Uncertainties in Nanometer-Scale Dimensional Metrology

J. S. Villarrubia
National Institute of Standards & Technology

0.2 µm × 0.2 µm

NIST

Overview

I.   Measurement issues

   A.   Probing errors

      Solution: Model-based metrology

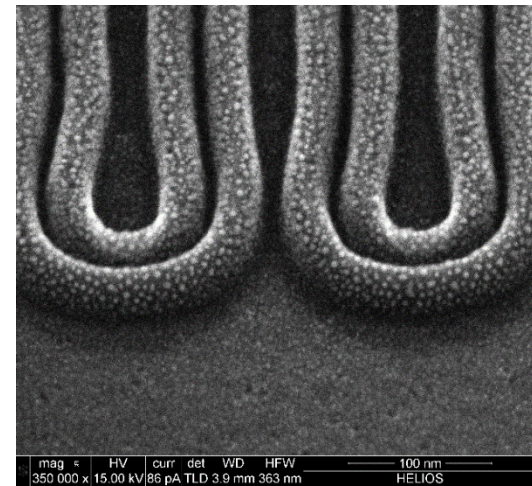   B.   Limited information in the signal

      Solution: Hybrid model-based metrology
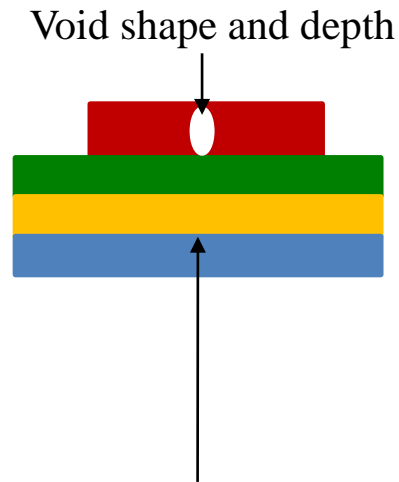
II. Uncertainty quantification issues

   A.   Modeling error

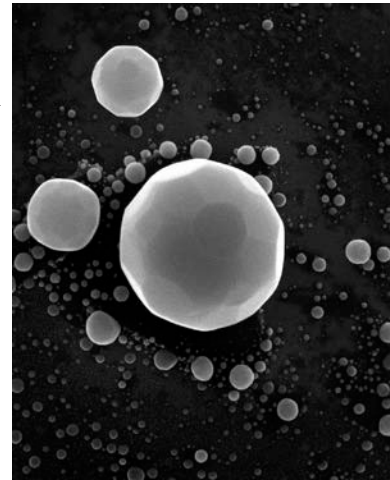   B.   "Methods divergence" in hybrid metrology
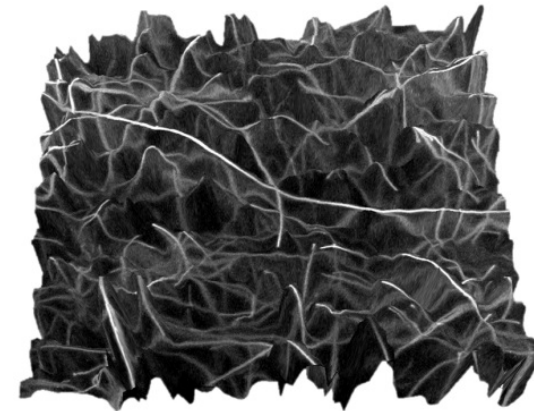
Nano-scale measurement requirements:

Width, height, wall angles of ~10 nm wide transistor gates and fins (for FinFETs) with 0.1 nm standard uncertainty.

Void shape and depth

Layer thickness, density, and composition

Size, shape, and orientation of nanoparticles (3 nm - 60 nm in this image)

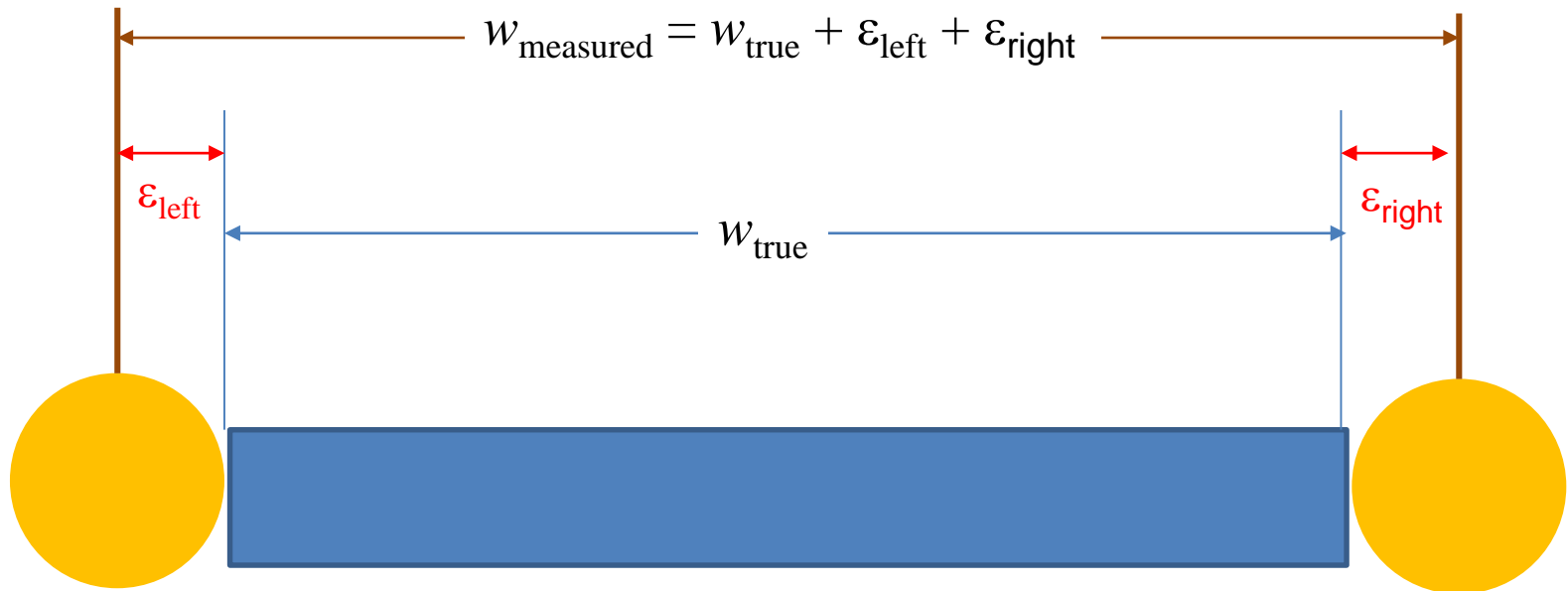Position & depth of carbon nanotubes in an epoxy matrix.

The tolerances are near (sometimes below) the spatial resolution limit of the measuring tool.

For such measurements, model-based metrology is essential.

NIST

MEASUREMENT ISSUE #1:
Probing errors

NIST

Probing errors for a mechanical probe

$$w_{\mathrm{measured}} = w_{\mathrm{true}} + \varepsilon_{\mathrm{left}} + \varepsilon_{\mathrm{right}}$$

$\varepsilon_{\mathrm{left}}$

w_{true}

$\varepsilon_{\mathrm{right}}$

Measurements of an object's width with a mechanical probe are subject to probing errors, unless you can correct for the probe size. Probe errors on left and right are the same sign and add.
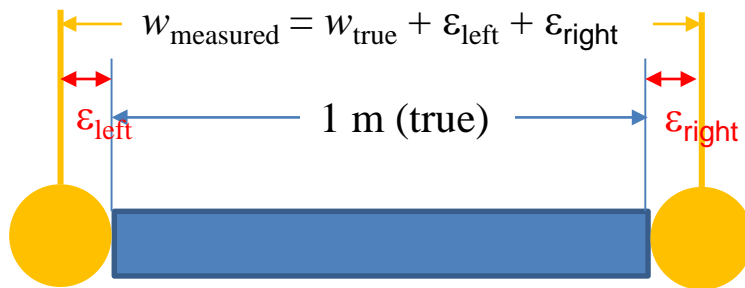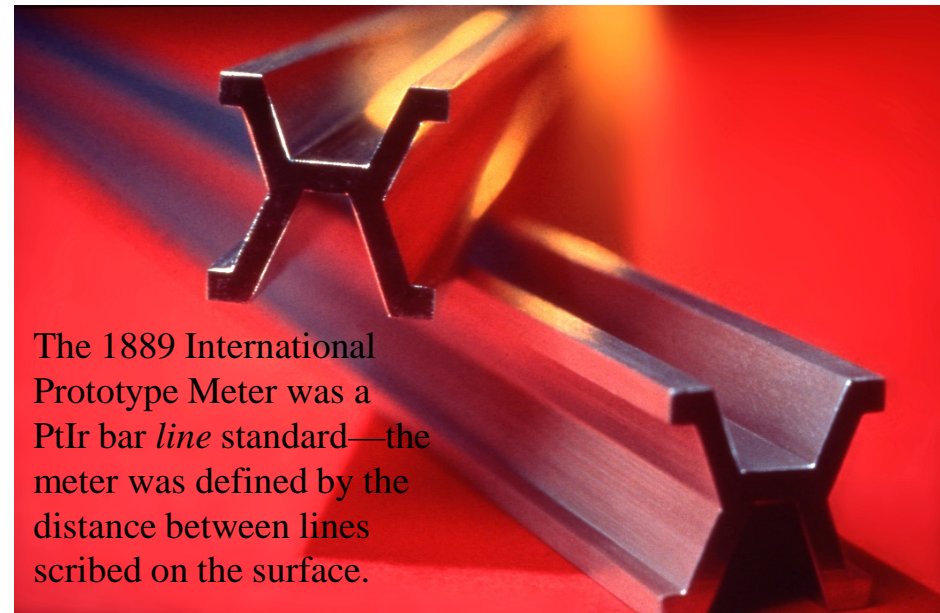
NIST

Probing errors for an optical probe

$$w_{measured} = w_{true} + \varepsilon_{left} + \varepsilon_{right}$$
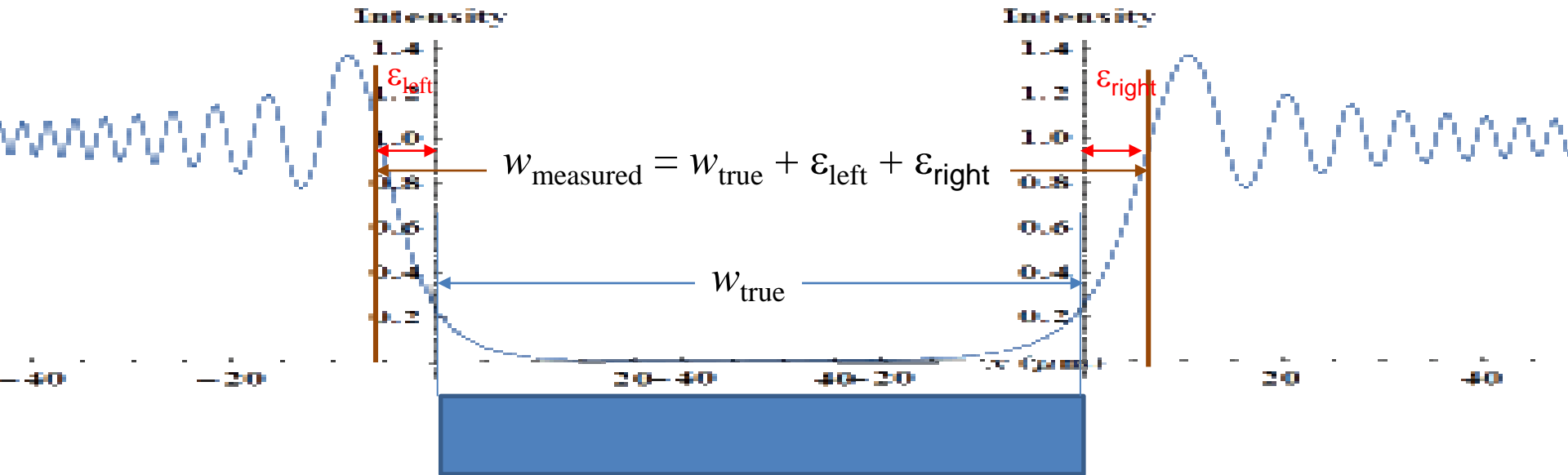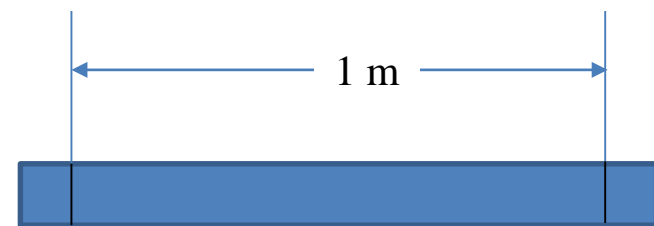
Optical measurements of the object's width are *still* subject to probing errors because of diffraction. If you guess the intensity that corresponds to the edge position, the error may be either positive or negative, but the errors on the two ends will still be additive.

