

Data Challenge

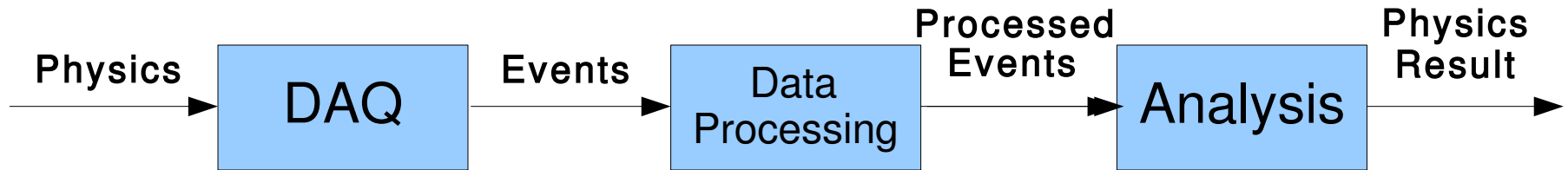
Jim Braun

2009 IceCube Science Advisory Committee Meeting

Data Challenge

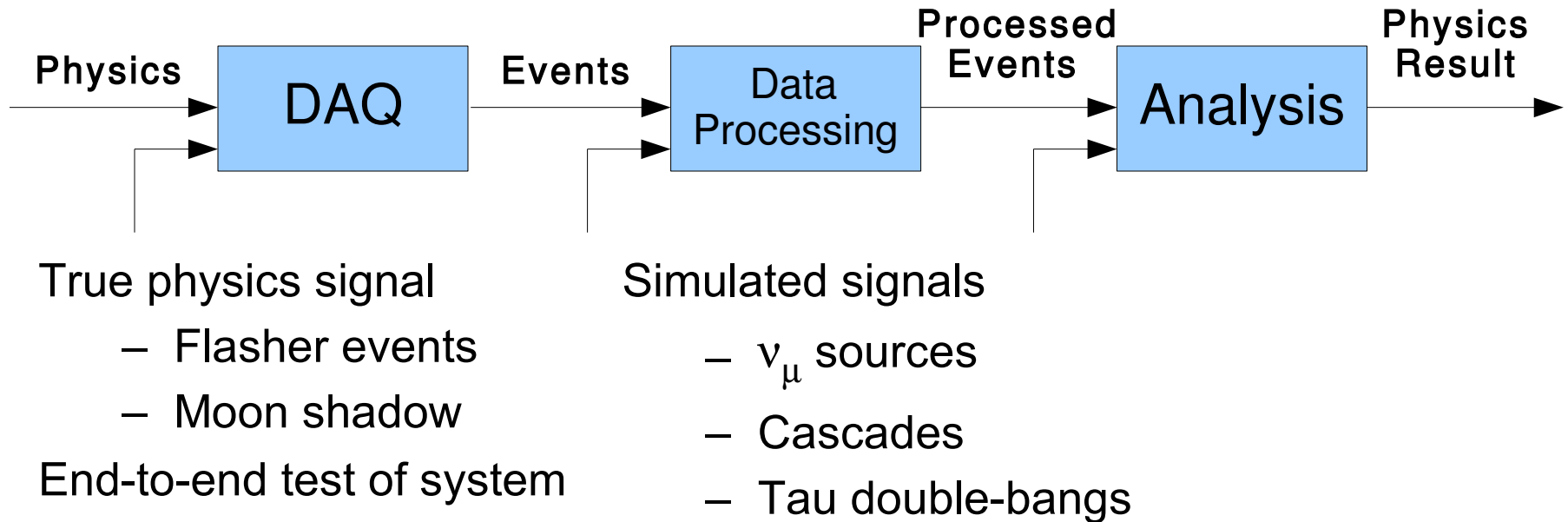
- **Most IceCube physics analyses lack a natural signal**
- **Data challenge** is the following generic process:
 - A simulated signal is added to the data
 - The data is analyzed, and the signal is either discovered at the expected level, or it is not
- **Data challenge** accomplishes the following:
 - Detects errors in DAQ, data processing, and analysis
 - Ensures those performing analysis understand their data and methods **before** physics results are obtained
 - Aids interpretation of physics results
 - Unbiased measurement of the performance of an analysis

What Can Go Wrong?



