Bayesian Analysis for Photolithographers

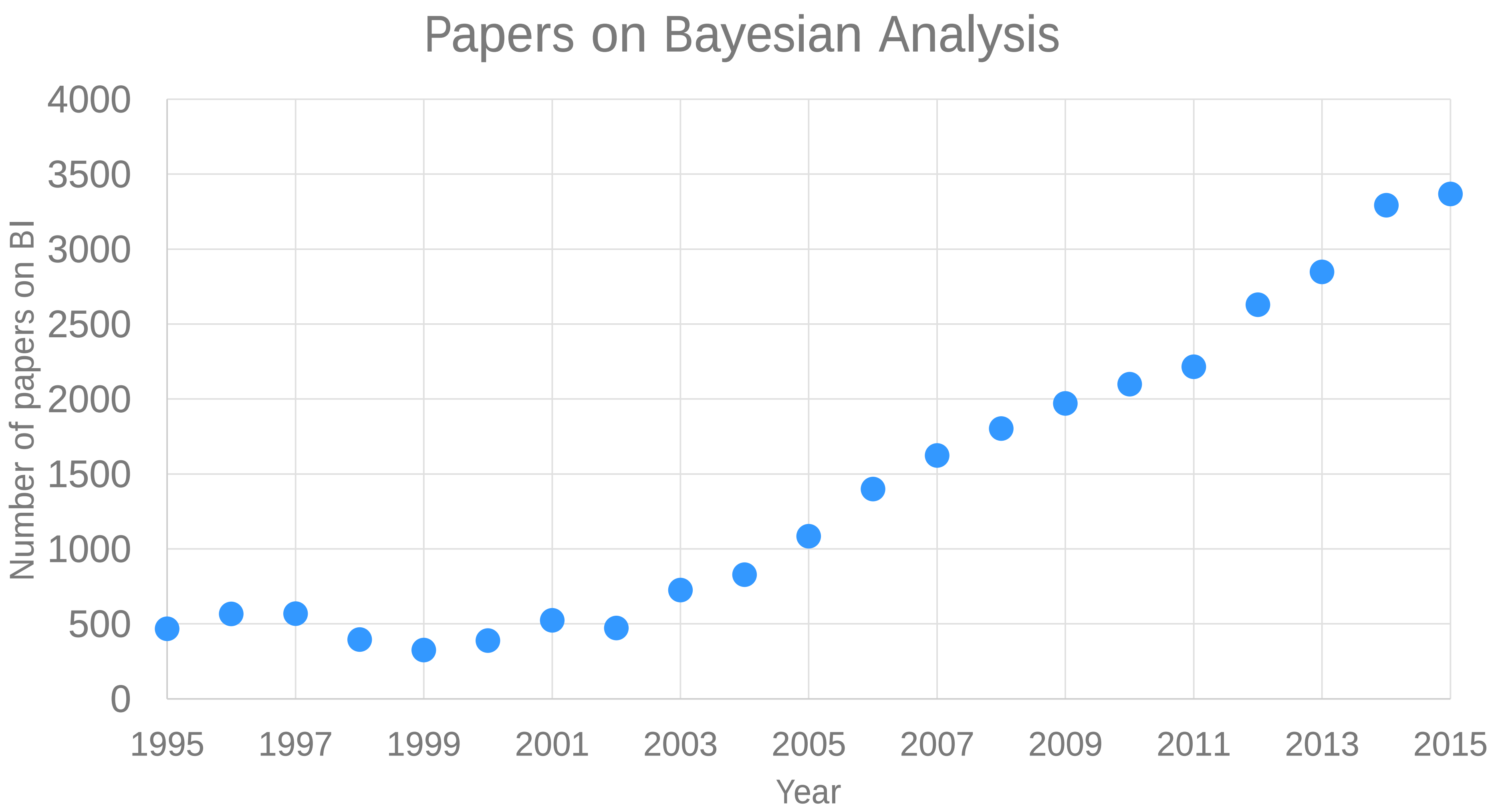
Andrew Burbine (RIT), David Fryer,
John Sturtevant, Bruce W. Smith (RIT)

February 24th, 2016

Overview

- Bayesian Analysis
  - Probabilistic view of the world
  - *A priori* knowledge
  - Linear model example
- Uncertainties in OPC model building
  - SEM CD measurements
  - Physical model parameters
- Prototype, cost functions & insights into the data

Bayesian Analysis on the Rise



$$p(\{y_i\}|\theta) = \theta^z(1-\theta)^{N-z}$$

Bayesian Analysis on a Bernoulli Trial Coin

- Determine the probability of heads for a coin

$$p(\{y_i\}|\theta) = \theta^z(1-\theta)^{N-z}$$

Bayesian Analysis on a Bernoulli Trial Coin

- Determine the probability of heads for a coin
- “Why? It’s probably fair.” ($\theta = 0.5$)

