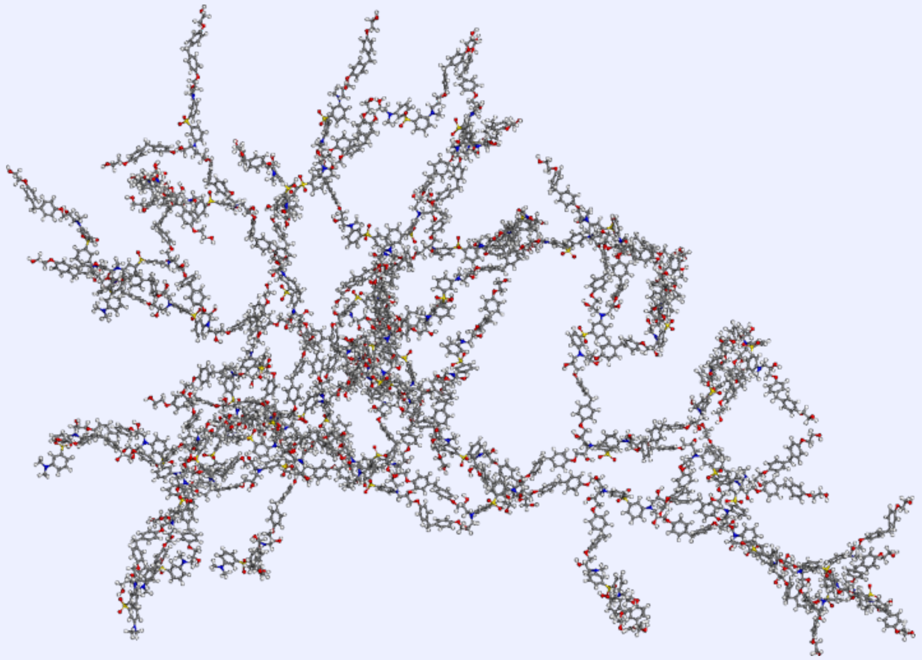
Uncertainty quantification of Molecular Dynamics Simulations for Crosslinked Polymers

Paul Patrone (IMA / NIST)
Andrew Dienstfrey (NIST Boulder)
Steve Christensen, Andrea Browning, Sam Tucker (Boeing)



Backstory:
Macro-economics of materials science

Advent of composites dramatically altered design space in aerospace engineering.

Example: Boeing 787

1st aircraft with majority carbon-composite structural components

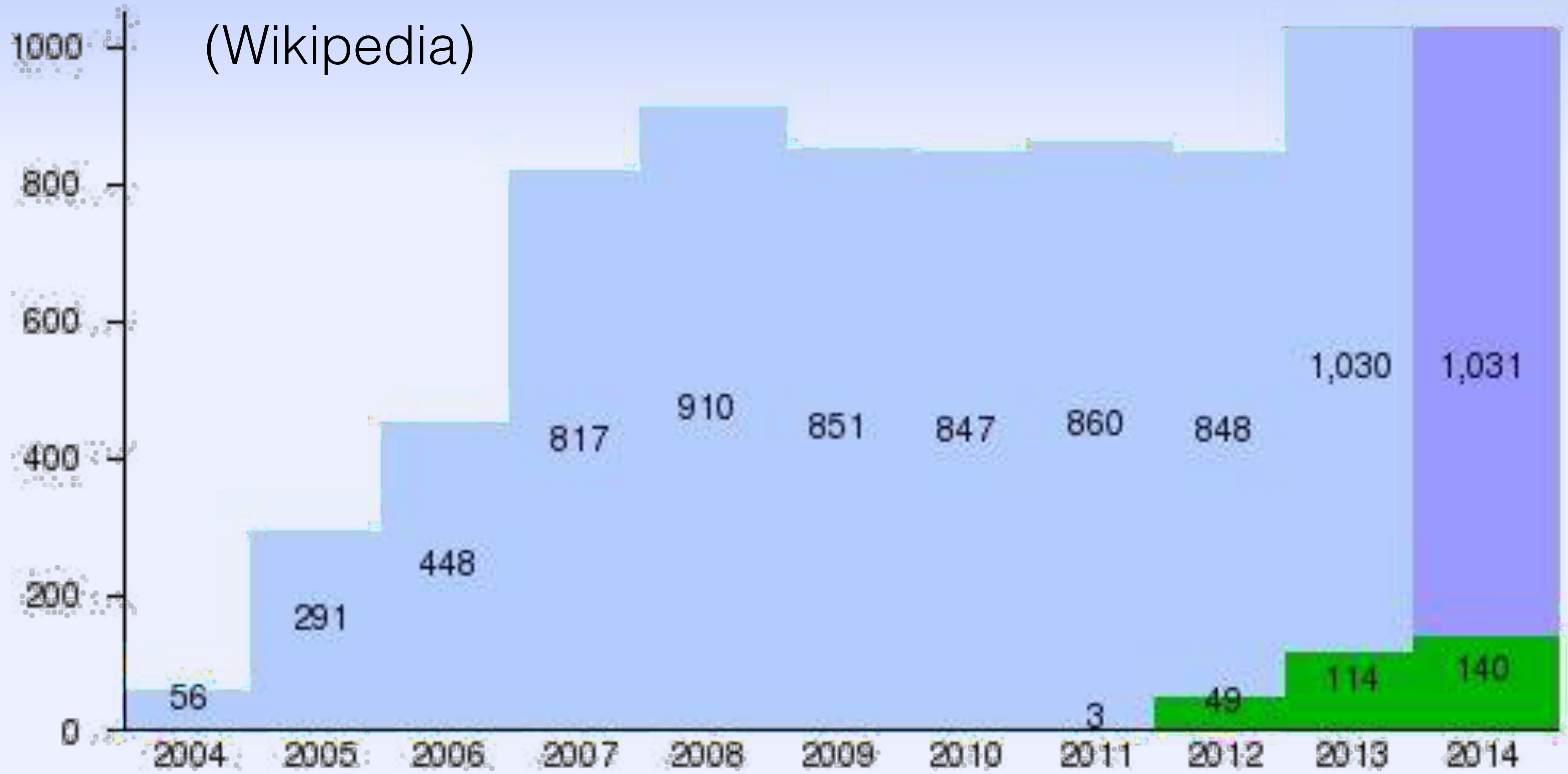
Lighter aircraft —> fuel savings (20%)

Scale of economics

(~ 1000 orders) x (~ $250 Million / order) = $250 Billion

Impact of advanced materials

Cumulative orders of 787 (blue) and deliveries (green)



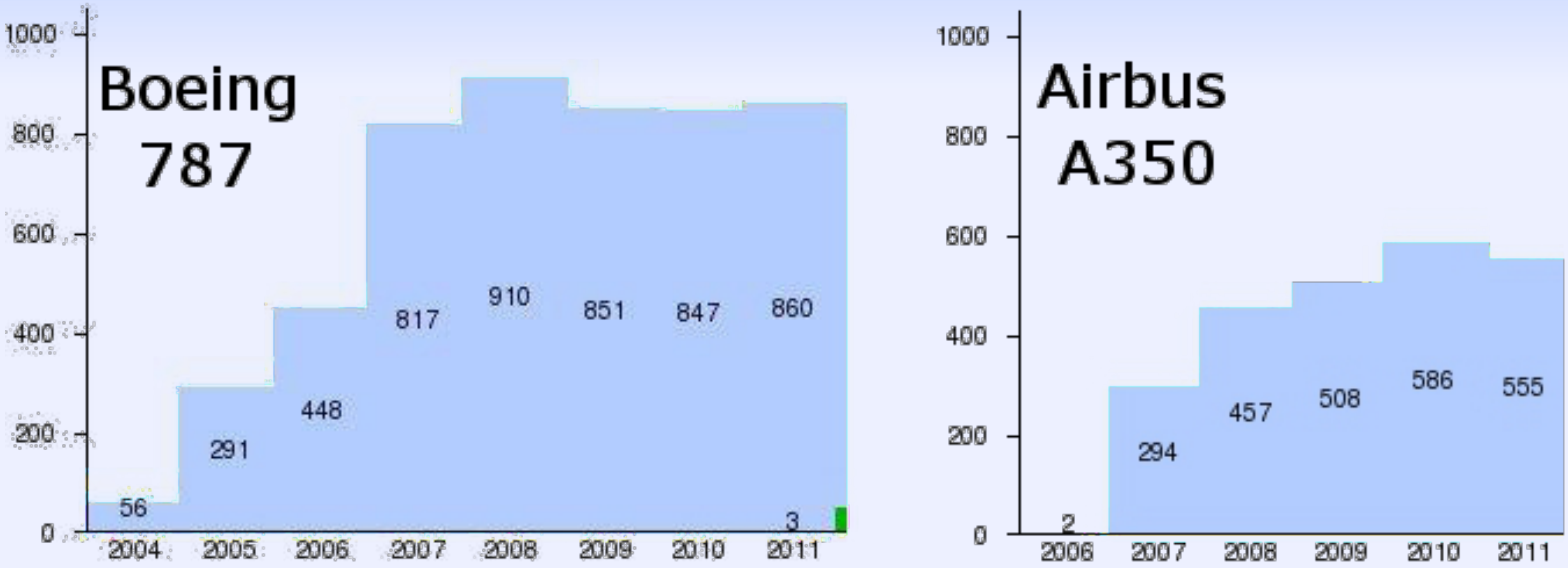
2006 Seattle Times headline

Airplane kingpin tells Airbus: Overhaul A350

"That's probably an $8 billion to $10 billion decision."

