Systematics in CMB Measurements

2022-08-12

Kirit S. Karkare

What are systematic effects/errors?

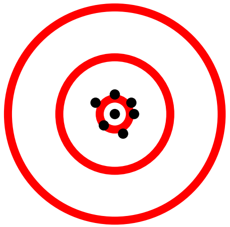
When reporting the results of a measurement, we generally include an error estimate - e.g. 1σ uncertainties.

That error can usually be broken down into a component that decreases when taking more data (*statistical/random*) and one that doesn't (*systematic*).

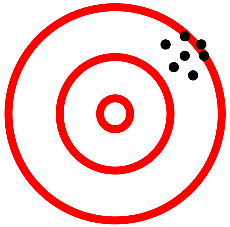
Ideally, the uncertainty in the measurement is dominated by the statistical term. If not, the experiment should be redesigned.

Corollary: If the measurement is very noisy, systematics don't matter much!

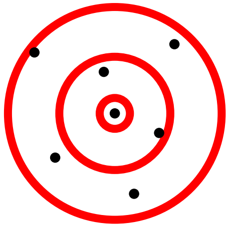
Classic Example: Dartboard



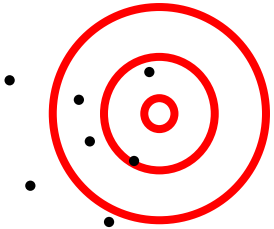
Statistical: small
Systematic: small



Statistical: small
Systematic: large

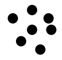


Statistical: large
Systematic: small

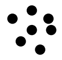


Statistical: large
Systematic: large

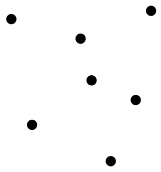
…but we usually don’t know the true value!



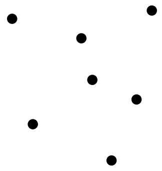
Statistical: small
Systematic: ?



Statistical: small
Systematic: ?



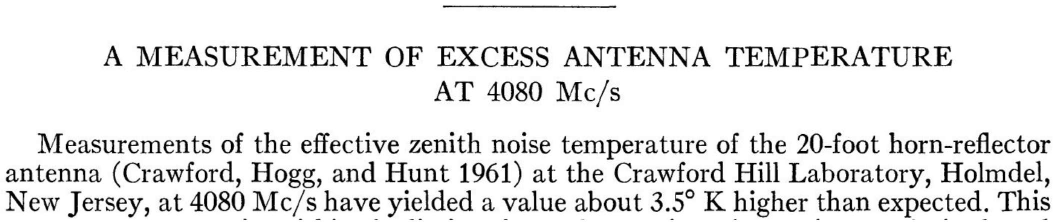
Statistical: large
Systematic: ?



Statistical: large
Systematic: ?

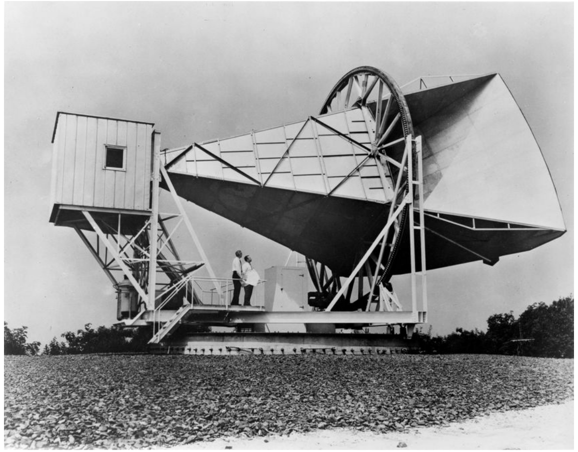
CMB Discovery: an exercise in systematics control

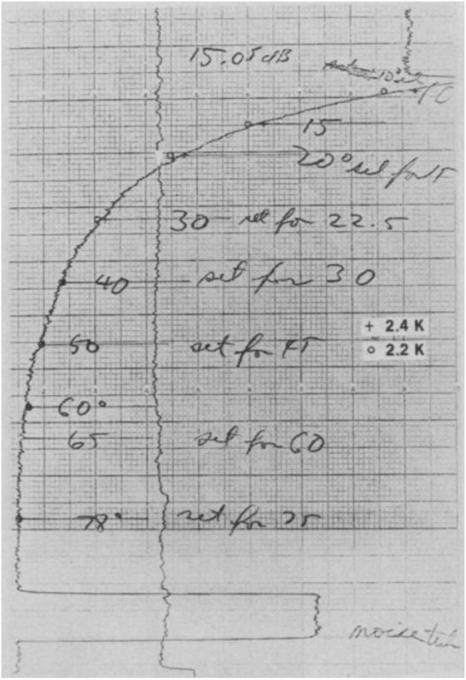
In 1965, using the 20 ft horn-reflector antenna at Bell Labs, Penzias and Wilson published an estimate of the zenith noise temperature obtained through a “skydip” (measuring the antenna temperature as a function of elevation, fitting the atmosphere, and subtracting it off).



A MEASUREMENT OF EXCESS ANTENNA TEMPERATURE AT 4080 Mc/s

Measurements of the effective zenith noise temperature of the 20-foot horn-reflector antenna (Crawford, Hogg, and Hunt 1961) at the Crawford Hill Laboratory, Holmdel, New Jersey, at 4080 Mc/s have yielded a value about 3.5° K higher than expected. This





Wilson 1978

CMB Discovery: an exercise in systematics control

They expected the zenith temperature to be the sum of the atmospheric contribution (2.3 K) and the radiation from the walls of the antenna and the ground (1 K). The measurement yielded ~7.3 K!

This started a year of investigations, including very careful absolute calibrations, beam measurements, cleaning the antenna, etc. The excess systematic load remained.

TABLE II — SOURCES OF SYSTEM TEMPERATURE

Source	Temperature
Sky (at zenith)	2.30 ± 0.20°K
Horn antenna	2.00 ± 1.00°K
Waveguide (counter-clockwise channel)	7.00 ± 0.65°K
Maser assembly	7.00 ± 1.00°K
Converter	0.60 ± 0.15°K
Predicted total system temperature	18.90 ± 3.00°K

the temperature was found to vary a few degrees from day to day, but the lowest temperature was consistently 22.2 ± 2.2°K. By realistically assuming that all sources were then contributing their fair share (as is also tacitly assumed in Table II) it is possible to improve the over-all accuracy. The actual system temperature must be in the overlap region of the measured results and the total results of Table II, namely between 20 and 21.9°K. The most likely minimum system temperature was therefore

$$T_{system} = 21 \pm 1°K.*$$

Previous measurements from Ohm indicating ~2.1 K excess temperature (and potentially an inflated statistical error)!

