Systematic Error: Good vs Bad Science

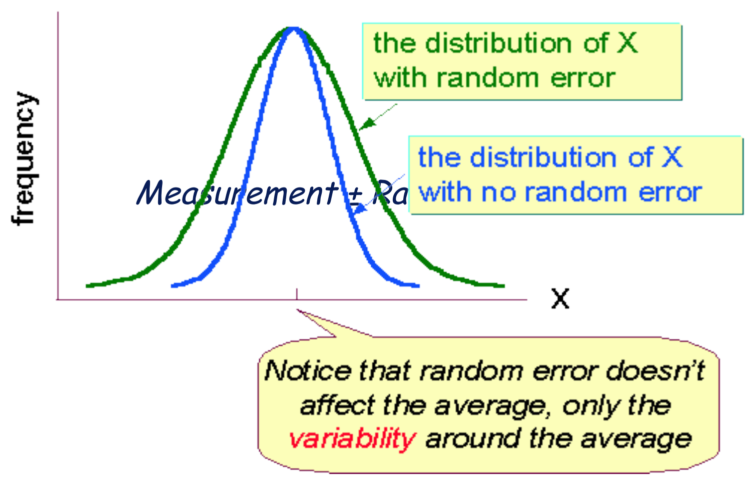
Tony Tyson

Physics Department

UC Davis

Random Errors

- ALWAYS **present.**

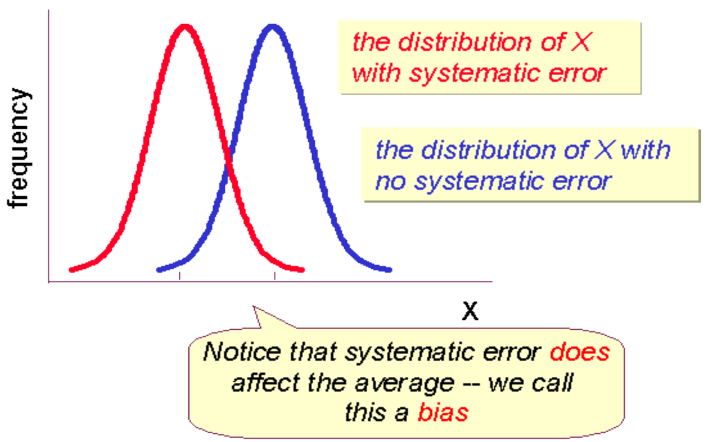


- Sources:
  - **Random operator errors**
  - **Random changes in experimental conditions**
  - **Noise in apparatus**
  - **Noise in Nature**

- How to minimize them?
  - **Take repeated measurements and calculate their average.**

Systematic Errors

- Are TYPICALLY present.



- Sources:
  - Instrumental, physical and human limitations.
    - » Example: Device is out-of calibration.

- How to minimize them?
  - Careful calibration.
  - Best possible techniques.
  - Discover and control them.

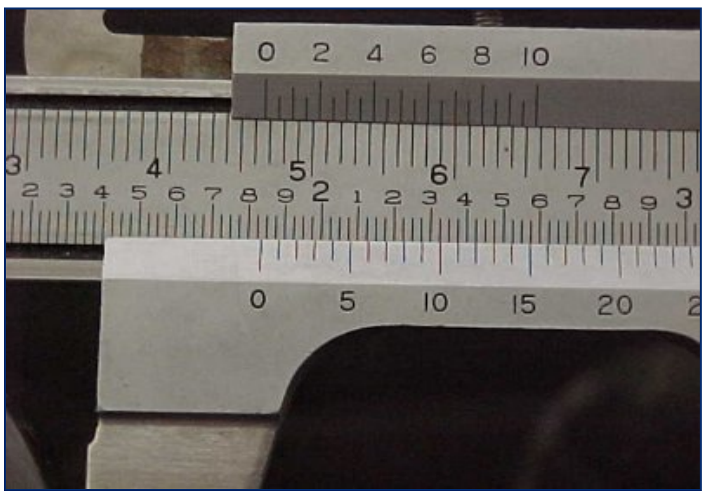
Precision and Accuracy in Measurements

- Precision

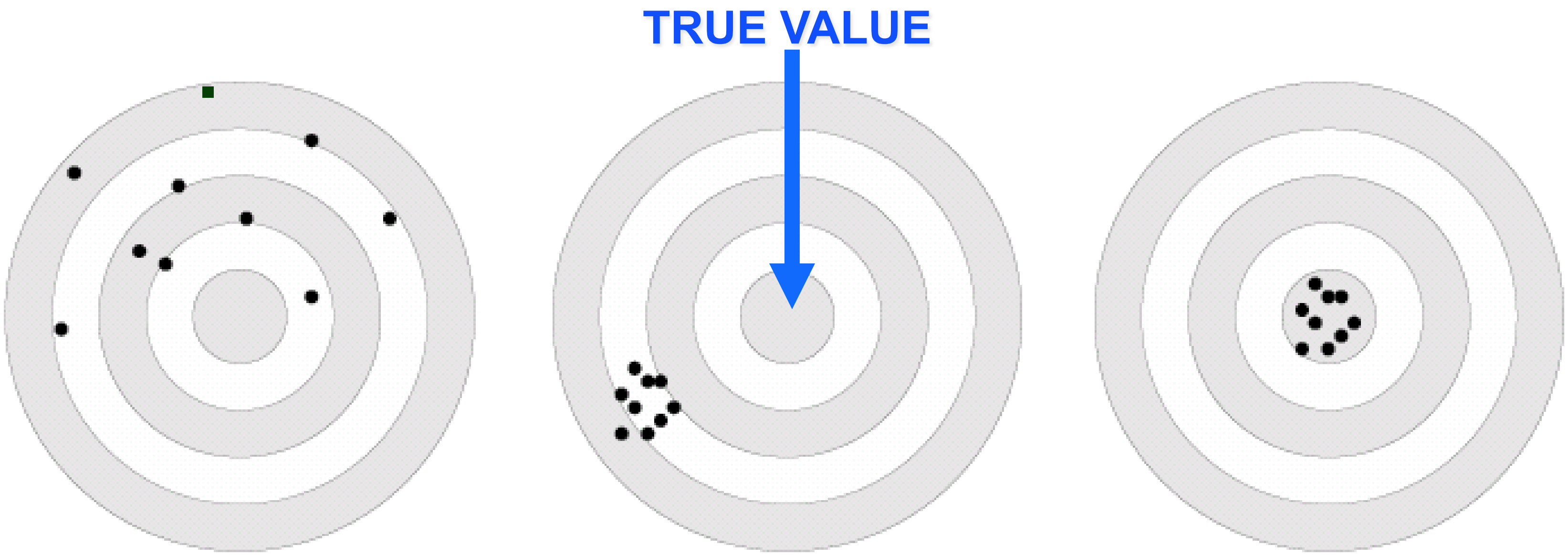
  How reproducible are measurements?

