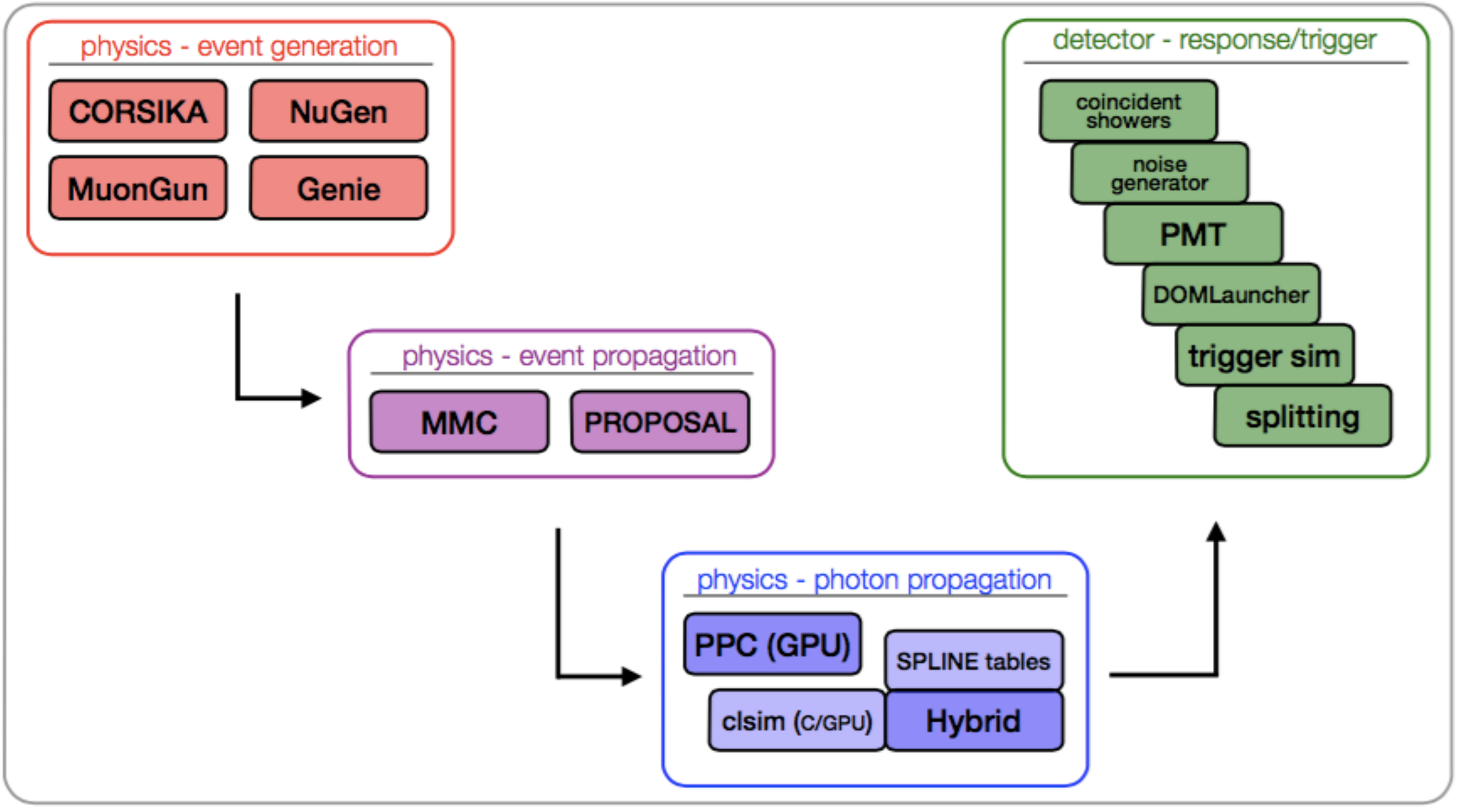
IceCube Simulation Software and Production

Alex Olivas
University of Maryland
(for the IceSim and SimProd Groups)