He temperature	4.22	
Calculated contribution from cold load waveguide	0.38	
Attenuator setting for balance	2.73	
Total cold load	7.33	
Atmosphere		2.3 ± 0.3
Waveguide and antenna loss		1.8 ± 0.3
Back lobes		0.1 ± 0.1
Total antenna		4.2 ± 0.7
Background		3.1 ± 1

Wilson 1978

Updated measurements from P&W with a lower-noise receiver and more accurate uncertainty estimates.

CMB Discovery: an exercise in systematics control

They expected the zenith temperature to be the sum of the atmospheric contribution (2.3 K) and the radiation from the walls of the antenna and the ground (1 K). The measurement yielded ~7.5 K!

This started about a year of investigations, including very careful absolute calibrations, beam measurements, cleaning the antenna, etc. The excess systematic load remained.

Moral: make every effort to understand and accurately report all sources of uncertainty, both statistical and systematic. It might get you a Nobel Prize!

Predicted total system temperature	18.90 ± 3.00°K

the temperature was found to vary a few degrees from day to day, but the lowest temperature was consistently 22.2 ± 2.2°K. By realistically assuming that all sources were then contributing their fair share (as is also tacitly assumed in Table II) it is possible to improve the over-all accuracy. The actual system emperature must be in the overlap region of the measured results and the total results of Table II, namely between 20 and 21.9°K. The most likely minimum system temperature was therefore

$$T_{\text{system}} = 21 \pm 1°\text{K}.^*$$

Previous measurements from Ohm indicating ~2.1 K excess temperature (and potentially an inflated statistical error)!

Total cold load	7.33
Atmosphere	2.3 ± 0.3
Waveguide and antenna loss	1.8 ± 0.3
Back lobes	0.1 ± 0.1
Total antenna	4.2 ± 0.7
Background	3.1 ± 1

Updated measurements from P&W with a lower-noise receiver and more accurate uncertainty estimates.

Systematics in CMB measurements

Many systematic effects can impact the thing we're ultimately trying to measure (usually cosmological parameters derived from a set of power spectra). It's often convenient to separate them into categories based on where they originate:

Astrophysical:

- How accurate is your model that explains the maps/power spectra?
- E.g. if the sky consists of CMB+dust+synchrotron and you ignore synchrotron, biasing the result

Instrumental:

- Do you understand everything about your instrument that *generated* the maps/power spectra?
- E.g. if your telescope has a sidelobe that contributes unmodeled structure to the maps

When designing an experiment we need to take both into account!

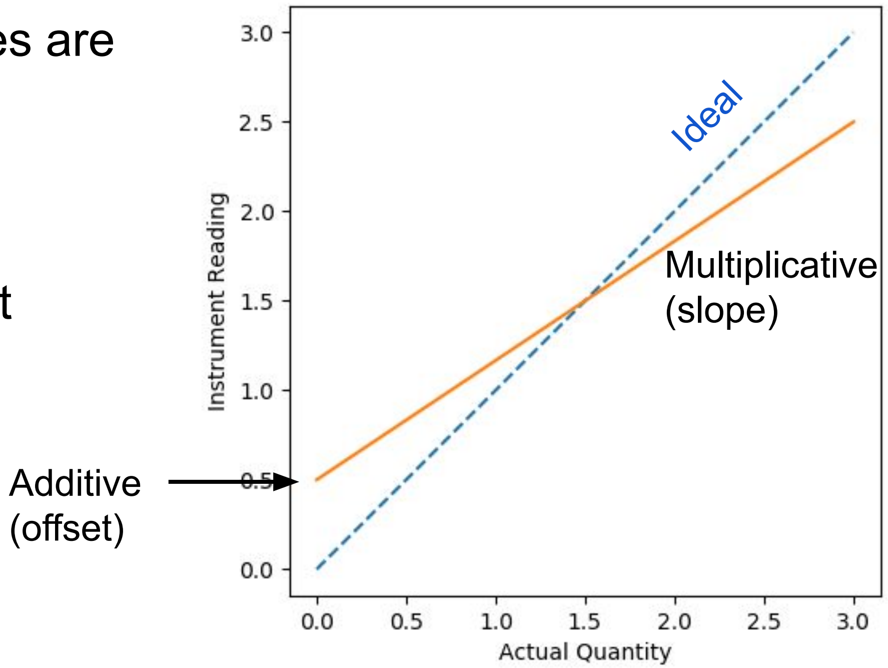
Common Instrumental Systematics

- Incomplete knowledge of beam profile (sidelobes, pol difference beams)
- Incomplete knowledge of passbands
- Polarization angles
- Crosstalk
- Electromagnetic interference
- Scan synchronous pickup: magnetic, electronic, optical
- Nonlinearities in detector response
- Detector time constants
- Cosmic ray hits
- Microphonics

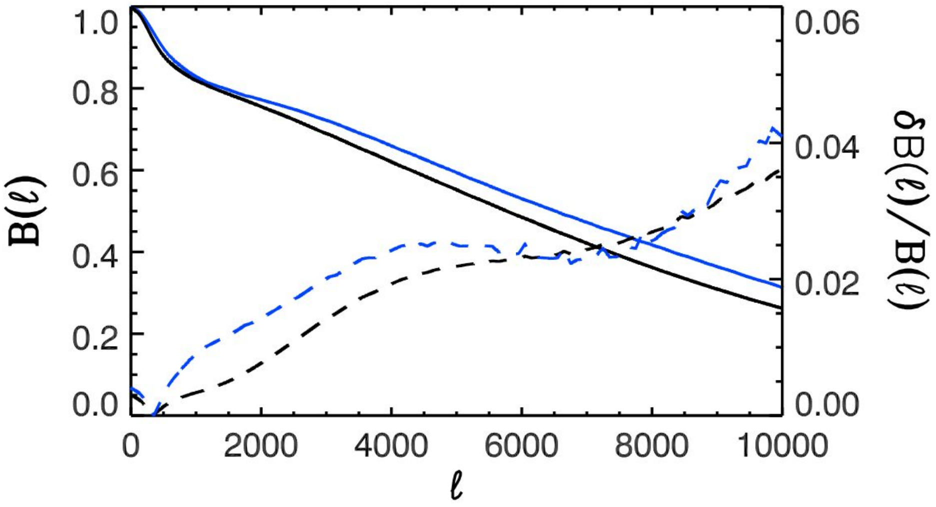
Classes of Systematics

Systematics can also be categorized by *how* they affect the data - two common classes are **additive** and **multiplicative**.