- System timing
 - DAQ hardware and software failures
 - Angular reconstruction
 - Energy reconstruction
 - Coordinate transformations
 - Data selection
 - Analysis methods
-
- We want to ensure the entire system is working properly from physics to analysis result

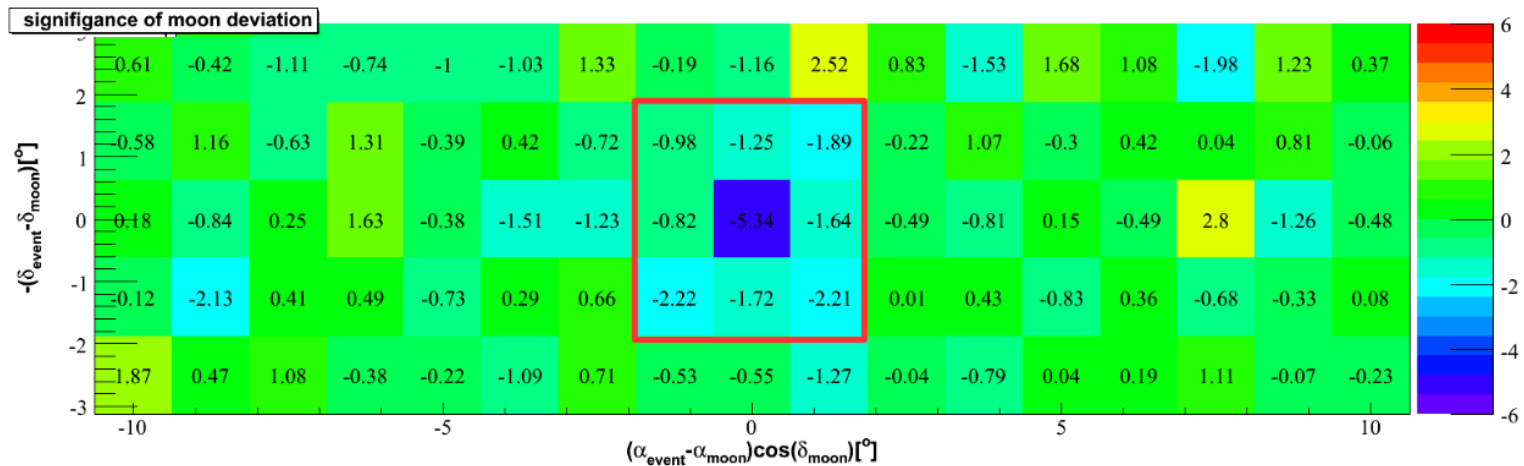
Levels of Data Challenge



- Data challenges at lower levels are more informative, but they are also more difficult

What We've Learned from the Moon

- The moon blocks cosmic rays and acts as a muon calibration source



- We observe the moon shadow with a significance consistent with the expected value ($5.34\sigma/6.1\sigma$)
 - Evidence DAQ and Data Processing are working properly
 - Focus on analysis-level data challenges