- Accuracy

  How close are the measurements to the true value.



accuracy and precision



not precise and not accurate	precise but not accurate	precise and accurate
large random and *systematic* errors	small random error, large *systematic* error	small random error, small *systematic* error

systematics

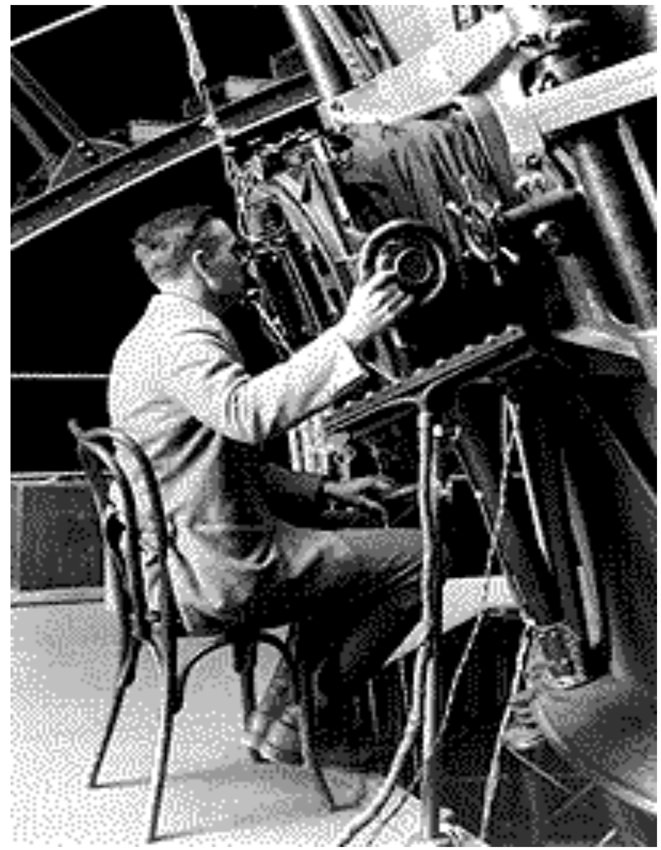
Particularly troubling today is that we don't fully know what we don't know

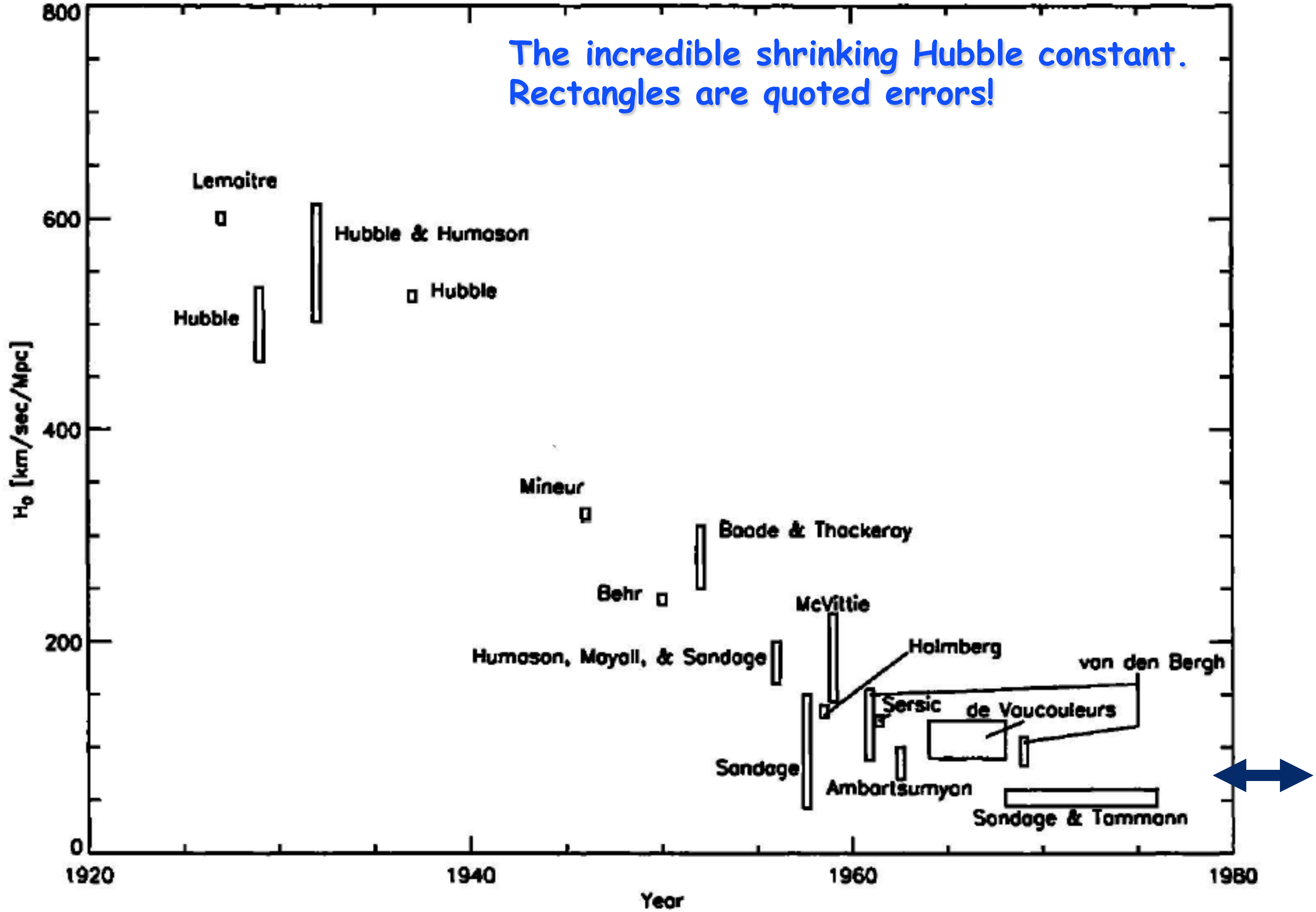
Testimony by Bert Ely to the Subcommittee on Financial Management, the Budget, and International Security of the Senate Committee on Governmental Affairs   July 21, 2003

