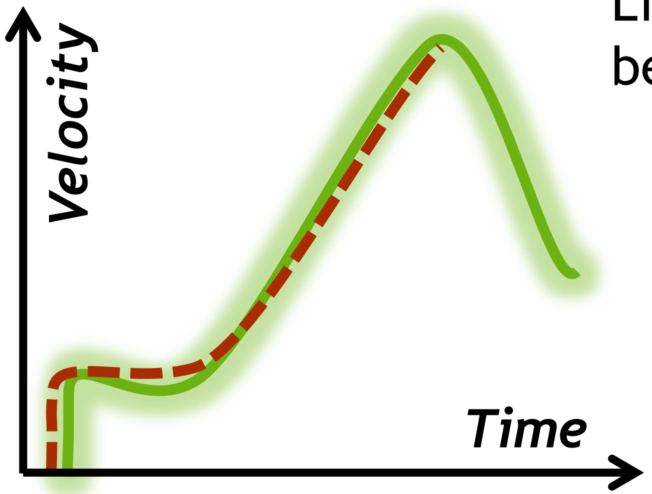


Experimental observable: Velocity of Pu/LiF interface

Standard statistical model:

$$V(t) = \eta(t, \boldsymbol{\theta}) + \epsilon(t)$$

Observed velocity | Hydrocode simulator | Experimental uncertainty



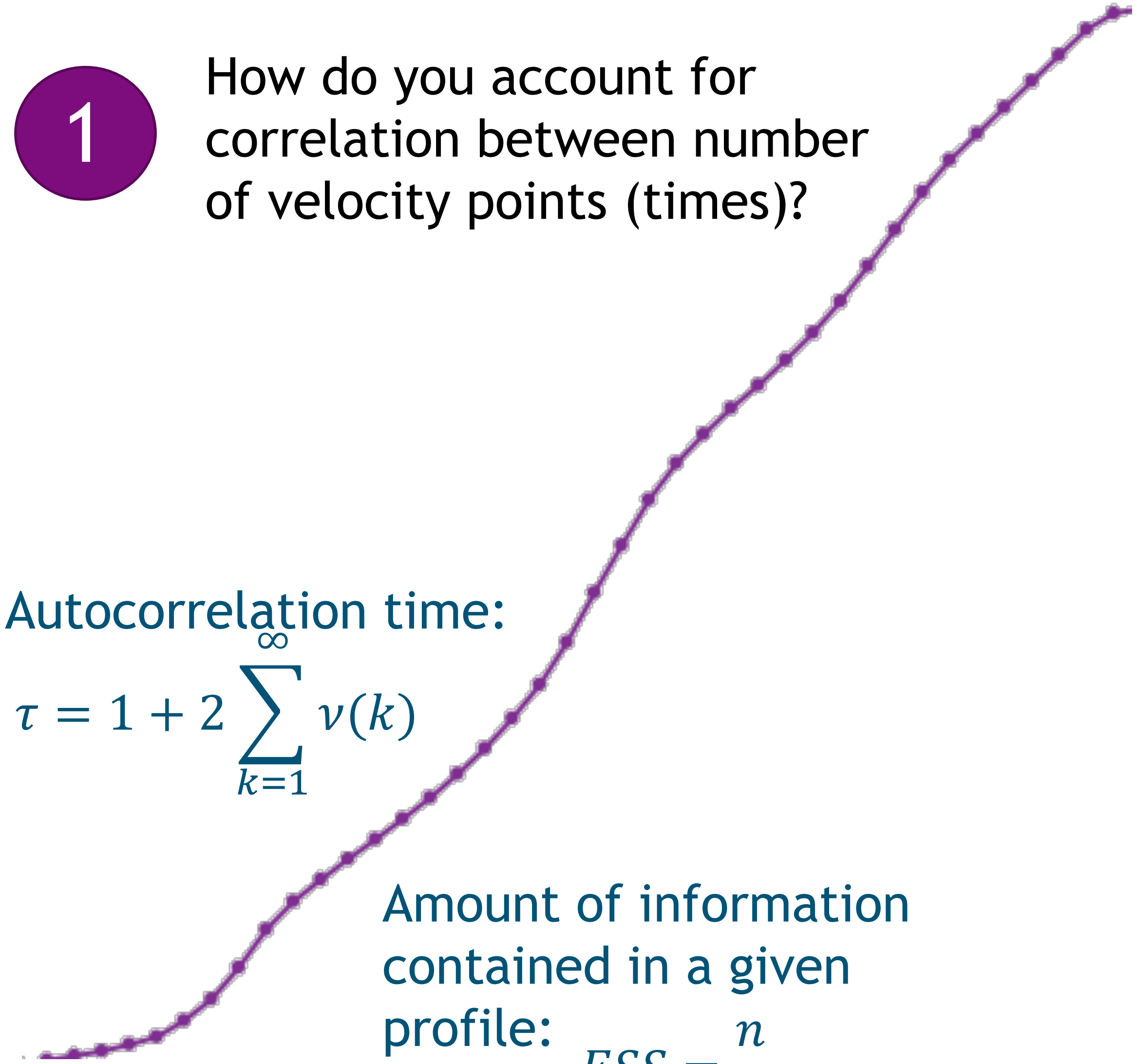
Likelihood reflects difference between simulation and experiment:

$$l(V|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_d|}} exp\{-\boldsymbol{r}^T\boldsymbol{\Sigma}_d^{-1}\boldsymbol{r}\}$$

Velocity uncertainty

Vector of residuals

Challenges unique to this calibration



1 How do you account for correlation between number of velocity points (times)?

Autocorrelation time:

$$\tau = 1 + 2 \sum_{k=1}^{\infty} \nu(k)$$

Amount of information contained in a given profile: $ESS = \frac{n}{\tau}$

2 How do you efficiently sample the posteriors using MCMC?

Solution: build a surrogate model to emulate the hydrocode

1. Run ~100,000 simulations sampling the parameter space to generate training data
   - Massively parallel Monte Carlo

2. Construct an emulator based on training data
   - We Gaussian Process (GP) surrogate

3. MCMC on the GP to sample posteriors
   - Usual metrics on chain mixing and convergence

Brown et al., *Journal of the Royal Statistical Society Series C*, **67**, 4 (2018)

Application focus

- Stagnation conditions for Magnetic Direct Drive Fusion experiments
- Analysis of pulsed power driven DMP experiments
- Analysis of MagLIF preheat experiments at NIF
- Z power flow data analysis
- Model calibration through focused physics experiments (plasma transport, non-linear instability growth, etc.)