Litho parallel

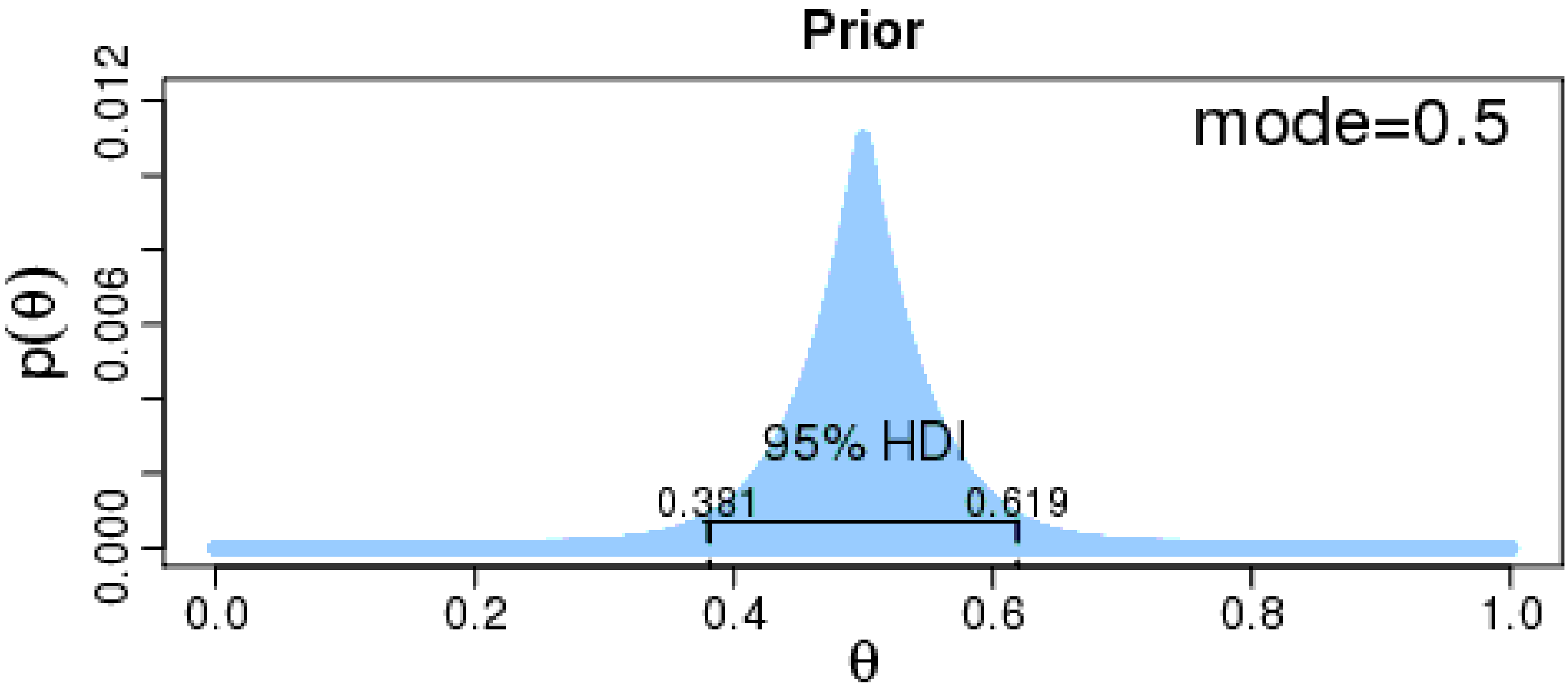
Determine the value of a mask parameter – absorber sidewall angle

Bayesian Analysis on a Bernoulli Trial Coin

- Determine the probability of heads for a coin
- “Why? It’s probably fair.” (θ = 0.5)
- You had *a priori* expectations about the experiment

https://izbicki.me/blog/how-to-create-an-unfair-coin-and-prove-it-with-math.html

Bayesian Analysis on a Bernoulli Trial Coin

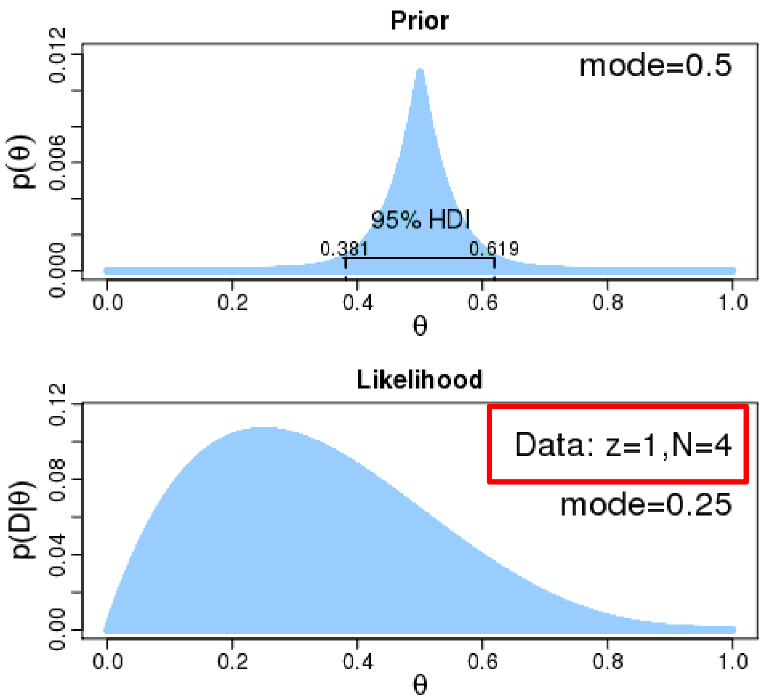


Some metrology data or the mask shop specifies:
SWA = 86°

Doing Bayesian Data Analysis, John K. Kruschke

$$p(\{y_i\}|\theta) = \theta^z(1-\theta)^{N-z}$$

Bayesian Analysis on a Bernoulli Trial Coin

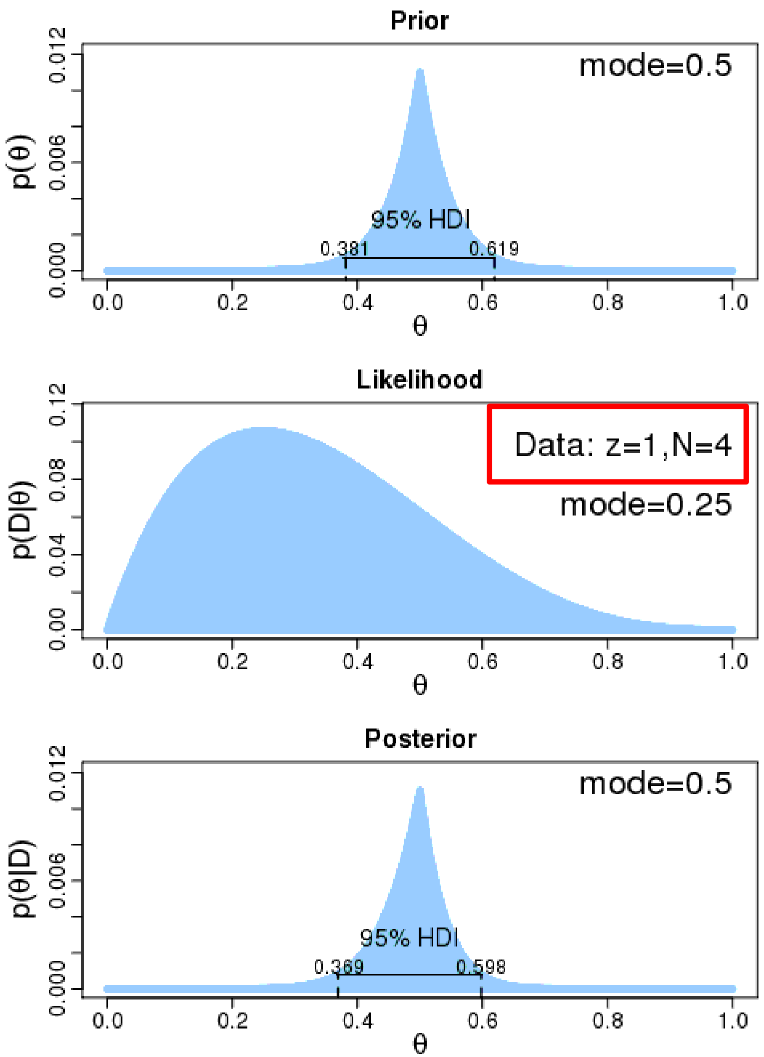


Running a few simulations suggests perhaps 86° isn't correct...

