PECOS

Predictive Engineering and Computational Sciences

Information-Theoretic Experimental Design for the Active Nitridation of Graphite

Gabriel Terejanu, Rochan Upadhyay, Kenji Miki

Institute for Computational Engineering and Sciences
The University of Texas at Austin

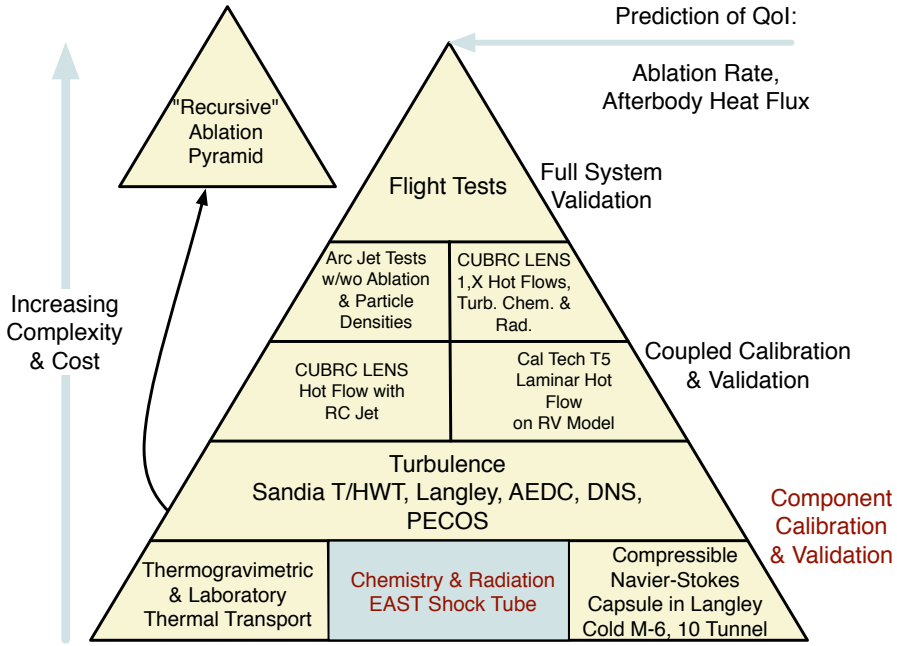
Acknowledgement: Jochen Marschall (SRI International)

DoE/NNSA: DE-FC52-08NA28615

Outline

1. Motivation
2. Sequential Experimental Design
3. Information-theoretic Interpretation
4. Proof of Concept
5. Graphite Nitridation
6. Conclusions

Motivation

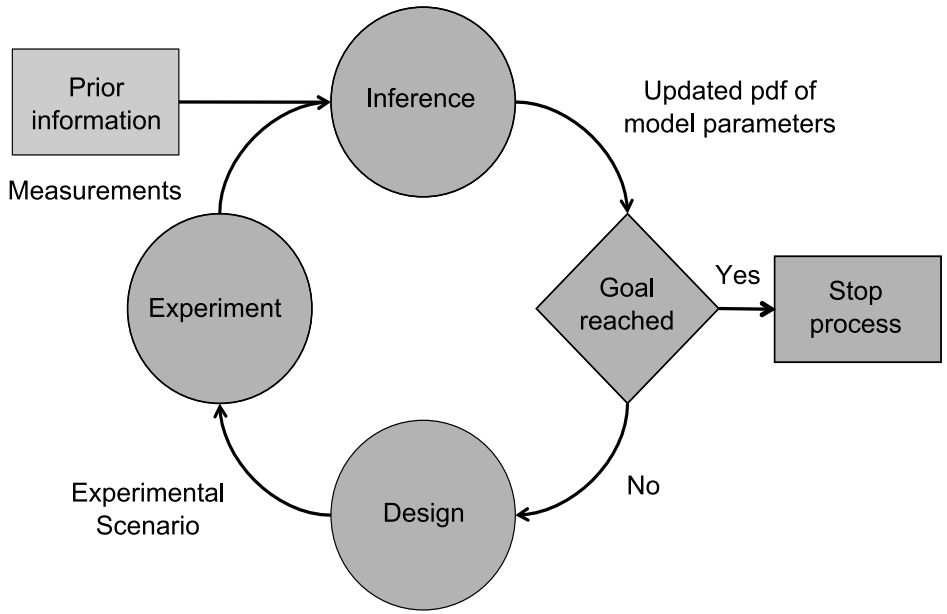


New Data: What is the experimental scenario that can provide the most informative measurements to calibrate our model? (Calibration)

Legacy Data: What can we say about the relation between the legacy data and our model? (Verification and Validation)

Sequential Experimental Design

$$\mathbf{r}(\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\xi}, \epsilon_s) = 0$$
$$\mathbf{d} = \mathbf{d}(\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\xi}, \epsilon_s)$$



1. Experiment: new data collected according to the strategy obtained in the previous design stage.
2. Inference: the newly obtained measurements are used to update the pdf of model parameters (Bayes rule).
3. Design: the best design is chosen by maximizing the expected Shannon information gain.

$$S_n = \{\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_2^*, \ldots, \boldsymbol{\xi}_n^*\}$$
$$D_n = \{\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \ldots, \hat{\mathbf{d}}_n\}$$

Note that since the next design depends on the previous experiments, it does not guarantee to find the optimal N designs among all possible combinations of N designs.