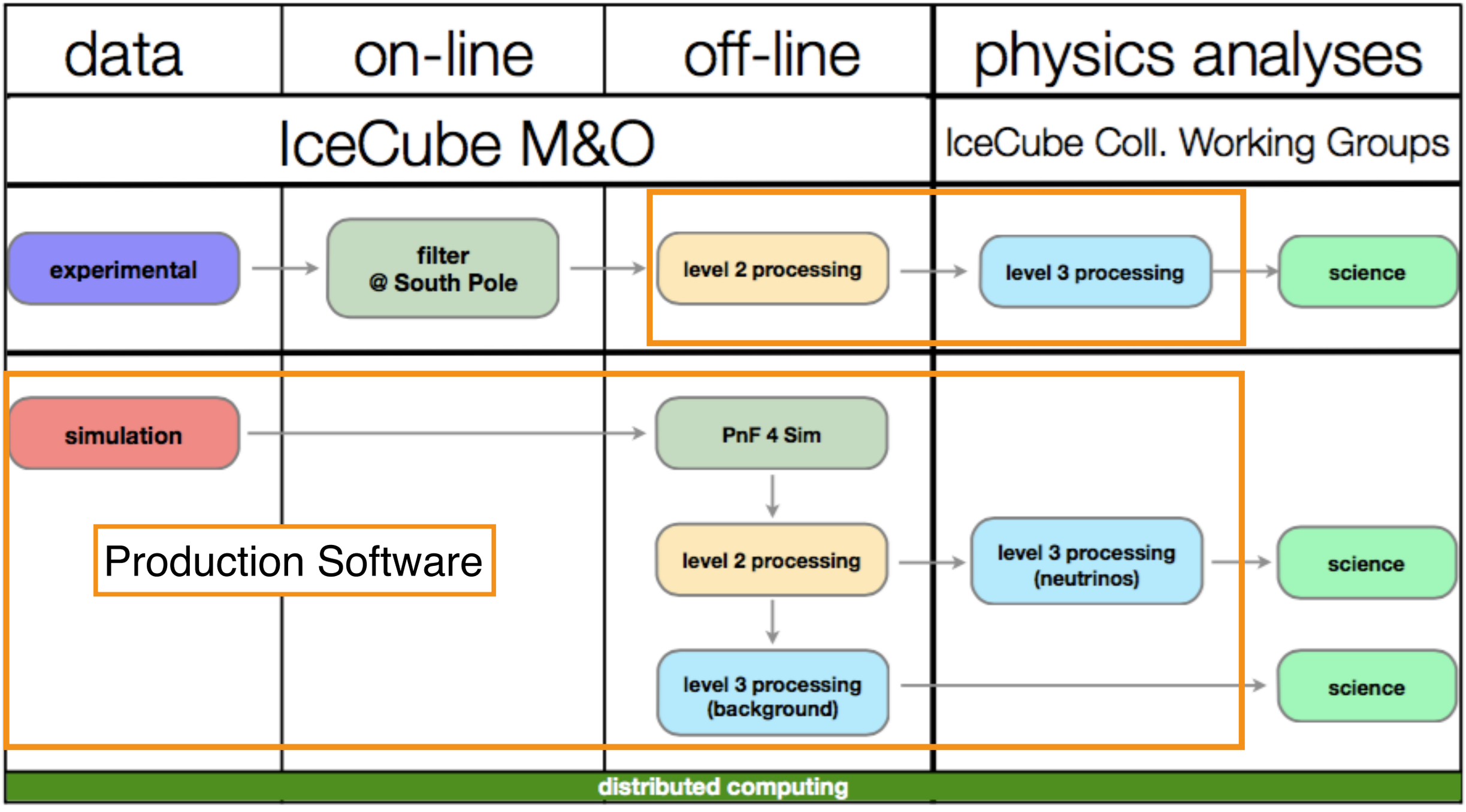
Talk Overview

- IceCube Physics Software Challenges
  - Impacts on analyses
  - Improving software through training
- Simulation Production Challenges
  - Simulation Data Usage and Requirements
  - GPU Simulations - Successes and Limitations
  - Generation of Systematic Datasets

simulation chain
modular software sequence



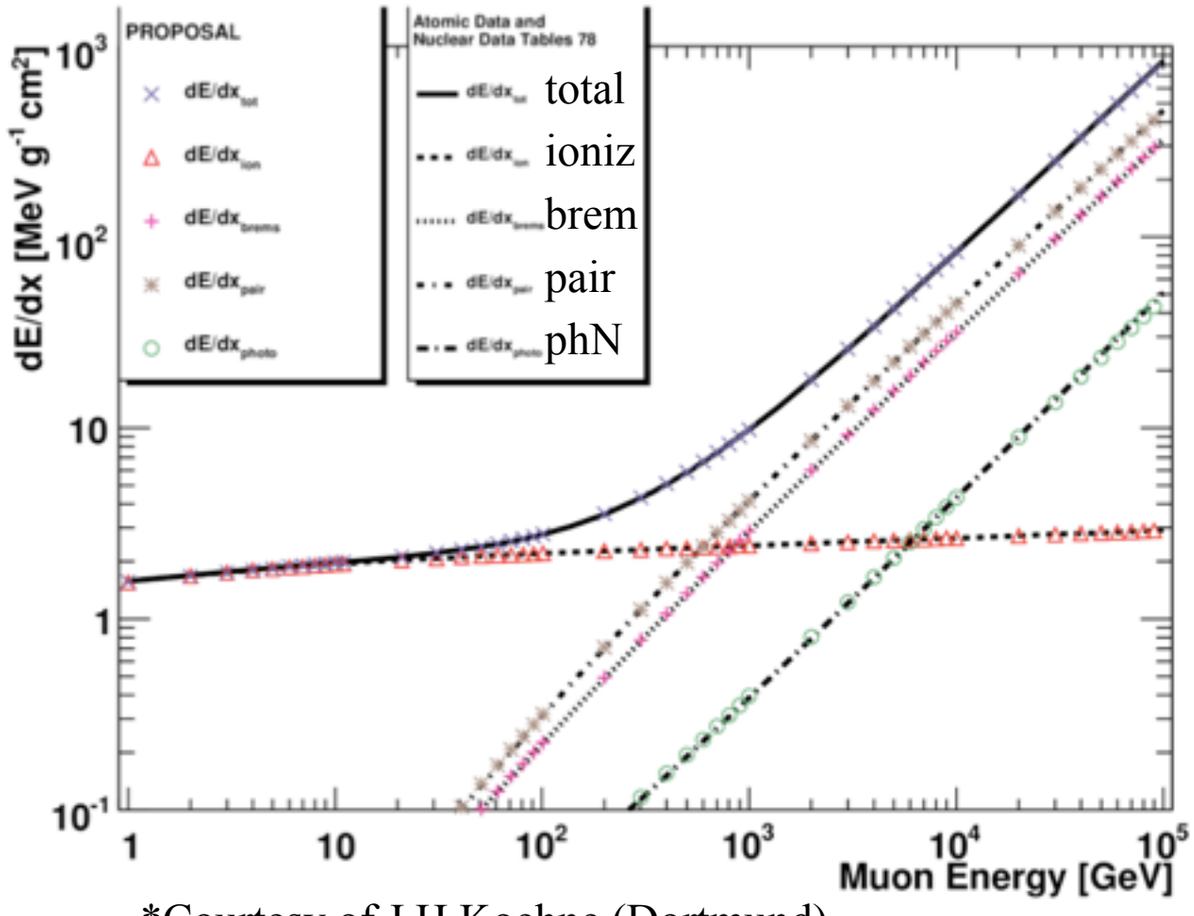
simulation chain
& simulation data processing