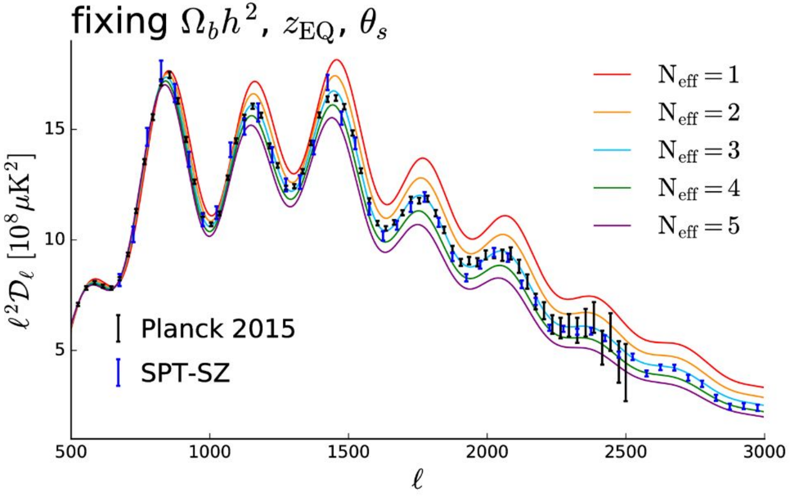
What you worry about is dictated by what you're trying to measure.



Example 1: Main Beam Profile at high ell

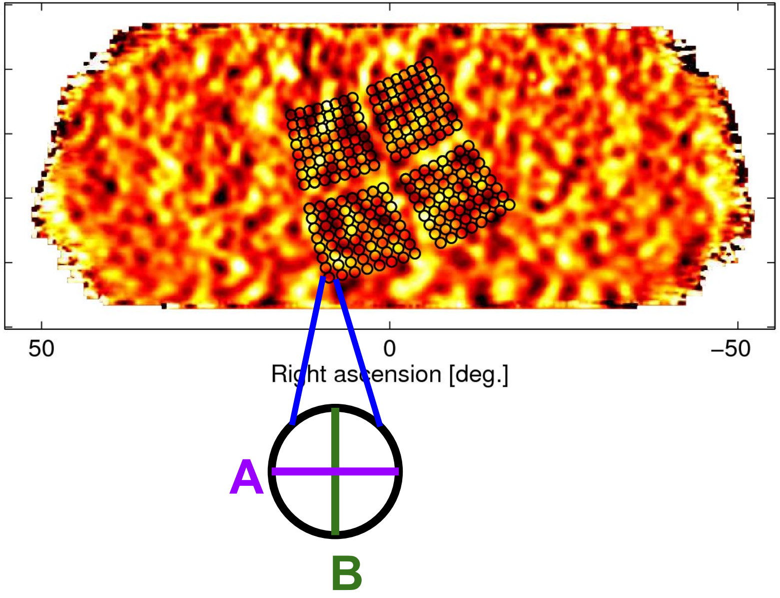


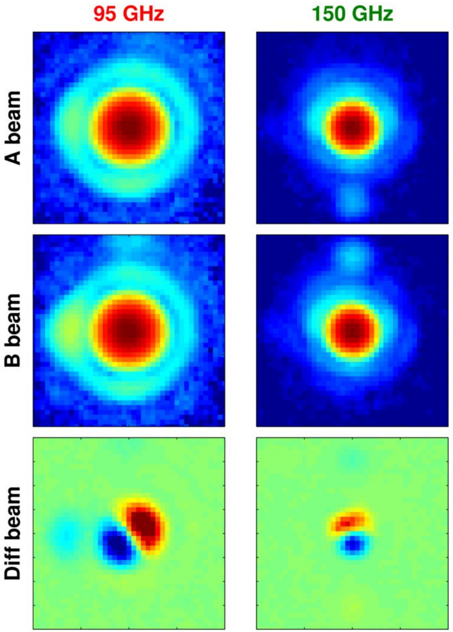
SPT-SZ beams (Shirokoff+ 2011)

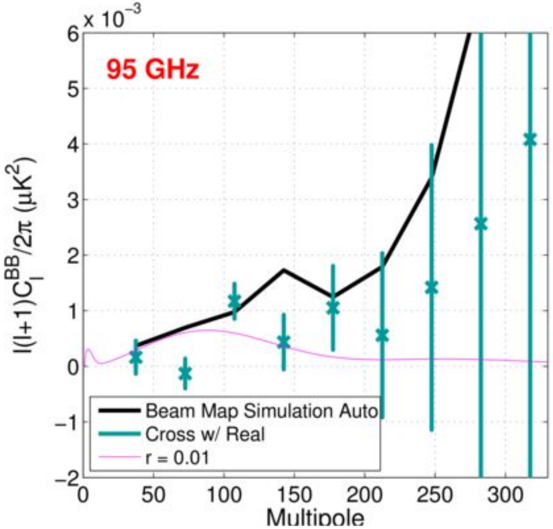


Dividing by the wrong beam profile would bias N_eff

Example 2: Pair Differencing for Polarization







BICEP/Keck XI 2019

Beam shape differences leak the CMB temperature into polarization, potentially biasing the measurement of *r*

Minimizing Systematics in Experimental Design

It's always better to design a systematics-immune experiment than to have to remove it in analysis!

Many aspects of modern CMB experiments are due to lessons learned the hard way over the last 50 years.

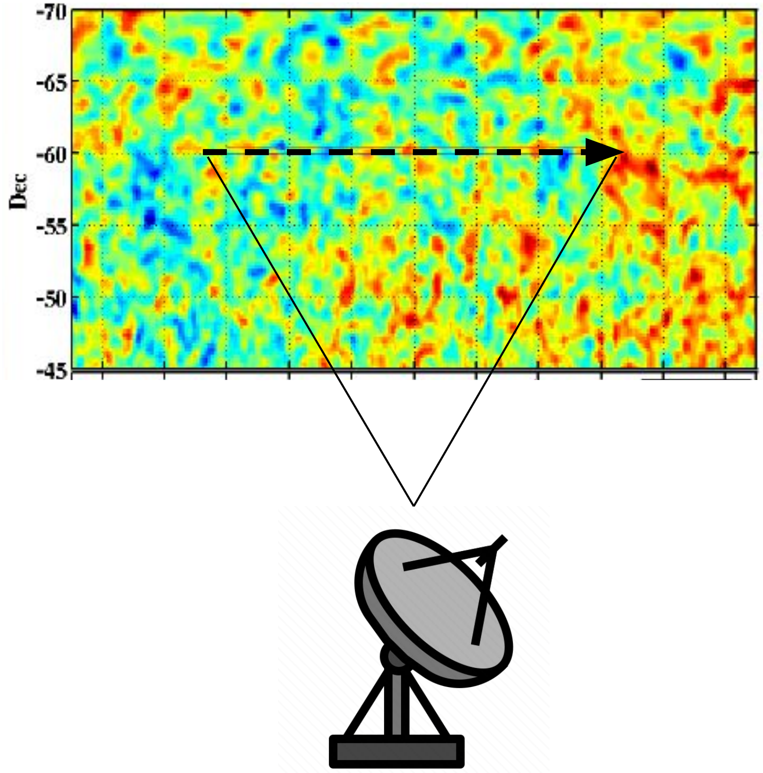
- Differential measurements
- Signal modulation
- Shielding optical elements, minimizing scattering and reflections
- Unobstructed optics
- Monolithic mirrors
- Regular calibration
- Observing strategy