In fact, probing error is a general phenomenon, due to the characteristic left/right symmetry of the edges.

A historical vignette on probing errors

The original 1799 Mètre des Archives was a Pt bar *end* standard—the meter was defined by the distance between the left and right ends of the bar.

1799: Mètre des Archives (Platinum Bar)

Source: http://en.wikipedia.org/wiki/History_of_the_metre

The 1889 International Prototype Meter was a PtIr bar *line* standard—the meter was defined by the distance between lines scribed on the surface.

$$w_{measured} = w_{true} + \varepsilon_{left} + \varepsilon_{right}$$

ε_{left} 1 m (true) ε_{right}

Width measurement

1 m

Displacement measurement

The international metrology community sidestepped probing errors in the meter definition when it switched from the original 1799 end standard (requiring a size measurement, with probing errors) to the 1889 line standard (displacement measurement, no inherent probing errors). This is great if you get to choose your measurand…
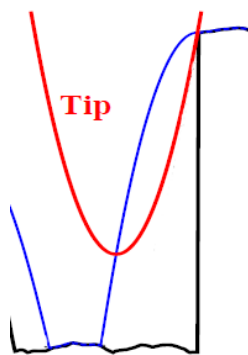
NIST

…but in real life the measurands choose us.

w
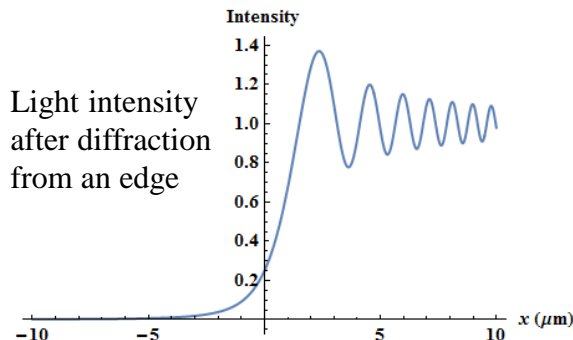
Gate

n   n

p
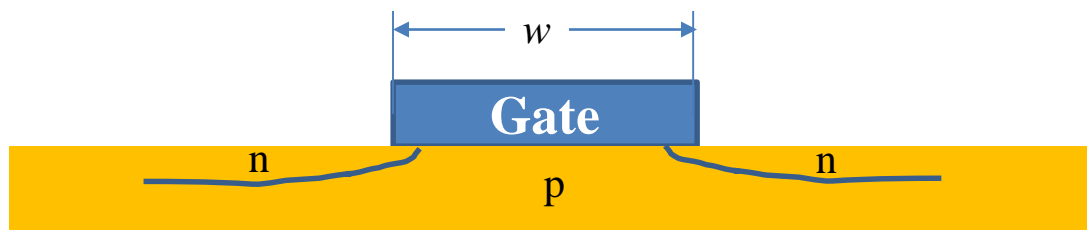
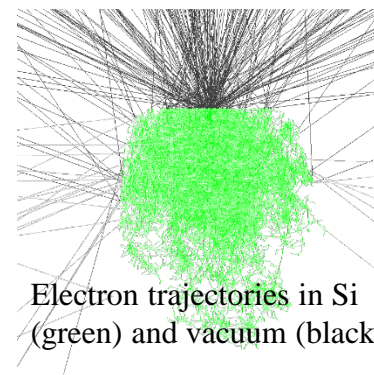Transistor
The gate width determines channel length

- If the technologically relevant property is a size, then we have to deal with probing errors.
- The magnitude of raw uncorrected probing errors is determined by the physics of the interaction. Typically:

Tip

For AFM, the tip size
$n \times 10$ nm

Light intensity
after diffraction
from an edge

Diffraction limited optics
λ: ~0.5 μm

Electron trajectories in Si
(green) and vacuum (black)

~800 eV SEM, electron scattering
Volume near the surface: ~few nm

These are all very large in magnitude compared to nanometrology's desired sub 1 nm accuracy.
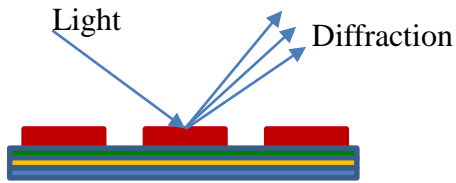We must correct probing error. To correct it, we must understand it.

NIST

Model-based metrology is…curve fitting

- When we can't measure a quantity, τ, directly, we can measure a signal that depends on the quantity
  - $\text{signal}(x) = \text{Model}(x; \tau)$
- Then we do a least squares fit that finds the best τ for our signal.
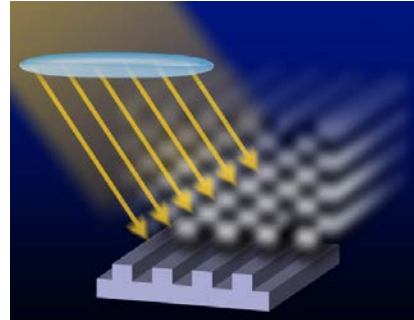

$$R(t) = R_0 e^{-t/\tau} + b$$
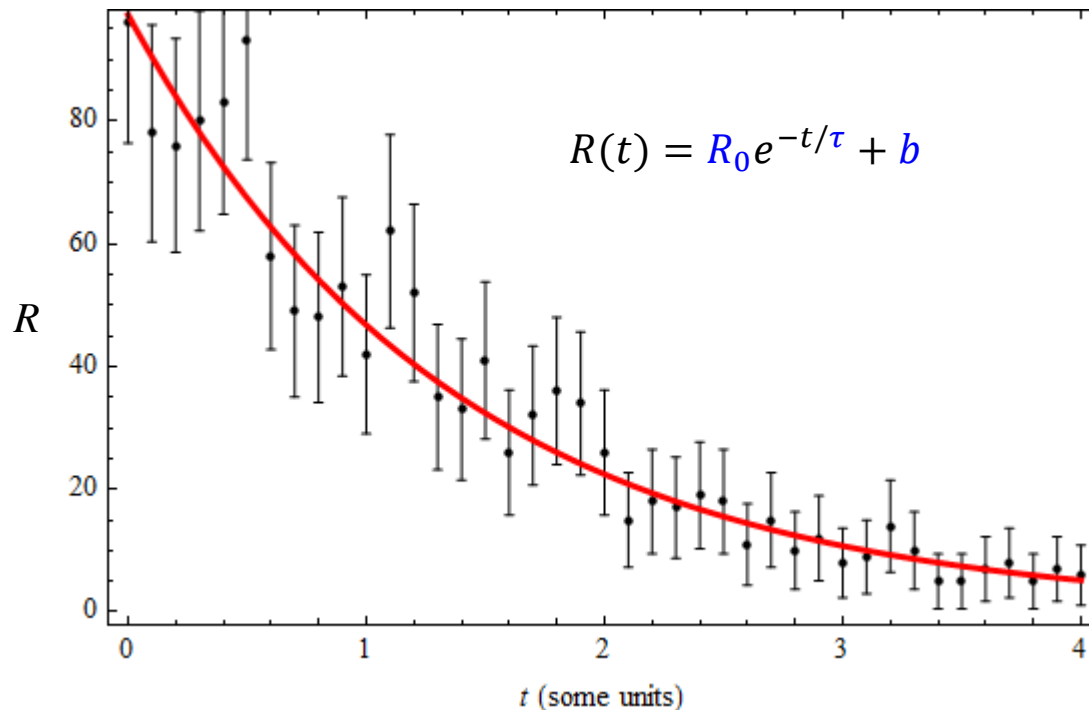
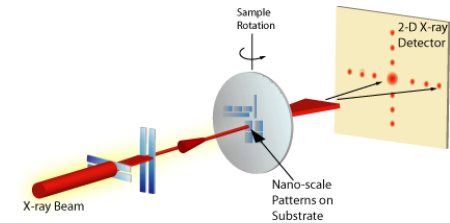# Examples of model-based dimensional nanometrology

Optical Scatterometry
Signal is scattered intensity vs angle or wavelength of incident light.

Optical Scatterfield Microscopy
Signal is set of intensity profiles of images, sampling the 3-D electromagnetic field above the finite period grating.

Critical Dimension Small Angle X-Ray Scattering (CD-SAXS)
X-rays transmit through a periodic grating. Signal is x-ray diffraction pattern.
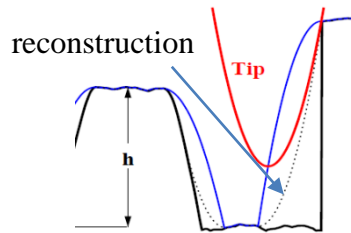
Atomic Force Microscopy (AFM)
with sample reconstruction using Mathematical Morphology
Signal is one or more AFM images.

Model-based Scanning Electron Microscopy
Signal is yield (secondary or backscattered) vs. position, energy, or angle.

Fitting schematic

Floating parameters (determined by the fit).

Fixed parameters (with associated uncertainties).

SEM parameters

SEM model output

$\{p_i\}$

signal($\boldsymbol{x}$) = Model($\boldsymbol{x}; \{p_i\}; \{a_i\}$)
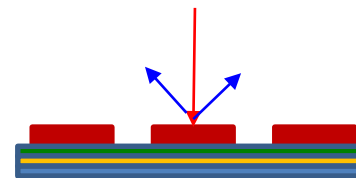
SEM Model
$\{a_i\}$