Doing Bayesian Data Analysis, John K. Kruschke

$$p(\{y_i\}|\theta) = \theta^z(1-\theta)^{N-z}$$

Bayesian Analysis on a Bernoulli Trial Coin

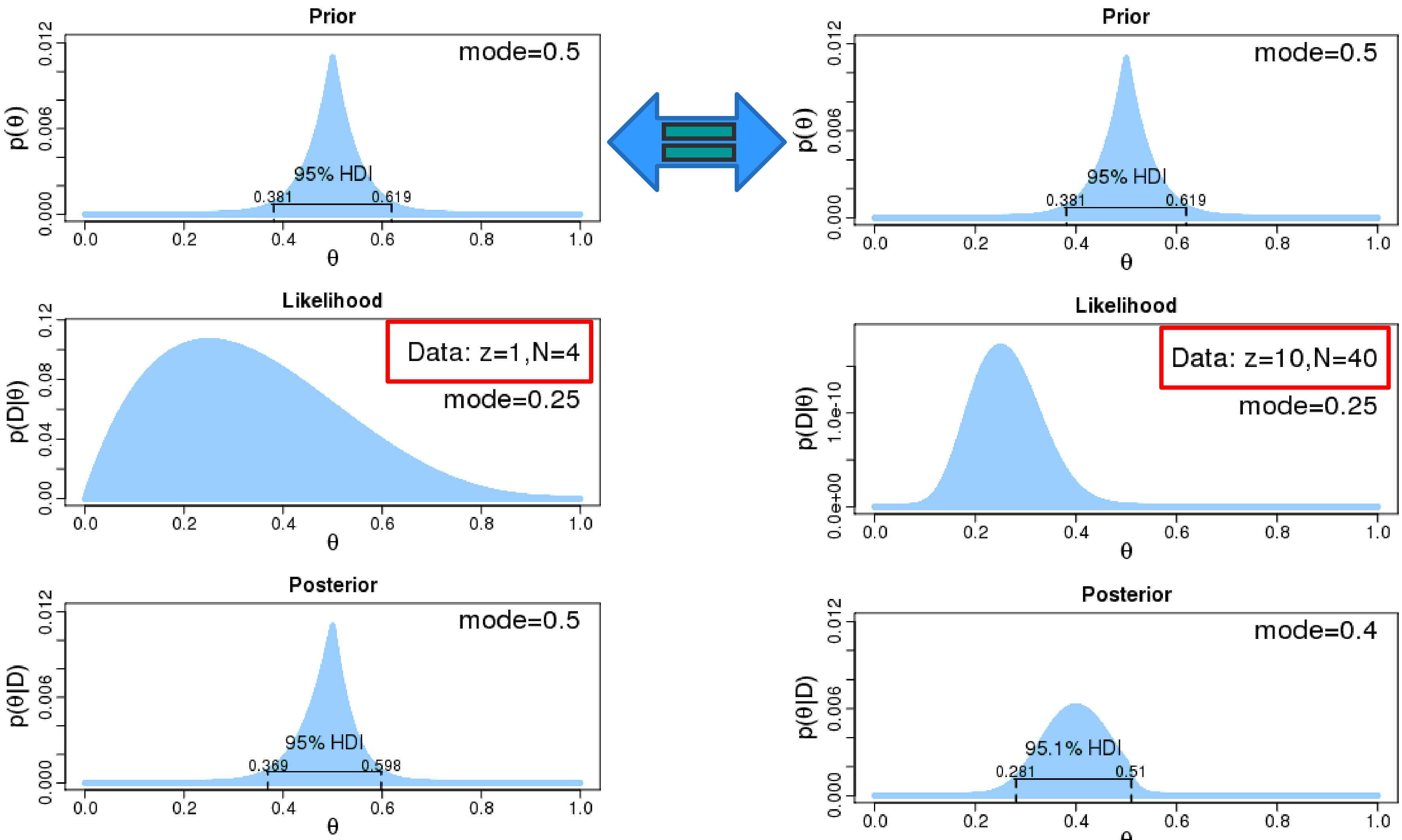


Still, 86° is more credible than the other values based on the information prior to the few simulations

Doing Bayesian Data Analysis, John K. Kruschke

$$p(\{y_i\}|\theta) = \theta^z(1-\theta)^{N-z}$$

Bayesian Analysis on a Bernoulli Trial Coin



Doing Bayesian Data Analysis, John K. Kruschke

Bayesian Analysis

Bayes' rule: $$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}$$

- The prior is the credibility of the parameters θ without the data D
- The likelihood is the probability that the data could be generated by the model with parameter value θ
- The evidence is the overall probability of the data according to the model, determined by averaging across all possible parameter values
- The posterior is the credibility of the parameters θ given the data D