Differential Measurements

Basic example: 4-wire resistance measurement removes lead wire resistance

Most CMB experiments are inherently differential! (vs. absolute)

- Attempting to measure *differences* in sky temperature/polarization from point to point
- A standard implementation of this is scanning across the sky patch at constant elevation
- Emission from the atmosphere is (roughly) constant, loading and calibration don't change much
- Polarization modulation

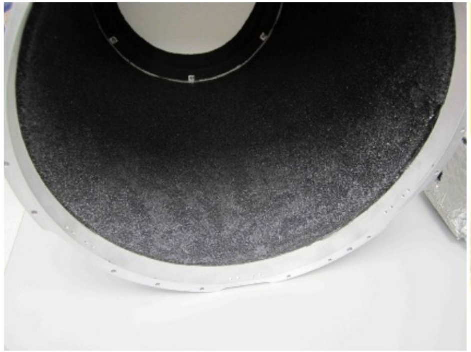


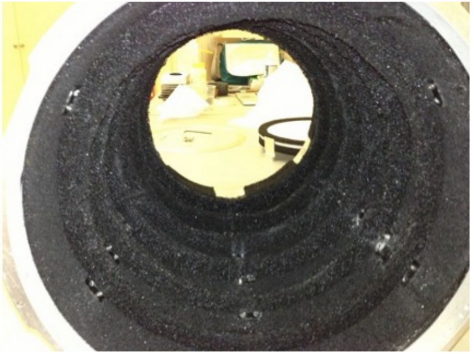
Reducing Sidelobes with Optics Tube Baffling

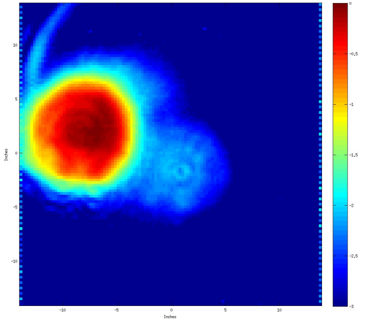
Pretty much every experiment has had an experience like this:

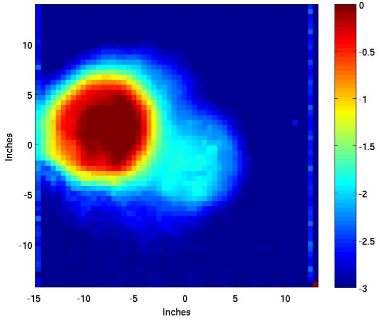
Buder+ 2014

No optics tube baffling: specular reflection leading to a “ring” sidelobe









With optics tube baffling: sidelobe removed!

Testing for Systematics

“Tennis racket” used to diagnose RFI susceptibility in mid-winter at South Pole

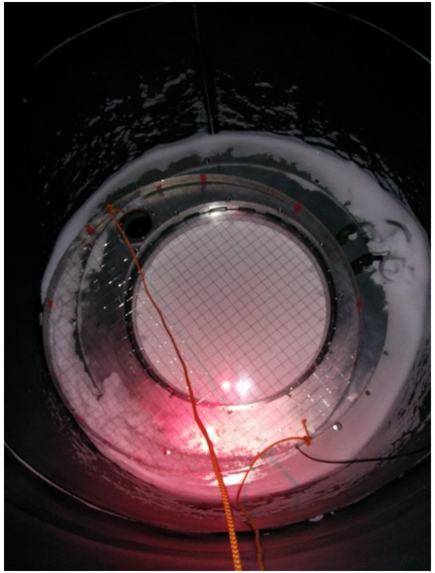


Photo: Sam Harrison

Sometimes it’s obvious that something is contaminating your data…

- After deploying BICEP3 in 2015, we noticed strong azimuth-synchronous signal that caused the SQUIDs to jump to different parts of the curve!
- Traced to RFI from the handheld radio system. Mitigation included:
  - Improved RF shielding around the cryostat and detector modules
  - Attenuating the radio signal power itself
  - Replacing the antenna with a directional one that reduced output towards the “Dark Sector”
- Why didn’t we see it before? The RFI was at ~450 MHz, and entering the optics tube before interacting with the readout system. Previous receivers had smaller-diameter optics tubes with higher waveguide cutoffs. BICEP3’s larger optics tube had a cutoff frequency of 340 MHz!
- *Lesson: unanticipated systematics are common, and often require post-hoc hardware solutions.* There’s a reason most experiments have an “engineering season.”

But: typical CMB timestreams look like noise! It’s only after averaging down over hours (or years) that we see the signal. How do we test for fainter systematics?

Testing for Systematics

Null tests (or “jackknives”) are the standard way to test for systematics.

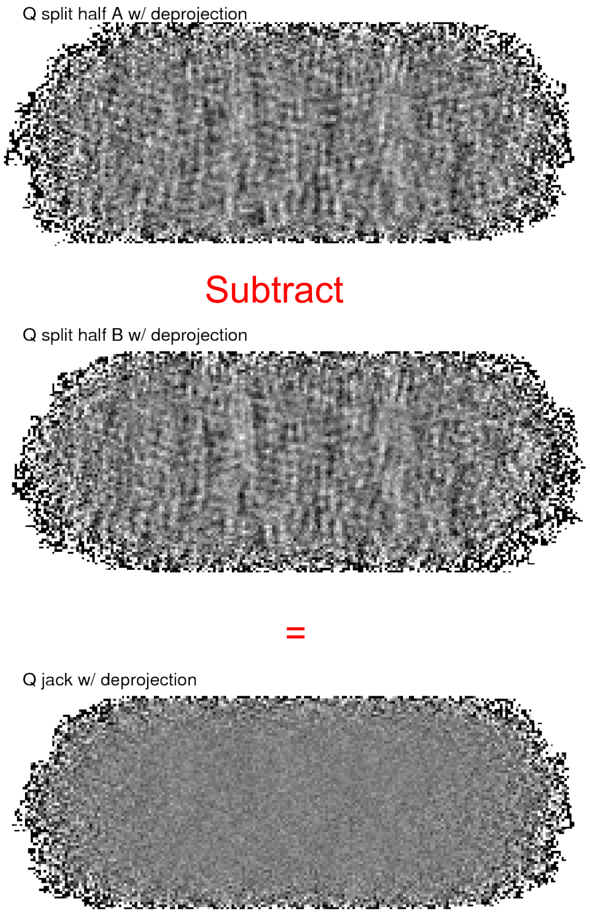
Divide your data into two halves, where the halves are chosen such that you expect a systematic to contaminate them at different levels.

Take the difference (usually at the map level):

- The sky signal should be removed, while systematics in one map but not the other should remain
- The difference map is then compared to the noise expectation.
- “Failing” a null test means there is structure in the difference map significantly greater than the noise

This also means that the science data can only probe systematics at or above the intrinsic noise level of the experiment!

Many analyses are “blind,” meaning that we don’t look at the real signal until the data pass null tests at a satisfactory level. This could encompass several rounds of cleaning data and calculating jackknives.