Inference Stage and Stopping Criteria

Sequential Bayesian Inference:

$$
\begin{aligned}
p(\boldsymbol{\theta}|D_n) &= \frac{p(D_n|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(D_n)} \\
&= \frac{p(\tilde{\mathbf{d}}_n, D_{n-1}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\tilde{\mathbf{d}}_n, D_{n-1})} \\
&= \frac{p(\tilde{\mathbf{d}}_n|D_{n-1}, \boldsymbol{\theta}, \boldsymbol{\xi}_n^*)p(D_{n-1}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\tilde{\mathbf{d}}_n|D_{n-1}, \boldsymbol{\xi}_n^*)p(D_{n-1})} \\
&\propto p(\tilde{\mathbf{d}}_n|D_{n-1}, \boldsymbol{\theta}, \boldsymbol{\xi}_n^*)p(\boldsymbol{\theta}|D_{n-1})
\end{aligned}
$$

Computational aspects: Multilevel MCMC (*Cheung and Beck 2008*) available in the statistical library QUESO (*Prudencio and Schulz 2001*)

When to stop the experimental process?

1. Desired precision in the estimation of model parameters has been reached

$$\det\left|\mathrm{E}[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \mid D_n]\right| \leq p_{const}$$

2. The reduction in the parameters uncertainty has slowed enough

$$\mathrm{H}\Big(p(\boldsymbol{\theta}|D_1)\Big) \geq \ldots \geq \mathrm{H}\Big(p(\boldsymbol{\theta}|D_n)\Big)$$

3. Indication exists that model predictions and experimental observations are in disagreement

$$\Big\{\mathrm{D}_{KL}\Big(p(\boldsymbol{\theta}|D_n)||p(\boldsymbol{\theta}|D_{n-1})\Big)\ ,\ \forall n\Big\}$$

Optimal Experimental Design Stage

Objective of the experiment is to learn about the model parameters $\boldsymbol{\theta}$ (*Lindley 1956*):

$$\boldsymbol{\xi}_n^* = \arg\max_{\boldsymbol{\xi}_n \in \boldsymbol{\Xi}} \underbrace{\int_{\mathcal{D}} U(\mathbf{d}_n, \boldsymbol{\xi}_n) p(\mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n) \mathrm{d}\mathbf{d}_n}_{\mathrm{E}_{\mathbf{d}_n}[U(\mathbf{d}_n, \boldsymbol{\xi}_n)]}$$

Predictive distribution of observables:

$$p(\mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n) = \int_{\Theta} p(\mathbf{d}_n | \boldsymbol{\theta}, \boldsymbol{\xi}_n) p(\boldsymbol{\theta} | D_{n-1}) \mathrm{d}\boldsymbol{\theta}$$

Utility function for our objective:

$$U(\mathbf{d}_n, \boldsymbol{\xi}_n) = \mathrm{H}\Big(p(\boldsymbol{\theta} | D_{n-1})\Big) - \mathrm{H}\Big(p(\boldsymbol{\theta} | D_{n-1}, \mathbf{d}_n)\Big)$$

Shannon's measure of information (*Shannon 1948*):

$$\mathrm{H}\Big(p(\boldsymbol{\theta} | D_{n-1})\Big) = -\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta} | D_{n-1}) \log p(\boldsymbol{\theta} | D_{n-1}) \mathrm{d}\boldsymbol{\theta}$$

Note: when the purpose of the experiment is other than learning about model parameters, i.e. accurate prediction of a quantity of interest, then other utility function may be defined (*Chaloner and Verdinelli 1995*).

Information-theoretic Interpretation

Optimal experiment is obtained by solving:

$$\boldsymbol{\xi}_n^* = \arg\max_{\boldsymbol{\xi}_n \in \boldsymbol{\Xi}} \mathrm{E}_{\mathbf{d}_n}[U(\mathbf{d}_n, \boldsymbol{\xi}_n)]$$

Information-theoretic interpretation (*Paninski 2005*, *Terejanu et al. 2011*):

$$\begin{aligned}
\mathrm{E}_{\mathbf{d}_n}[U(\mathbf{d}_n, \boldsymbol{\xi}_n)] &= \int_{\mathcal{D}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}, \mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n) \log \frac{p(\boldsymbol{\theta}, \mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n)}{p(\boldsymbol{\theta} | D_{n-1}) p(\mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n)} \mathrm{d}\boldsymbol{\theta} \mathrm{d}\mathbf{d}_n \\
&= \mathrm{D}_{KL}\left(p(\boldsymbol{\theta}, \mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n) \,\middle\|\, p(\boldsymbol{\theta} | D_{n-1}) p(\mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n) \right) \\
&= \mathrm{I}(\boldsymbol{\theta}; \mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n)
\end{aligned}$$

Mutual Information:

- quantifies the reduction in uncertainty that knowing either variable provides about the other (*Cover and Thomas 1991*)
- provides a measure of statistical dependence between the two random variables, (connection with copula functions, see *Calsaverini and Vicente 2009*)

$$\boldsymbol{\xi}_n^* = \arg\max_{\boldsymbol{\xi}_n \in \boldsymbol{\Xi}} \mathrm{I}(\boldsymbol{\theta}; \mathbf{d}_n | D_{n-1}, \boldsymbol{\xi}_n)$$

Estimating Mutual Information from Samples

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2)$$

Kazachenko-Leonenko estimator for differential Shannon entropy:

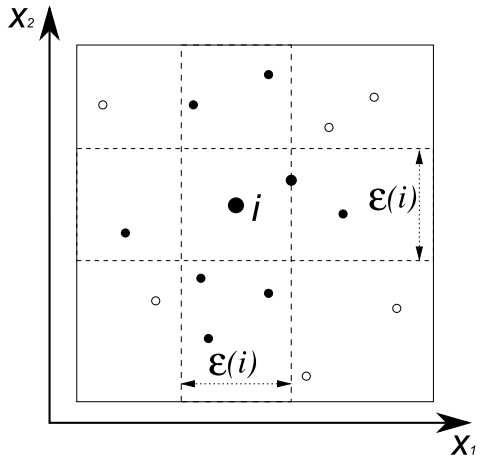
$$H(X) = -\int p(x) \log p(x) \mathrm{d}x \approx -\frac{1}{N} \sum_{i=1}^{N} \hat{\log} p(x_i)$$

$$p(x) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i)$$

Estimate $\log p(x_i)$ using the distance, $\epsilon(i)$, from x_i to its $k$nn (using L-infinity):

$$H(X) \approx -\psi(k) + \psi(N) + \frac{d}{N} \sum_{i=1}^{N} \log \epsilon(i)$$

Kraskov estimator for mutual information:



Different distance scales → different biases which would not cancel. Do not have to fix k when estimating entropy of marginals.

$$I(X_1; X_2) \approx \psi(k) - \mathrm{E}\Big(\psi(n_{x_1} + 1) + \psi(n_{x_2} + 1)\Big) + \psi(N)$$

Note: small k yields a small bias and large variance, and a large k yields a large bias and a small variance

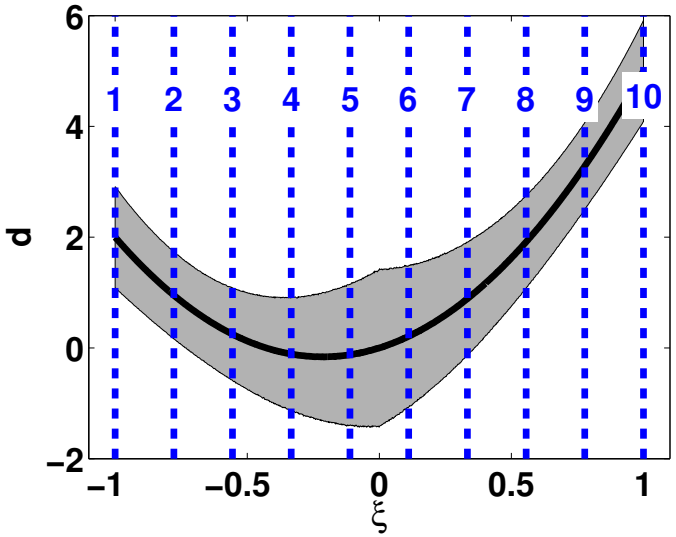
Proof of Concept

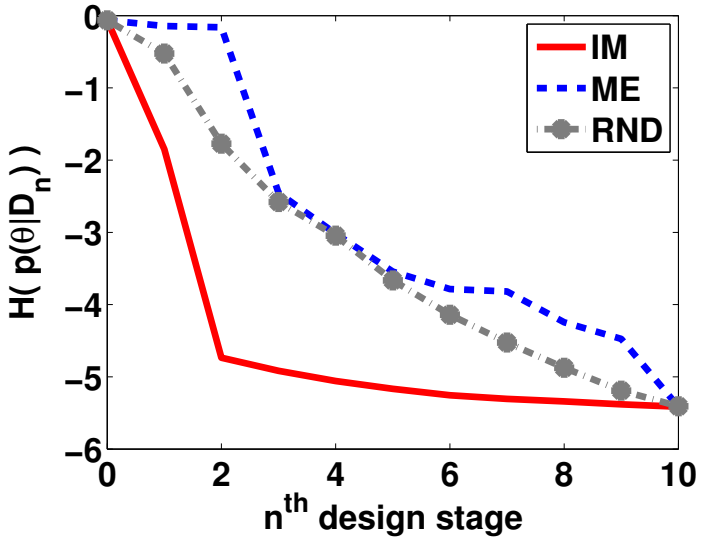
Consider the **model**:

$$
\begin{aligned}
d &= \theta_1 \xi + \theta_2 \xi^2 + f(\xi)\epsilon \ , \ \epsilon \sim \mathcal{N}(0,1) \\
f(\xi) &= 0.5(1.1 - |\xi|) \ , \ \xi \in [-1, 1]
\end{aligned}
$$

Measurements generating using:

$$\theta_1 = 1.5, \quad \theta_2 = 3.5$$

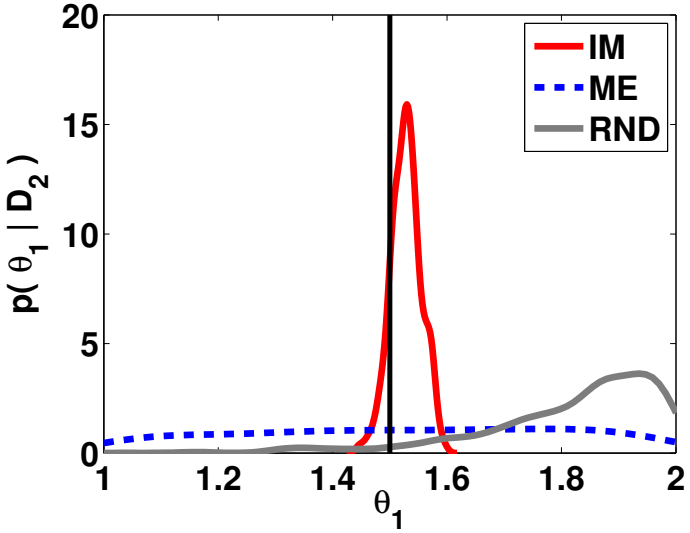


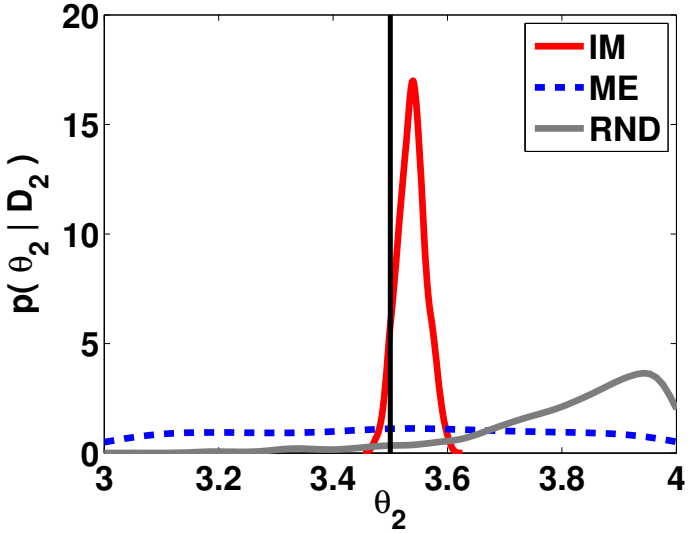


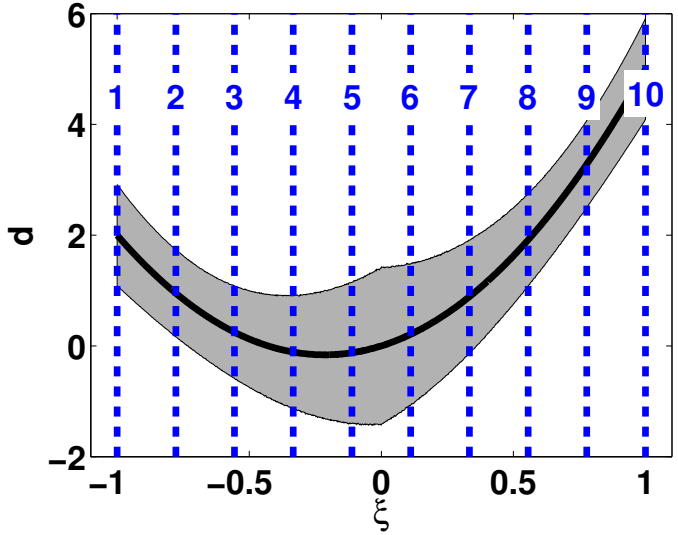
Comparison of design strategies:

- IM - Information Maximization sampling
- ME - Maximum Entropy sampling (*Sebastiani and Wynn 1997*)
- RND - Random sampling