Example:
Measurements of expanding universe

Vesto Slipher

Edwin Hubble





Trimble (1996) PASP 108, 1073

Systematics: catch-22

The difficulty is this: if we understand the systematic we can correct for it, but if we don't understand the systematic we won't think of it at all or our error estimate will be wrong.

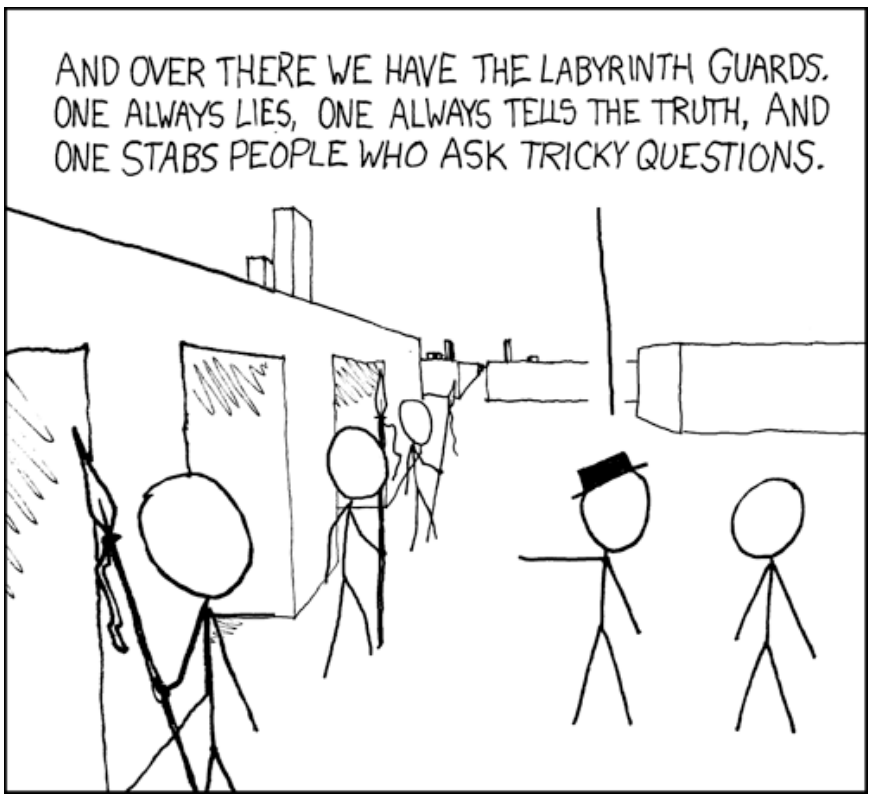
It is only at the edge of understanding where systematic errors are meaningful: we understand enough to realize it might be a problem, but not enough to easily fix it.

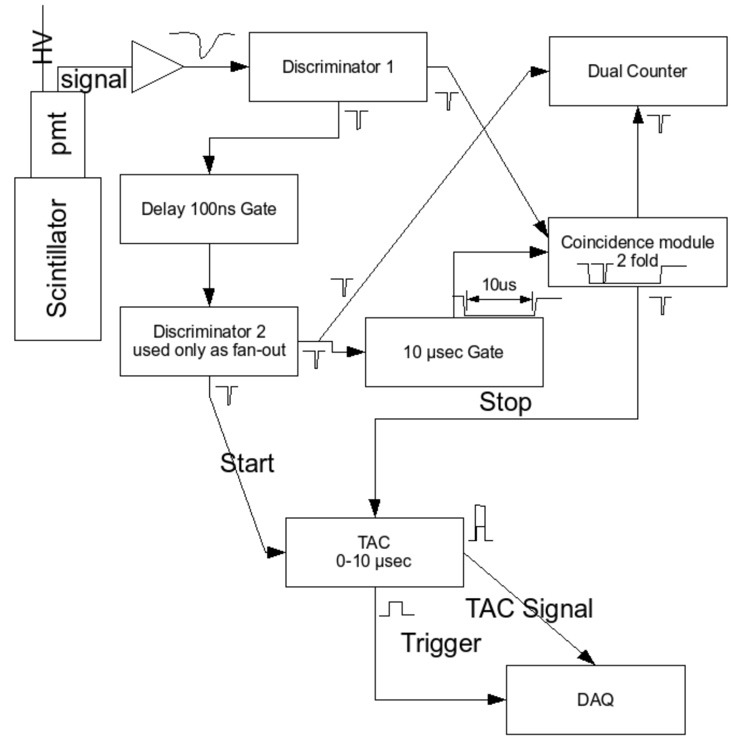
How can we find systematic errors?

Calibrate everything.

Do *experiments* on our Experiment.

Logical deduction.