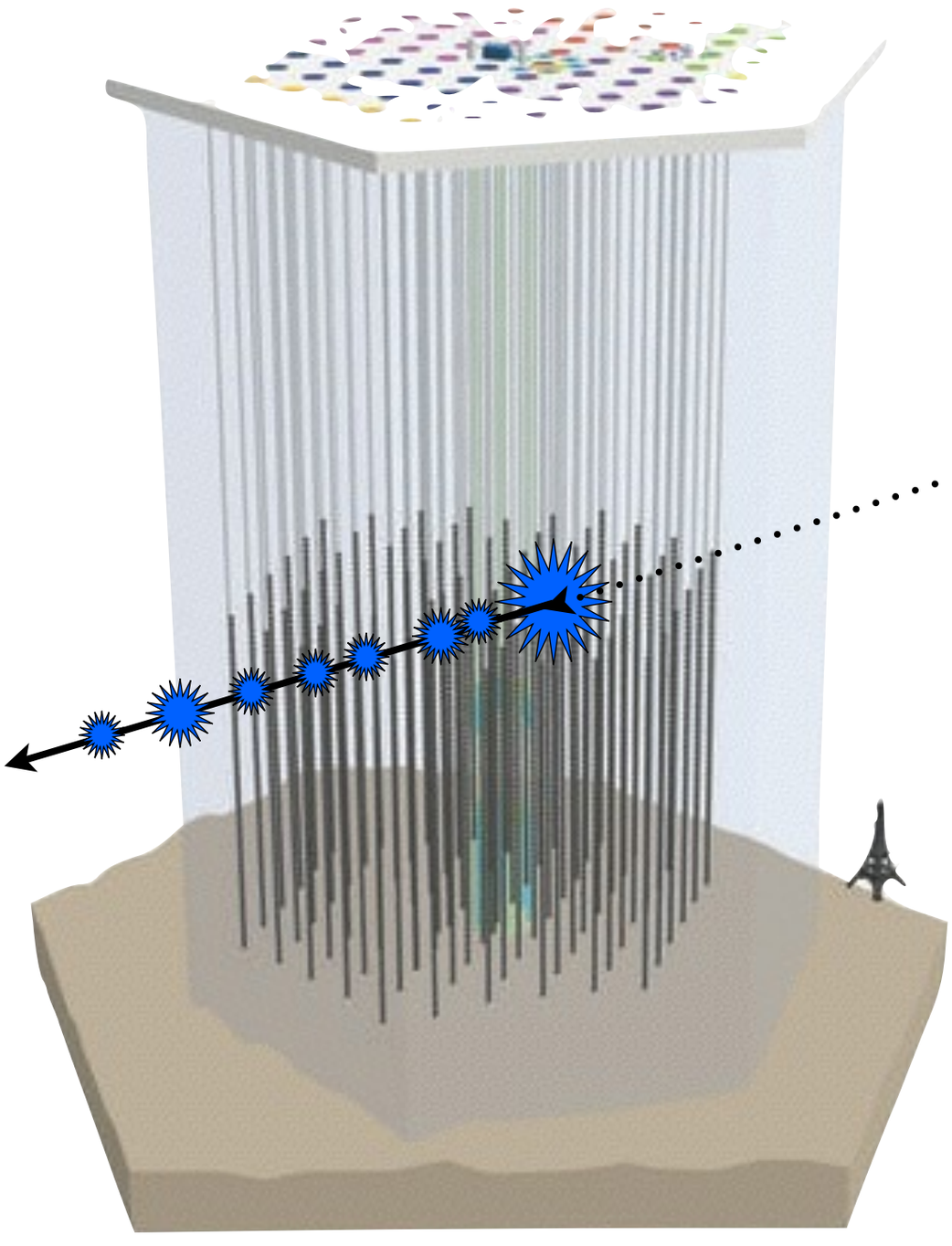
Lepton Propagation

PROPOSAL (C++) - **PR**opagator w/ **O**ptimal **P**recision and **O**ptimized **S**peed for **A**ll **L**eptons (Dortmund and Bochum)

- Largest contribution to Monte Carlo data
- Two schemes to reduce data size (J. van Santen - UW Madison)
  - Combining cascades (30% reduction)
  - **Storing RNG state (on par with data - 3x reduction)**
  - **Allows for more efficient use of disk space**
  - **Starting in 2016 production**



*Courtesy of J.H.Koehne (Dortmund)