Proof of Concept: After calibration

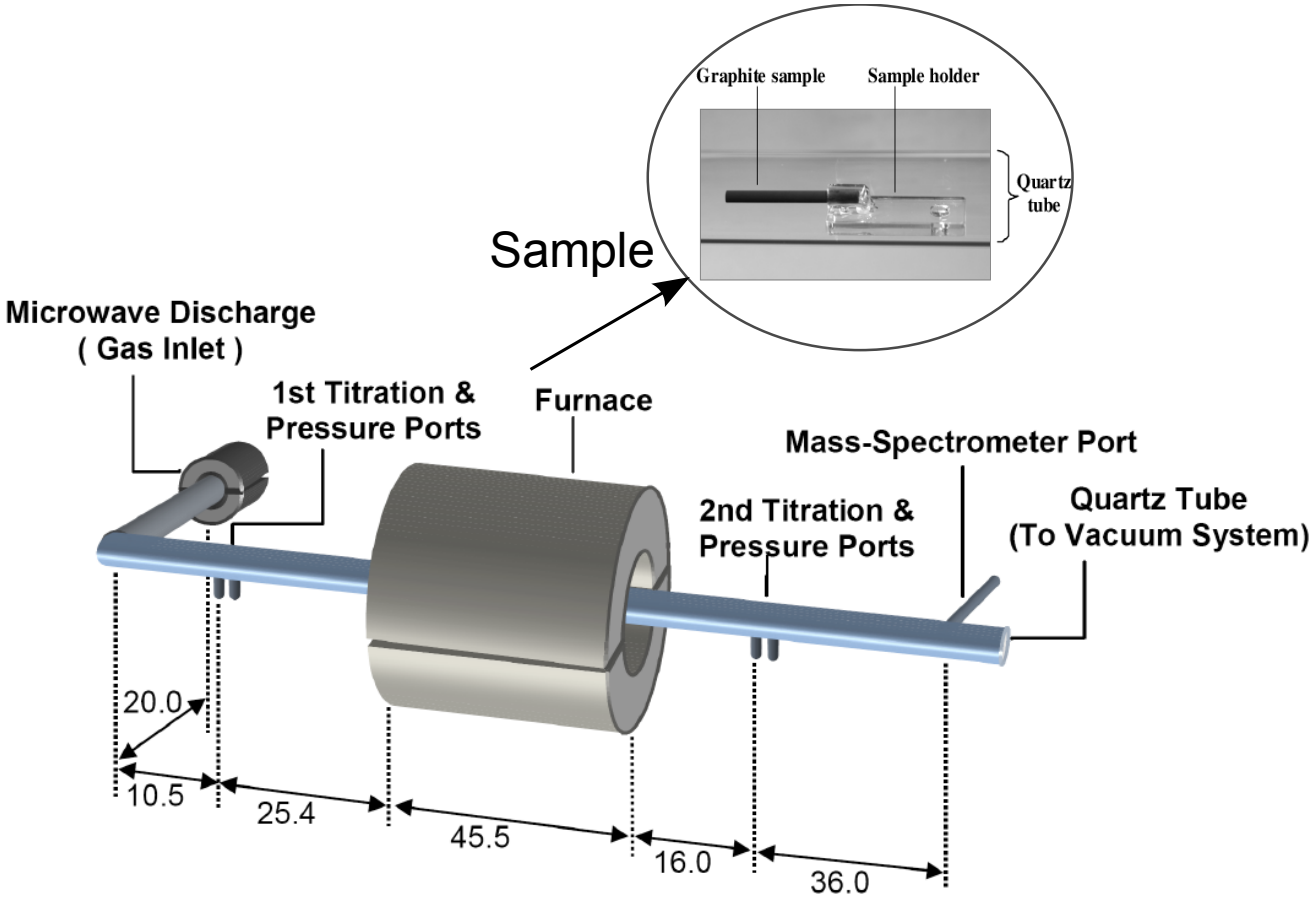






Strategy	Sequence
ME Sampling	5 6 7 4 1 3 8 9 2 10
IM Sampling	1 10 9 7 2 8 4 3 5 6
RND Sampling	6 1 8 5 10 9 4 2 3 7

SRI Experimental Setup



$F_{N_2,in}$	P_1	$\mathcal{X}_{N,1}$	Δt	L_S	Δm_C	P_2

Graphite Nitridation Model

Hagen-Poiseuille flow model:

$$P\frac{dP}{dz} = -\frac{128\dot{m}_{N_2}\mu_{N_2}RT(z)}{\pi d_{eff}{}^4 W_{N_2}}$$

Mean velocity profile:

$$v(z) = \frac{4.0\dot{m}_{N_2}}{\rho(z)\pi d_{eff}{}^2}$$

Ideal gas density profile:

$$\rho(z) = \frac{P(z)W_{N_2}}{RT}$$

N-atom concentration profile:

$$\frac{d(v(z)C_N)}{dz} = -\frac{\gamma_N(T)\bar{v}_N(T)C_N}{d_{eff}} - 2k(T)C_{N_2}C_N^2$$

Sutherland's model for viscosity:

$$\mu_{N_2} = \mu_{ref,N_2}\left(\frac{T_{ref}+S_{N_2}}{T+S_{N_2}}\right)\left(\frac{T}{T_{ref}}\right)^{3/2}$$

Reaction rate for gas phase recombination:

$$k(T) = A_{NN}exp\left[-\frac{E_{a,N-N}}{RT}\right]$$

Reaction coefficient with the wall:

$$\gamma_N(T) = \gamma_N(T_{ref})\left(\frac{T}{(T_{ref})}\right)^{\alpha}$$

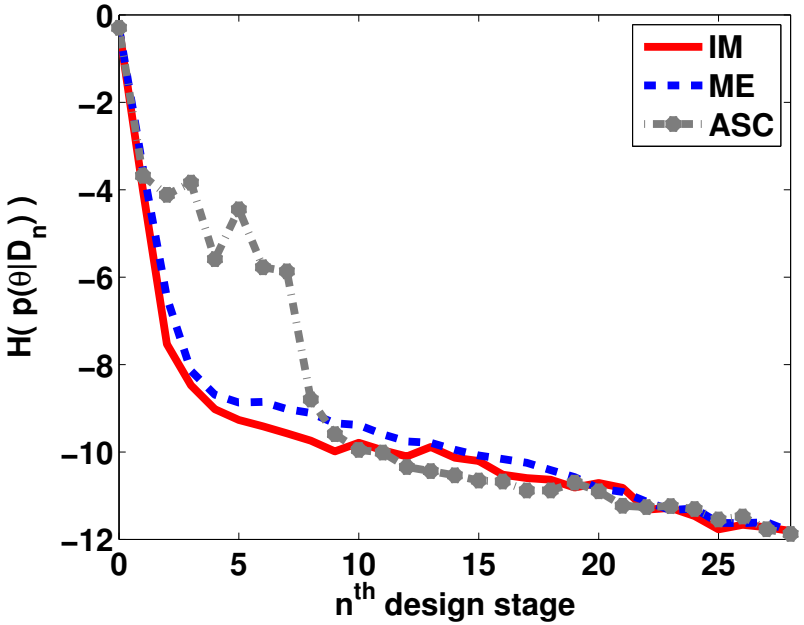
Thermal velocity:

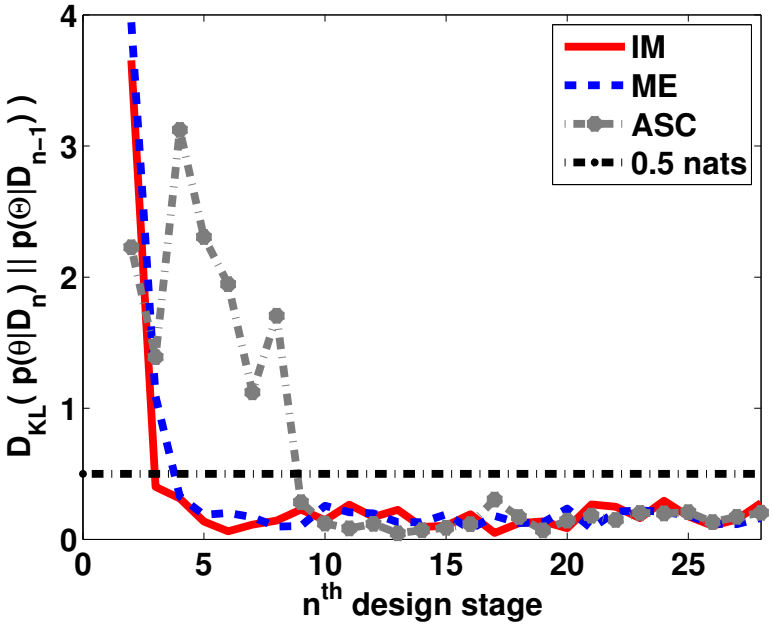
$$\bar{v}_N(T) = \sqrt{\left(\frac{RT}{2\pi W_N}\right)}$$