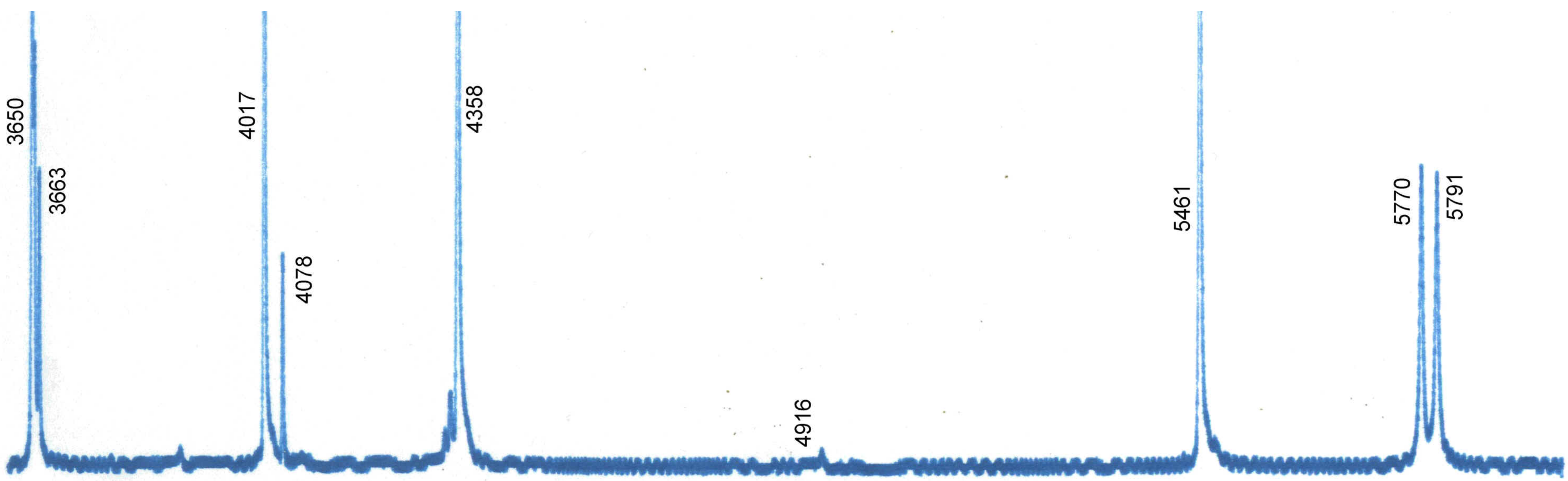


Logical process of elimination

Calibration

Your instrument reading

0 **50** **100**



Avoiding Systematics

The best prevention of systematic error is good experiment design.

How can we robustly attack this problem in an existing experiment or observation?

A mix of calibration, simulations and exploratory tests.

Simulations can teach us where sensitivity to systematics are. We may then explore these avenues; search for the signature of each systematic, isolate it, understand it, and gain control of it.

In practice, for each experimental field it is a kind of "art" which demands familiarity with the likely systematics. It is the responsibility of the experimentalist to probe for systematics and of the theorist to allow for them.

Healthy skepticism

- Be skeptical of your own work
- Test relentlessly for systematics
- Avoid early press conferences

A Result of Unexplored Systematics: *Pathological science*

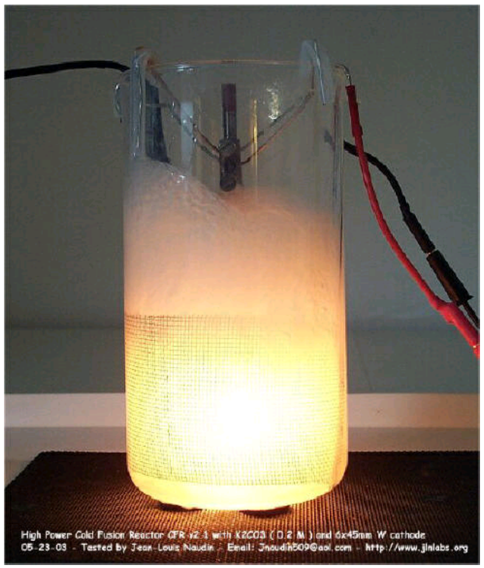
Not fraud

Well intentioned, enthusiastic scientists are led astray

Examples abound in every field of science

Example: Cold fusion

- **Pons and Fleischman claimed bench-top fusion using a palladium battery**
- **Before doing a control experiment, and before peer review, they held a press conference**



“Cold fusion” has since been debunked.

Features of Pathological Science

- The maximum effect is produced by a barely perceptible cause, and the effect doesn't change much as you change the magnitude of the cause.

- The effect only happens sometimes, when conditions are just right, and no one ever figures out how to make it happen reliably. The people who can do it are unable to communicate how they make it happen to the people who can't.

- The effect is always close to the limit of detectability.

- There are claims of great accuracy, well beyond the state of the art or what one might expect.

- Fantastic theories contrary to experience are suggested. Often, mechanisms are suggested that appear nowhere else in physics.

- Criticisms are met by ad hoc excuses thought up on the spur of the moment.

Irving Langmuir 1953 see: *Physics Today* Oct. 1989

Some common mistakes

Poor experiment design

Not testing for systematics (control)

Ignoring sample selection effects (bias)

Bad statistics: assume wrong distribution (tails!)

Failure to repeat the experiment using different sample with same physics

Trick

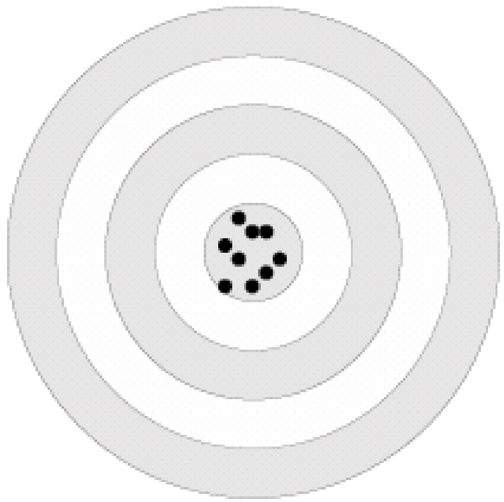
You are trying to measure hopelessly small SIGNAL

Suppose you suspect your experiment has systematic error (drift, false signal...)

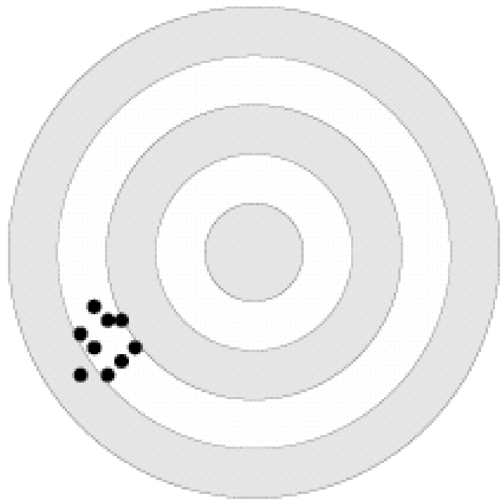
Somehow arrange to turn the SIGNAL off and on

Result: SIGNAL <u>without</u> systematic error!