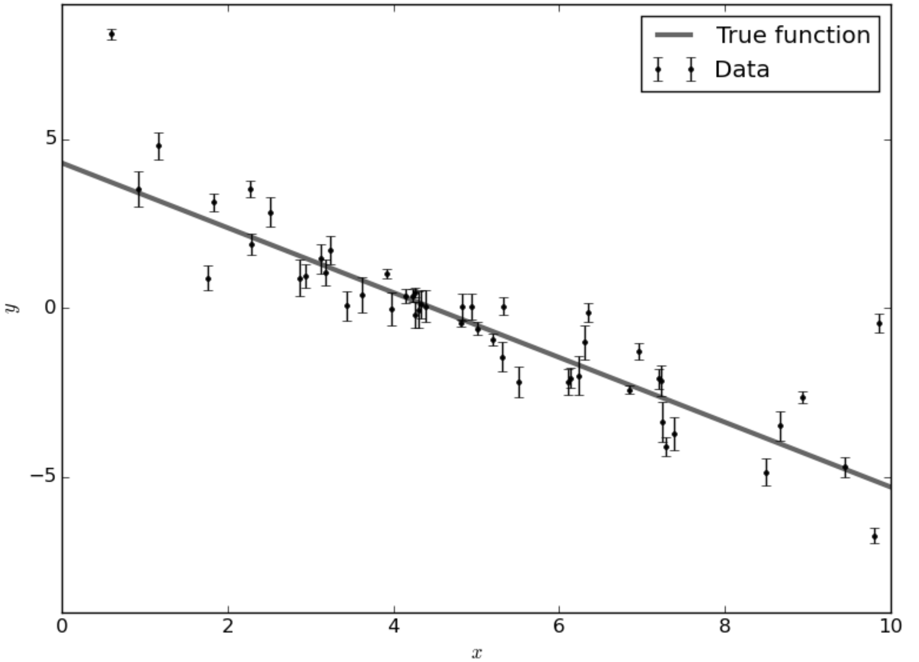
Analysis on a Simple Linear Model

- x by y linear model
  - Underestimated uncertainties in the error of y
- Generate 50 data points
- Fit a line to the data
- Seeking to find the values for the coefficients in model:
  - $y = mx + b \pm \epsilon(f)$
- We have many choices for obtaining such parameters...

CD measurements have uncertainty associated with their values (measurement variability, site accuracy)

$$m_{true} = -0.9594$$
$$b_{true} = 4.294$$
$$f_{true} = 0.534$$

$$y = m_{true}x + b_{true} + |f_{true}y| \cdot U(0,1)$$



http://dan.iel.fm/emcee/current/user/line/

Applying Bayes Rule

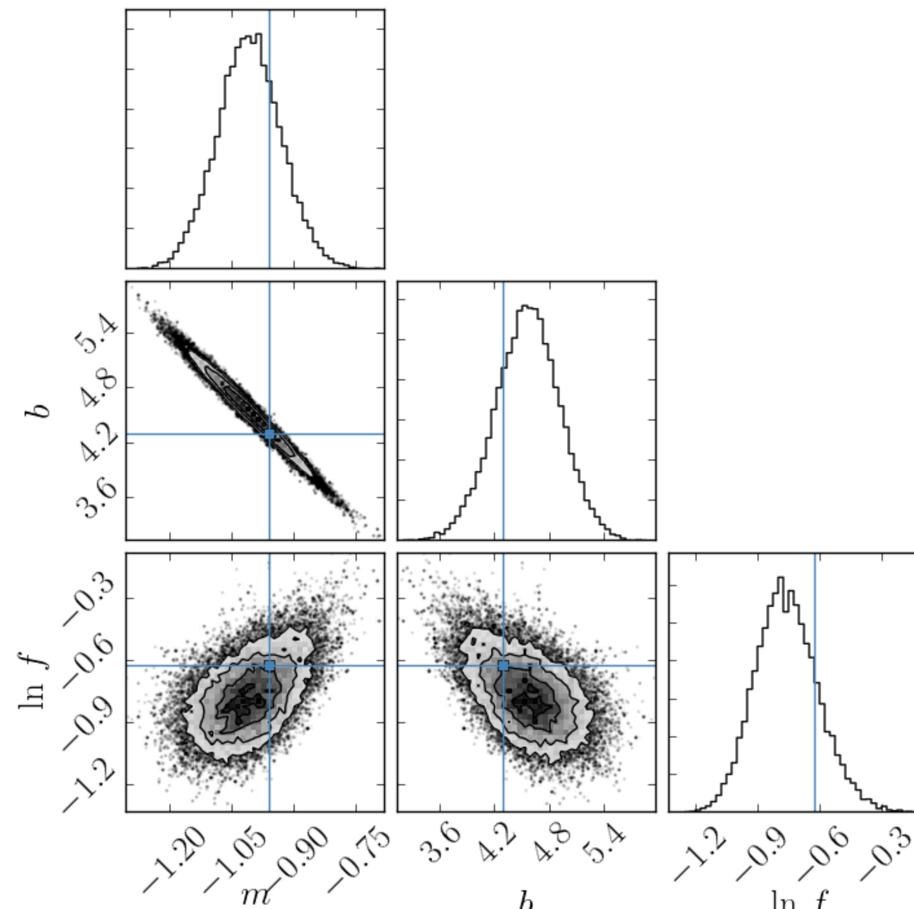
- In order incorporate our prior knowledge and generate posteriors for the parameters in the model, we must evaluate:

$$\mathrm{p}(m,b,\sigma,f|D) = \frac{p(D|m,b,\sigma,f)\,p(m,b,\sigma,f)}{\iint\iint p(D|m,b,\sigma,f)\,p(m,b,\sigma,f)\,dm\,db\,d\sigma\,df}$$

- Intractable integral for any real world problem
  - — Conjugate priors exist for few modeling problems
  - — Empirical solutions only for complex models
- We must computationally explore the parameter space
  - — Increased computational power and algorithms in the past decade have enabled Bayesian Analysis

Markov Chain Monte Carlo

- MCMC methods are a class of algorithms designed to explore parameter spaces
- Move proposals
  - Some simple, some complex
  - Accepted or rejected based on certain criterion
- Empirical posterior distributions are generated which indicate highest credibility parameters