Mass loss of carbon:

$$\Delta M_C = (\Delta t\pi d_{sample}W_C)\left(\int_{z=z_s}^{z_s+L_s} C_N(z)\sqrt{T(z)}\beta_N dz\right)\sqrt{\frac{R}{2\pi W_N}}$$

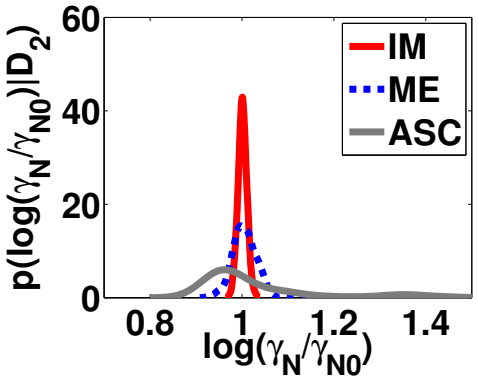
Graphite Nitridation - Simulated Measurements



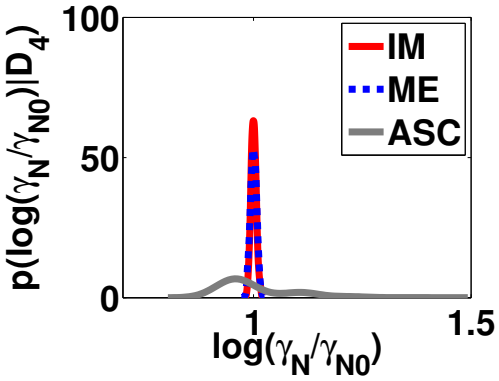


- IM - Information Maximization sampling
- ME - Maximum Entropy sampling
- ASC - The design sequence is given by the order of the measurements presented in Ref. *Zhang, Pejakovic and Marschall 2009*

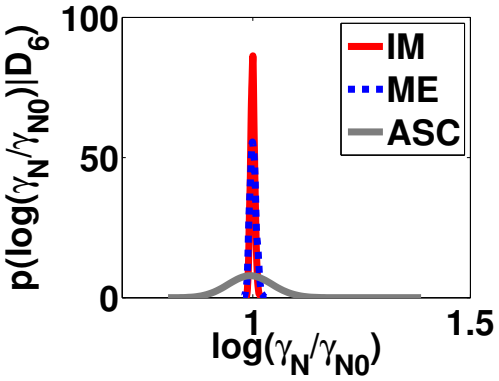
Simulated Measurements: After calibrations