Background Generation

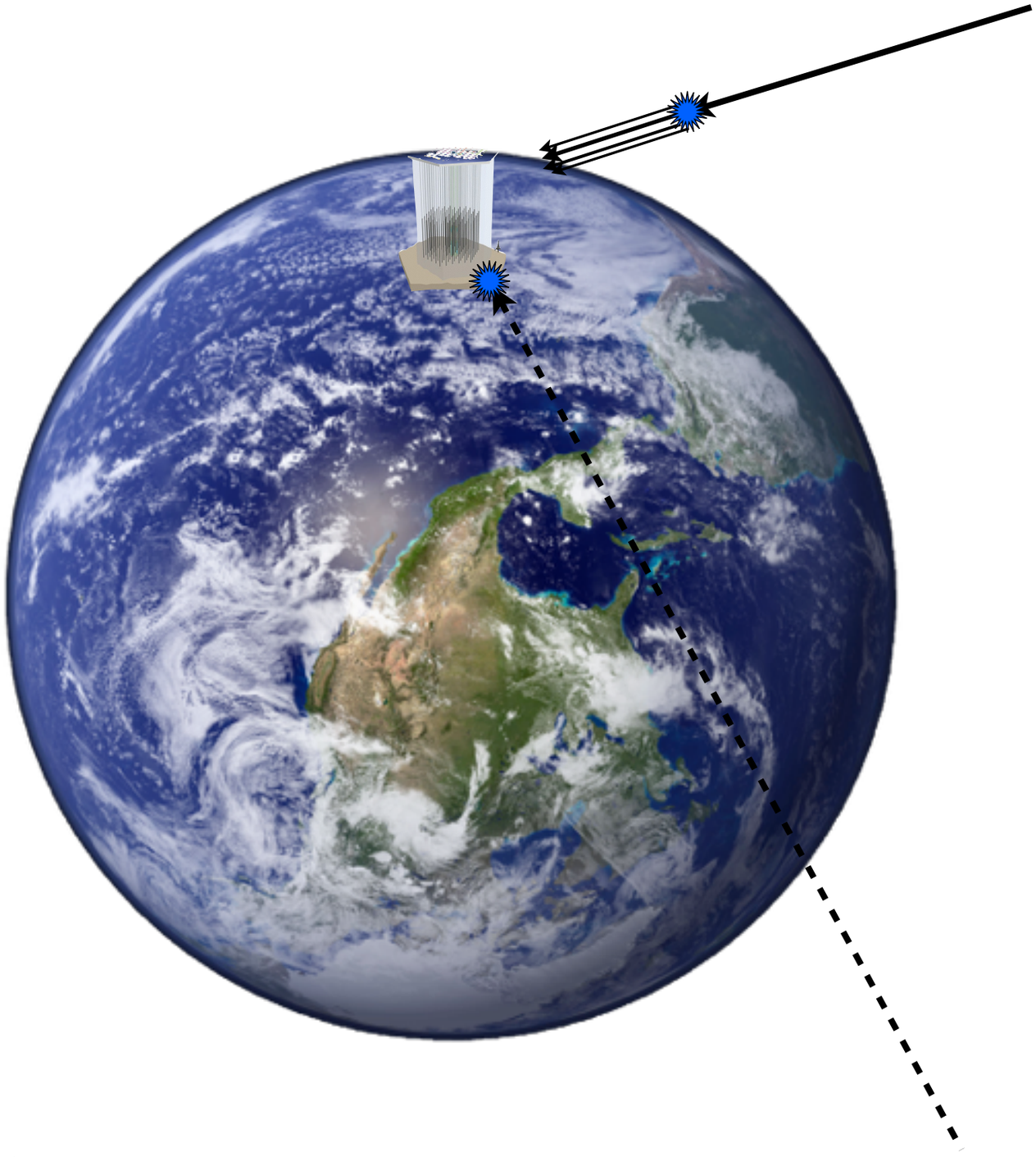
Cosmic Ray Showers

CORSIKA

- Full shower simulation
- Resource intensive (both disk and CPU)
- **Very difficult to produce enough background to meet several analyses needs, such as diffuse, neutrino oscillations, and exotics.**

MuonGun - 2x speed-up

- Samples single muons from fits to CORSIKA distributions
- Somewhat reasonable estimation of the background for some analyses
- Few people trust it as a perfect replacement for CORSIKA
- Doesn't include muon bundles, which limits its use for many analysis.



Background Generation

IceCube	primary target (2 livetime-months/det-configuration)				secondary target (1 livetime-year/det-configuration)			
	CPU (days)	GPU (days)	L2 size (TB)	L3 size (TB)	CPU (days)	GPU (days)	L2 size (TB)	L3 size (TB)
CORSIKA	134,000	19,000	58	17	802,000	112,000	348	102