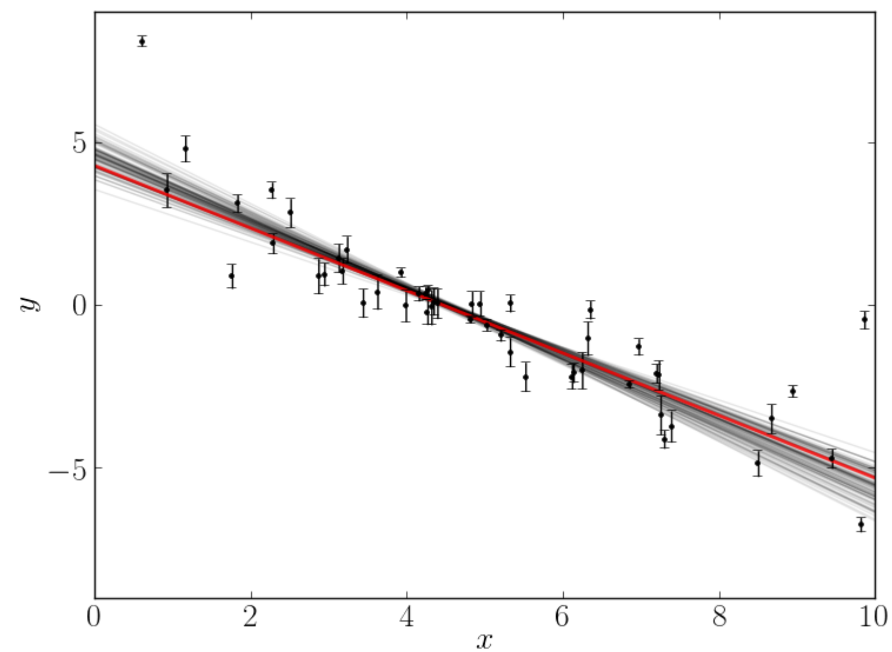
http://dan.iel.fm/emcee/current/user/line/

Model Parameters with Uncertainty

- Using the AIES sampler, 50,000 models are explored
- Posteriors (previous slide) are generated
- From the posteriors:
  - 95% highest density intervals
  - *Maximum a posteriori probability* (MAP)
- Accuracy of the MLE result + uncertainty in parameters
  - Used uninformative priors

$$m_{true} = -0.9594 \qquad m_{bi} = -1.009^{+0.077}_{-0.075}$$
$$b_{true} = 4.294 \qquad b_{bi} = 4.556^{+0.346}_{-0.353}$$
$$f_{true} = 0.534 \qquad f_{bi} = 0.463^{+0.079}_{-0.063}$$

$$y = m_{true}x + b_{true} + |f_{true}y| \cdot U(0,1)$$

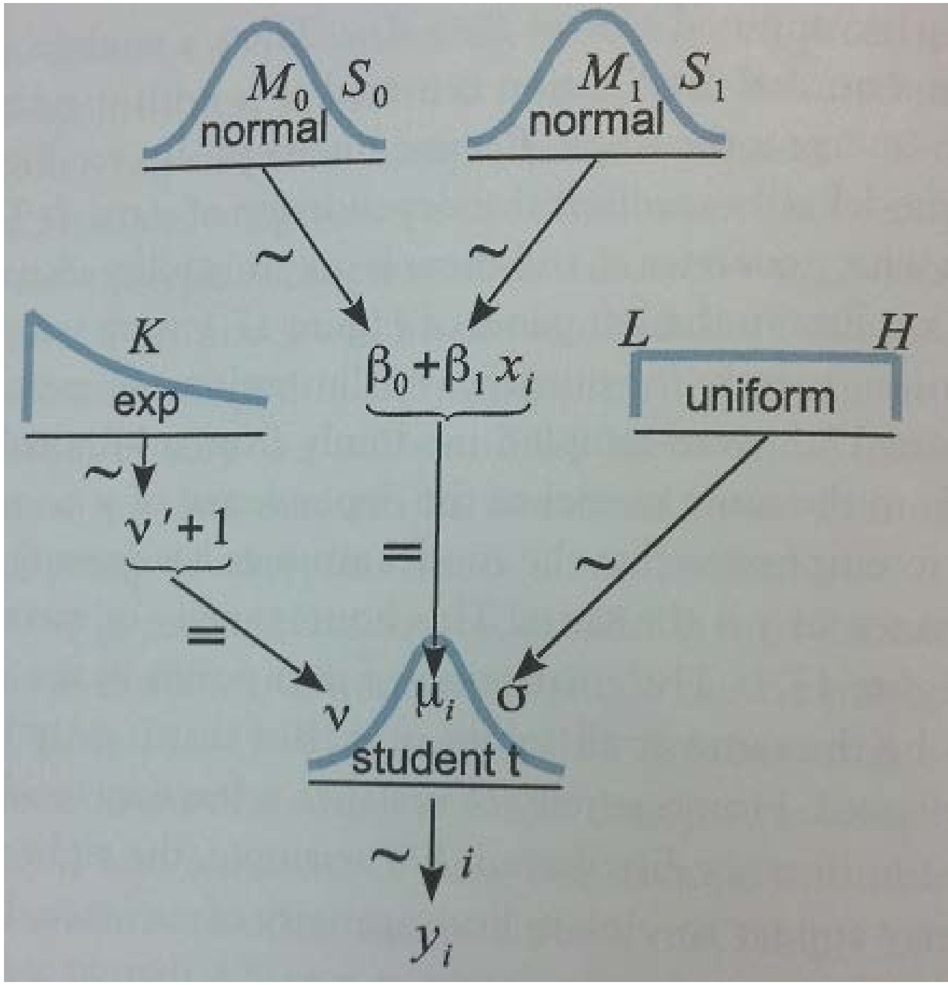


http://dan.iel.fm/emcee/current/user/line/

Bayesian Model Hierarchy

- Add to the prior knowledge by building a set of dependencies
- Model to predict baseball hitter's batting average
  - Past performance indicates odds
  - Expect first baseman to have higher BA
  - Position average informs player average