Overcoming systematics: *Chop*

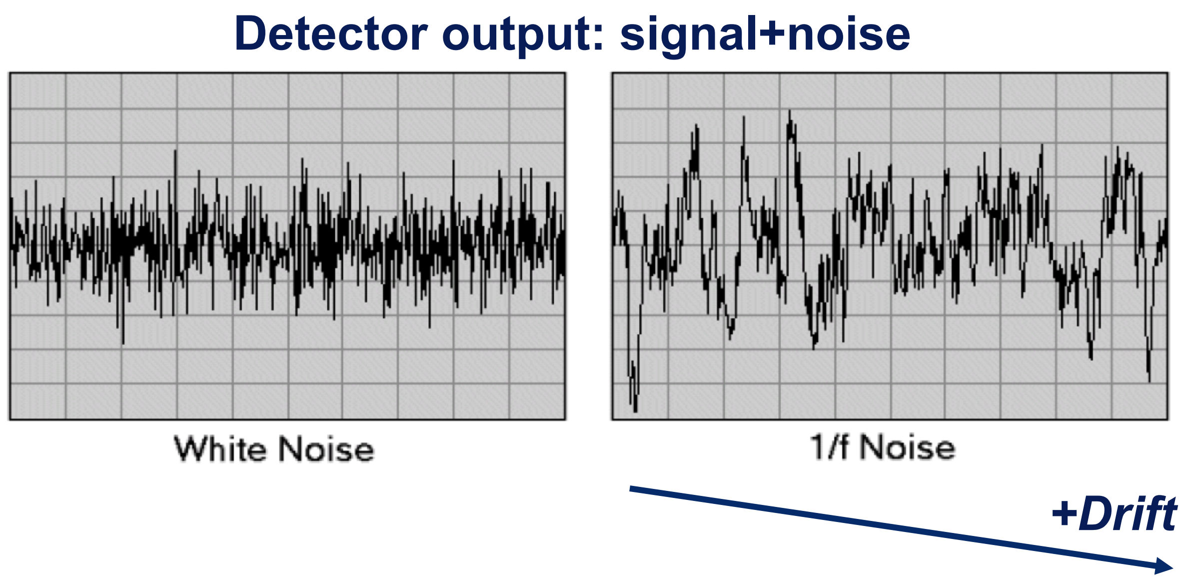


Overcoming systematics: *Chop*



Suppose your signal is at zero frequency and smaller than the noise

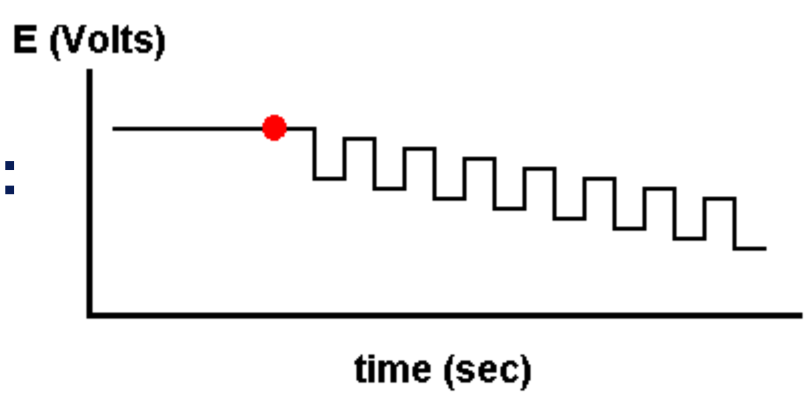
Signal ↕



Random error (noise)

Systematic error

CHOP SIGNAL:



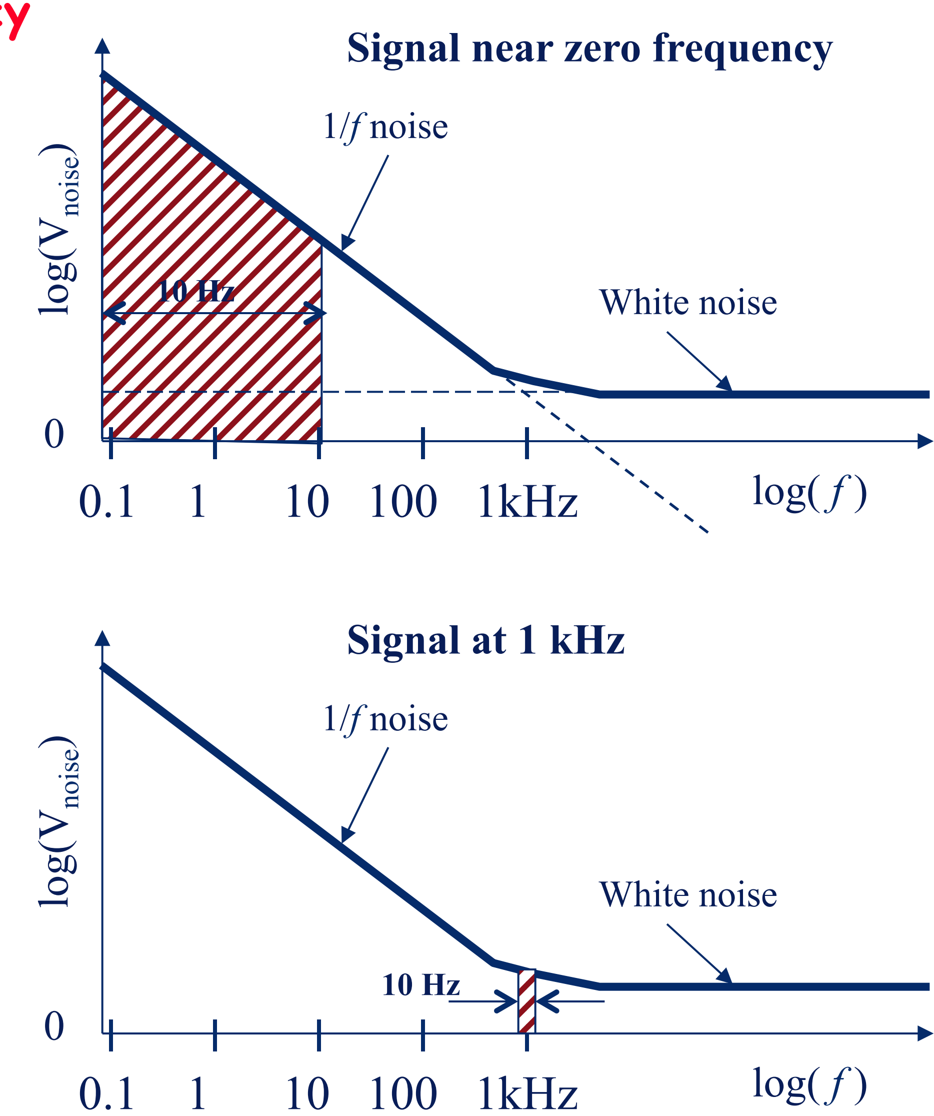
Signals and noise

Many systems have more noise at low frequency

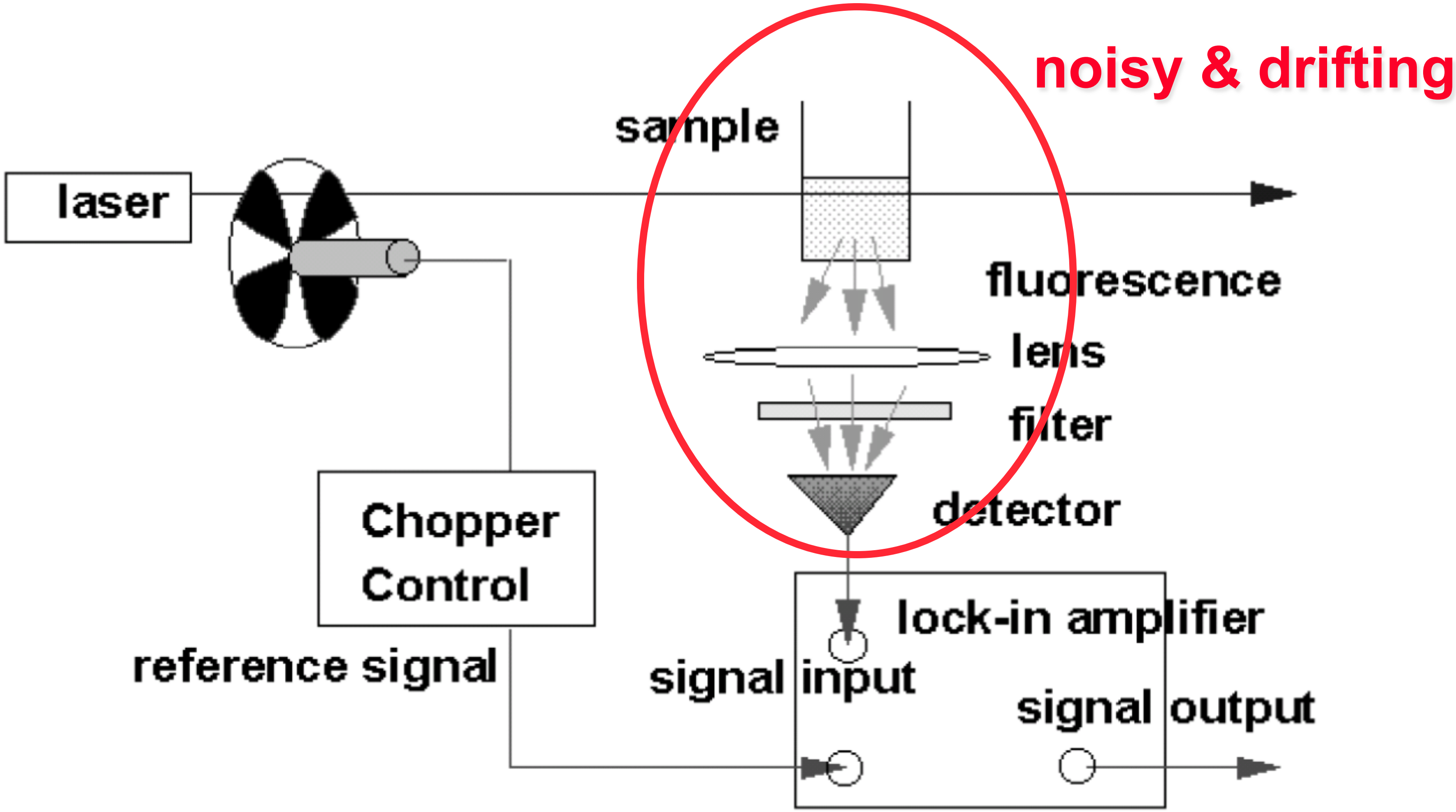
Frequency dependence of noise

- Low frequency ~ 1 / *f*
  - example: temperature (0.1 Hz) , pressure (1 Hz), acoustics (10 -- 100 Hz)
- High frequency ~ constant = white noise
  - example: shot noise, Johnson noise, spontaneous emission noise
- Signal/Noise ratio depends strongly on signal freq
  - worst at DC, best in white noise region
- *Problem: most signals at DC or at low frequency*
- *Solution: chop, thus moving signal to high (chop) frequency*

Total noise in 10 Hz bandwidth:



Phase-sensitive detection



A microwave system with low antenna side-lobes

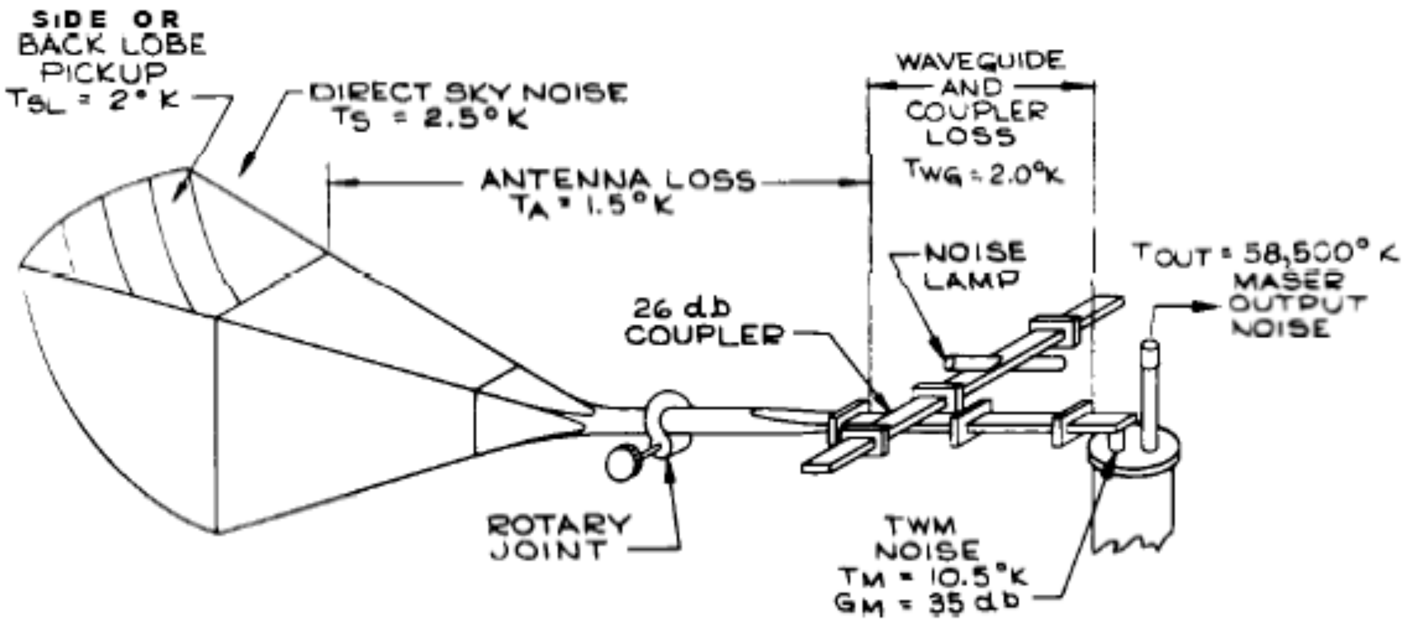


Fig. 6 A diagram of the low noise receiver used by deGrasse, Hogg, Ohm and Scovil to show that very low noise earth stations are possible. Each component is labeled with its contribution to the system noise.

CMB Discovery missed

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

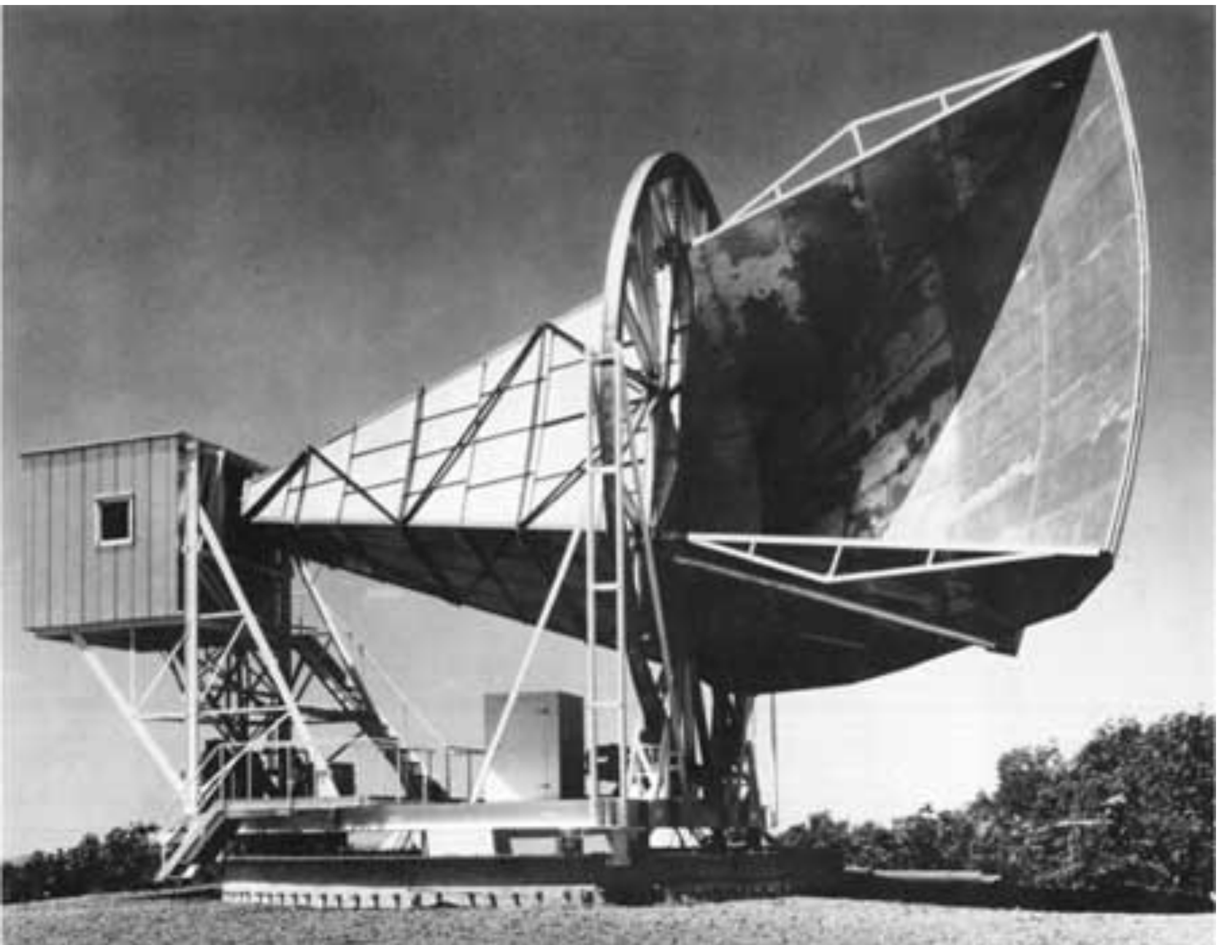
the temperature was found to vary a few degrees from day to day, but the lowest temperature was consistently 22.2 ±2.2°K. By realistically assuming that all sources were then contributing their fair share (as is also tacitly assumed in Table II) it is possible to improve the over-all accuracy. The actual 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

$$T_{system} = 21 \pm 1°K.*$$

The inference from this result is that the "+" temperature possibilities of Table II must predominate.