Impact of advanced materials

Cumulative orders of 787 and A350



2006 Seattle Times headline

Airplane kingpin tells Airbus: Overhaul A350

"That's probably an $8 billion to $10 billion decision."

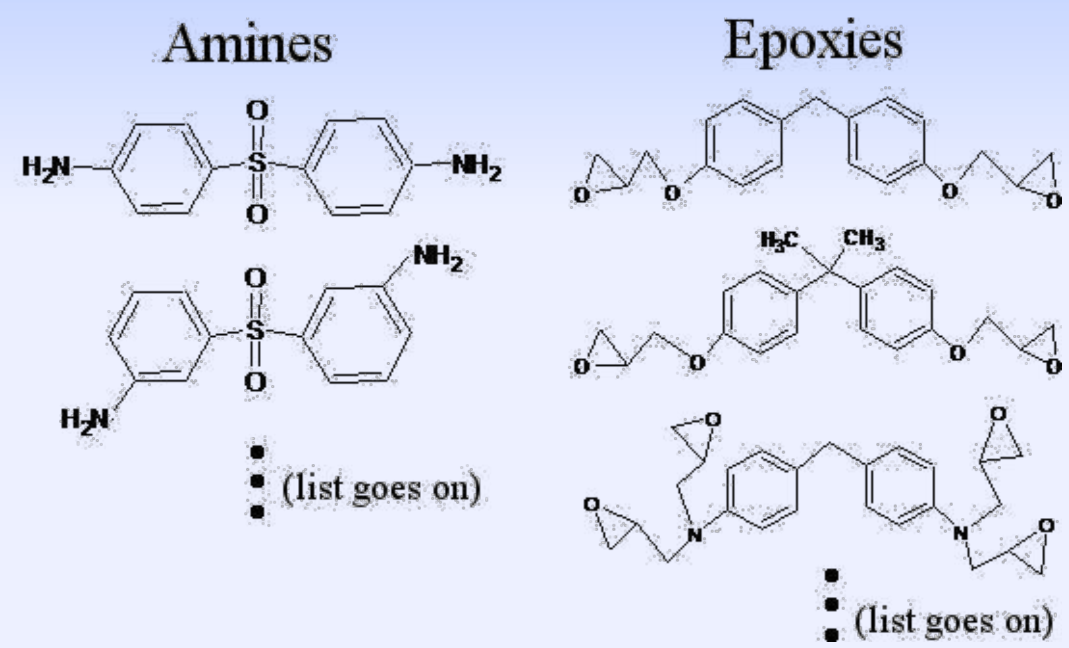
Accelerating market insertion: materials by design

Assume

"Design space" of ingredients

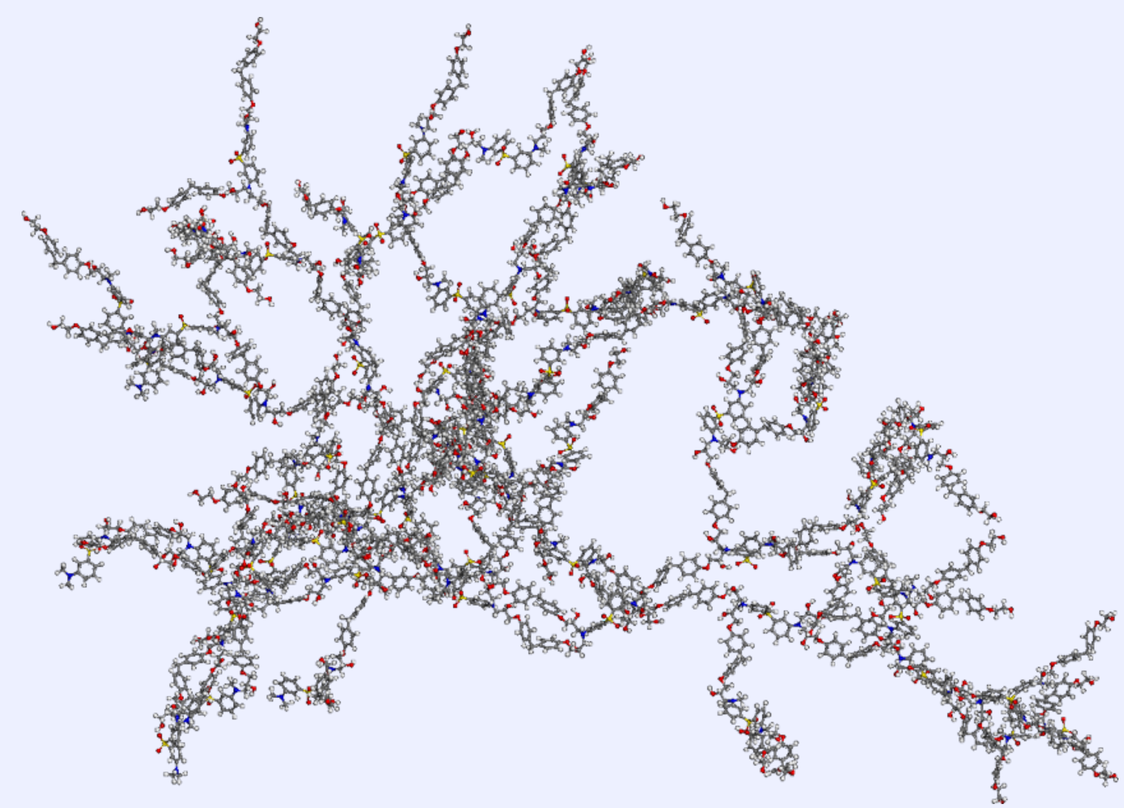
Finite simulation resources

(Very) Few experiments



Goal

Find chemistry with,
e.g. highest T_g



Roles of UQ in modeling workflows

Verification

Check math,
remove bugs

Does data look
like I expect?

Data of sufficient quality
to make predictions?

Compute uncertainties
arising **from within model**.

Validation

Calibrate model

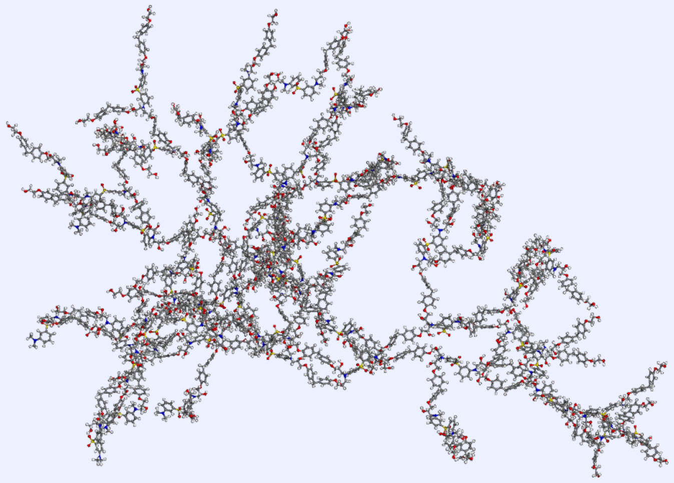
Estimate uncertainties
arising from ….

calibration parameters

missing physics

model form error

Otherwise assume model is valid at this stage



Test "real-world" predictive power

Roles of UQ in modeling workflows

Verification

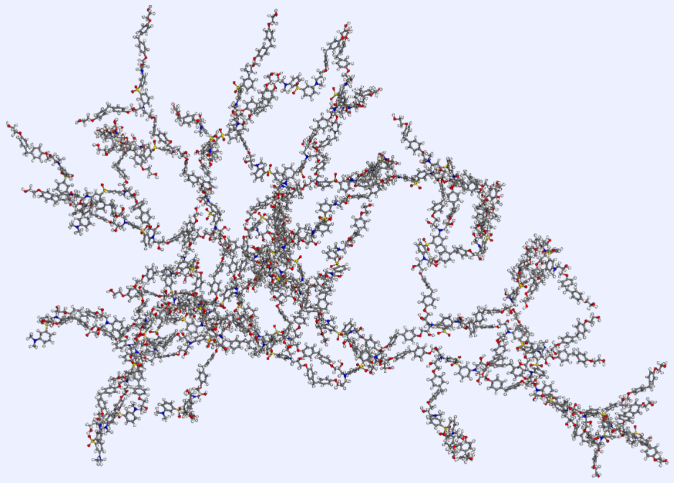
Check math,
remove bugs

Does data look
like I expect?

Data of sufficient quality
to make predictions?

Compute uncertainties
arising **from within model**.