SEM data

χ^2 minimization
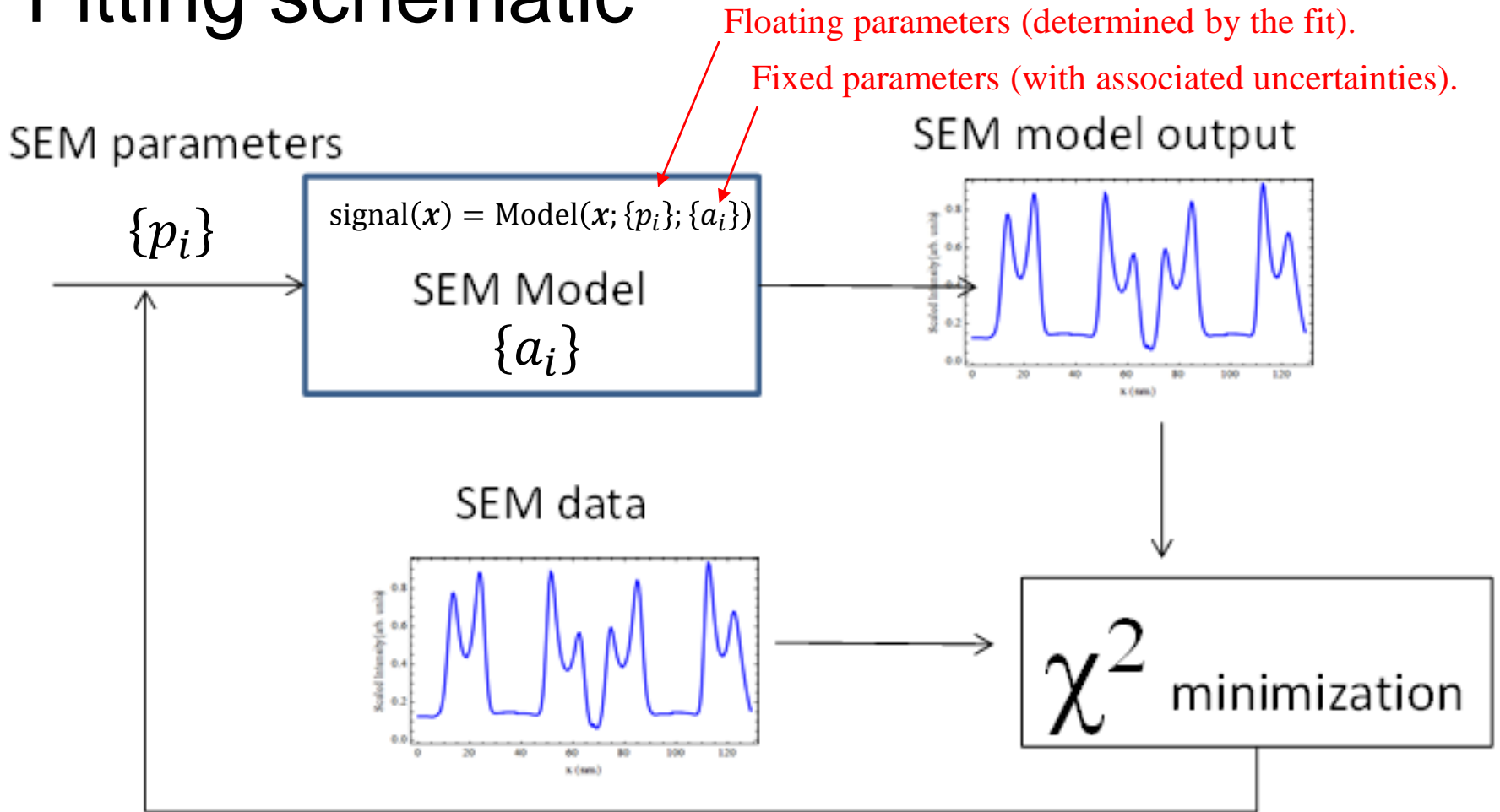
Generalized least squares (Aitken, 1934)

$$\chi^2 = [\text{Signal}(\boldsymbol{x}) - \text{Model}(\boldsymbol{x}; p_i; a_i)]^T V^{-1} [\text{Signal}(\boldsymbol{x}) - \text{Model}(\boldsymbol{x}; p_i; a_i]$$
$$= R^T V^{-1} R \quad (V \text{ the covariance matrix, } R \text{ the residual vector})$$

NIST

This is what it looks like for SEM metrology...
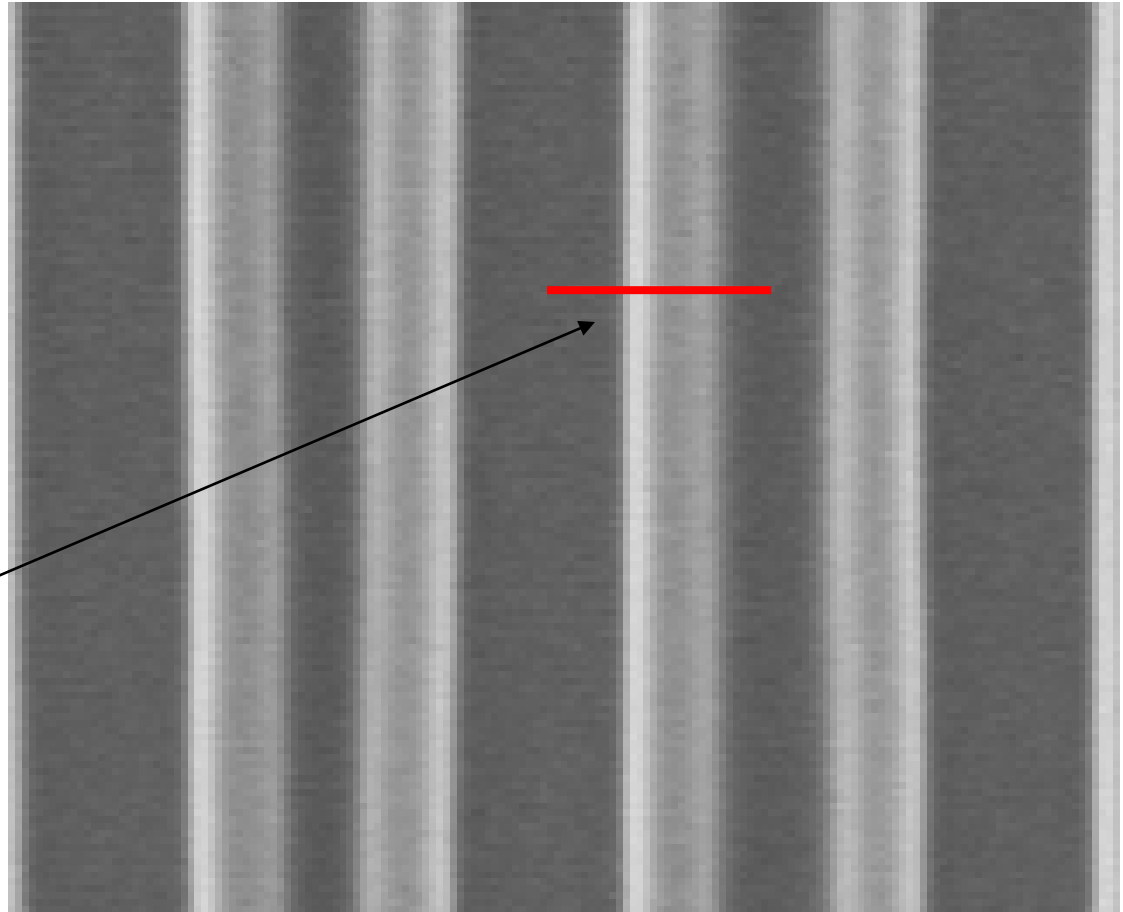
The signal is an image like this.

The next slide shows a graph of intensities along the red path.

NIST

This is what it looks like for SEM metrology...

Signal: Intensity (red data points & error bars)…

…is fit by our model, giving us the black best fit curve…

…which is associated (by the model) with this geometrical shape and position (green curve).

Then we can assign values (width, corner positions, angles, radii) based on the geometrical curve instead of the intensity curve.
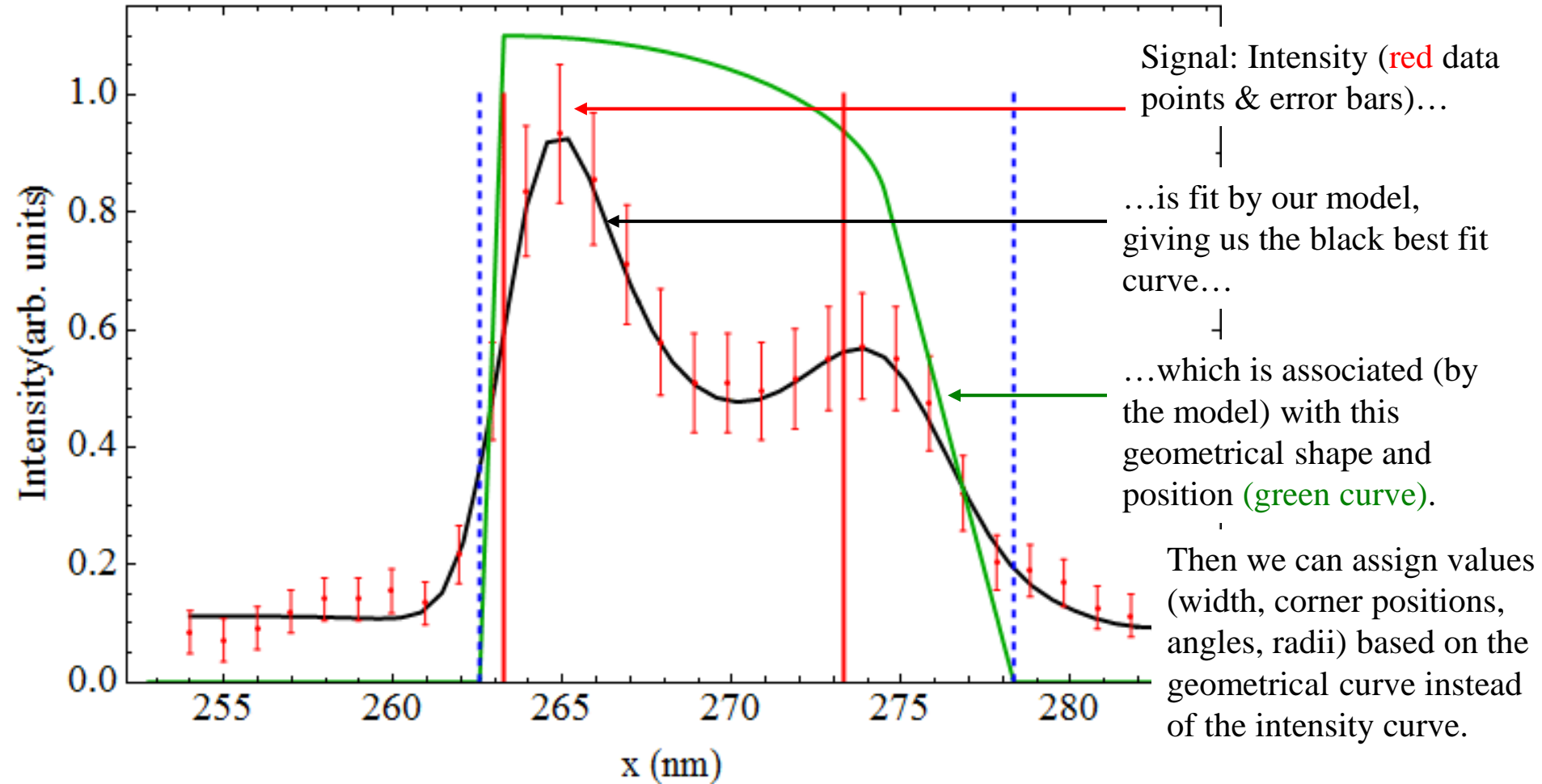
Prerequisites: (1) Signal. (2) Model.

Why do we think this model-based approach is an improvement?

- It gives us more information about our sample: shape in addition to size
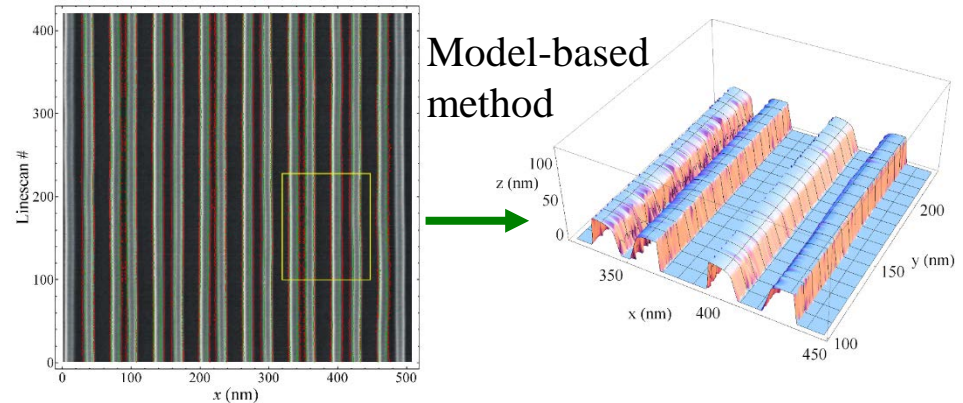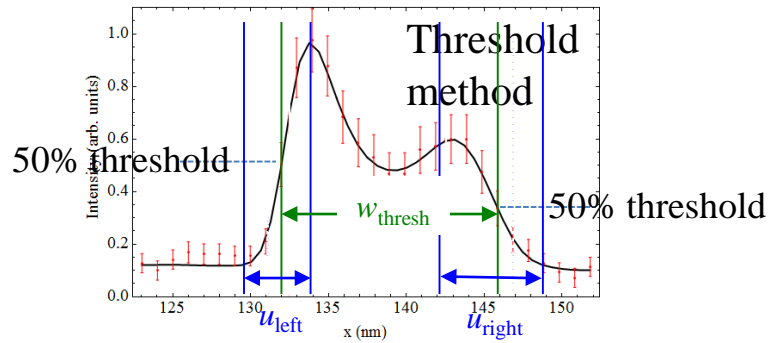
- It uses more of they physics we know; intuitively it should be more accurate.
- It is sensitive: small changes in shape from best fit produce significantly worse χ^2.
- In comparisons, agreement is much better than threshold method

Model-based (blue) vs. TEM:
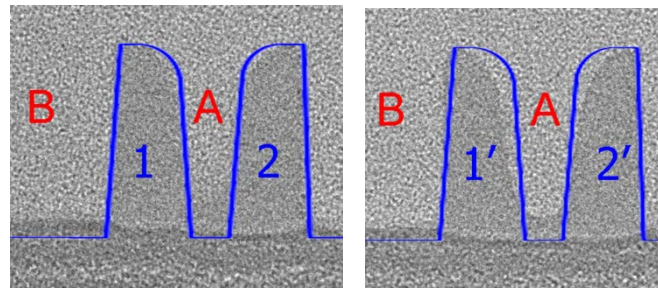
J.S. Villarrubia et al, Ultramicroscopy **154** (2015) p. 15.