Fig. 8 An excerpt from E. A. Ohm's article on the Echo receiver showing that his system temperature was 3.3K higher than predicted

Discovery of the CMB



Chop between sky and a cold load:

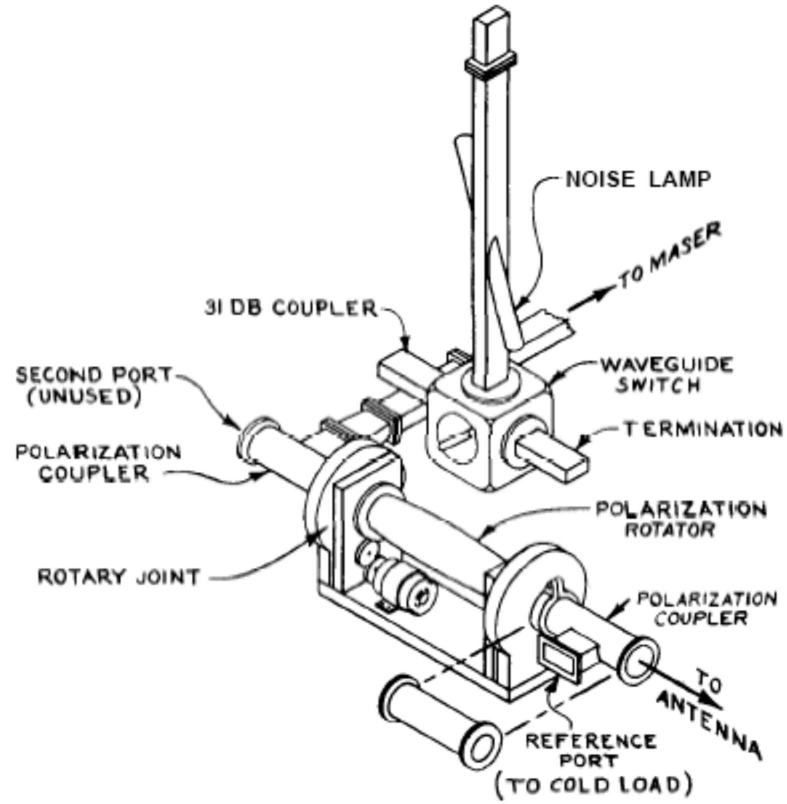


Chart recording

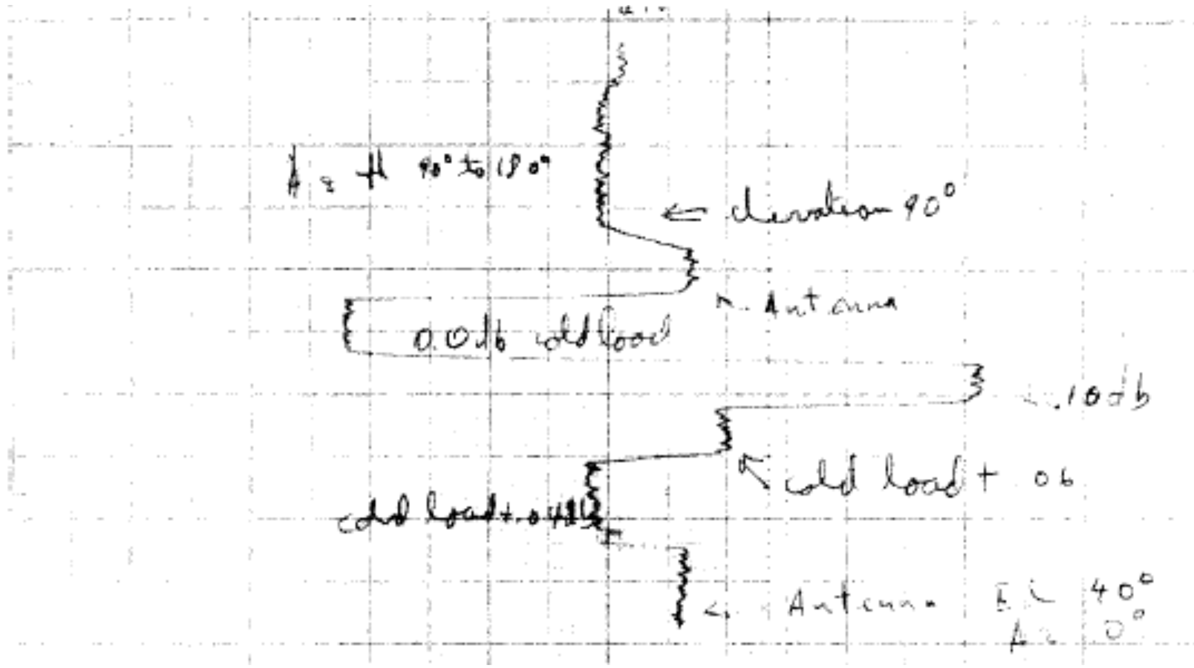


Fig. 9 The first measurement which clearly showed the presence of the microwave background. Noise temperature is plotted increasing to the right. At the top, the antenna pointed at 90° elevation is seen to have the samt noise temperature as the cold load with 0.04 db attenuation (about 7.5K). This is considerably above the expected value of 3.3K.

Quoting errors

Fourth Test of General Relativity: New Radar Result

Irwin I. Shapiro,* Michael E. Ash,† Richard P. Ingalls,‡ and William B. Smith†
Massachusetts Institute of Technology, Cambridge, Massachusetts 02319

and

Donald B. Campbell, Rolf B. Dyce, Raymond F. Jurgens,§ and Gordon H. Pettengill‖
Arecibo Observatory, Arecibo, Puerto Rico
(Received 15 March 1971)

New radar observations yield a more stringent test of the predicted relativistic increase in echo times of radio signals sent from Earth and reflected from Mercury and Venus. These "extra" delays may be characterized by a parameter λ which is unity according to general relativity and 0.93 according to recent predictions based on a scalar-tensor theory of gravitation. We find that $\lambda = 1.02$. The formal standard error is 0.02, but because of the possible presence of systematic errors we consider 0.05 to be a more reliable estimate of the uncertainty in the result.