Otherwise assume model is valid at this stage



Today's focus on verification

Helps modelers to be precise about what they mean

Improves reproducibility

Streamlines validation

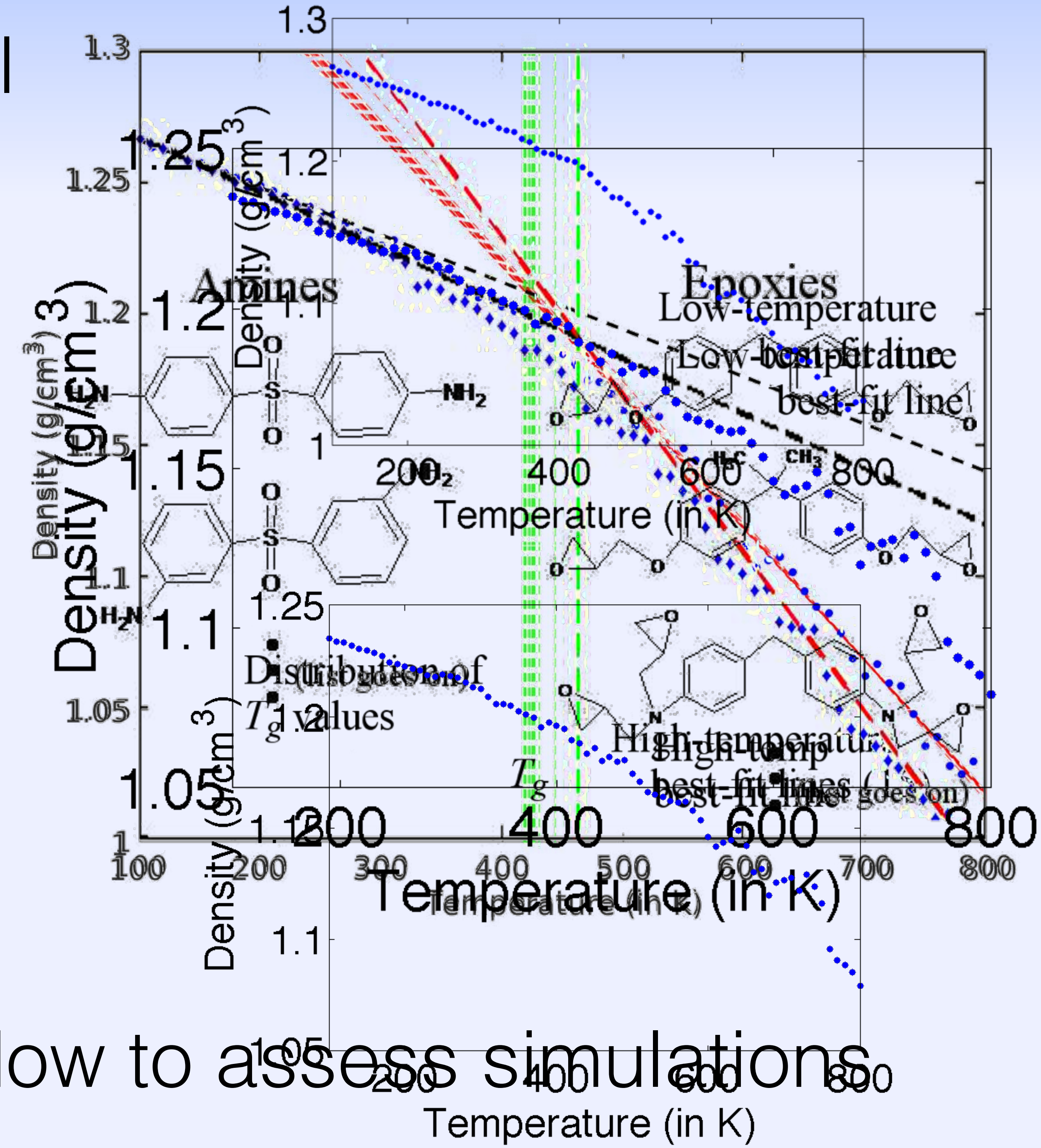
Some complicating issues for Tg

Incomplete list

1. Can we extract meaningful T_g from simulated data

2. How to combine data?

3. How to work within non-analytic design space?



Hardened & verified workflow to assess simulations

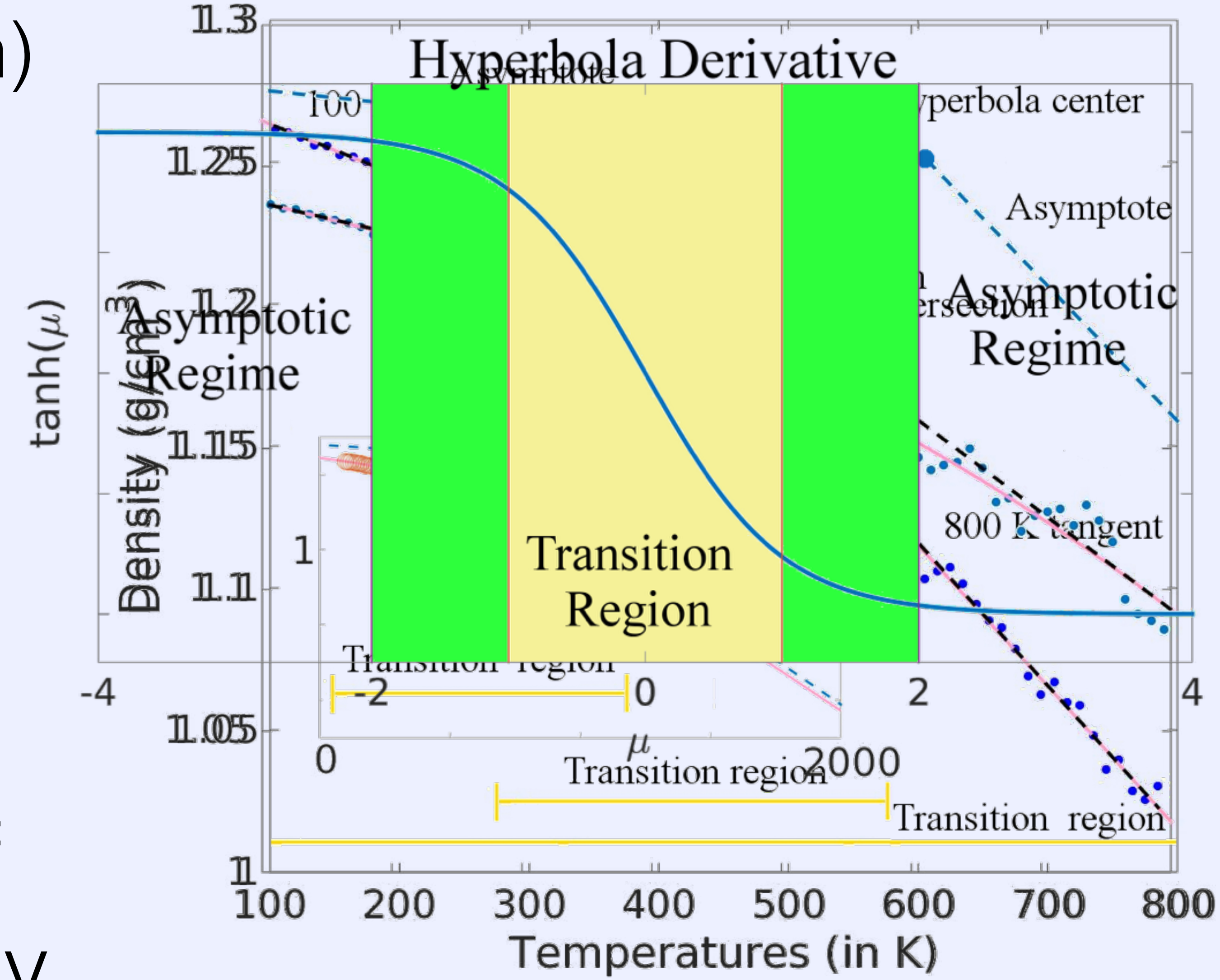
Assessing ability to extract T_g

Consistency with underlying definitions

T_g defined as hyperbola center
(same as asymptote intersection)

Automatically finds
“asymptotic regimes”