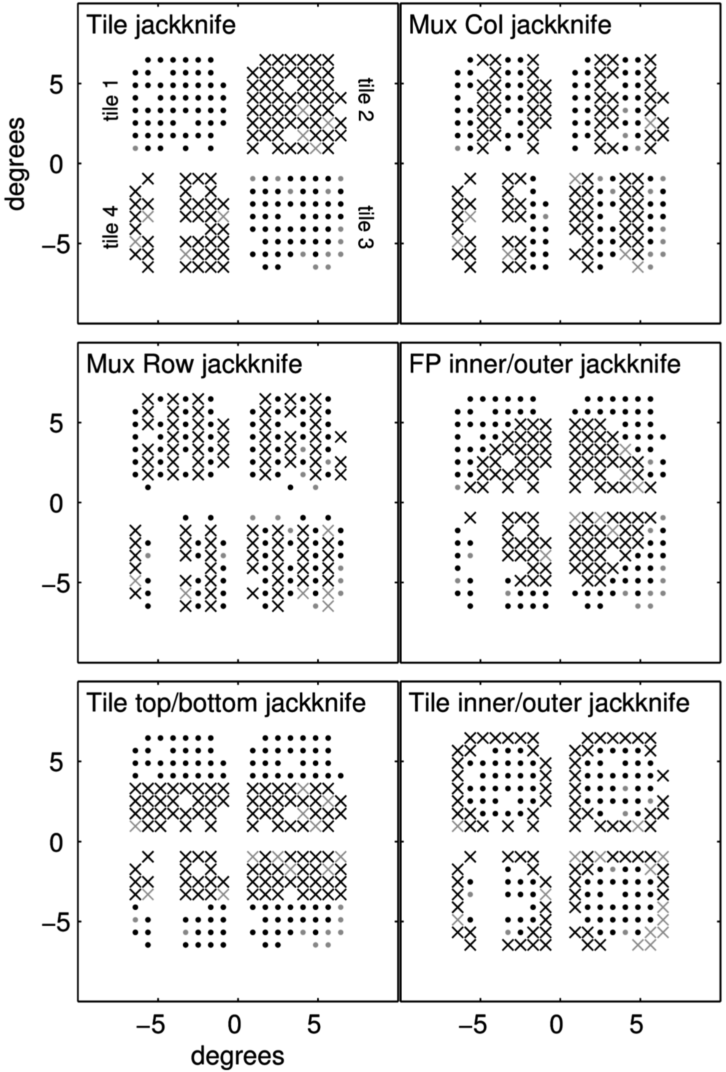
BICEP/Keck III 2015

Typical Null Tests

- Temporal (long-term changes in instrument/environment)
- Scan direction (time constants, transfer function)
- Azimuth or elevation (ground pickup)
- Moon/Sun up/down (sidelobes)
- Boresight angle (polarization)
- Detector tests
  - Wafer splits (bandpasses, etc.)
  - Wafer inner/outer (fabrication effects)
  - Focal plane inner/outer (optics)
  - Mux row/column (readout, crosstalk)

Jackknives can also be chosen to *enhance* a systematic effect (e.g. if a systematic cancels in the real map).

BICEP/Keck III 2015



Null Test Statistics

For each null map we typically calculate the PTE (probability to exceed) - the probability that you would exceed the observed χ^2 if the instrument+noise model is correct.

For a set of bandpowers **d** and a bandpower covariance matrix **C** (derived from signal+noise sims):

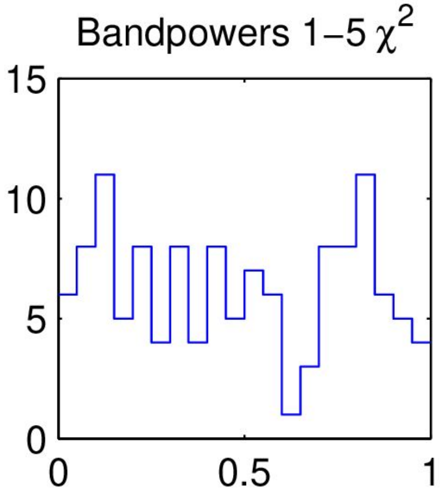
$$\chi^2 = \mathbf{d}\ \mathbf{C}^{-1}\ \mathbf{d}^T$$

To get the PTE, compare to a χ^2 distribution with degrees of freedom = number of bandpowers. In practice we compare to the distribution of large number of simulations, which can take into account complications like bandpower correlations.

A very **low** PTE could indicate excess signal above the noise, potentially pointing to systematics.

A very **high** PTE could indicate overestimated error bars.

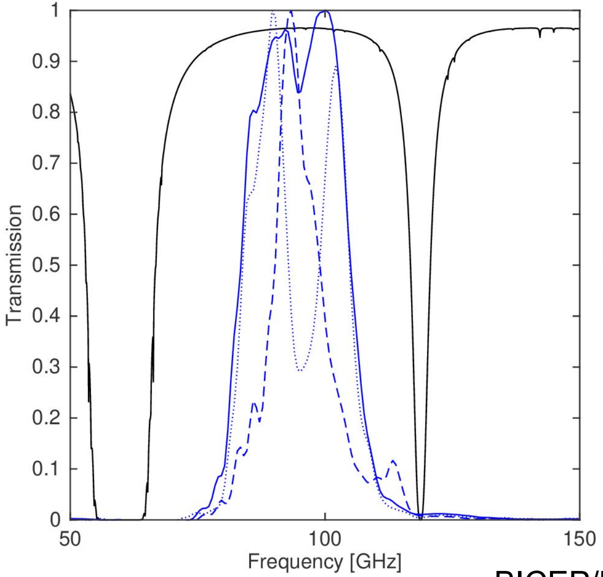
We often compute PTEs for many null maps, and inspect a histogram of PTE values with the expectation that they are consistent with a uniform distribution. In a set of 20 uncorrelated PTEs you might expect one with PTE < 0.05.



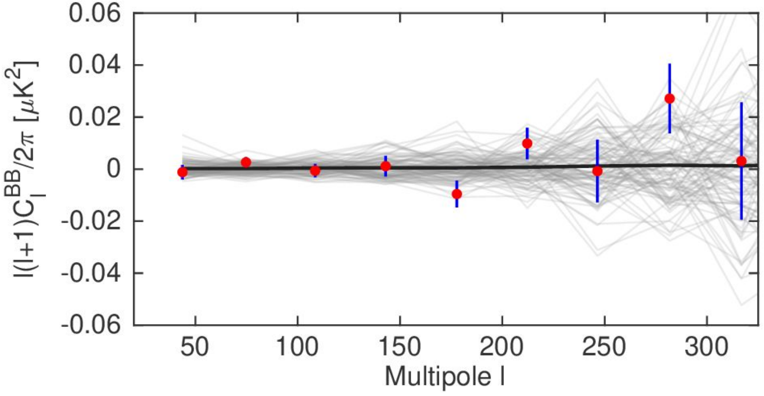
BICEP/Keck XV 2022

Example: BICEP3 Spike/Dip Jackknife

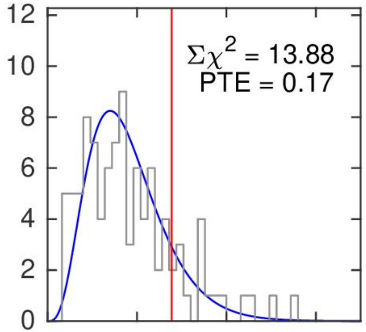
In the 2016 season we found that some FTS spectra showed “spikes” and “dips,” which we traced to reflections between a few delaminating low-pass edge filters and detector wafers. Was there a noticeable systematic associated with these bandpass differences?