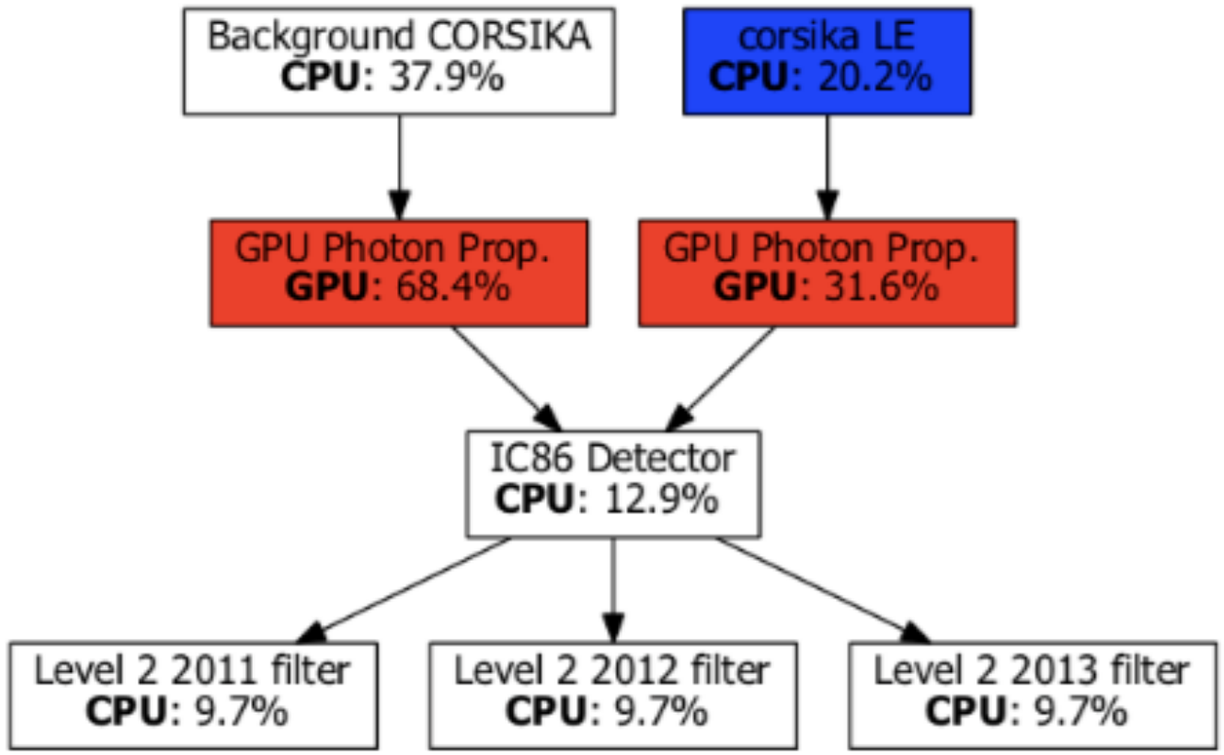
Site	GPUs	Production Status
UW-Madison	158	In production
U. Alberta	43	In production
DESY-Zeuthen	58	Testing
U. Maryland	96	Deploying
NERSC	32	Testing
RWTH-Aachen	42	Pending
Mainz	48	Pending
Wuppertal	48	Pending
Dortmund	?	Pending
Copenhagen	?	Pending

	real time livetime production	
	CPU cores	GPU cores
required	2,200	310

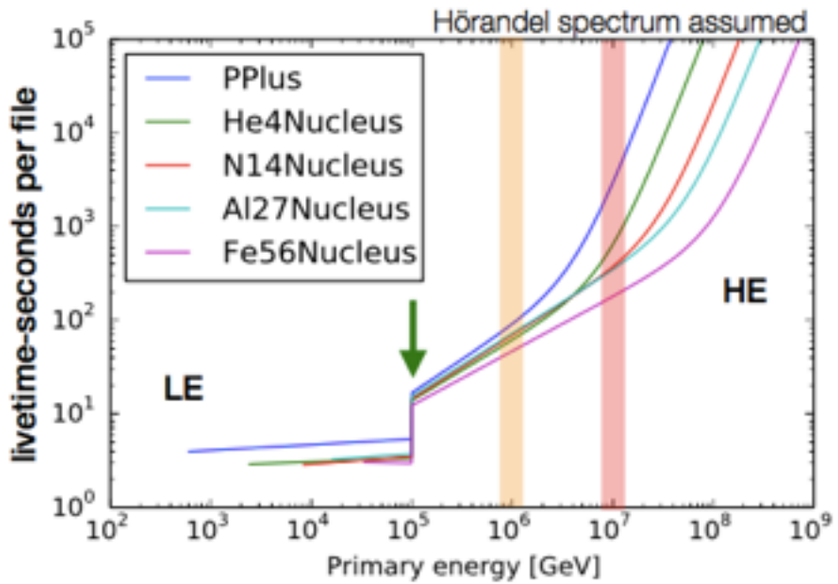
IceCube	target (per det-configuration)		
	CPU (days)	GPU (days)	L2 size (TB)
neutrinos	27,000	7,000	10