Because it performs better and makes sense, we intuitively expect the uncertainties to be smaller than those of the threshold method.

MEASUREMENT ISSUE #2:
Limited information in the signal

NIST

There's limited information in the signal
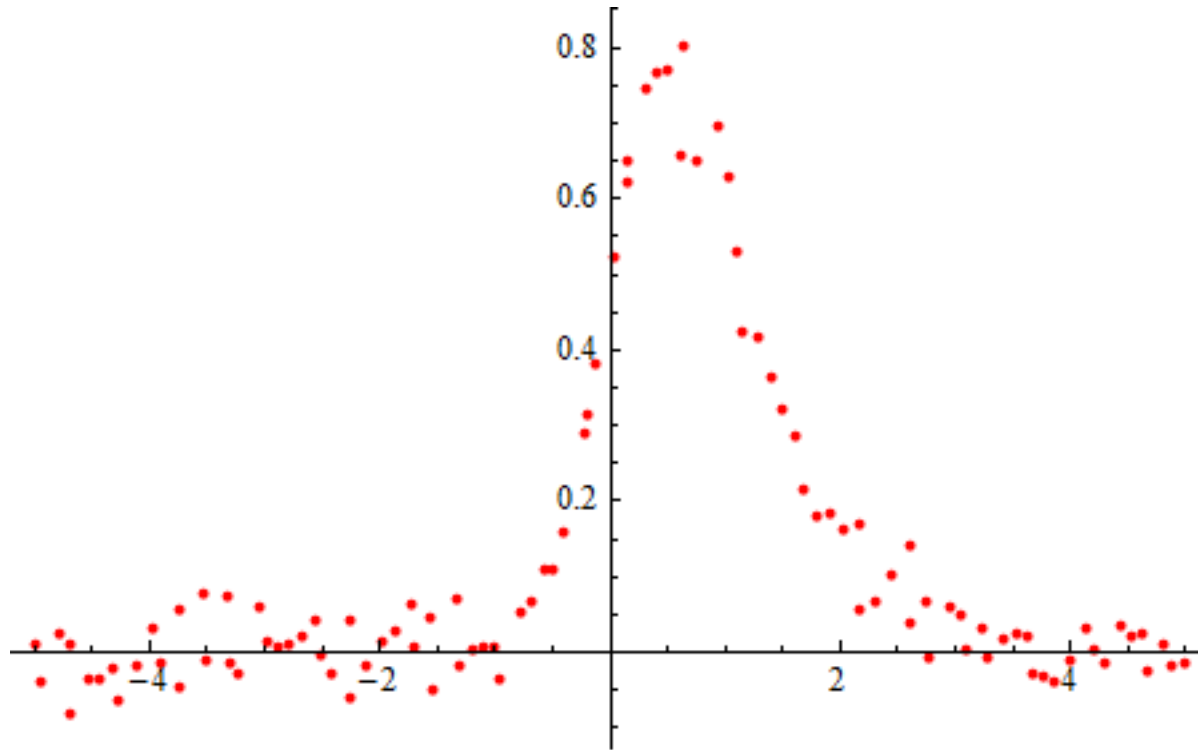
Suppose this is our signal. How many parameters can we justifiably extract from it?

There's limited information in the signal
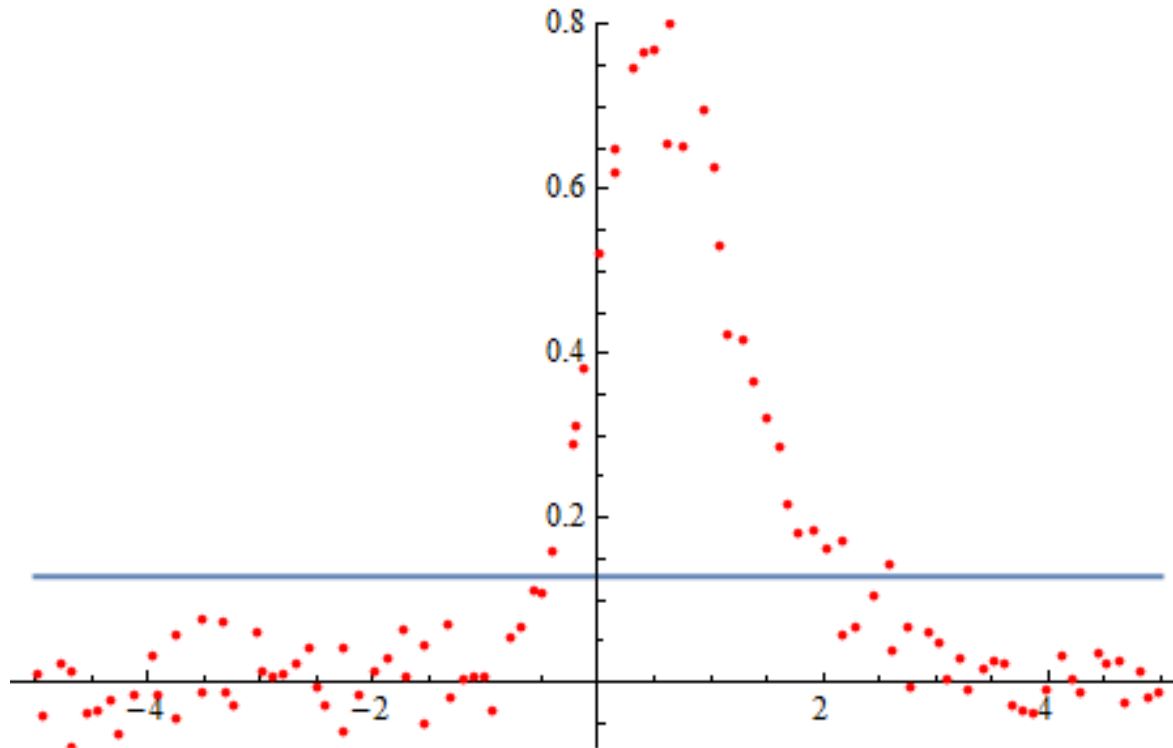
Fit with a single background parameter.

There's limited information in the signal
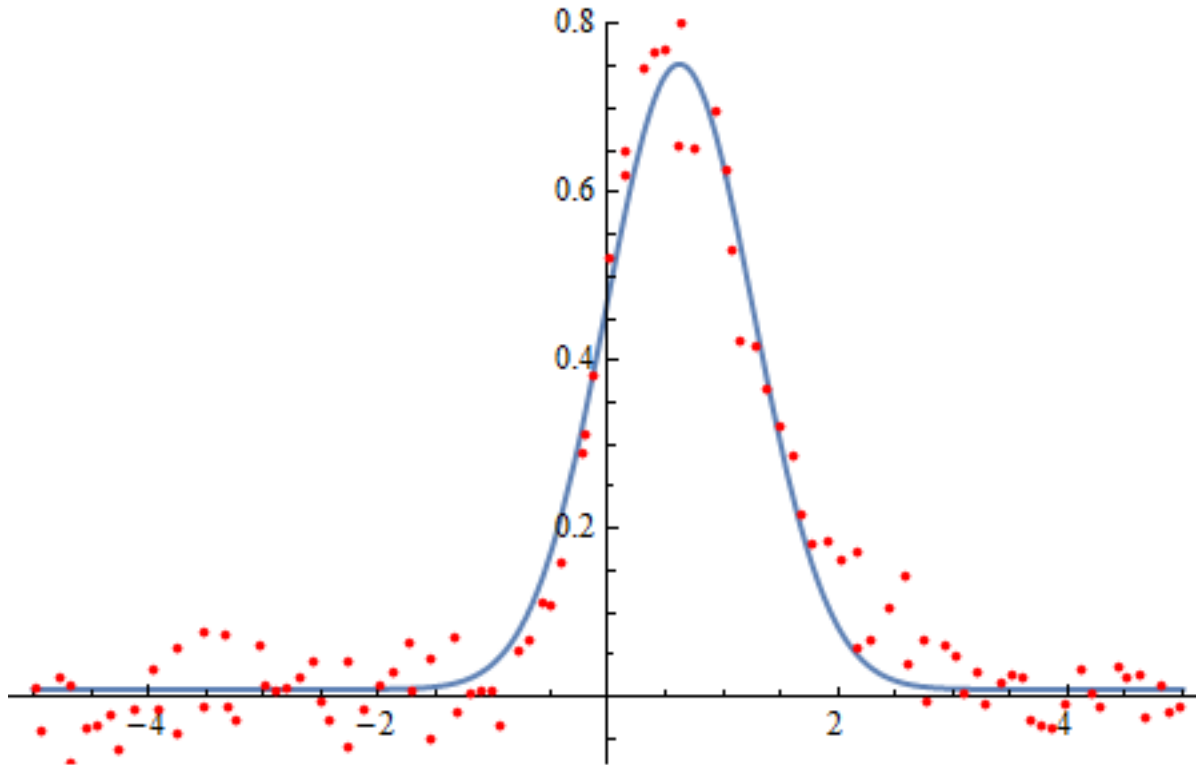
Fit with background + a symmetrical peak (position, height, & width).

NIST

There's limited information in the signal
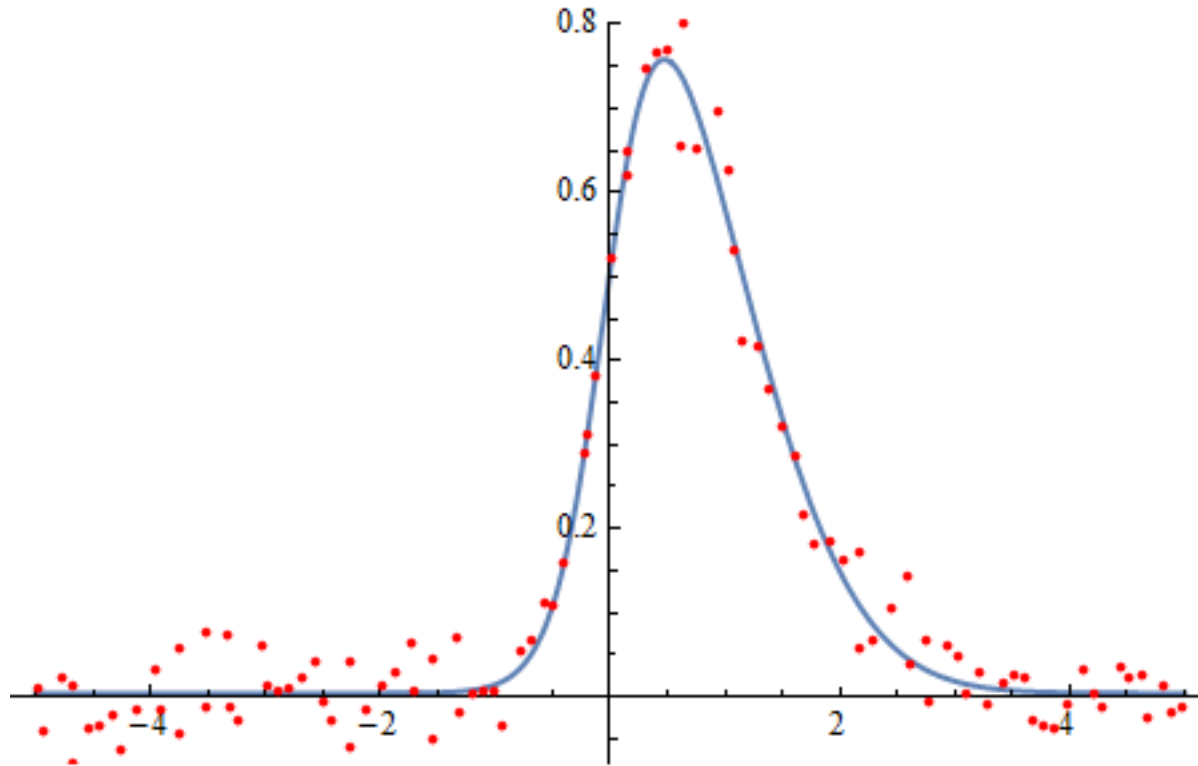
Same as before but with an additional asymmetry parameter.