Spike/Dip Jackknife bandpowers



Sum of squared devs



BICEP/Keck XV 2022

In the next season we replaced the delaminating filters and the FTS spectra cleaned up!

Removing Systematics

How you remove a systematic depends on how well you understand it!

- A well-characterized effect could just be subtracted at the timestream or map level
- If we knew the form but not the amplitude, we could fit for a template and subtract (**deprojection**)
- If we knew the amplitude but not the form, we could calculate bandpowers of the systematic and debias
- If there is substantial uncertainty on both the form and amplitude, little argument for direct removal. Run simulations to determine how important this could be. (Often the case for systematics close to or lower than the noise level of the experiment)

Deprojection

We often know that a systematic exists at some level, and we can parametrize it, but its amplitude is uncertain.

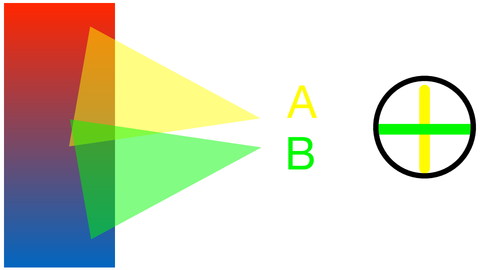
- Form a *template* of the expected signal when making maps
- Regress the template against the real data and subtract off the best fit

Deprojection allows the data to “choose” the systematic amplitude, and renders the maps insensitive to systematics of that form. (It’s just a filter.)

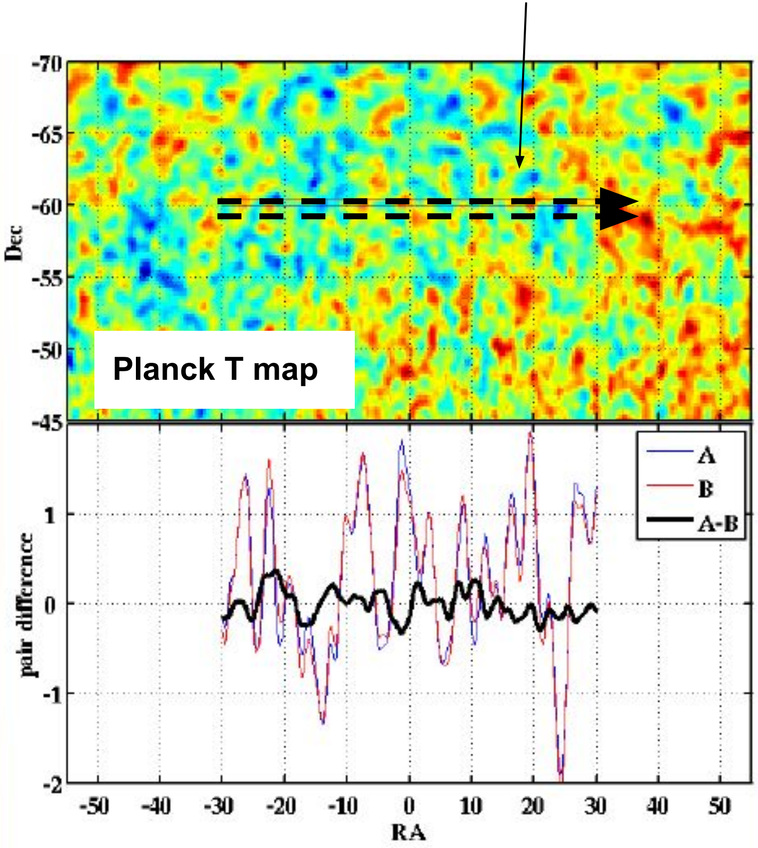
It also removes real sky signal, so we need to include it in simulations and transfer function calculations!

Can consider this procedure “marginalizing” over the systematic.

Deprojecting Differential Pointing

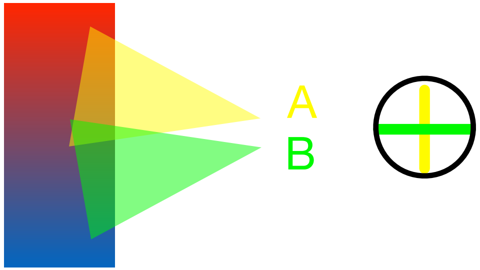


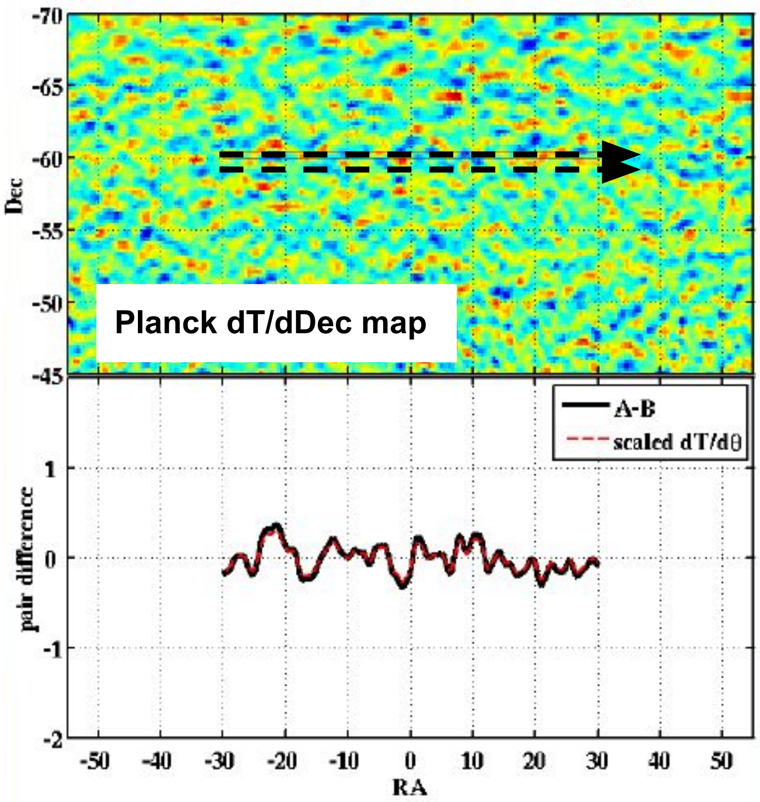
A and B's paths along the sky as the telescope scans



In a pair differencing experiment, a pointing mismatch between the orthogonally polarized detectors can inject a false polarization signal from the bright unpolarized sky.