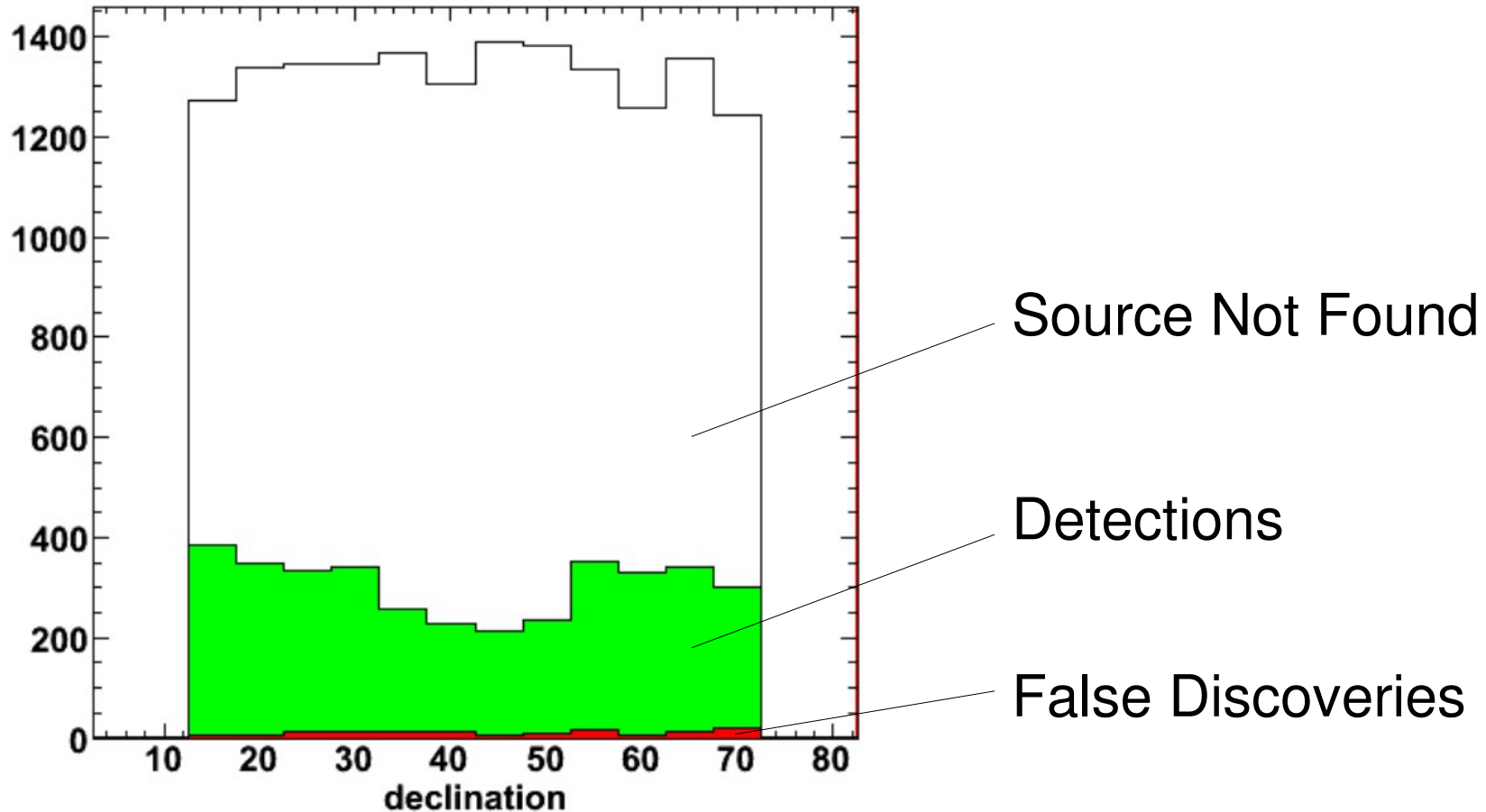
The FDR Data Challenge

- New neutrino point source method developed (FDR)
 - Advertised as powerful, but difficult to understand
- Need to assess the method in an unbiased way
 - Measure performance (Does the method work as advertised?)
 - If the method works, convince others
- Data challenge is an ideal strategy to accomplish this

The FDR Data Challenge

- Evaluate the method using a data challenge
 - Create ~17,000 sky maps with randomized AMANDA data and sources of various strength added
 - Send the set of maps to be analyzed
 - Receive coordinates for any sources discovered
 - Compare coordinates of sources with key: Sources listed within 5° of true location are considered detections; the rest are false discoveries.
- Software tools make this easy