That's it. Beyond those 5 parameters, there's no more information in this signal.

But we often try anyway. Why?

- Unlike my constructed example, in real life:
  - We don't know how many parameters nature will really permit.
  - We may have the right number of parameters but the wrong functional form, therefore less than perfect fit, and so think that more parameters are justified to address the imperfections.
- We have limited information, but "unlimited" measurement needs.

What happens when we try to extract too much information?

- We may just start fitting noise: we get meaningless parameter values.

- Very commonly, we get increasingly large correlations between parameters.

- Simple example of correlation: An increase in beam size and an increase in sample feature size have a similar effect on the observed signal. The 2 parameters are correlated.

- More subtly, when there are many parameters, it is common for some combination of them to be able to mimic some combination of others.

χ^2 contour plot near best fit

If our estimate for x0 changes

…the best-fit value of σ also changes.

An example of an ordinary level of correlation.

σ (nm)

x0 (nm)

NIST
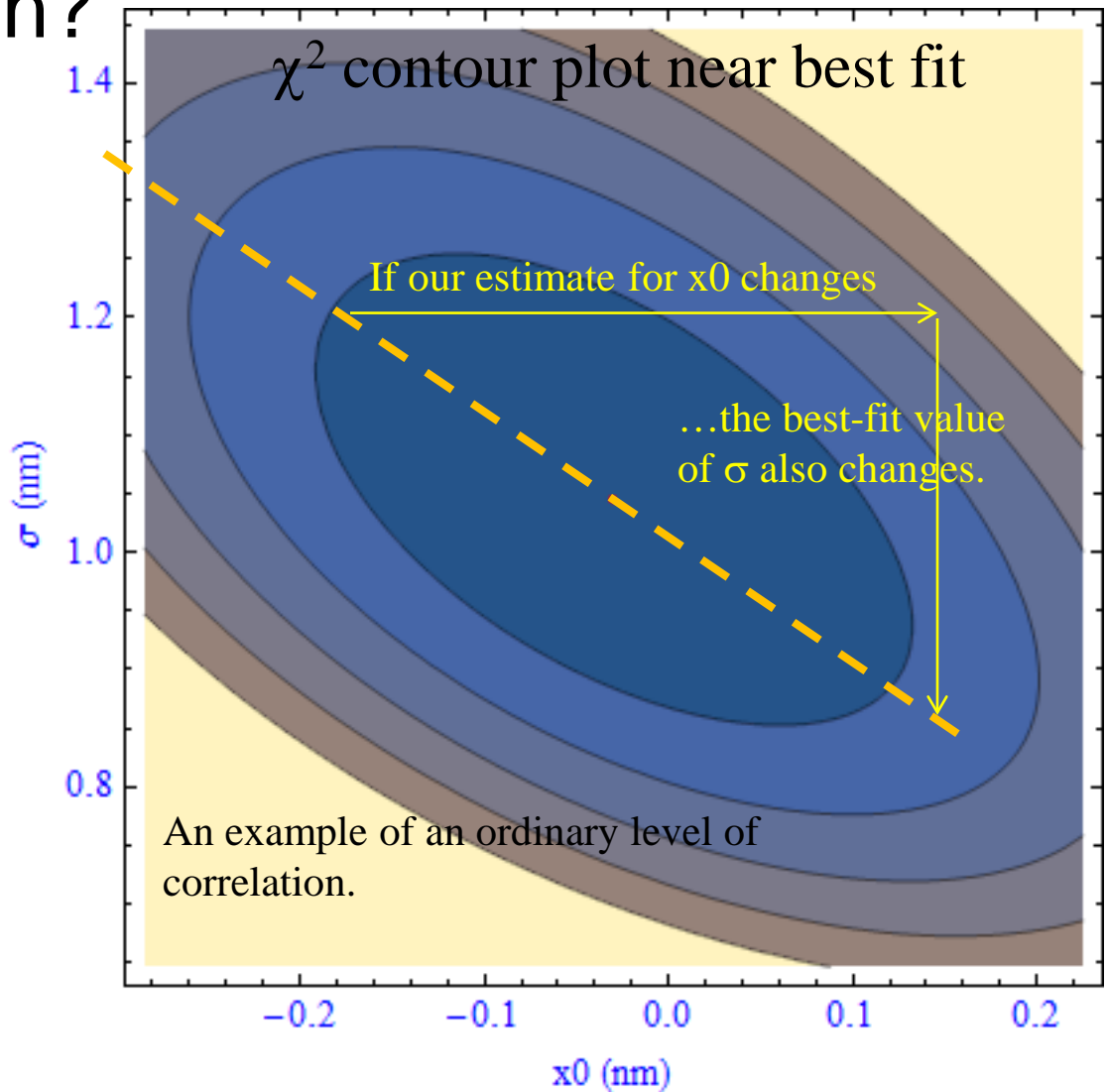
When there are correlations, the uncertainty interval looks like this:

Information from an independent source helps a lot.

u_1

Technique A

Technique B

u_2
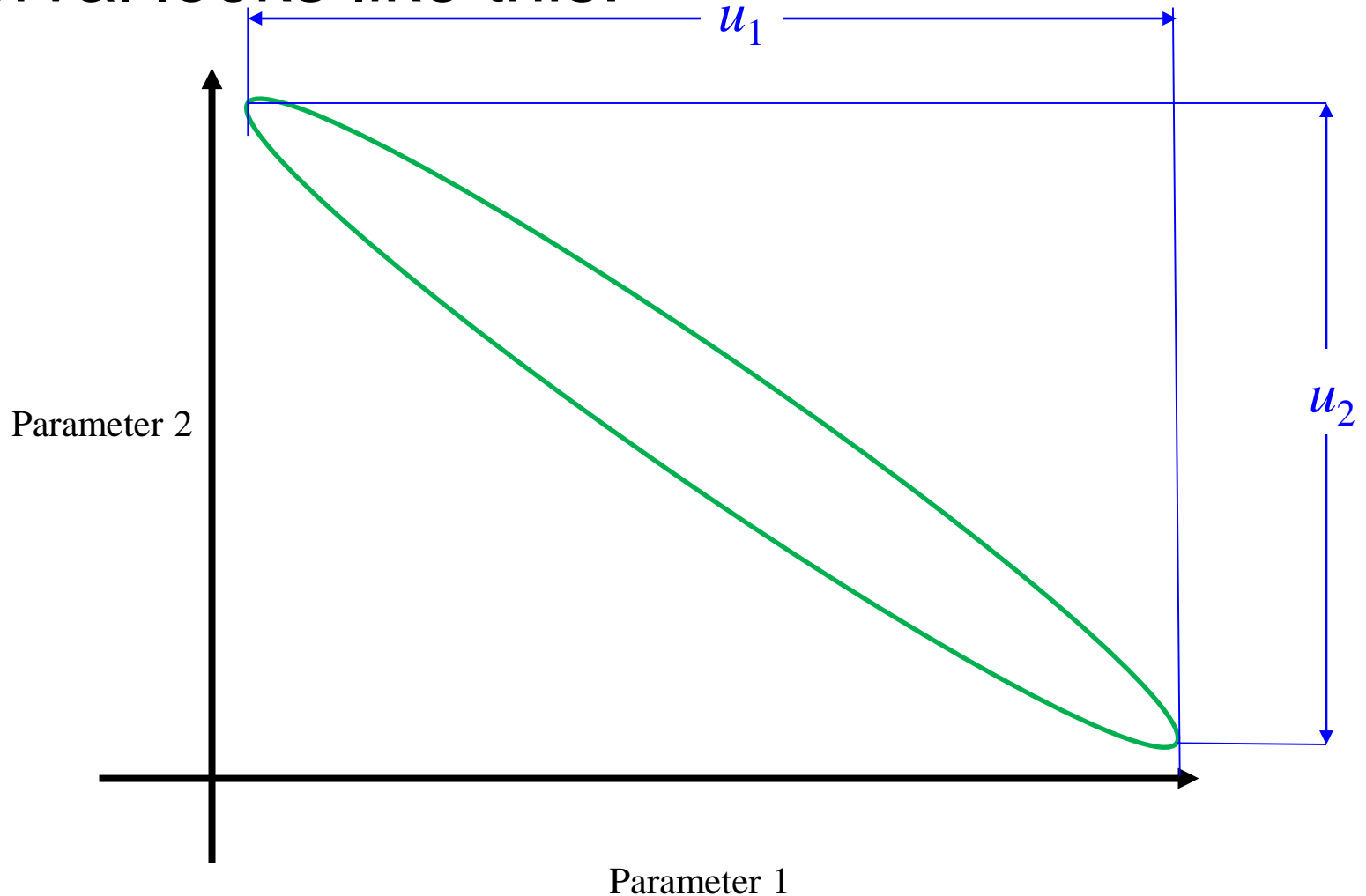
Parameter 2
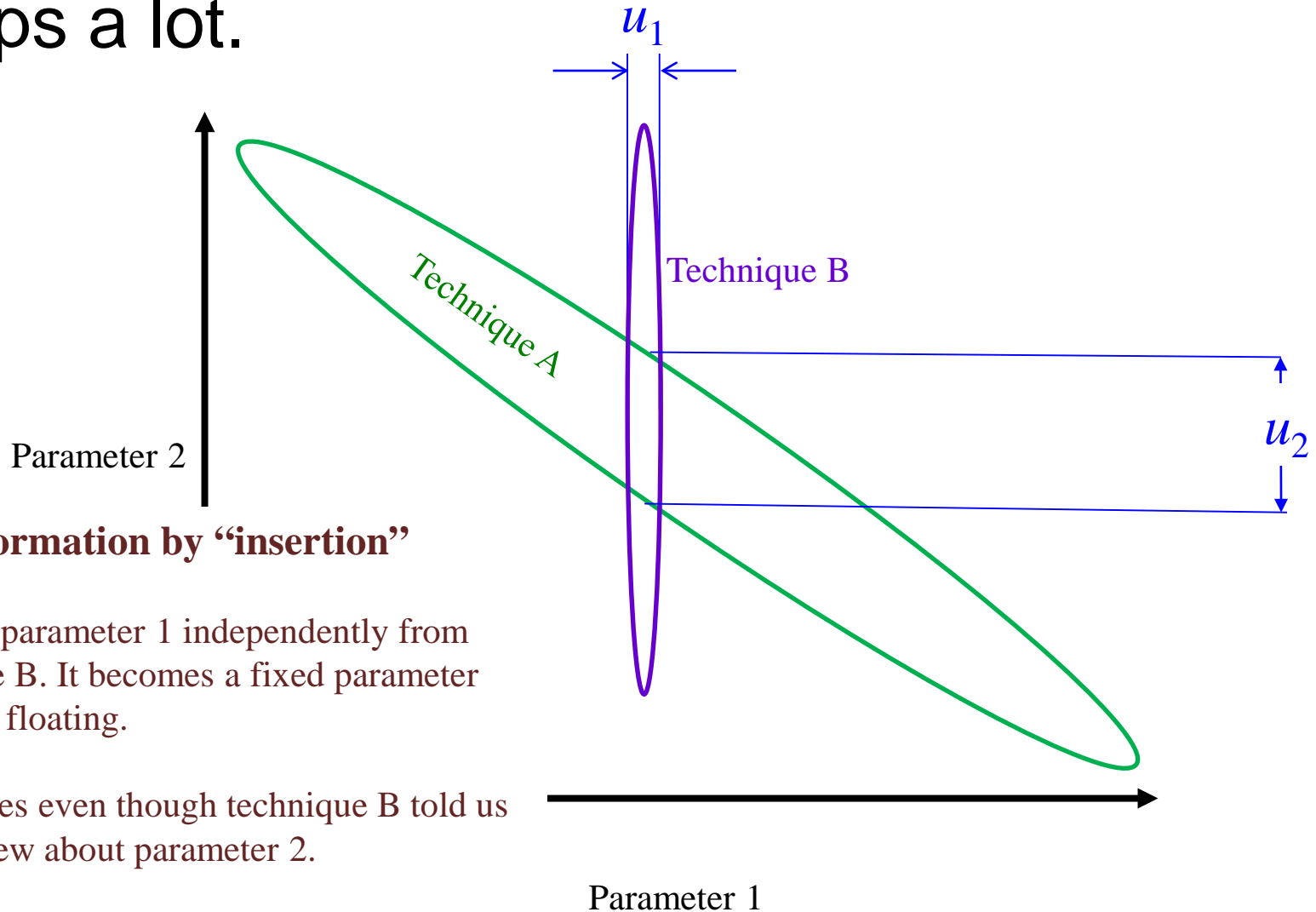
Parameter 1

New information by "insertion"

We know parameter 1 independently from Technique B. It becomes a fixed parameter instead of floating.

u_2 improves even though technique B told us nothing new about parameter 2.