Data inconsistent with T_g if
asymptotic regimes far away



Assessing ability to extract T_g

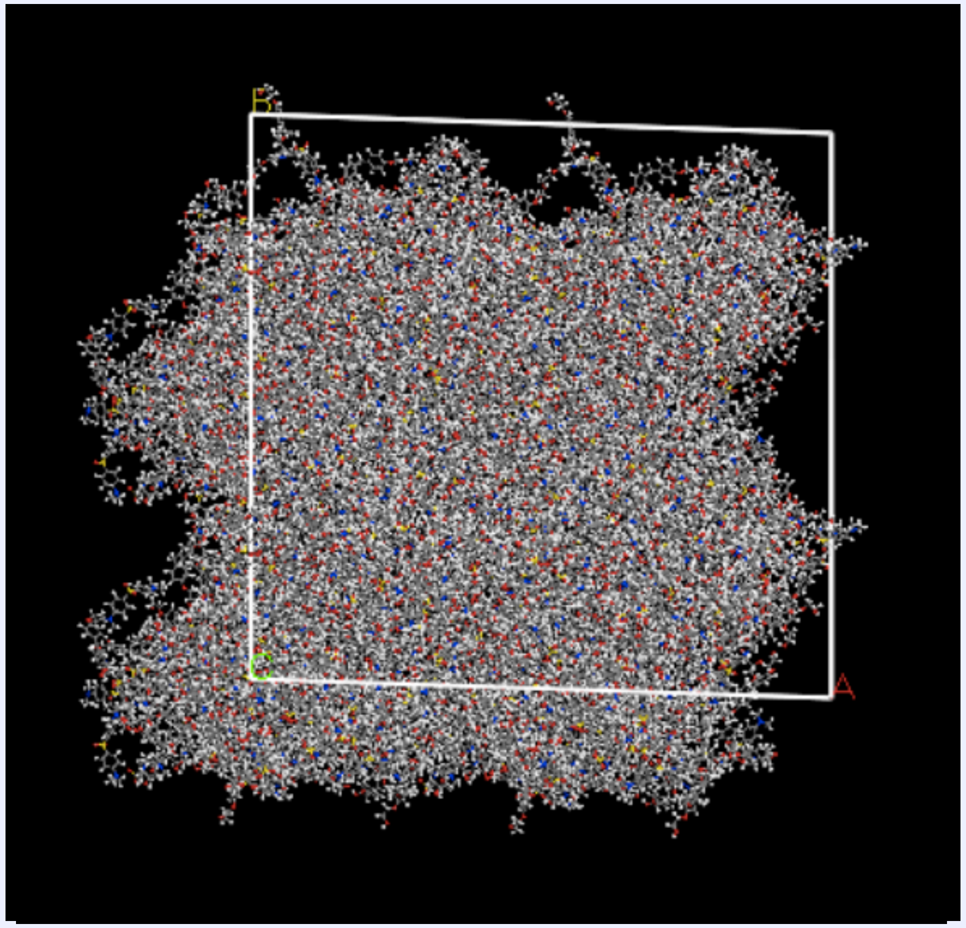
Convergence to bulk limit

An industry oxymoron:

high-throughput, bulk-scale, atomistic-detail MD

This is not bulk…?

How do we know?



Observations from statistical mechanics

As # of particles $N \to \infty$

1) measurable quantities are independent of N
2) variances of measurable scale as $1/N$

Hyperbola fit **(non-linear)**

Analytically: $T_g = H(T, \rho)$

density data

temperatures

$$As\ N \to \infty,\ T_g \approx H(T, \bar{\rho}) + \delta\rho(N) \cdot \nabla_\rho H \Big|_{(T, \bar{\rho})} + O\left(\|\delta\rho\|^2\right)$$

bulk mean

Hyperbola fit approximately linear

$1/N$ fluctuations

Observations from statistical mechanics

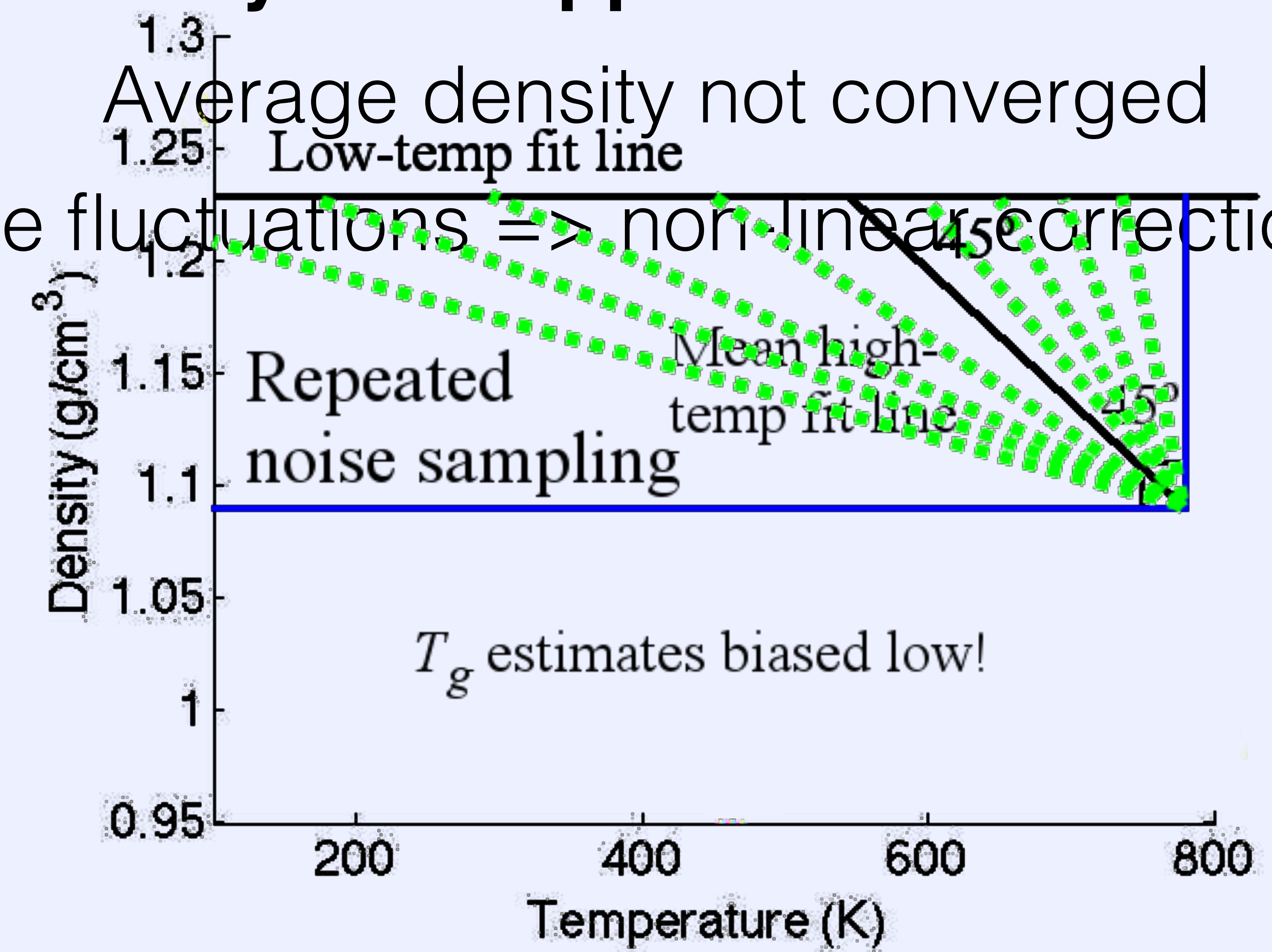
$$As\ N \to \infty,\ T_g \approx H\left(T, \bar{\rho}\right) + \delta\rho\left(N\right) \cdot \nabla_\rho H \Big|_{(T,\bar{\rho})} + O\left(\|\delta\rho\|^2\right)$$