Deprojecting Differential Pointing





That leakage signal is proportional to the spatial derivatives of the underlying temperature sky.

We can form a *deprojection template* by sampling the derivative map at the pointing center, regress the data against the template, and remove.

Calibration

Many systematic effects can be probed using calibration measurements (some of which may be necessary for the experiment to be done in the first place).

- Bandpasses
- Beam profiles
- Polarization angles
- Detector gains
- Time constants

These measurements are commonly used in specialized simulations to determine the potential effect of systematics on the parameter of interest.

When doing this, we also need to be aware of noise and systematics in the calibration measurement itself!

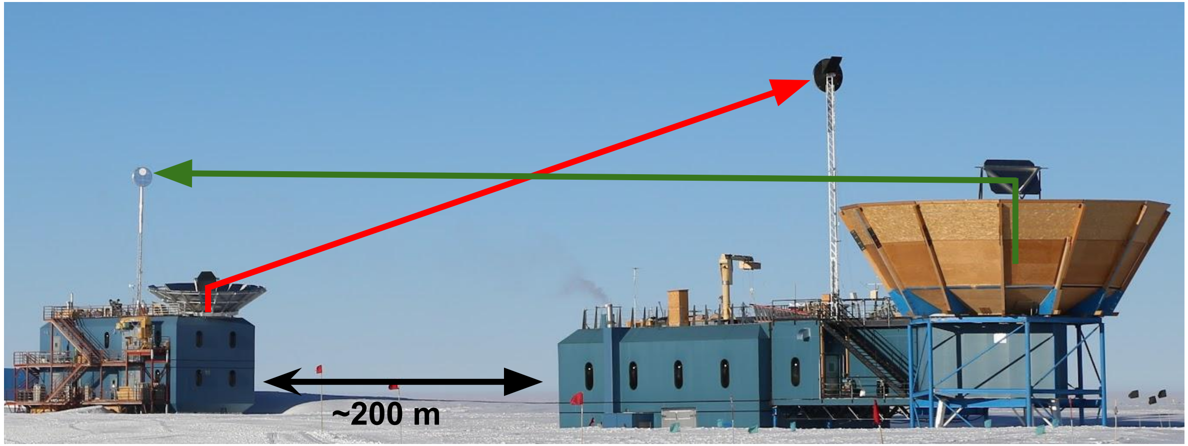
Can also use calibration measurements to verify deprojection coefficients.

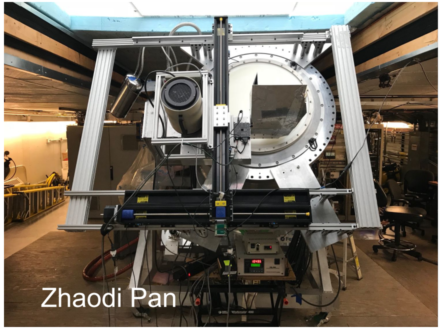
Calibration Hardware

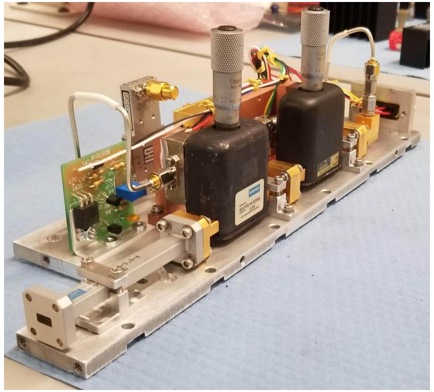
Things to measure	**Hardware**
Main beams	Near-field beam mapper; Large thermal chopper; Far-field flat mirror; Mast
Sidelobes	Far-field flat mirror; Mast; Amplified microwave source (+ rotating polarized stage)
TQU beams/pol angles	Far-field flat mirror; Mast; Amplified microwave source (+ rotating polarized stage); Sparse wire grid, dielectric sheet
Bandpasses	Fourier Transform Spectrometer; Optical filters; Narrowband source
Time constants	Amplified microwave source (+ rotating polarized stage)
Optical efficiency	Aperture-filling thermal load
Magnetic sensitivity	Helmholtz coil
Vibration sensitivity	Accelerometers
RFI sensitivity	RF sources
Pointing	Star camera
Precipitable water vapor	Water Vapor Radiometer
RFI environment	RF monitoring

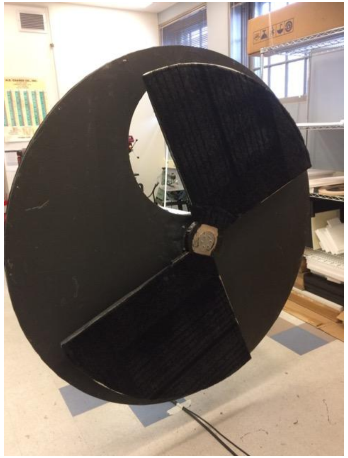
For CMB-S4 SATS

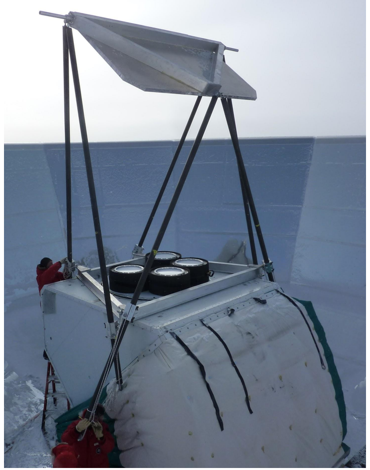
Calibration Hardware











Example: Higher-order beam shape mismatch

Mismatched beam shapes can leak temperature into polarization, biasing the *r* measurement. We deproject out the leading-order beam difference modes, but higher-order leakage remains.

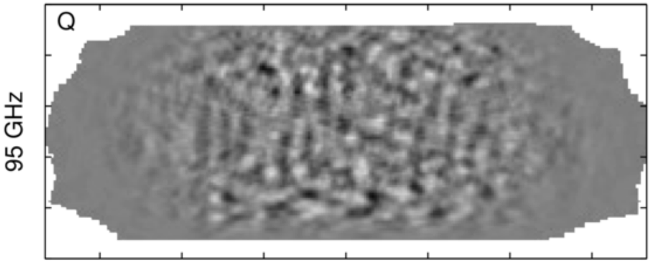
Using high-fidelity beam maps (of over 10,000 beams in the experiment), we propagate this leakage through a full-scale simulation and calculate the additive contamination expected in our CMB maps.