But it's increasingly difficult to find such a "Technique B."

NIST

Information from an independent source helps a lot.

u_1

Technique A

Technique B

u_2
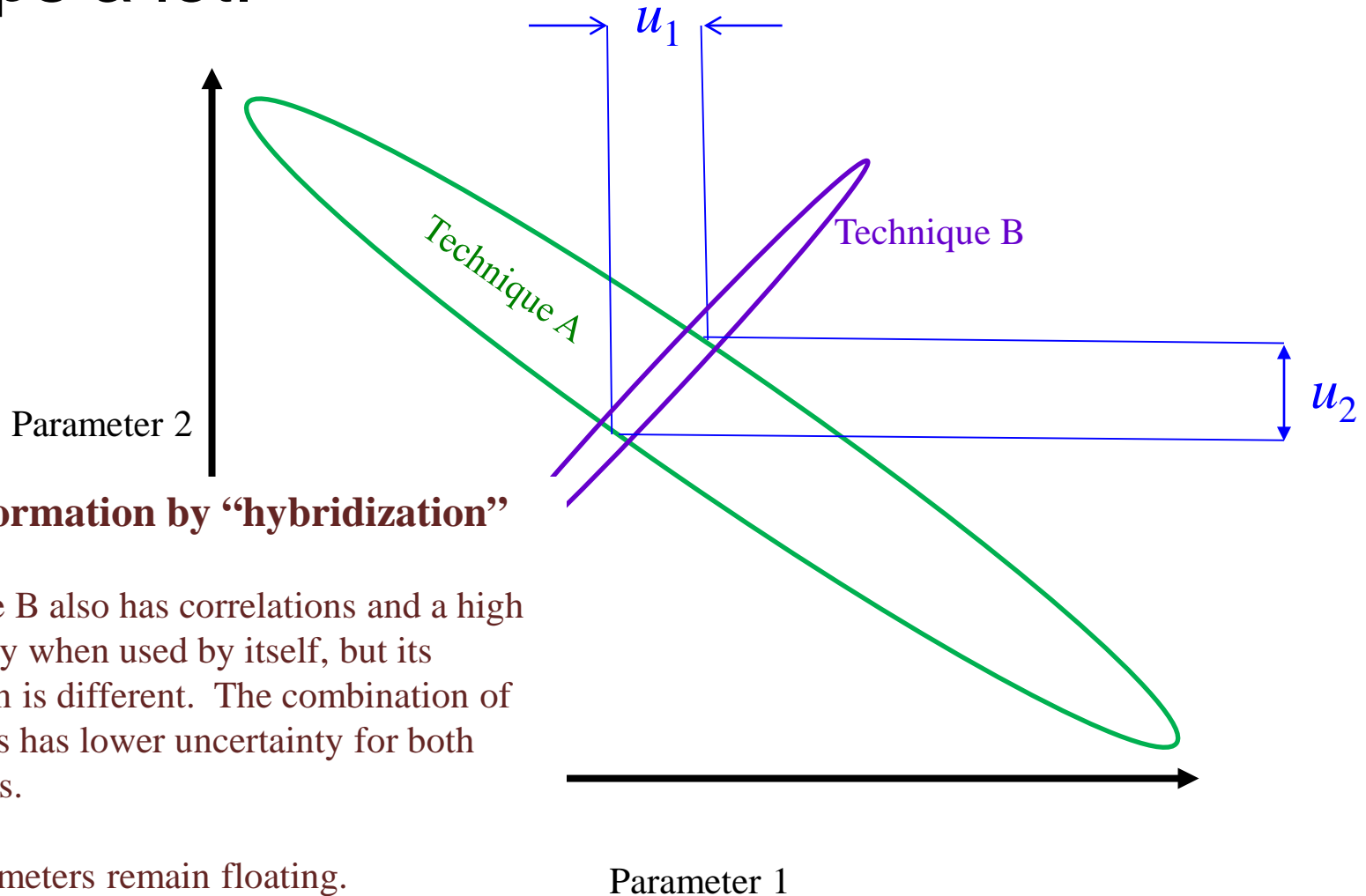
Parameter 2

New information by "hybridization"

Technique B also has correlations and a high uncertainty when used by itself, but its correlation is different. The combination of techniques has lower uncertainty for both parameters.

Both parameters remain floating.

Parameter 1

We sometimes call this "Parallel Regression."

NIST

Hybrid Metrology by Parallel Regression fitting schematic

Hybrid Model

Hybrid Output

h_i → **SEM Translator** → S_i → **SEM Model** →

h_i → **Optical Translator** → O_i → **Optical Model** →

h_i → **CDSAXS Translator** → X_i → **CDSAXS Model** →

h_i

$$\chi^2 = R^T V^{-1} R$$

$$\chi^2 = \begin{bmatrix} R_A^T & R_B^T \end{bmatrix} \begin{bmatrix} V_A & \mathbf{0} \\ \mathbf{0} & V_B \end{bmatrix}^{-1} \begin{bmatrix} R_A \\ R_B \end{bmatrix}$$

Hybrid data

→ **Difference measure: χ^2**

Minimization/ Parameter Adjustment

NIST
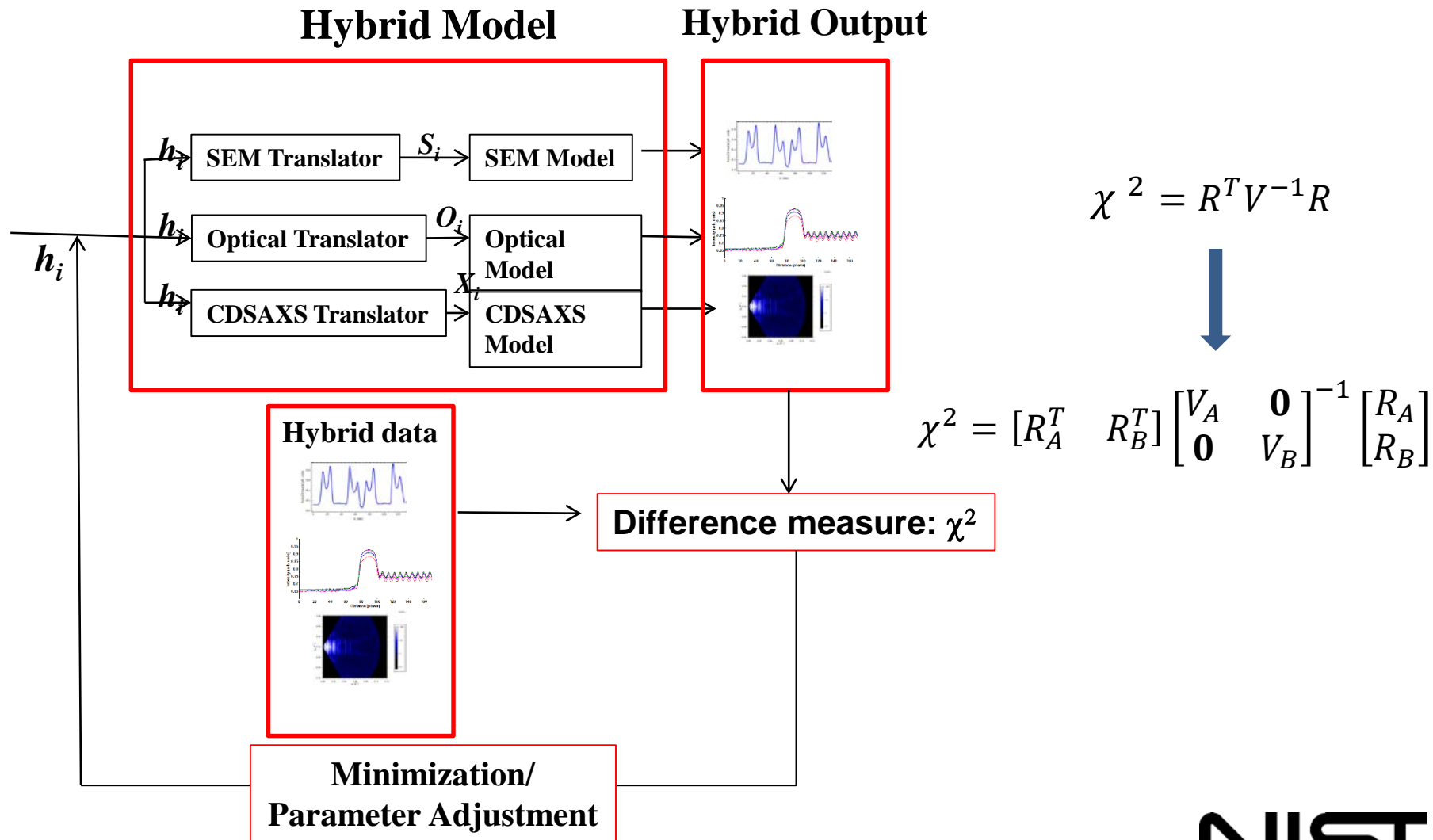
UNCERTAINTY QUANTIFICATION ISSUE #1: MODELING ERROR

Modeling error
1. Is likely our biggest error, but
2. There's no good way to estimate its associated uncertainty.

NIST

What do I mean by model uncertainty?

- The theory that relates signal to parameters could be wrong.

The parameter of interest

$$\text{signal}(x) = \text{model}(x; \{w, p_i\}; \{a_i\})$$

Suppose the functional (or algorithmic) form of this model is uncertain.