	target (12 months / det-config)	
	CPU cores	GPU cores
required	74	20

simulation plan
optimization of cpu/gpu resources



- generate shower events to **photo-electron level** and split into IC86-1,2,3

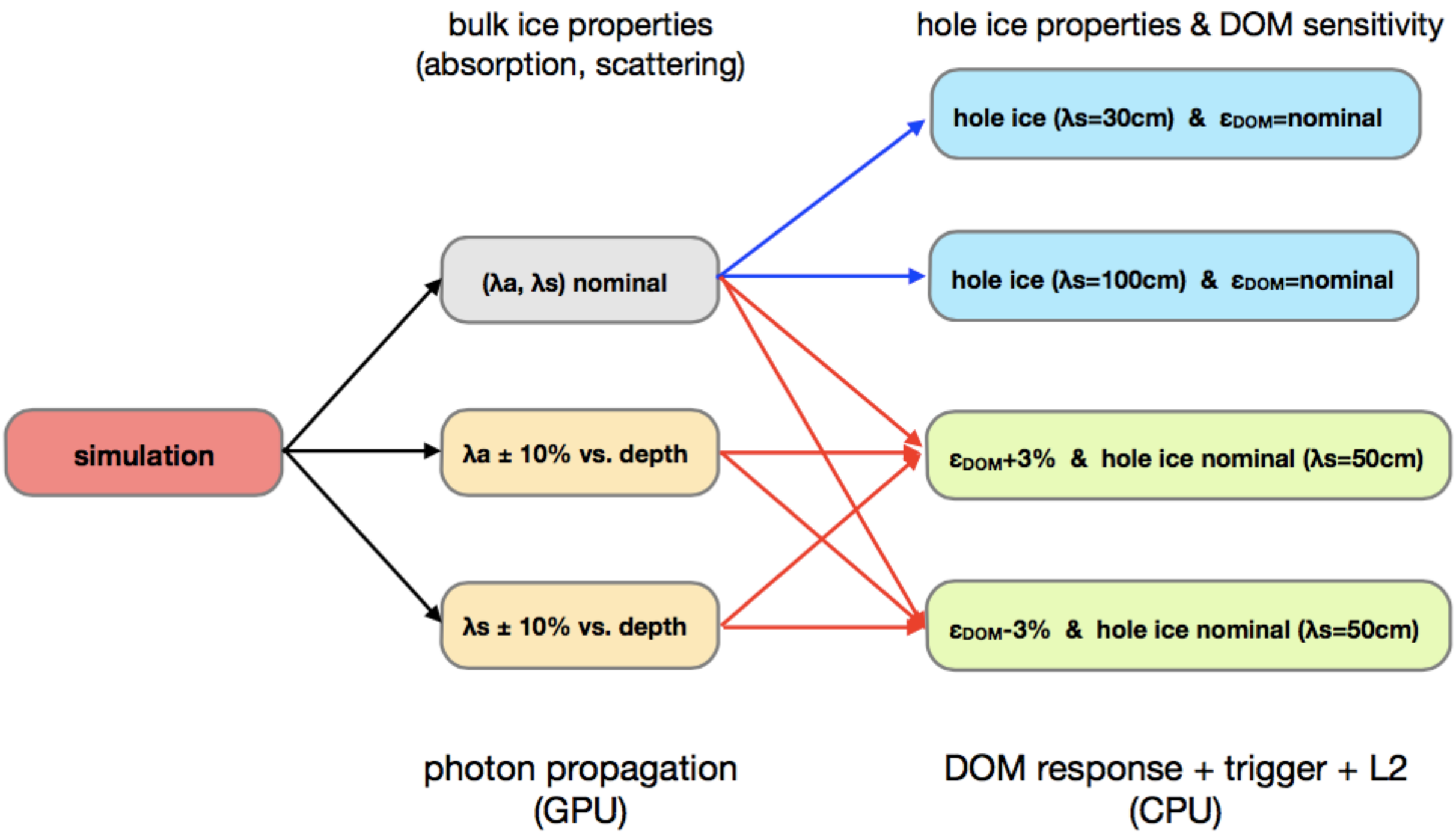


lose statistical independence of events

penalizes science for IC86-1,2,3

Need 516 GPU cores to get 1 year livetime for 4 seasons in 4 years

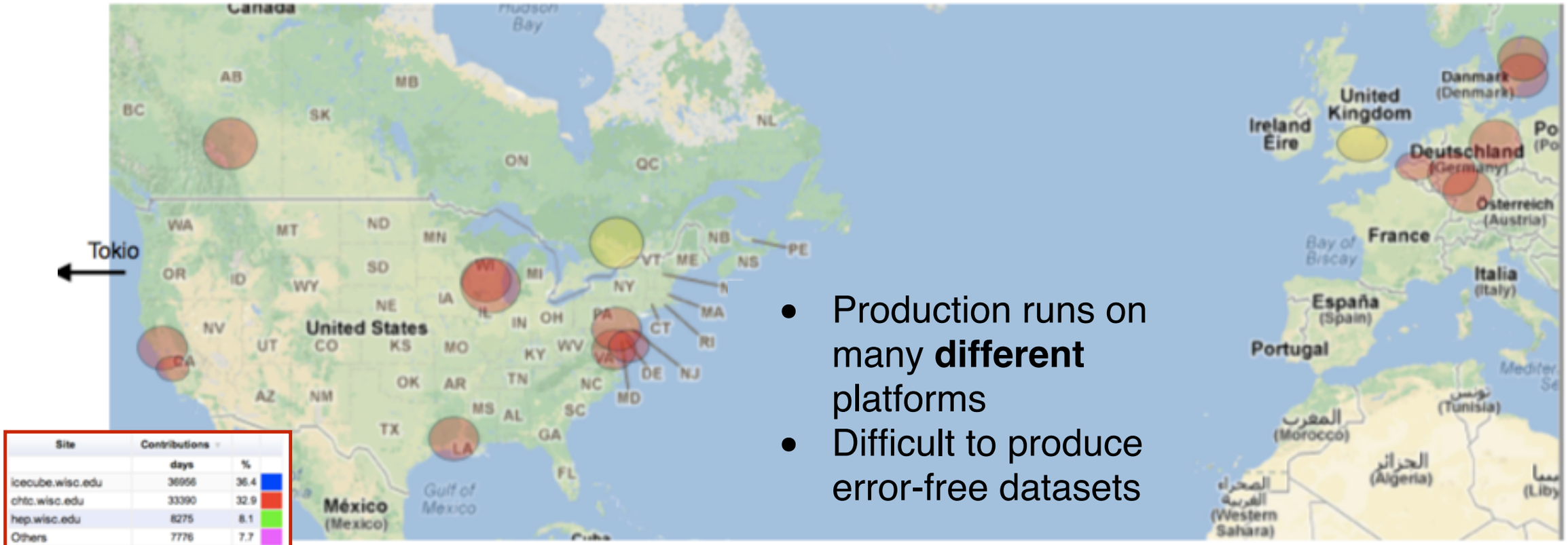
Systematic Datasets



simulation plan
systematics datasets

- to study effects of systematic variations of physics parameters
  - high energy **hadronic interaction models** in CORSIKA
  - **neutrino cross section**: CSMS in neutrino-generator
  - **neutrino production with CORSIKA**: interaction models / primary composition
  - **MuonGun** parameterized simulation for cascade systematics
  - **significant impact on manpower & computing requirements**