Large fluctuations => non-linear correction bias

Two ways this approximation can fail

Average density not converged

Large fluctuations => non-linear correction bias

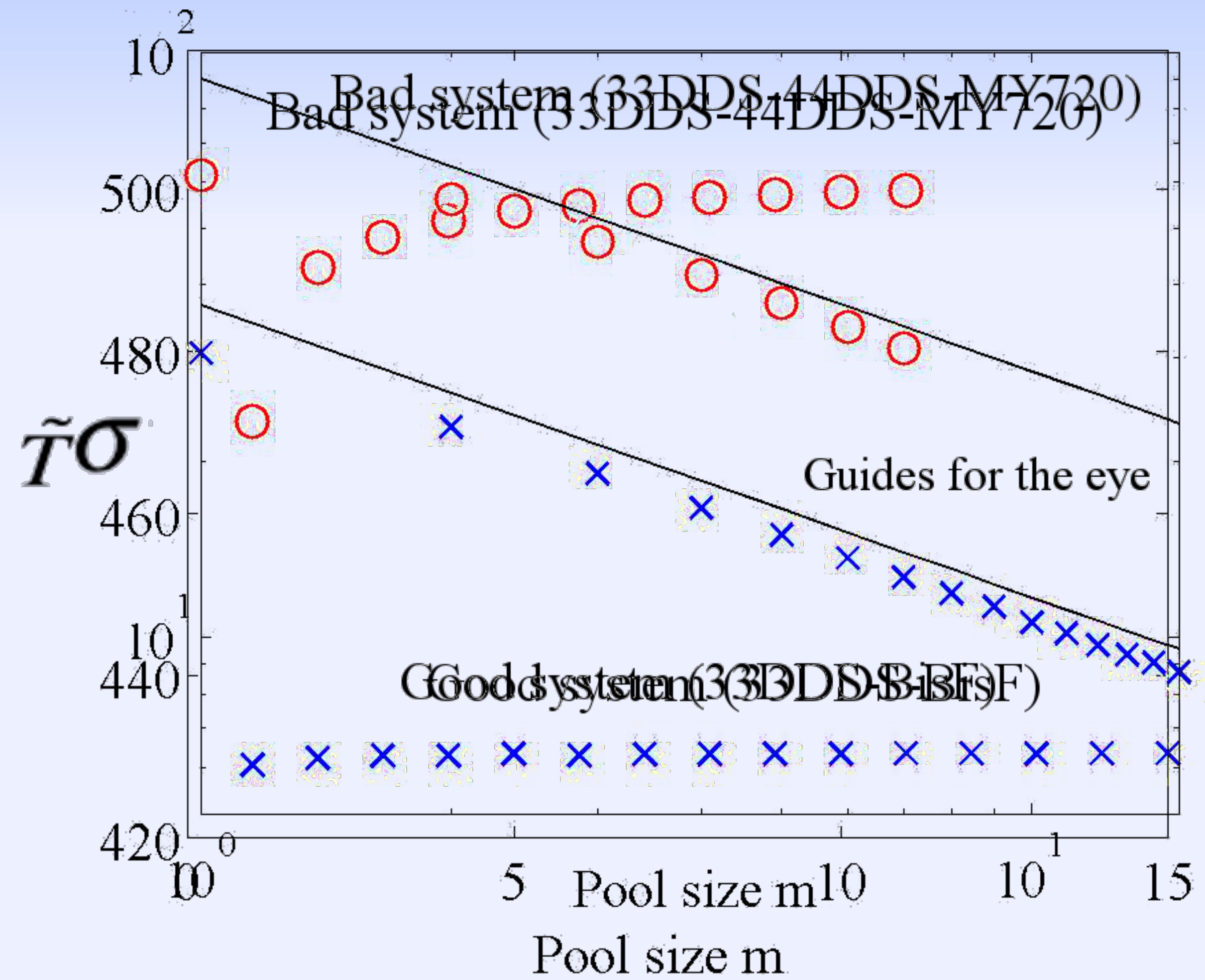


Assessing ability to extract T_g

Is hyperbola fit biasing results?

Test for bias (pooling)

Construct $\binom{M}{m}$ average $T_{g,i}$ from every combination of m data sets chosen from a total of M



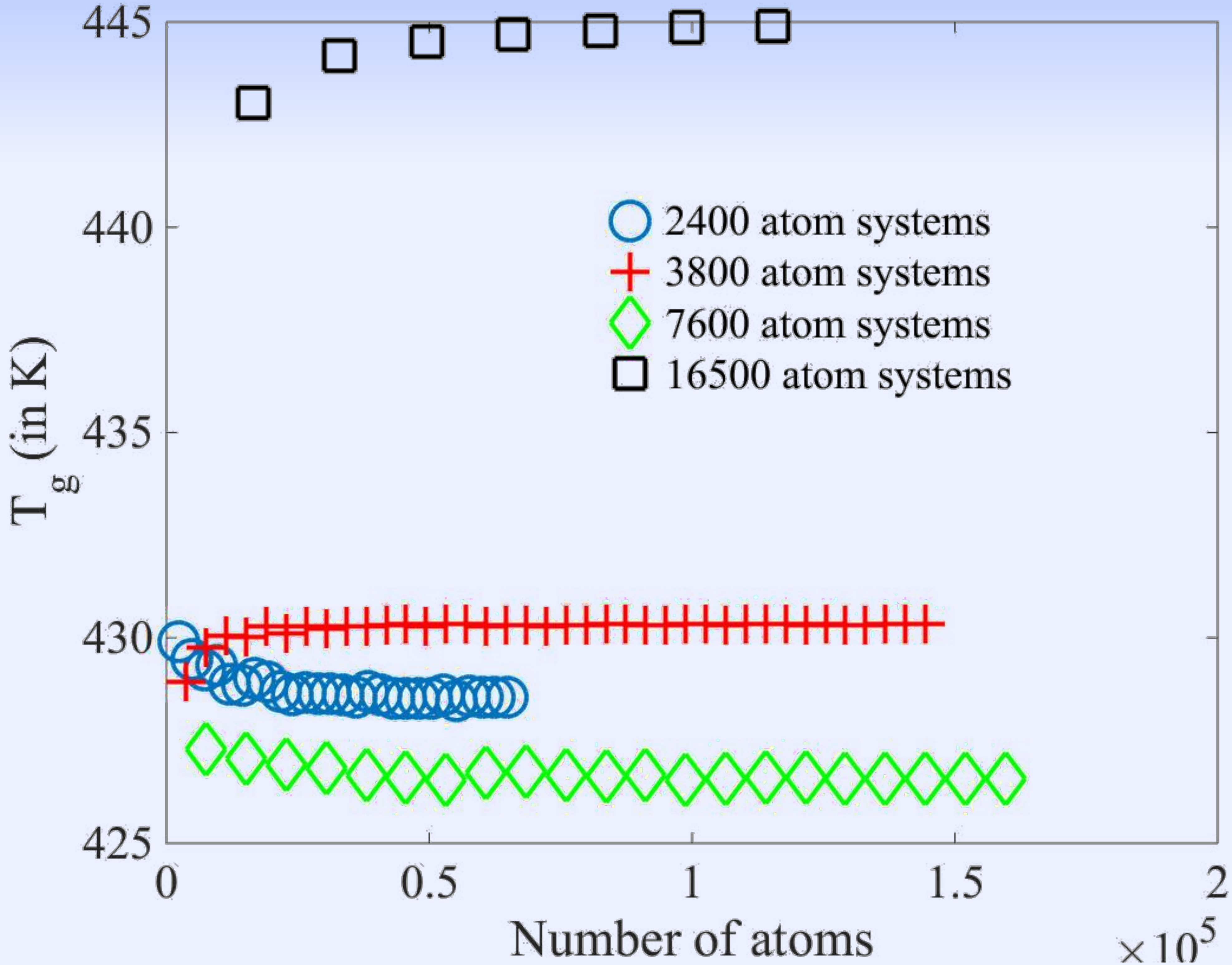
$$\tilde{T} = \binom{M}{m}^{-1} \sum_i T_{g,i} = \text{constant}$$

$$\sigma^2 = \frac{1}{M-m} \sum_i \left(T_{g,i} - \tilde{T}\right)^2 \propto \frac{1}{m}$$

IF linearity holds

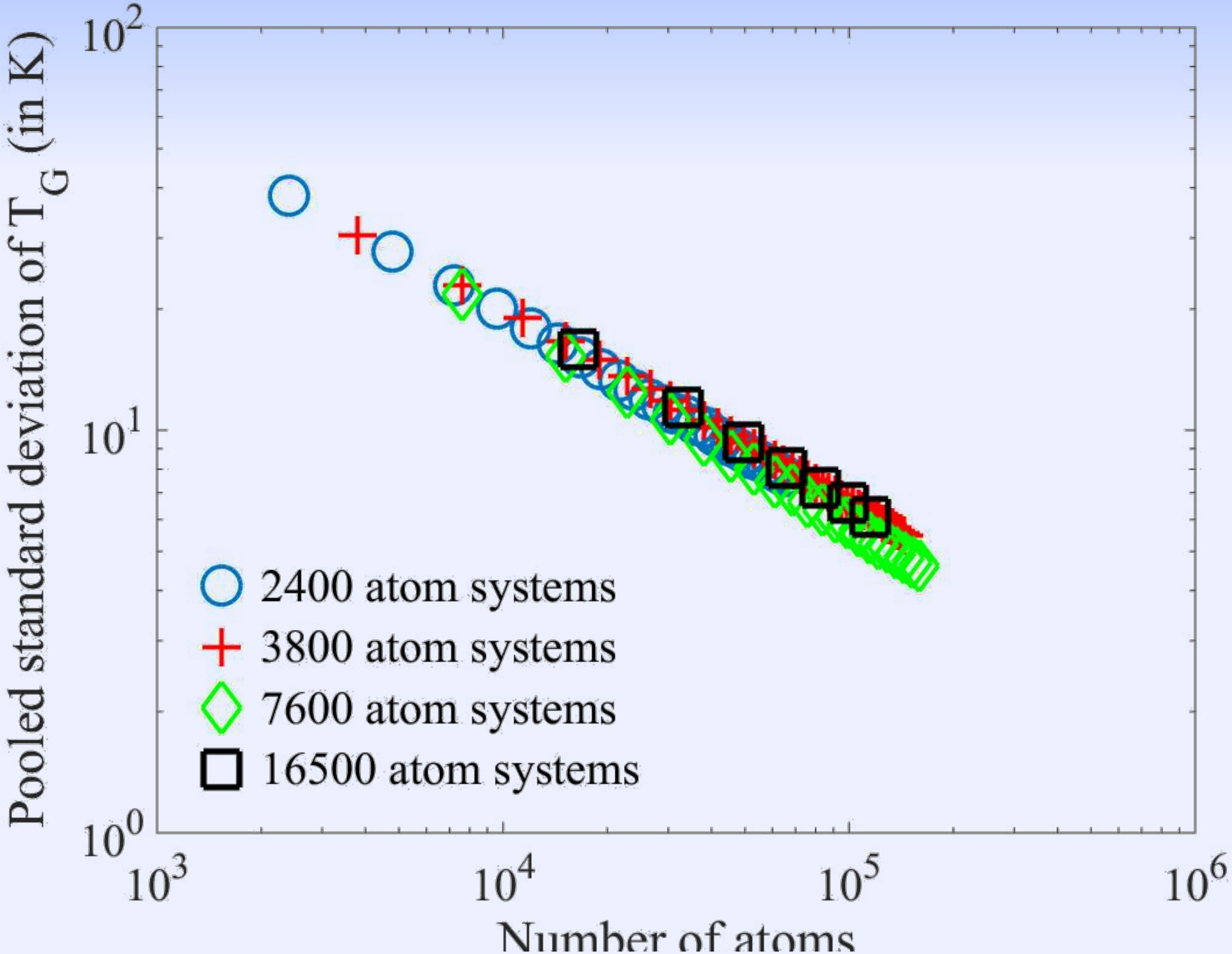
Assessing ability to extract T_g

Is average density converged?