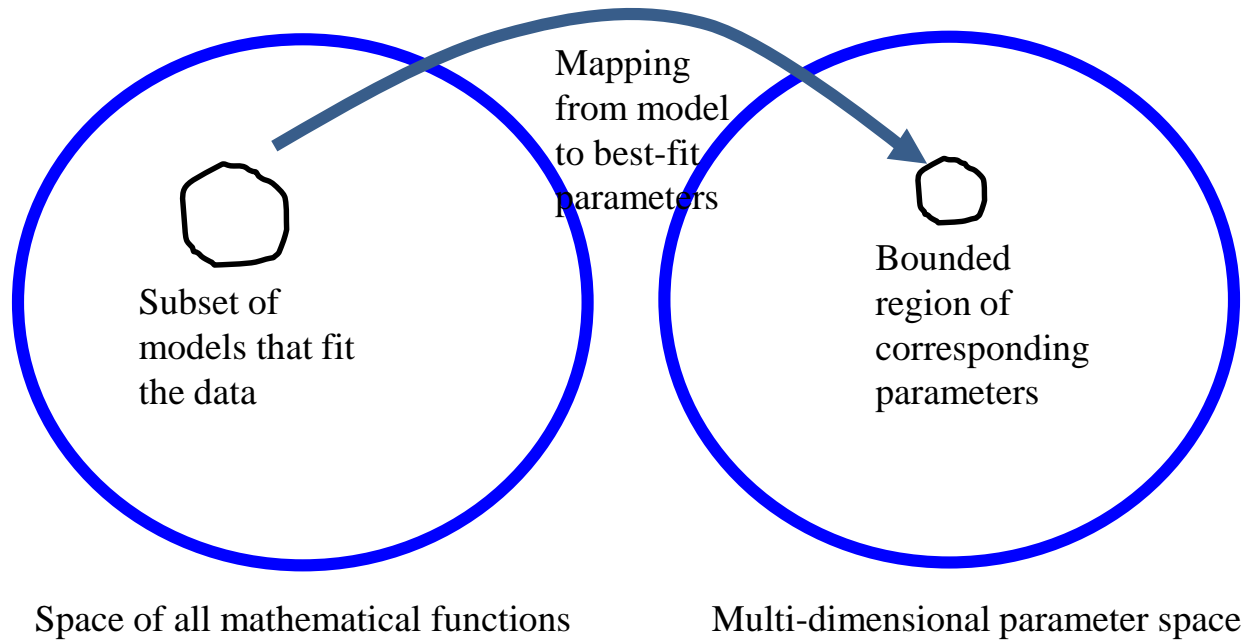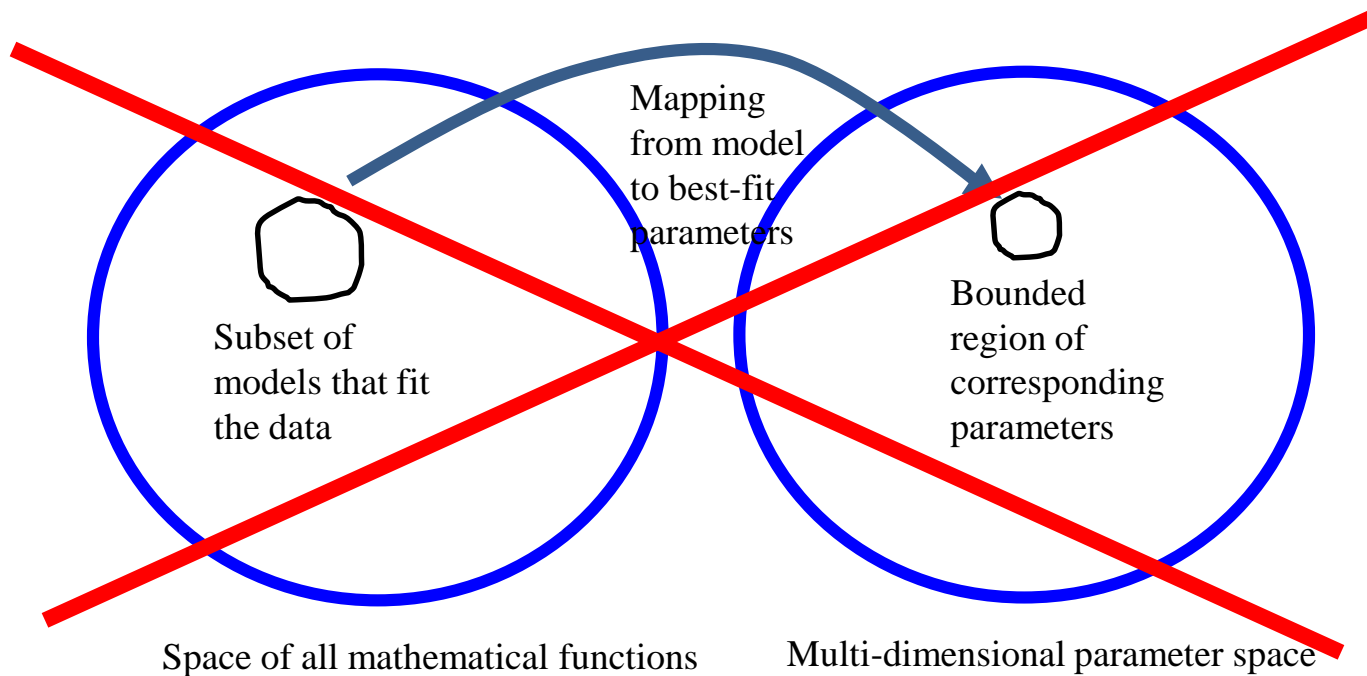- Examples from SEM modeling:
  - It is possible that our model omits some relevant but unknown scattering phenomena.
  - Generally available tables for elastic (electron-nucleus) scattering do not extend below 50 eV, where almost all of the secondary electron (SE) cascade happens. We must extrapolate.
  - At energies relevant for SE, quantum mechanical exchange is likely important, but there is no agreement on how to deal with it.
  - The best inelastic (SE generation) scattering model is based on dielectric function theory (DFT), but it requires ELF data that are available only for some materials. For other materials, we must use phenomenological models that we expect to have larger errors.
  - Even within DFT, the dielectric function of a real material is generally written as an expansion in terms of Lindhard free-electron dielectric functions. But Lindhard's is not the only theory for this. Suppose we used Mermin dielectric functions?
  - The available ELF data are usually optical data mainly sensitive to bulk losses. ELF should differ near the surface (e.g., due to surface plasmons)—and electrons will be more sensitive to these—but we don't presently account for this.

NIST

We might think: this is not a problem. If the model is wrong, we'll know because it will not fit the data…

Mapping from model to best-fit parameters

Subset of models that fit the data

Bounded region of corresponding parameters

Space of all mathematical functions

Multi-dimensional parameter space

NIST

…but it's not so

Mapping from model to best-fit parameters

Subset of models that fit the data

Bounded region of corresponding parameters
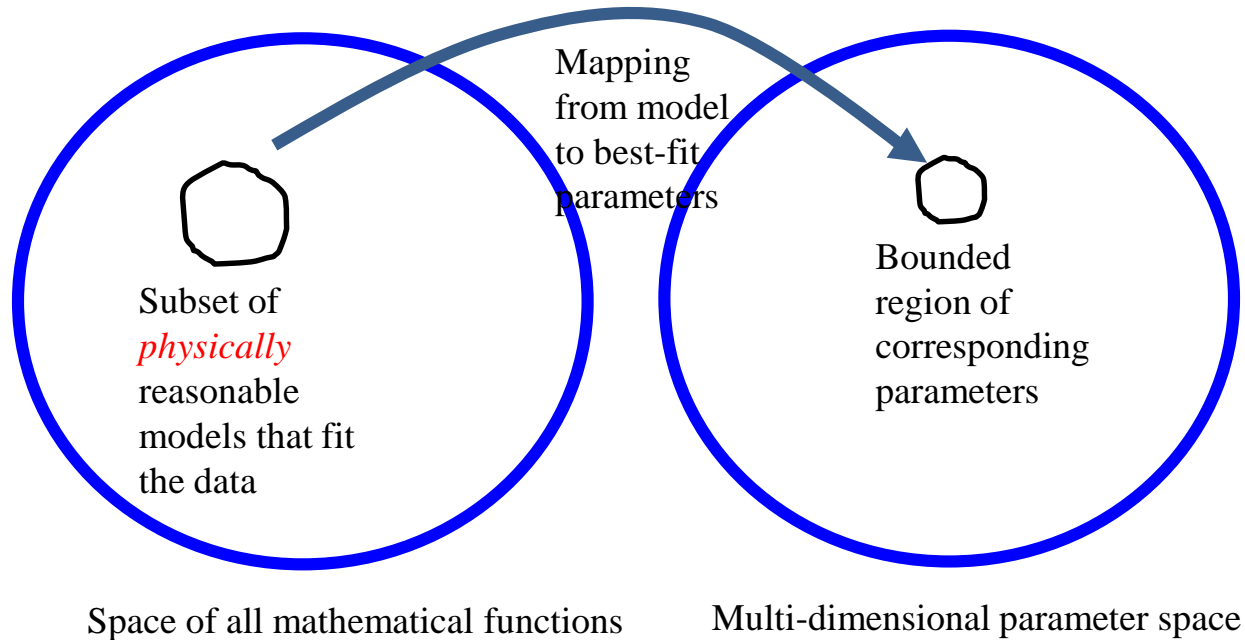
Space of all mathematical functions

Multi-dimensional parameter space

Mathematics is too rich. It is trivial to come up with alternative functions that fit the data but with different values of the parameters. The difference between parameters is, in fact, unbounded (by mathematics).

NIST

We still hope for this:

Mapping from model to best-fit parameters

Subset of *physically* reasonable models that fit the data

Bounded region of corresponding parameters

Space of all mathematical functions

Multi-dimensional parameter space

However, what *is* this subset of physically reasonable models?

An example:

Model 1 with the red shape predicts the red signal.

Model 2 with the black shape predicts the black signal.

Red and black *signals* agree quite well, but the shapes differ by ~2 nm in top and bottom edge position.

Villarrubia & Ding, J. Micro/Nanolith. MEMS MOEMS **8** (2009) 033003
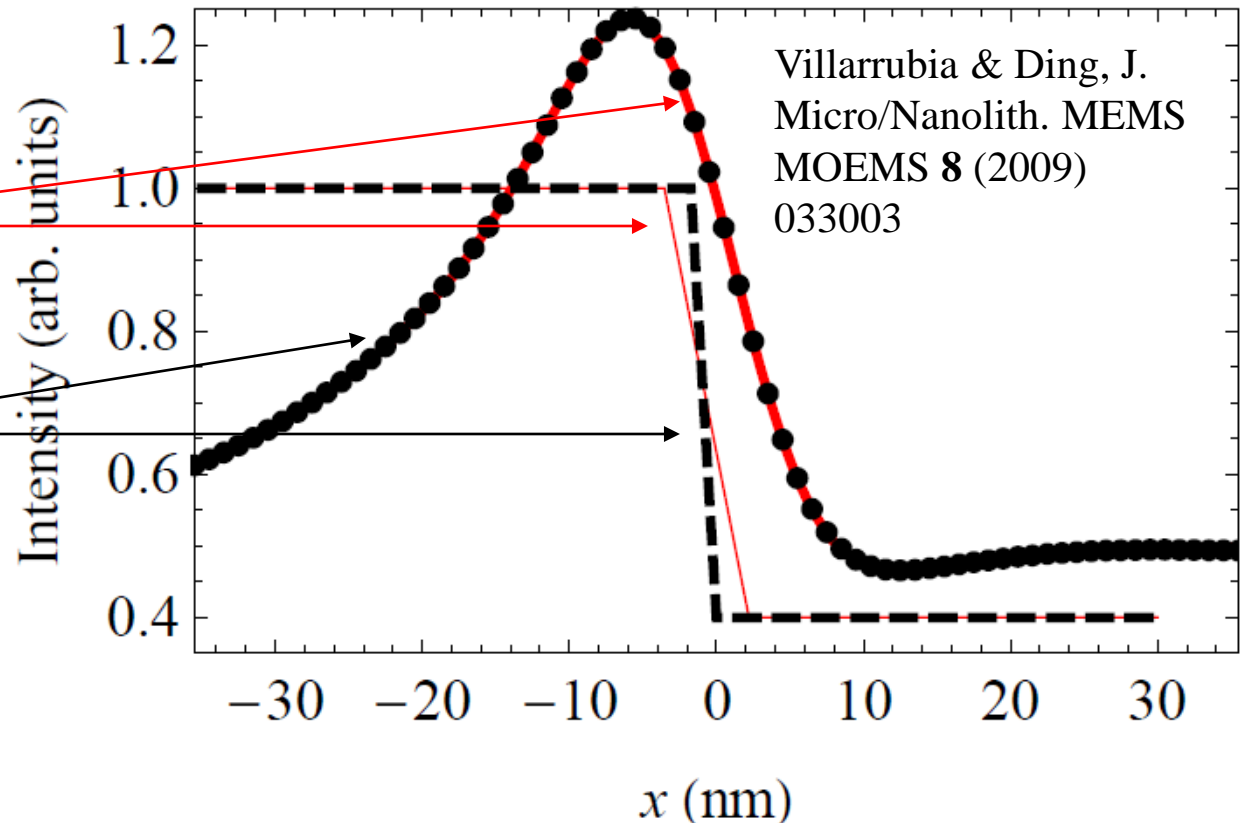
To determine the right parameter values, it's not enough that the model should fit.