Simulation Production

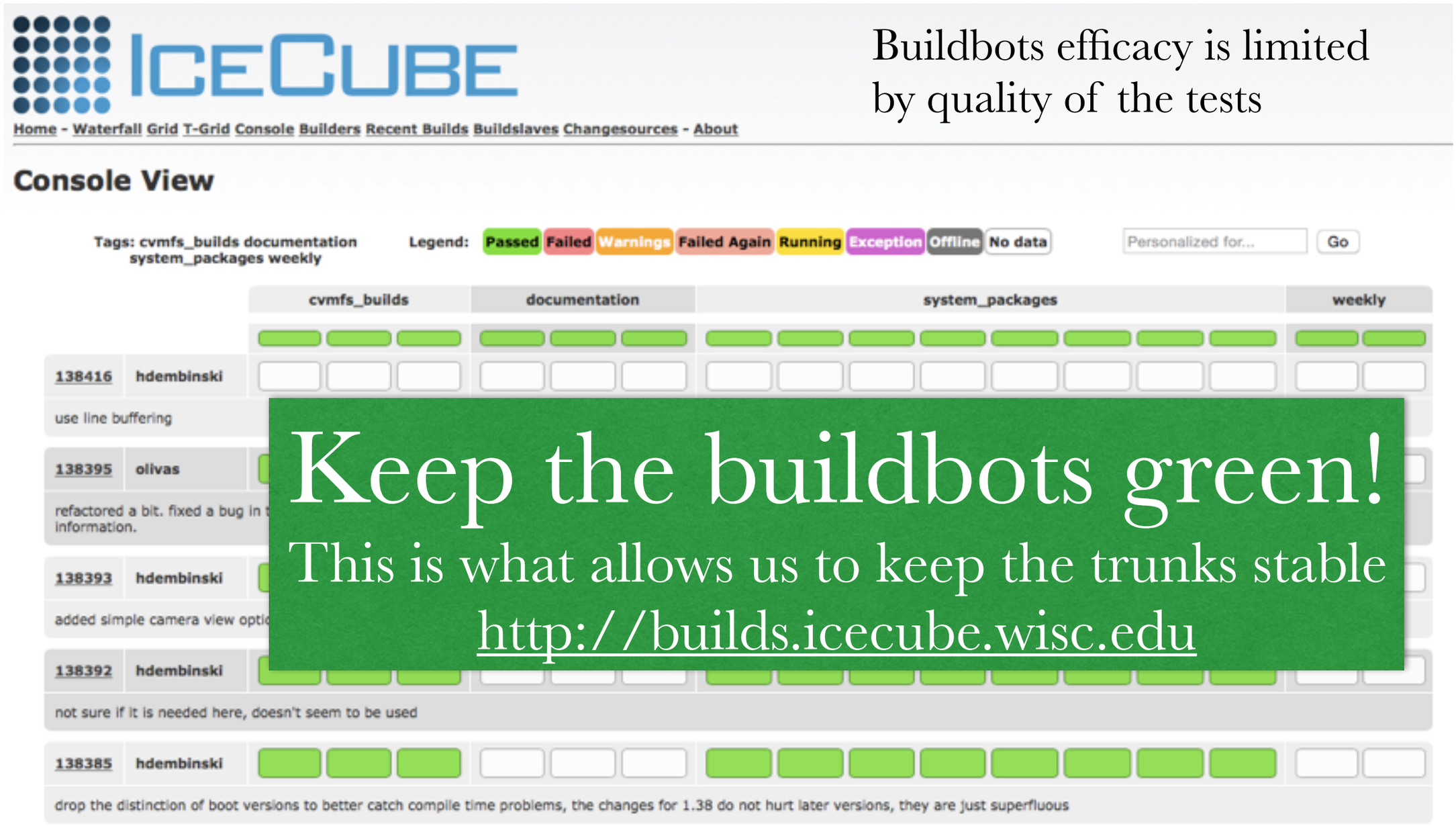


- Production runs on many **different** platforms
- Difficult to produce error-free datasets

Site	Contributions	
	days	%
icecube.wisc.edu	36956	36.4
chtc.wisc.edu	33390	32.9
hep.wisc.edu	8275	8.1
Others	7776	7.7
cs.wisc.edu	2330	2.3
sandhills.hcc.unl.edu	2283	2.3
t2.ucsd.edu	108	0.1
nut.bu.edu	70	0.1
CMSAF.MIT.EDU	42	0
usatlas.bnl.gov	11	0
batlab.org	5	0
genetics.wisc.edu	3	0
uchicago.edu	2	0
campuscluster.illinois.edu	624	0.6
lmcg.wisc.edu	571	0.6
private.pa.umd.edu	545	0.5
crane.hcc.unl.edu	347	0.3
tusker.hcc.unl.edu	343	0.3
red.hcc.unl.edu	207	0.2
ifh.de	192	0.2
phys.uconn.edu	1968	1.9
unl.edu	1303	1.3
rcac.purdue.edu	1180	1.2
agit2.org	1048	1
mwt2.org	783	0.8

Aachen (Germany)	Mainz (Germany)	NPX4 - Madison (USA)
WestGrid - Alberta (Canada)	OSG (USA)	Ghent (Belgium)
Bartol (USA)	PSU - Pennsylvania (USA)	EGI (Germany)
Brussels (Belgium)	SweGrid (Sweden)	UC-Irvine (USA)
DESY (Germany)	Tokyo (Japan)	Bochum (Germany)
Dortmund (Germany)	UMD - Maryland (USA)	Wuppertal (Germany)
LONI - Southern U. (USA)	CHTC/GLOW UW-Madison (USA)	PDSF/Carver/Dirac LBNL (USA)

Software Development Tools



Buildbots efficacy is limited by quality of the tests

Keep the buildbots green!