2D structures expected to have more uncertainty than pitch structures



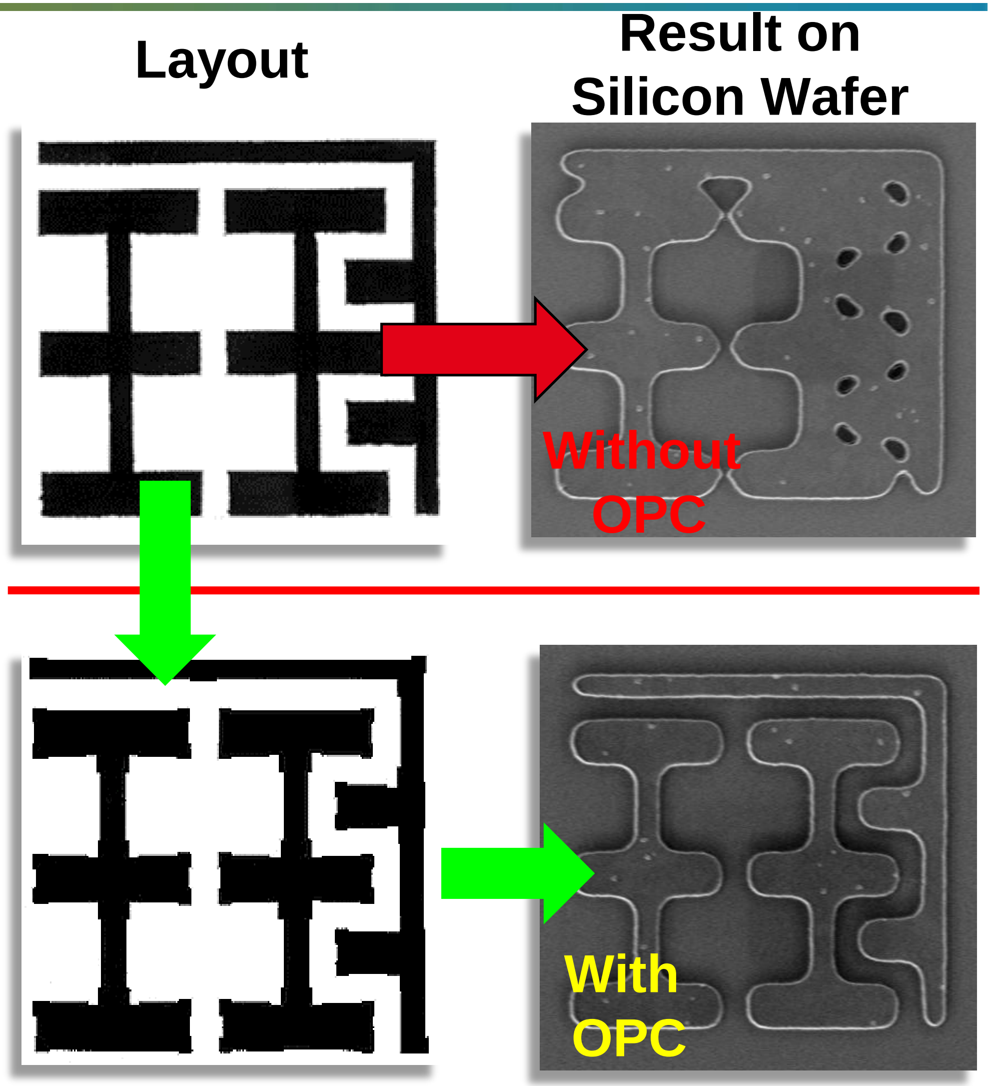
Doing Bayesian Data Analysis, John K. Kruschke

Optical Proximity Correction

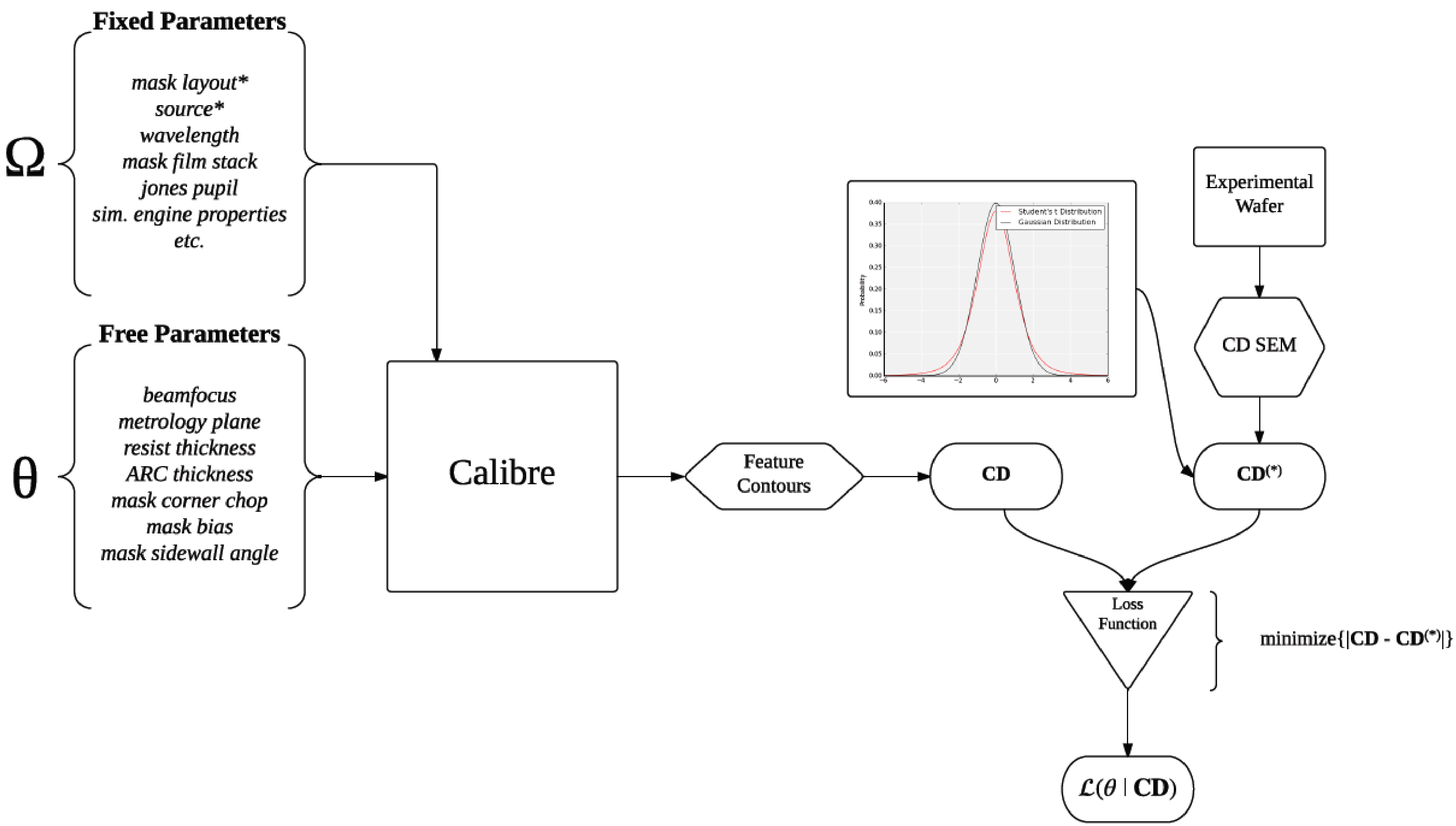
Pushing the limits of lithographic printability