# UNCERTAINTY QUANTIFICATION ISSUE #2: "Methods divergence" in hybrid metrology

This is a reappearance of issue #1 (modeling error) in a new context:

1. In hybrid metrology each method has its own model. If the models are erroneous (issue #1), each method is biased in its own way. (Methods diverge.)
2. This renders false a necessary assumption for combining measurements: that they're measuring the same thing.

NIST

Symptoms of methods divergence
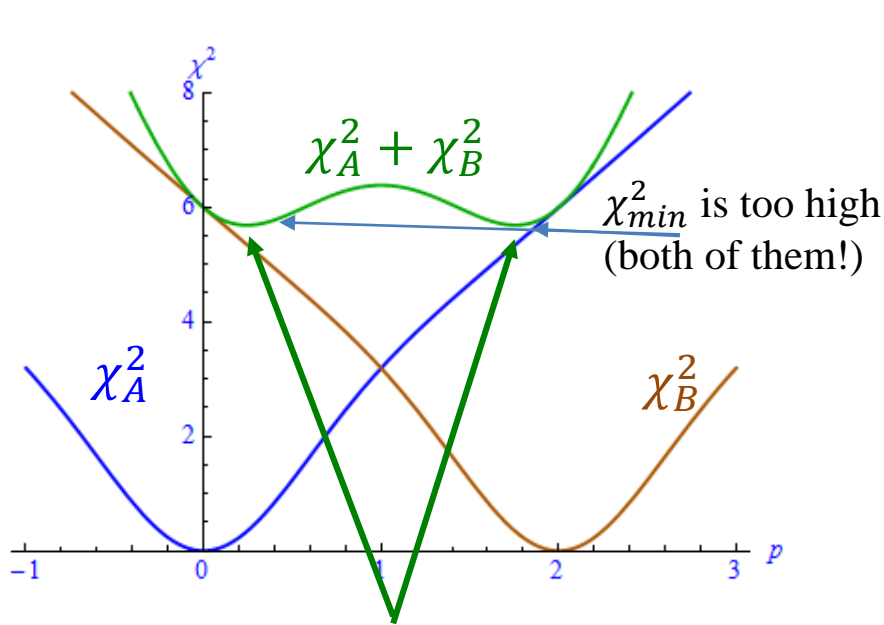
In extreme cases, we get a double minimum in the combined χ^2.
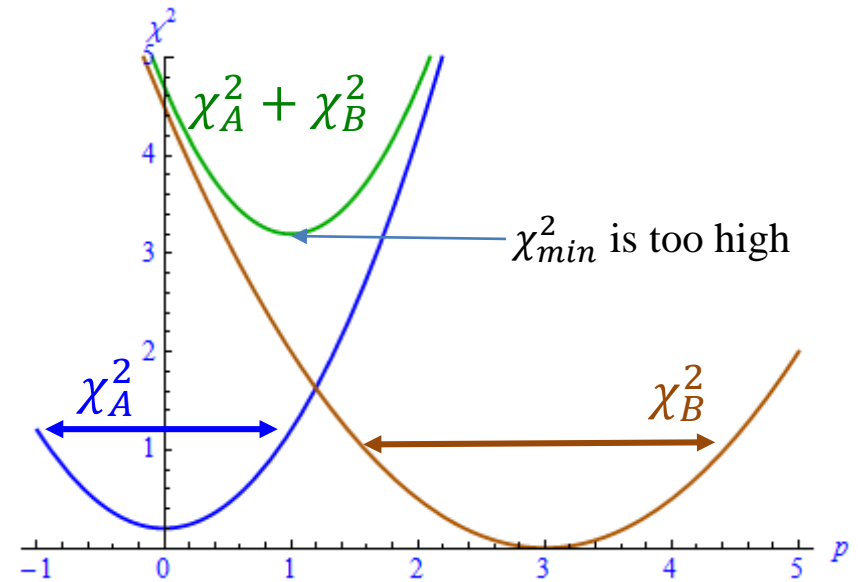
Even when there is a single minimum, the uncertainty intervals of the component minima do not overlap.

In both cases, $\chi^2_{min} \gg$ # degrees of freedom. In such a case we know the combined model and at least 1 individual model are wrong.

This isn't just a theoretical problem

- It's not uncommon in particle size measurement to observe 10% - 20% differences between sizes measured by dynamic light scattering vs. SEM or differential mobility analysis.

- In recent hybrid measurements, differences from SEM look like this:
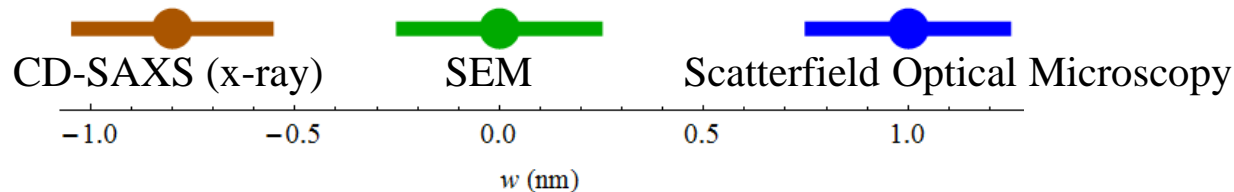
- This clustering within ~1 nm represents about a factor of 5 or more improvement over previous non-model-based methods.
- But the fact that uncertainties don't overlap means we are not entitled to the usual: "Assuming the model is correct, these are our uncertainties…"
- We know the model is *in*correct. Intuitively, we'd like to think that a *slightly* incorrect model generates a *slightly* incorrect measurement, but
  - It does not seem to be strictly mathematically true (e.g., if model incorrectness is judged by failure to fit)
  - If it is scientifically true (i.e., within the subset of scientifically plausible models), we don't know what that relationship is.

Summary/Conclusions

- In nanometer-scale metrology we face very tight ("unlimited") measurement needs. In this environment we benefit by
  - Understanding the measurement process in detail (model-based metrology) to make maximum use of information in the signal.
  - Enriching the signal by combining complementary measurements (hybrid metrology).
- If the metrologist does his job, he makes random errors small.
  - In this case the remaining errors are dominated by modeling error,
  - the uncertainties are dominated by "unknown unknowns," "dark uncertainty,"…
  - …and methods divergence causes hybrid models to fail to fit by a statistically significant amount. (Statistics is exquisitely sensitive when noise approaches 0.)
- In the long run, modeling is the job of the scientist
  - Math does not bound the uncertainty. Possible errors in the parameters are unbounded within the set of *mathematically* possible models (i.e., the set of all functions that fit the data).
  - The set of possible models has to be further constrained by physical/scientific considerations.
  - Time to solution is order of a research-project-lifetime.
- In the meantime, manufacturers will not shut down their $Billion fabs' production lines.

What do we tell them about the uncertainty of the measurements?

NIST

The validation problem

- Validation is best when we have measured points in a neighborhood of the parameters of interest.
- In our case (parameters describe a shape) the parameter space is multidimensional.
- Populating a multidimensional neighborhood requires many measurements.
- The measurand is the true shape. Good reference techniques are rare, and generally restricted to special samples.
  - E.g., samples thin enough for TEM
  - These may occupy a different part of the parameter space than our interesting industrial samples.
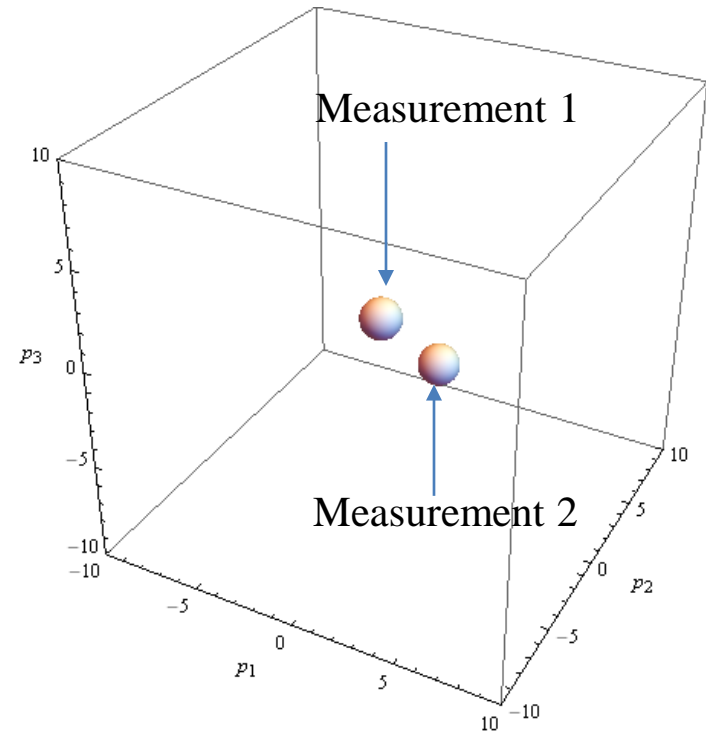
Probably in this circumstance our uncertainty is actually rather large. It is not the size of the uncertainty that is the issue.

It would be nice to have a plausible statistical model and associated formalism that tells me what the uncertainty is when validations are sparse (and which reduces to the expected result when they are dense).

Example alternative function

Space of all mathematical functions   Multi-dimensional parameter space
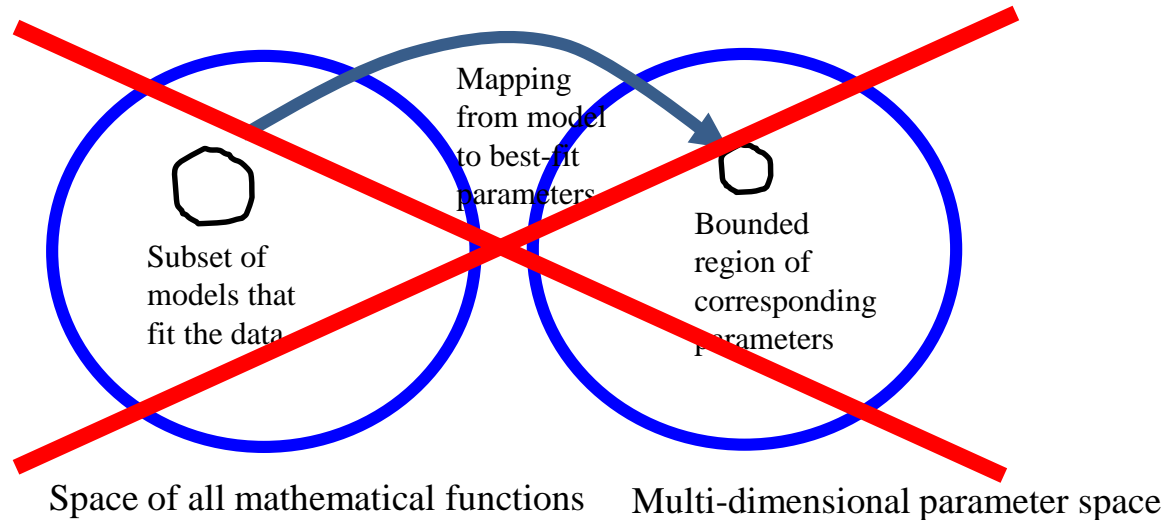
Suppose $f(x; p)$ fits the data. Find a different function, $g(x; q)$, that also fits but with $p \neq q$.

Let $g(x; q) = f(x; h(q))$. As long as $h(q) = p$ then it's obvious that $g(x; q)$ fits the data because $g(x; q) = f(x; p)$. But h can be any function whatsoever. E.g., it could be any offset: $h(q) = p + p_0$, however large in magnitude, or any nonzero multiple, $m\,p$, however large or small.

Most of these would be physically implausible—but that's the point. The criterion for rejecting them is physical, not mathematical/statistical.

NIST