This is what allows us to keep the trunks stable

http://builds.icecube.wisc.edu

Coming Soon : Full chain tests of simulation through L2 and L3 will part of weekly stress tests.

Software Challenges

IceCube has a diverse physics program.

- Analyses require significant resources dedicated to systematic datasets.
- Also want to investigate future extensions.

Requires flexible and extensible software, which is often not what you get from students and postdocs who are not only have little training in the core languages, but have no training or experience in design and software engineering.

If we're going to continue with the model where a significant fraction of production code is developed by students and postdocs, we need to invest in training to ensure efficient use of our computing resources.

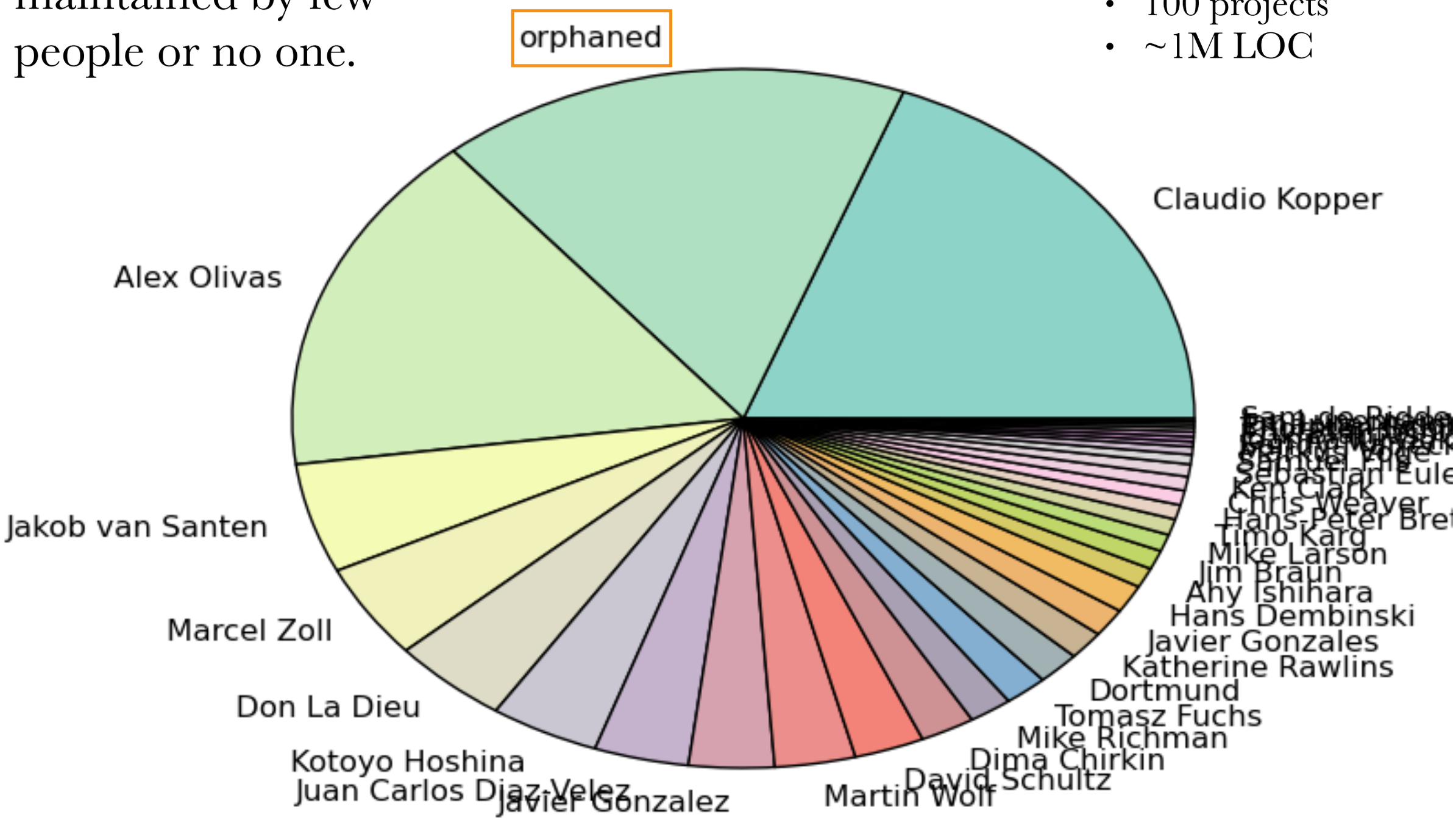
State of Physics Software

Spring 2015

Large fraction of production code maintained by few people or no one.

Production Code

- 100 projects
- ~1M LOC



IceCube Software Strike Team

Problem

We have a lot of code to maintain by physicists who traditionally have little to no training in software engineering.

Solution

Invest in training collaborators and form a team who's service to IceCube is to work on software problems.

Organize monthly code sprints to work on the most urgent problems.

Bootcamps

- June 15-19, 2015 - Strike Team Training covering advanced C++ and python.
- October 12th - Intro to IceCube Software
- October 17th-18th - C++ Best Practices

Code Sprints

- July - Simulation Production Testing I
- August - Simulation Production Testing II
- September - Simulation Production Testing III
- October - Simulation Release Preparation
- November - Reconstruction Release Preparation

Conclusion

Software and production problems impacting physics :

- Low statistics in CORSIKA background.
- Analyzers have to work around the low statistics problem

Buggy simulation datasets

- Wastes computing resources
- Wastes students and postdocs time, potentially delaying graduation and/or publication of results.