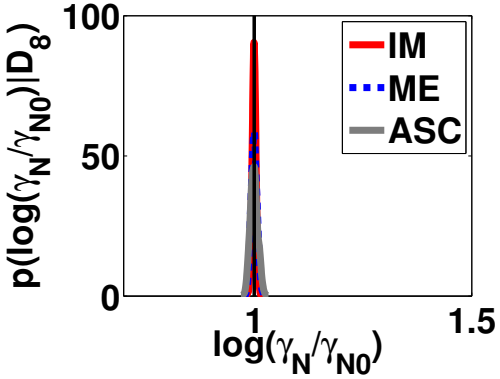
meas #2



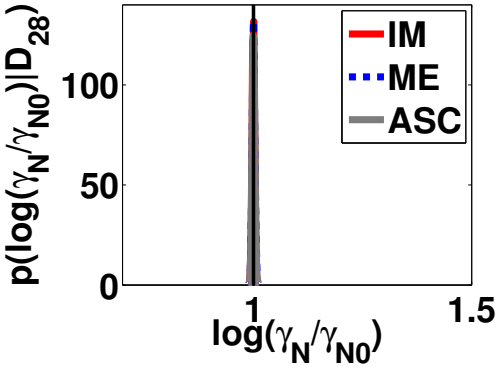
meas #4



meas #6

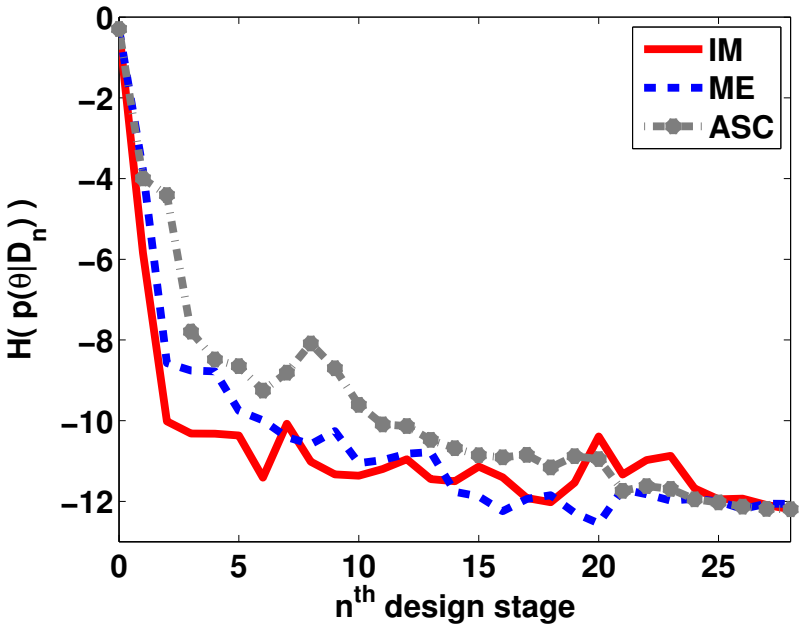


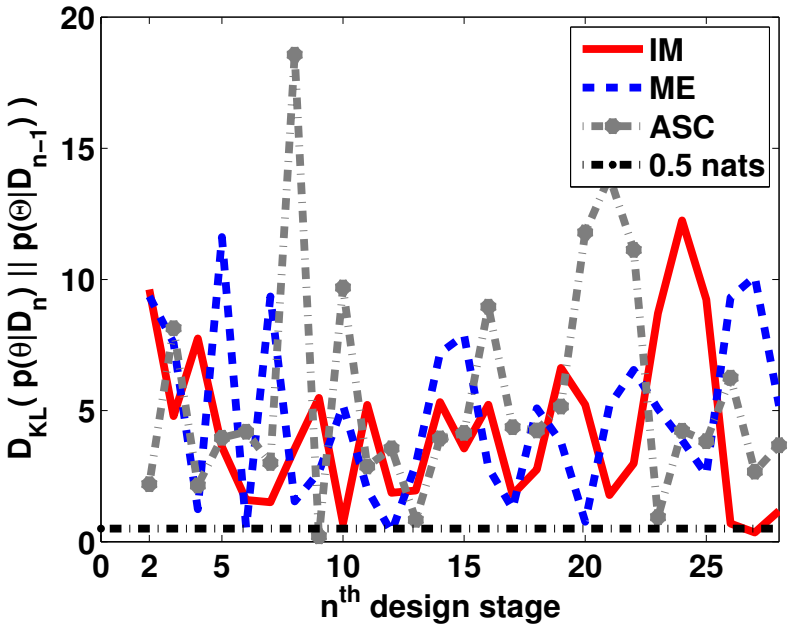
meas #8



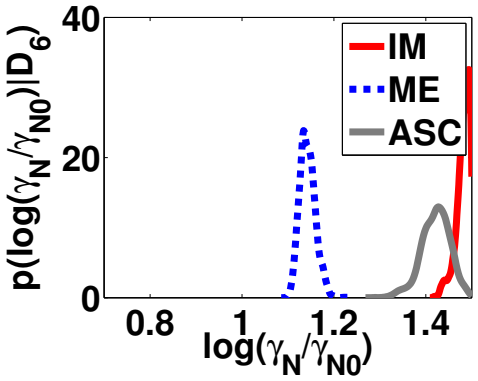
meas #28

Real Measurements: Carbon Nitridation

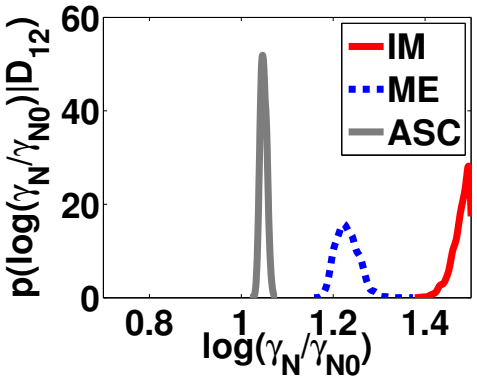




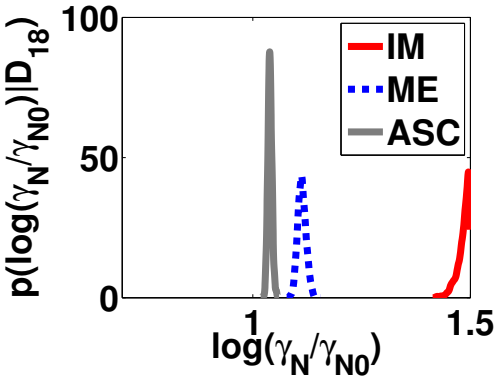
Real Measurements: After calibrations



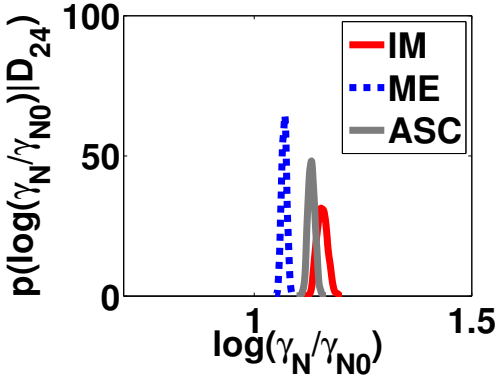
meas #6



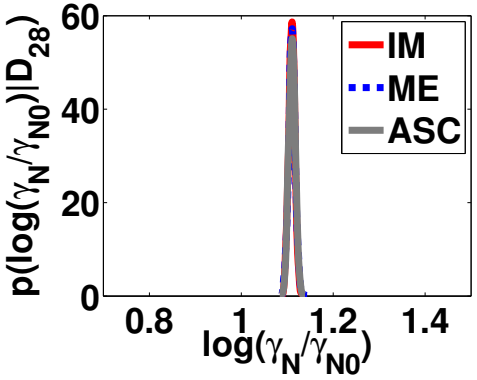
meas #12



meas #18



meas #24



meas #28

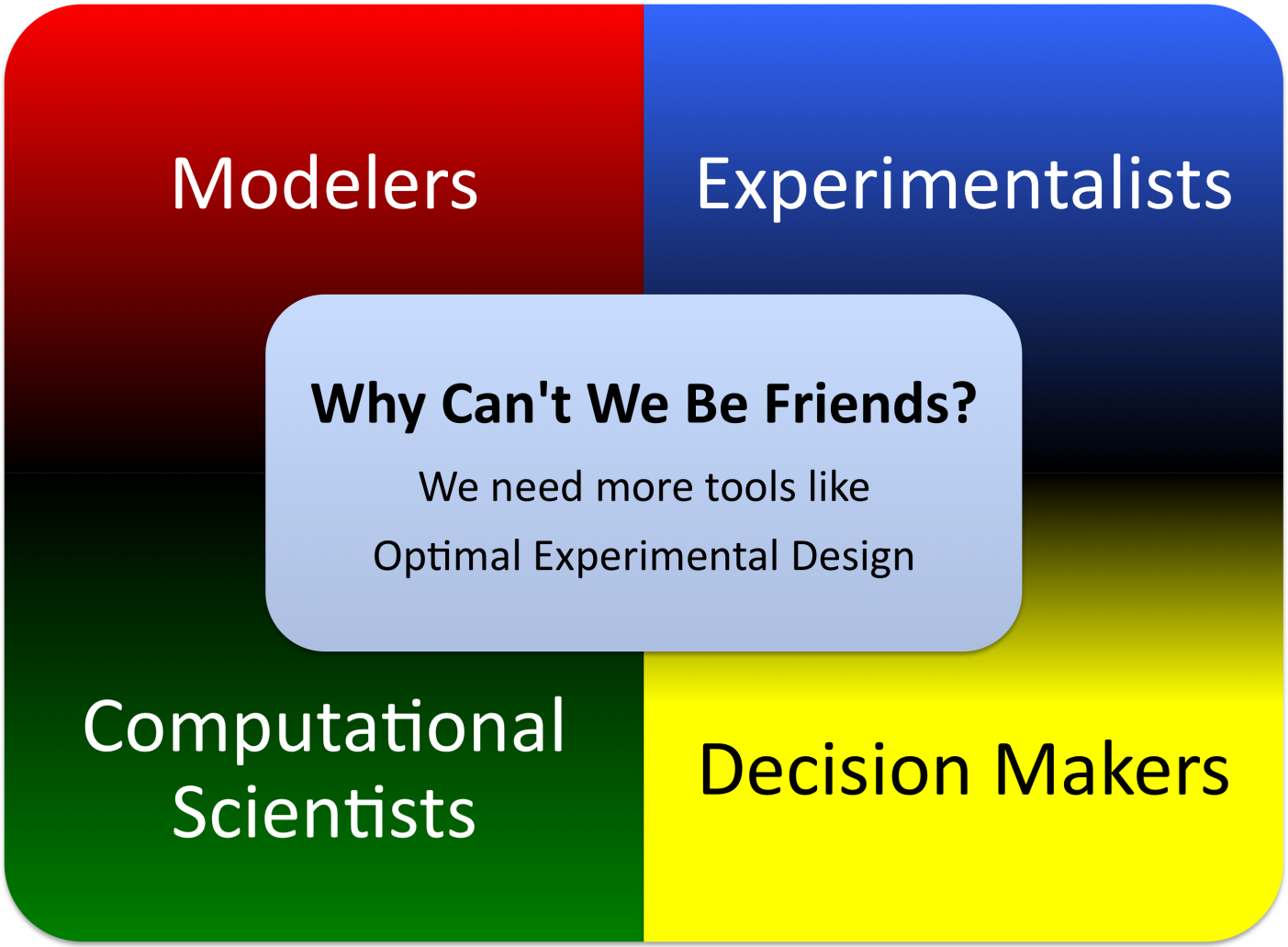
FLAG: Conflicting information:

- model + probabilistic assumptions have limited explanatory power
- different biases exist in the measurements
- all of the above

TODO: Experiment Verification and Model Verification and Validation.

Conclusions

- Bayesian experimental design equivalent with an **information-theoretic sensitivity analysis**.
- Optimal design provides the **highest statistical dependence** between model parameters and observables.
- **Information maximization sampling** is more general that the maximum entropy sampling and more efficient that random sampling.
- **Decreasing trend in entropy** indicates that the uncertainty of model parameters is reduced with additional measurements.
- The trend in the Kullback-Leibler divergence may signal the existence of **conflicting information** between model predictions and measurements.


Modelers
Experimentalists
Why Can't We Be Friends?
We need more tools like
Optimal Experimental Design
Computational Scientists
Decision Makers

Thank you!

Questions?

References

1. L. Zhang, D.A. Pejakovic, J. Marschall, D.G. Fletcher, "*Laboratory Investigation of Active Carbon Nitridation by Atomic Nitrogen*", AIAA Paper 2009-4251, AIAA, 2009

2. Gabriel Terejanu, Rochan R. Upadhyay, Kenji Miki, "*Bayesian Experimental Design for the Active Nitridation of Graphite by Atomic Nitrogen*", Experimental Thermal and Fluid Science, 2011, under review, arXiv:1107.1445v1 [physics.data-an]