Assessing ability to extract T_g

Is average density converged?



Assessing ability to extract T_g

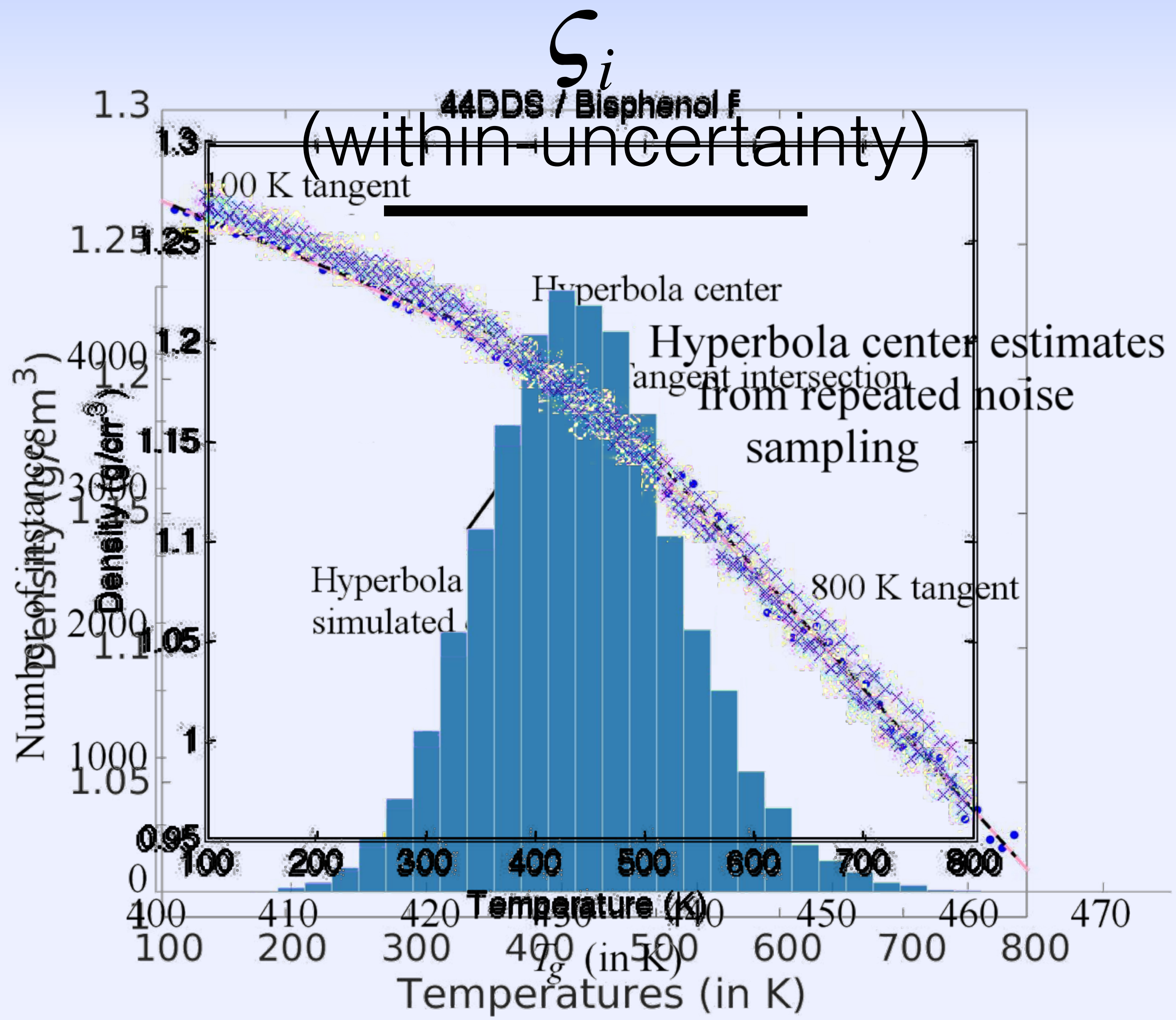
Did we extract a “precise” T_g value from the fit?

Noise affects fit, & hence our T_g estimate

Noise model for residuals

$$\rho(T) = \bar{\rho}(T) + \eta$$

Sample noise η & fit hyperbola to yield new T_g



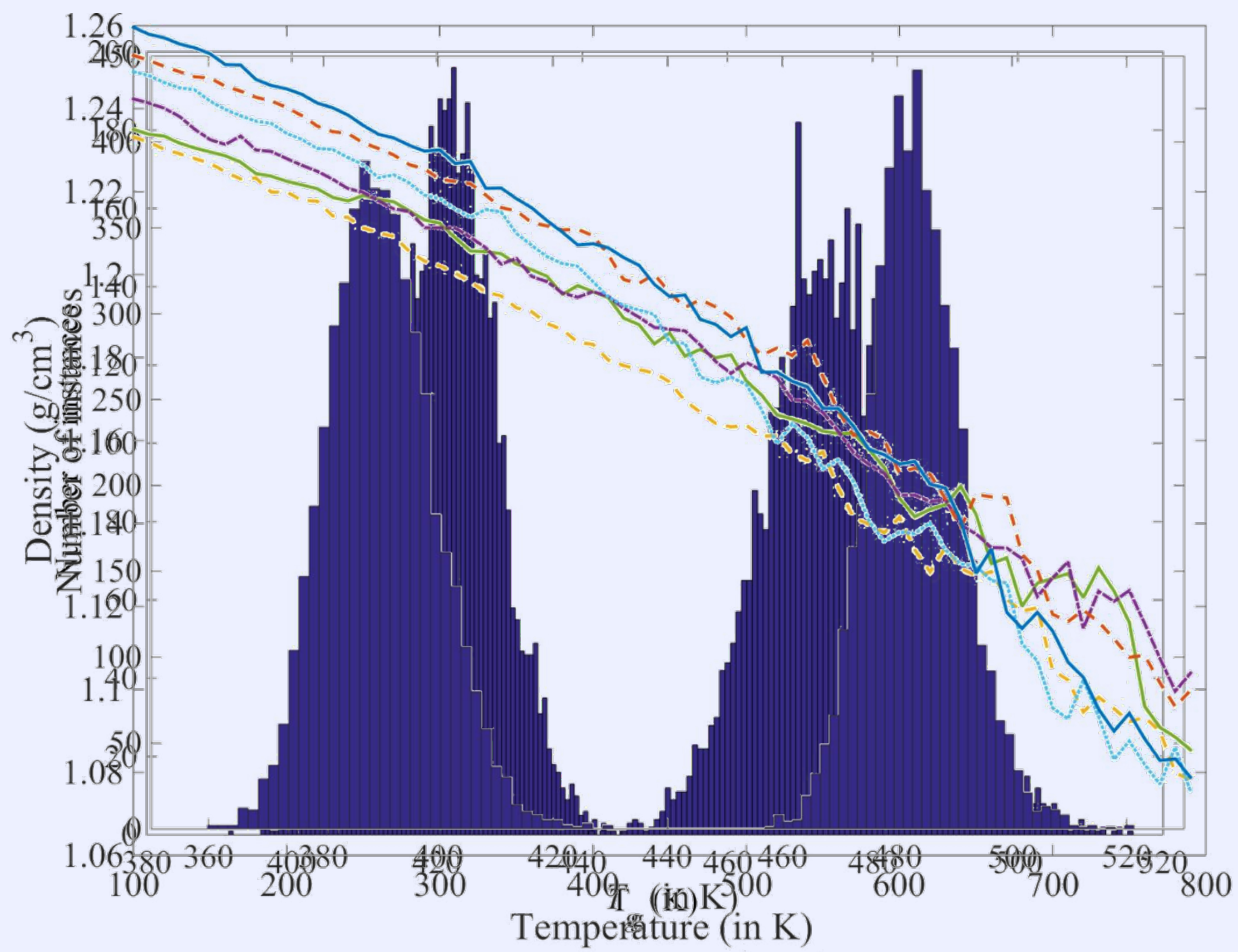
Combining data

Should all data sets be treated equally?