Optical Proximity Correction

- Because optical lithography is performed near the physical limits of its manufacturing systems, advanced techniques, such as optical proximity correction (OPC) are needed to achieve the design intent
- On the right there are two examples of going from the layout (mask) to exposure in photoresist
- The edges of the layout are modified, algorithmically, so that the actual printed design is closer to the original layout than it would otherwise be



Photolithographic Model Parameters & Uncertainty

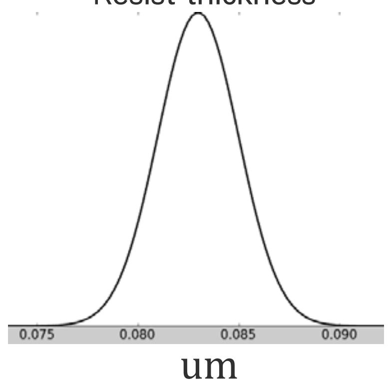


Assumptions and Distributions

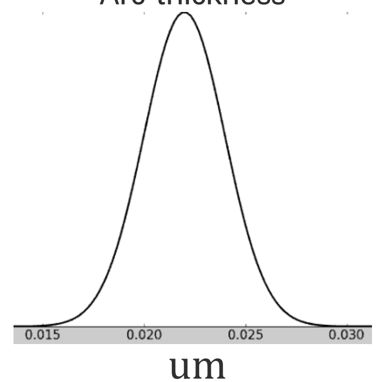
Physical parameters

Normally distributed

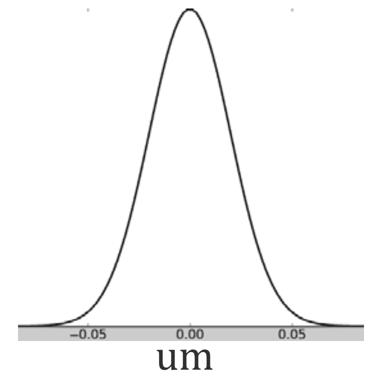
Normal (μ=given, σ=1nm)
Resist thickness