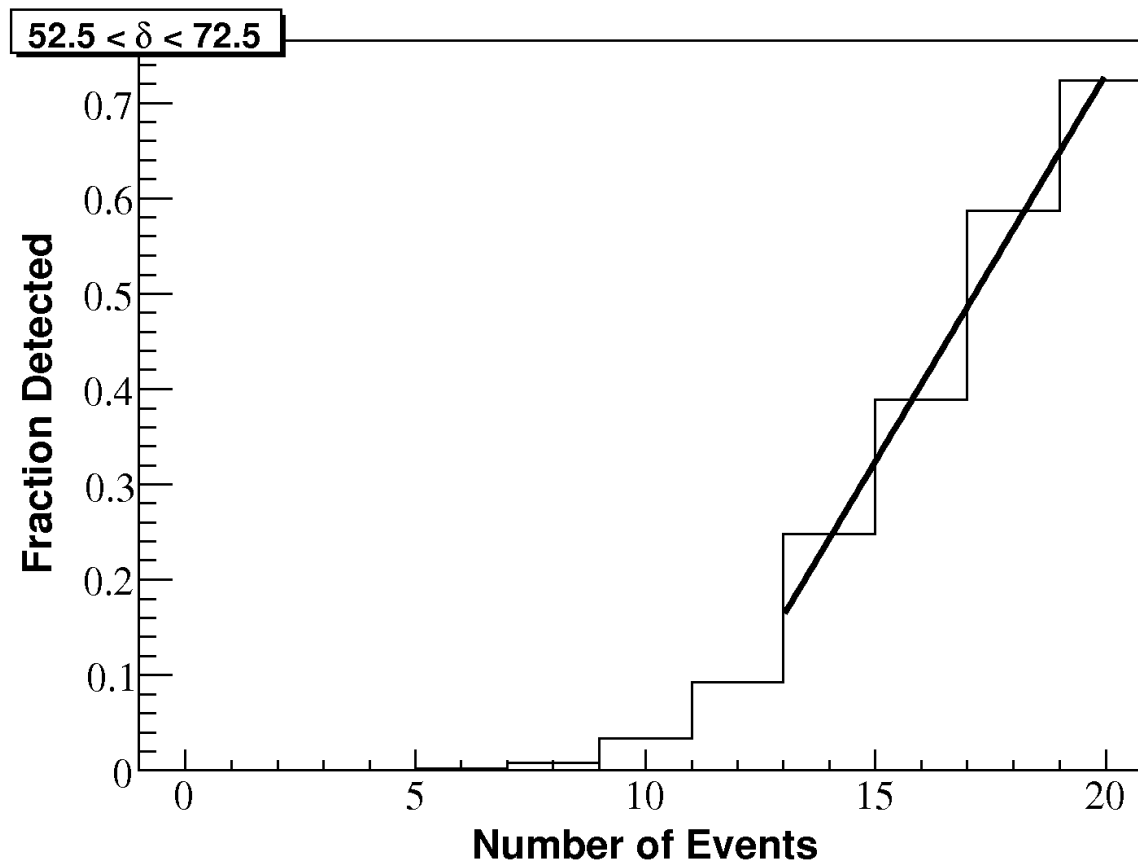
FDR Data Challenge Results



- 0.3% of sky maps contain false discoveries (Expected: < 1%)

The FDR Data Challenge

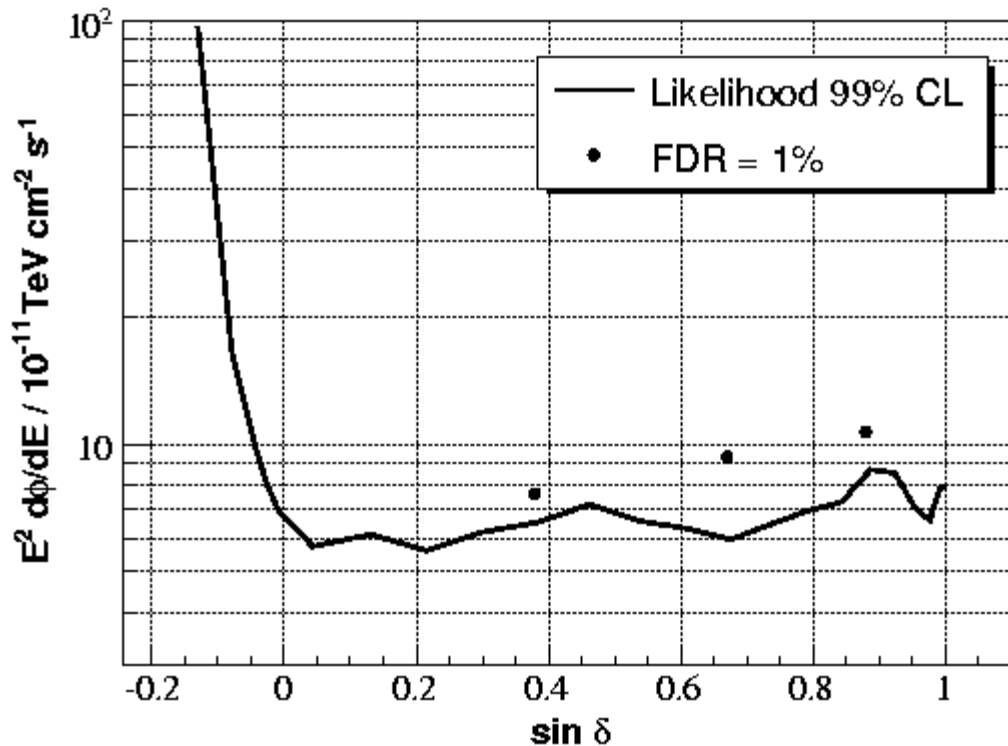
- Detection depends on simulated source strength



- Requires ~17 added signal events to have a 50% chance of detecting a source.

The FDR Data Challenge

- Convert 50% discovery probability event thresholds to flux and compare against the baseline method



- FDR data challenge caught a problem:
 - Data had been assumed to be pure atmospheric neutrinos, but actually contains 5% cosmic ray muon background
 - Tighter cuts remove the background and make the method consistent

Future Plans

- Use flasher events as a cascade data challenge
- Further analysis of the moon shadow
 - Systematic pointing errors
 - Measure muon point spread function
- Strongly encourage data challenges for most analyses when feasible

Summary

- IceCube is committed to using data challenges to ensure understanding of analyses and physics results
- Observation of the moon shadow has provided an end-to-end system test for muons
- FDR data challenge is a successful example of an analysis-level data challenge
- Flasher events will provide an end-to-end test for cascades