Two simulations may yield different within-uncertainties

Worse, predictions may not overlap

How do we account for missing physics?



Combining data

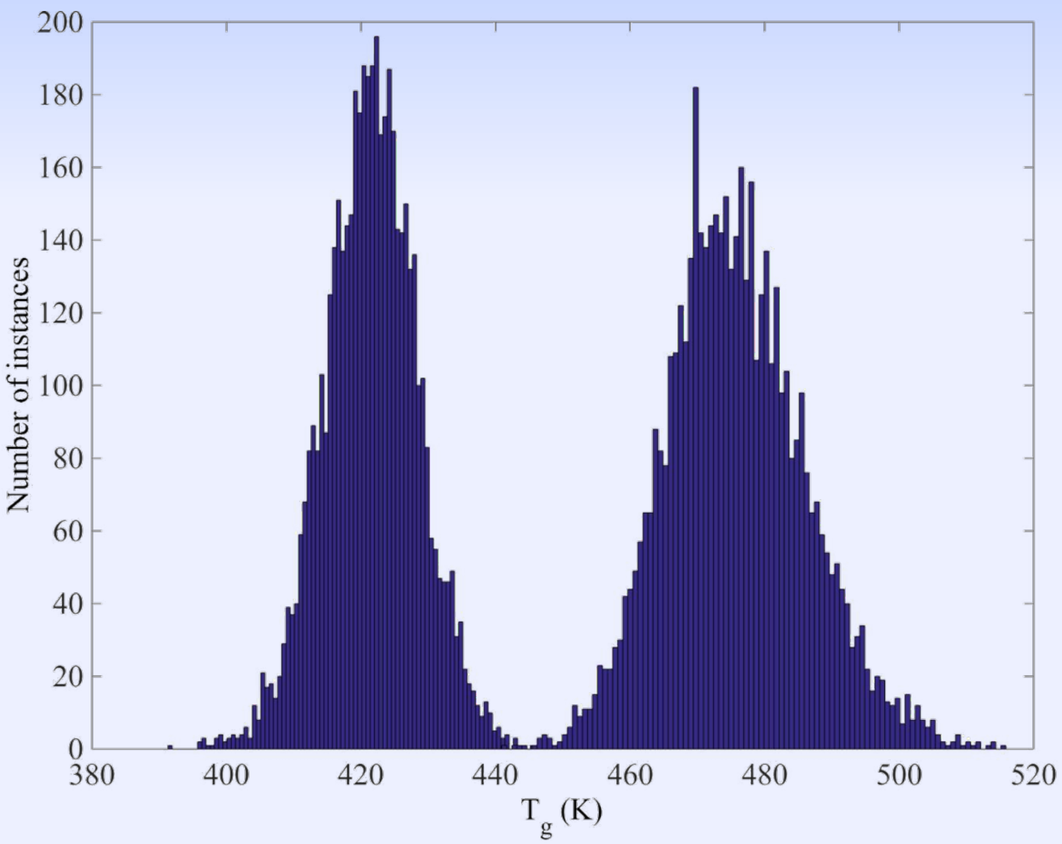
Should all data sets be treated equally?

Weighted-mean statistic model:

T_g from ith simulation

$$\tau = \left[\frac{1}{\sum_i y^2} \right]^{-1} \sum_i \frac{T_{g,i}}{y^2}$$

uncertainty from
under-modeled physics

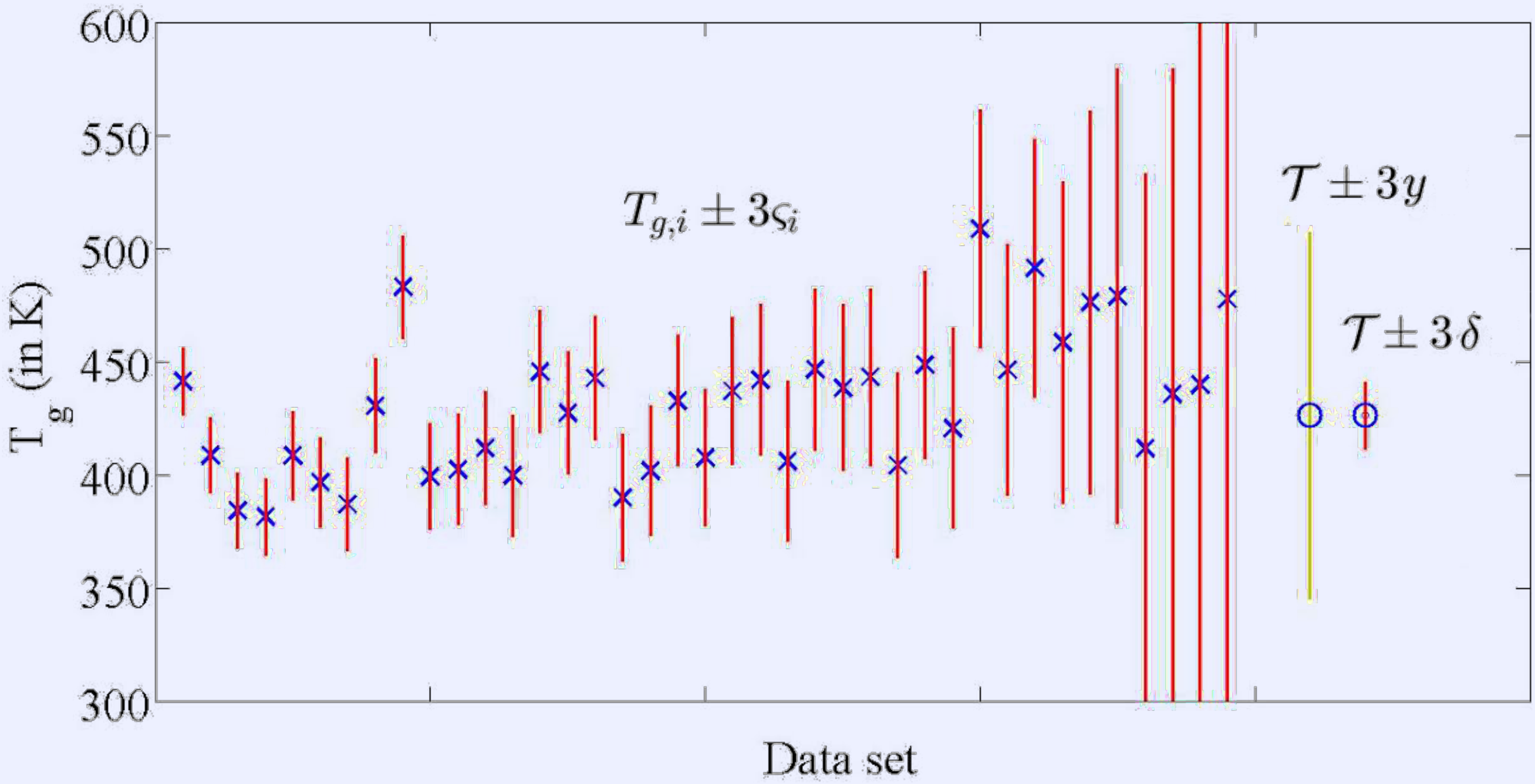


de-weights "imprecise"
& overconfident T_i

Solve for y using maximum likelihood analysis (MLE)

Combining data

Final uncertainty estimate: $\delta^2 = \left[\frac{1}{\sum_i y^2 + \varsigma_i^2} \right]^{-2} \sum_i \frac{\left(T_{g,i} - \tau\right)^2}{\left(y^2 + \varsigma_i^2\right)^2}$