Normal (μ=given, σ=1nm)
Arc thickness

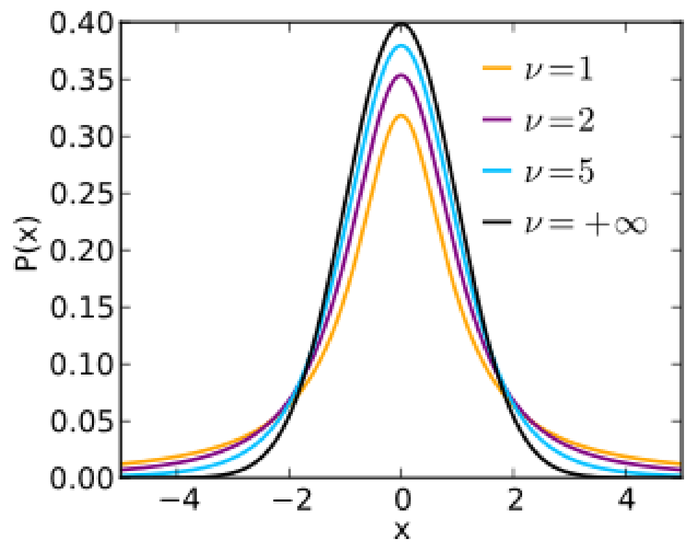


Normal (μ=0, σ=1nm)
Mask bias

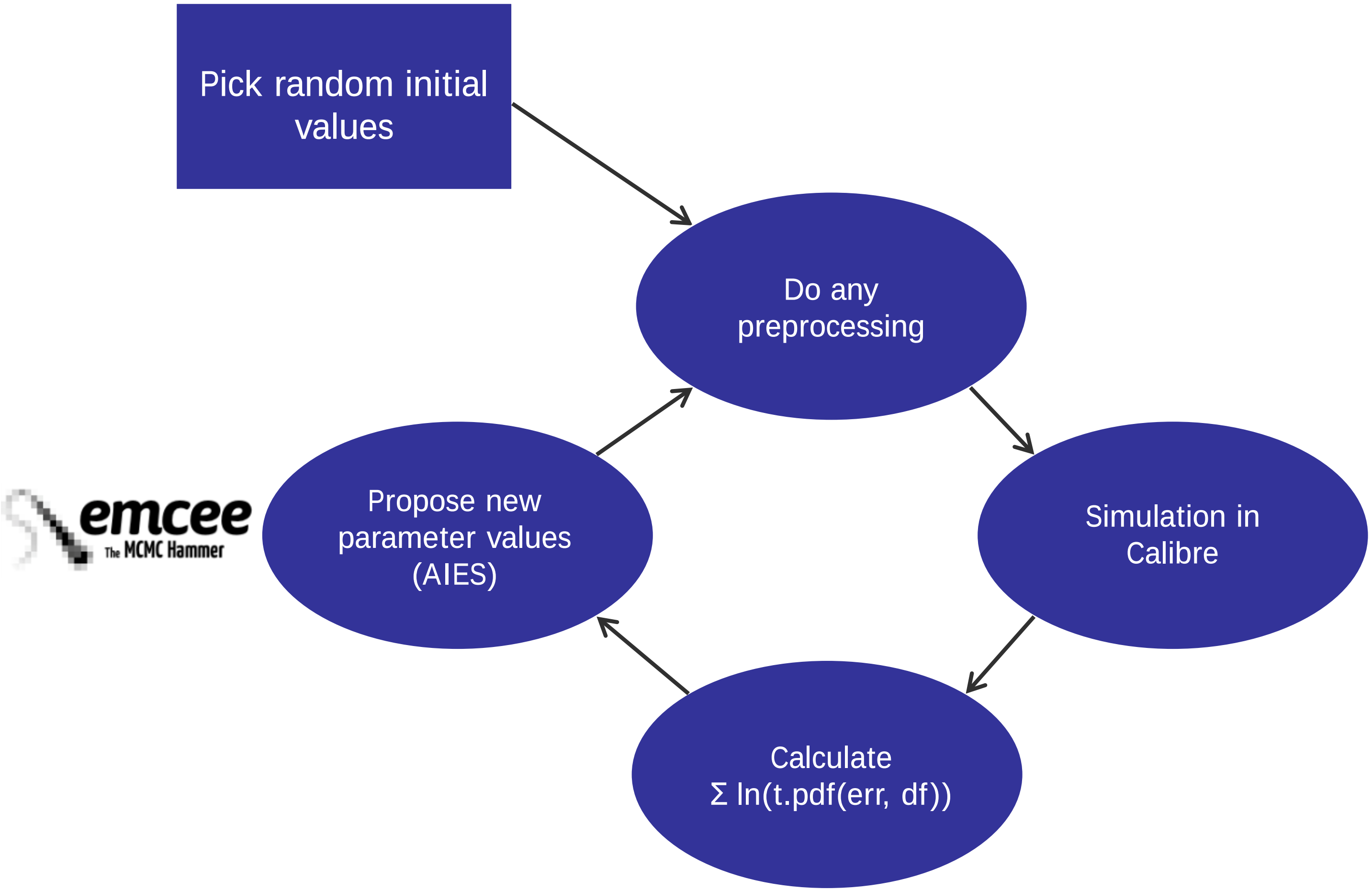


CD Measurements

Student's t-distributed

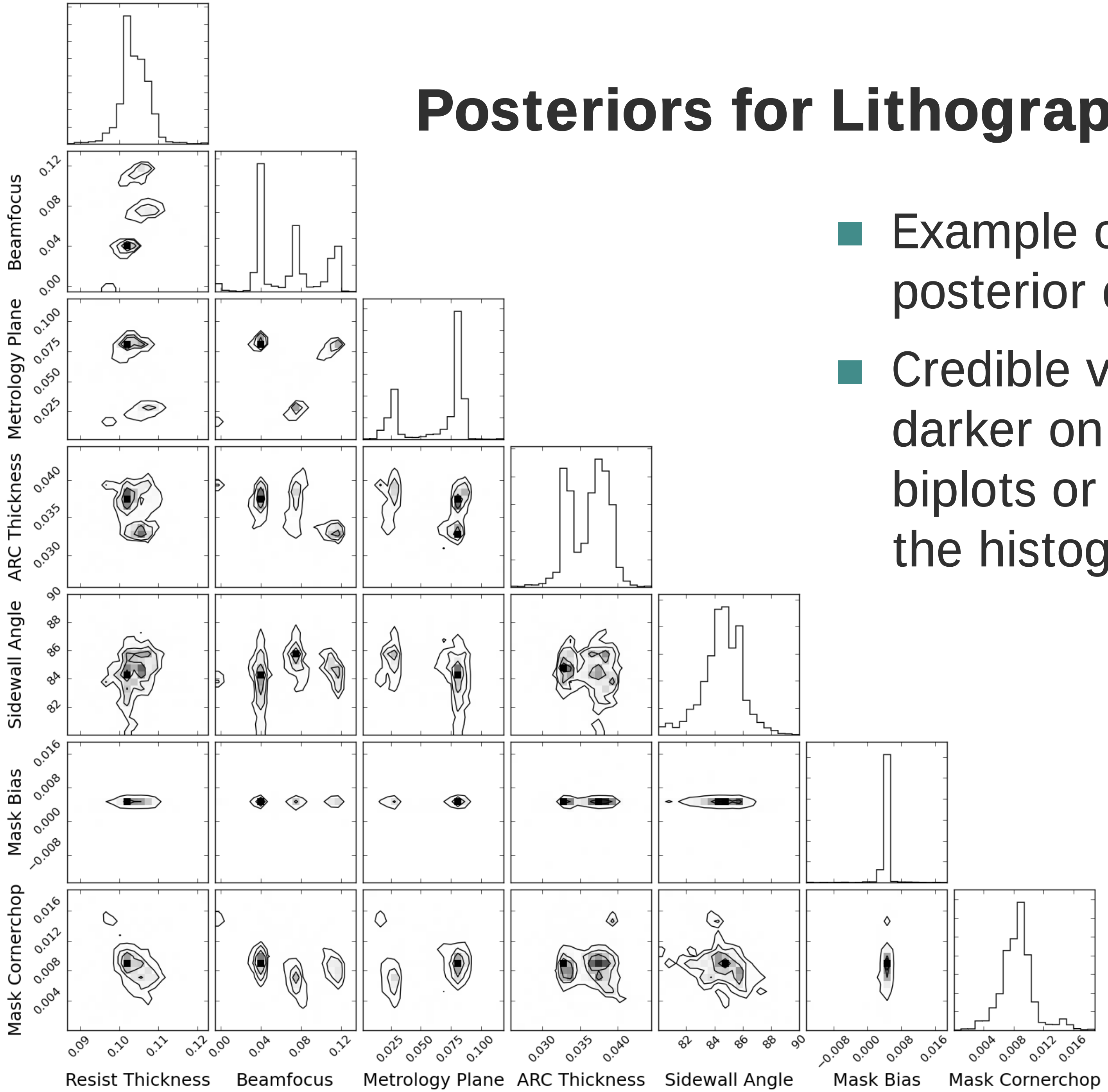


Workflow (Prototype)



http://dan.iel.fm/emcee/current/

Posteriors for Lithographic Models



- Example of parameter posterior distributions
- Credible values are darker on the 2D biplots or the peaks of the histograms

Posterior Predictive Check

- Train model with a subset of CD measurements
  - 100 measurements
- Evaluate model on larger set of CD measurements
  - 1000 measurements
- This is a ‘sanity check’ on the parameter values
- Evaluate against *traditional* (non-Bayesian) methods
  - Trained on the *full* dataset

- Best traditional RMSe
  - **4.11**
- Best Bayesian RMSe
  - **3.97**

Conclusions

- Markov Chain Monte Carlo methods are effective at optimizing many parameters in parallel
  - This is particularly important for the hefty full-chip OPC models used in manufacturing, which can take a significant amount of computational time to build and evaluate
- Uncertainties in modeling can be captured
  - Wafer CD data and model parameter information is imperfect
  - Acknowledging this under the Bayesian framework is beneficial
- Evaluation over traditional methods
  - Bayesian method with expanded parameter set finds more accurate models than traditional methods
  - Key feature: models are trained on only a subset of the data, giving them a very small over-fit risk

Mentor Graphics®

www.mentor.com