After accounting for uncertainties in the beam measurement, we estimate a bias on the tensor-to-scalar ratio in the BK18 results of

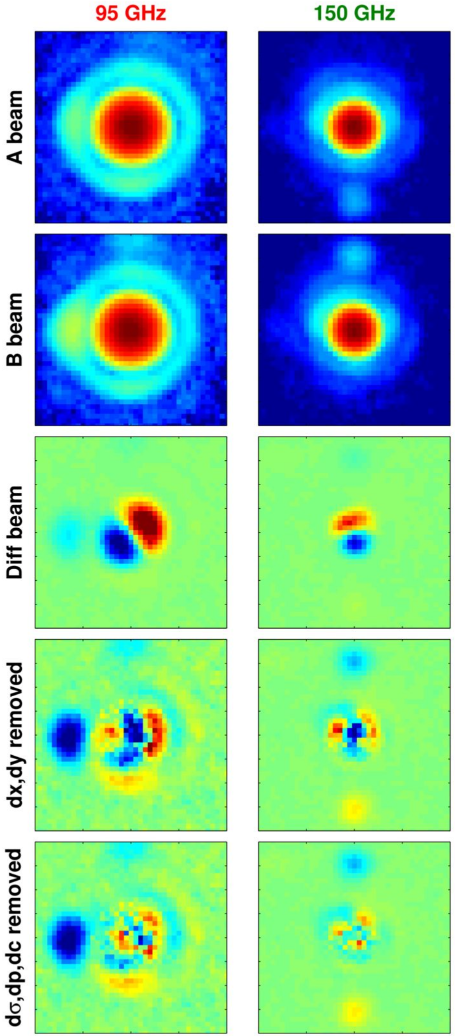
Δr = 0.0015 ± 0.0011 …which is much smaller than the statistical uncertainty **σ(r) = 0.009**

So at this point, differential beam systematics do not significantly affect the science results.

Predicted 95 GHz
T→P leakage



BICEP/Keck XI 2019



Auto vs Cross Spectra

Power spectra are traditionally calculated by correlating a map with itself ("auto spectrum."). If a map consists of signal + noise, this schematically comes out to:

MxM = (S+N) x (S+N) = SxS + 2(SxN) + NxN which needs to be "noise debiased."

We can also correlate a map with a different map containing the same signal but different noise:

$M1$x$M2$ = (S+$N1$) x (S+$N2$) = SxS + Sx$N1$ + Sx$N2$ + $N1$x$N2$ so the noise bias disappears.

Similarly, if one map contains additive systematics and the other doesn't, the cross-spectrum should be unbiased.

This can be done internally (e.g., maps taken at different azimuth ranges) or externally (different experiments should have independent systematics).

You can also take the cross spectrum of a predicted systematic with your real maps to see if they exist in the data!

Reporting Systematics

It’s common to report the statistical and systematic uncertainties separately: x = 1.0 ± 0.2 (stat) ± 0.1 (syst)

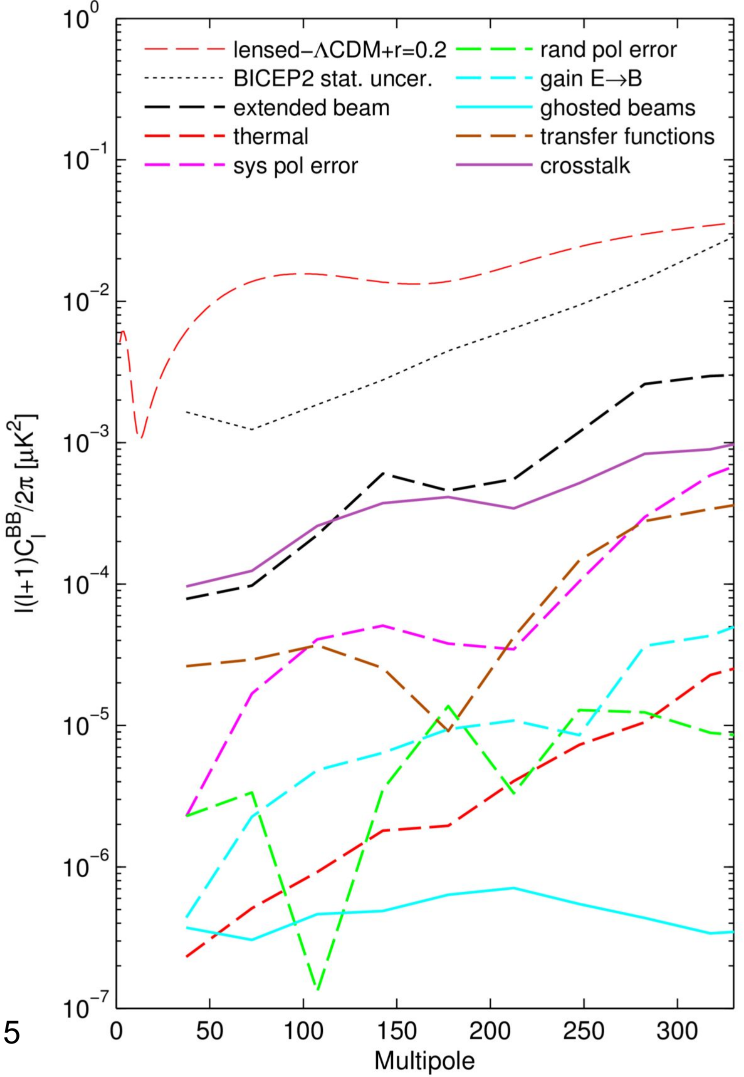
CMB experiments often break down the systematic uncertainty by contribution.

TABLE 4
INSTRUMENTAL SYSTEMATICS

Systematic	Characteristic r
Crosstalk	$\simeq 3.2 \times 10^{-3}$
Beams (including gain mismatch)	$< 3.0 \times 10^{-3}$
EMI	$\lesssim 1.7 \times 10^{-3}$
Cross polar response	$\lesssim 10^{-3}$
Eetector transfer functions	$< 5.7 \times 10^{-4}$
Systematic polarization angle error	$< 4.0 \times 10^{-4}$
Gain variation $E \rightarrow B$	$< 5.3 \times 10^{-5}$
Random polarization angle error	$\lesssim 5.0 \times 10^{-5}$
Thermal fluctuations	$< 1.2 \times 10^{-5}$
Ghost beams	$\simeq 7.2 \times 10^{-6}$
Scan synchronous contamination	$\lesssim 1 \times 10^{-8}$
Total	$\simeq (3.2-6.5) \times 10^{-3}$

NOTE. — The comparable characteristic r of BICEP2’s statistical uncertainty is $r = 3.1 \times 10^{-2}$.

BICEP/Keck III 2015



References

Bob Wilson's Nobel Lecture

BICEP/Keck III 2015 (Systematics paper)

BICEP/Keck XI 2019 (Beam Systematics)

A parting thought: Systematics are a part of life. Dealing with them will always be painful and take longer than you expect. So don't hate the systematics - hate the system!