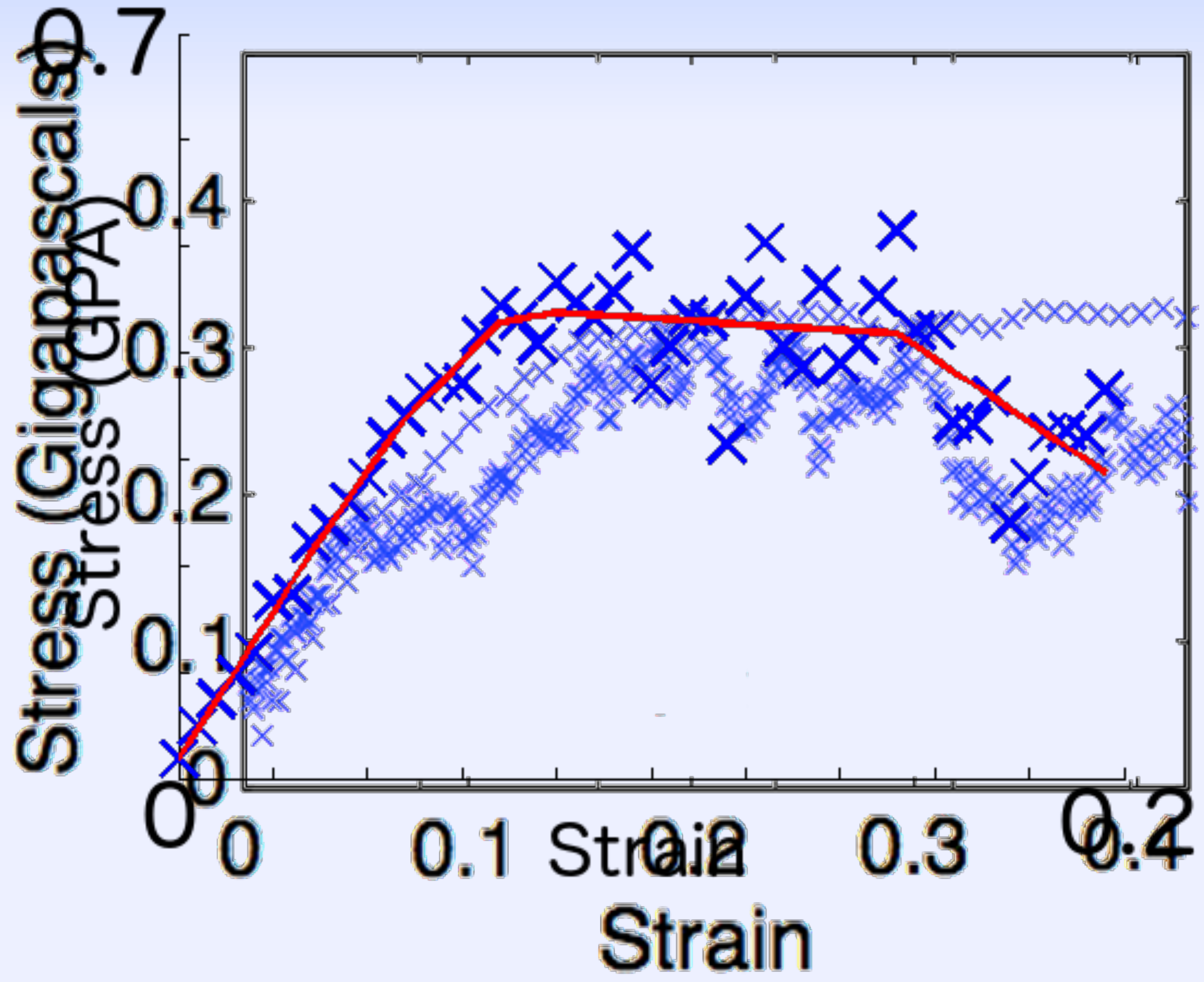


Open problems: yield strain

Strain at which material no longer resists a load

Identified as maximum of stress-strain curve

How do we deal with noisy data?



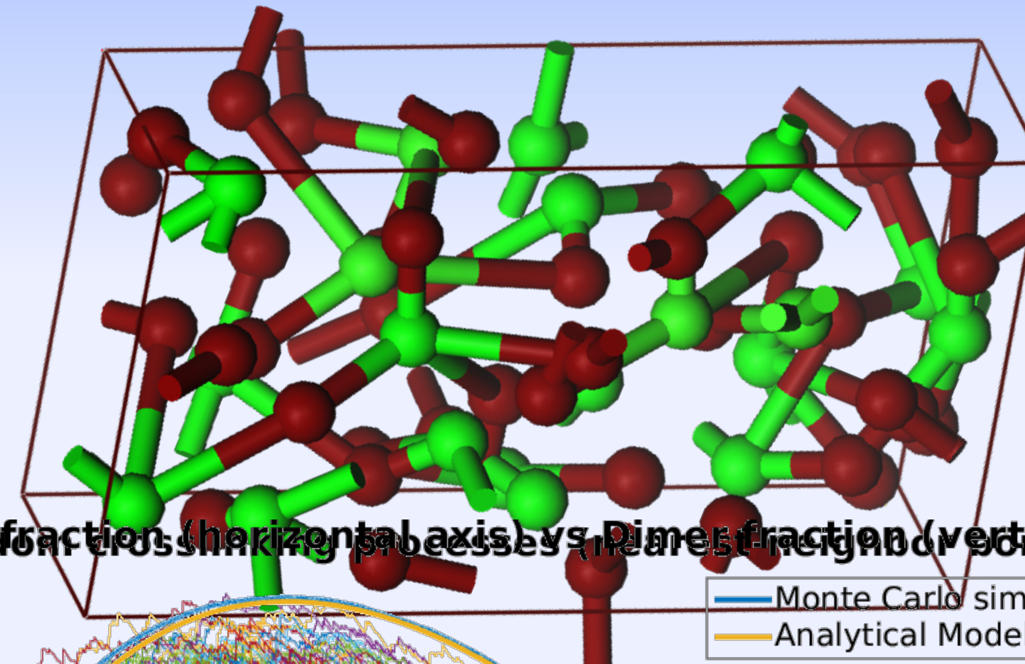
Analysis using convex functions.

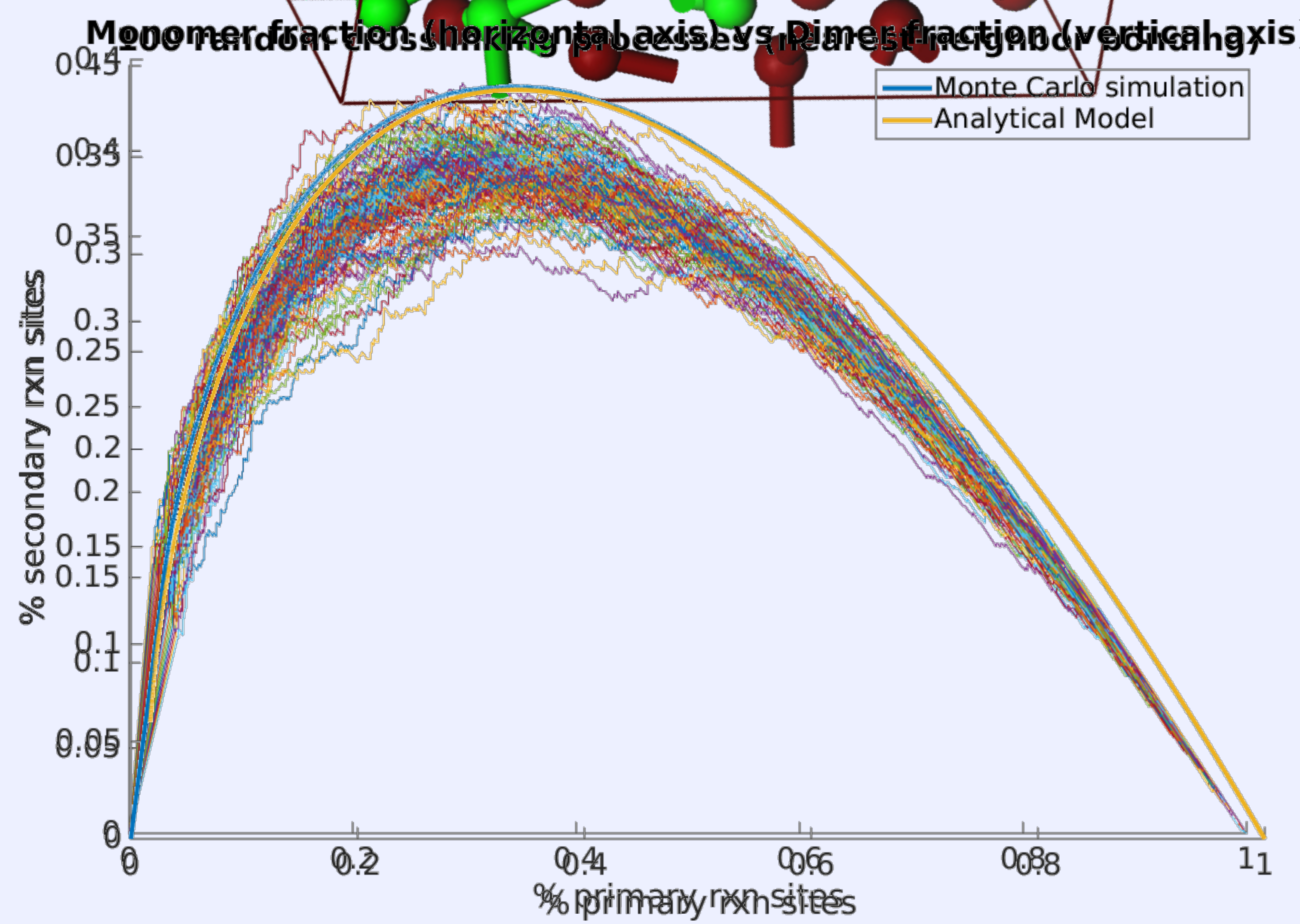
Open problems: understanding statistics of "realistic" crosslinked networks

What is mean number of edges at a given vertex?

Depends on x-link algorithm: e.g. random bonding, nearest neighbor….

Analytical (probabilistic) models to describe simulated predictions





Conclusions

MD is driving development of materials & other disruptive technologies

UQ can help industry assess usefulness of their simulations

Lots of open problems