Antimatter

Using High Throughput Computing to Study Very Rare Processes

Anirvan Shukla
Department of Physics, University of Hawaii at Manoa
OSG AHM 2020, September 1

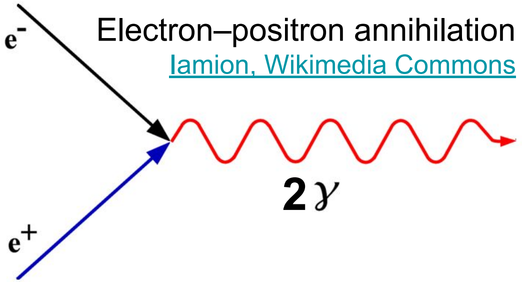
What is antimatter?

All matter is made of particles like protons, electrons etc.

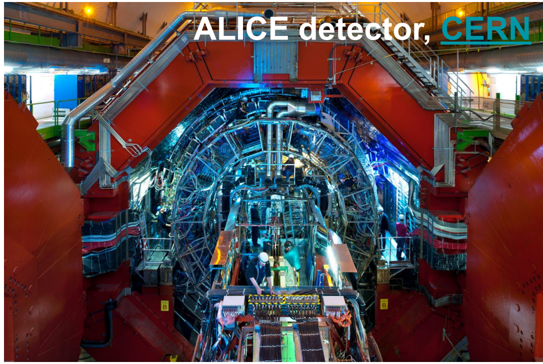
- Each elementary particle has a corresponding antiparticle.
- They have same mass, but differ in other properties like electric charge etc.
- If a particle collides with its antiparticle, there is mutual annihilation.
- Natural sources of antimatter are cosmic rays and some types of radioactive decays.

They can be produced artificially as well!

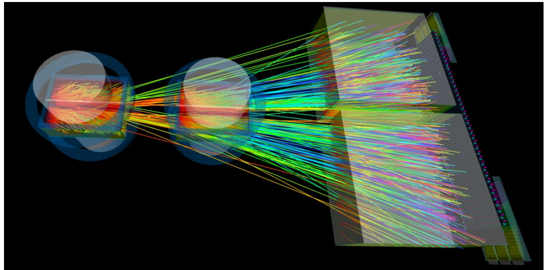
- Experiments like ALICE, NA61 etc. at CERN, Geneva, and other high-energy particle accelerators produce measurable amounts of antimatter for studying the fundamental theories of our universe.
- Antimatter in cosmic rays can be excellent messengers for dark matter as well.



Electron–positron annihilation
Iamion, Wikimedia Commons



ALICE detector, CERN



An example collision recorded by the NA61/SHINE experiment.

What can antimatter tell us about dark matter?

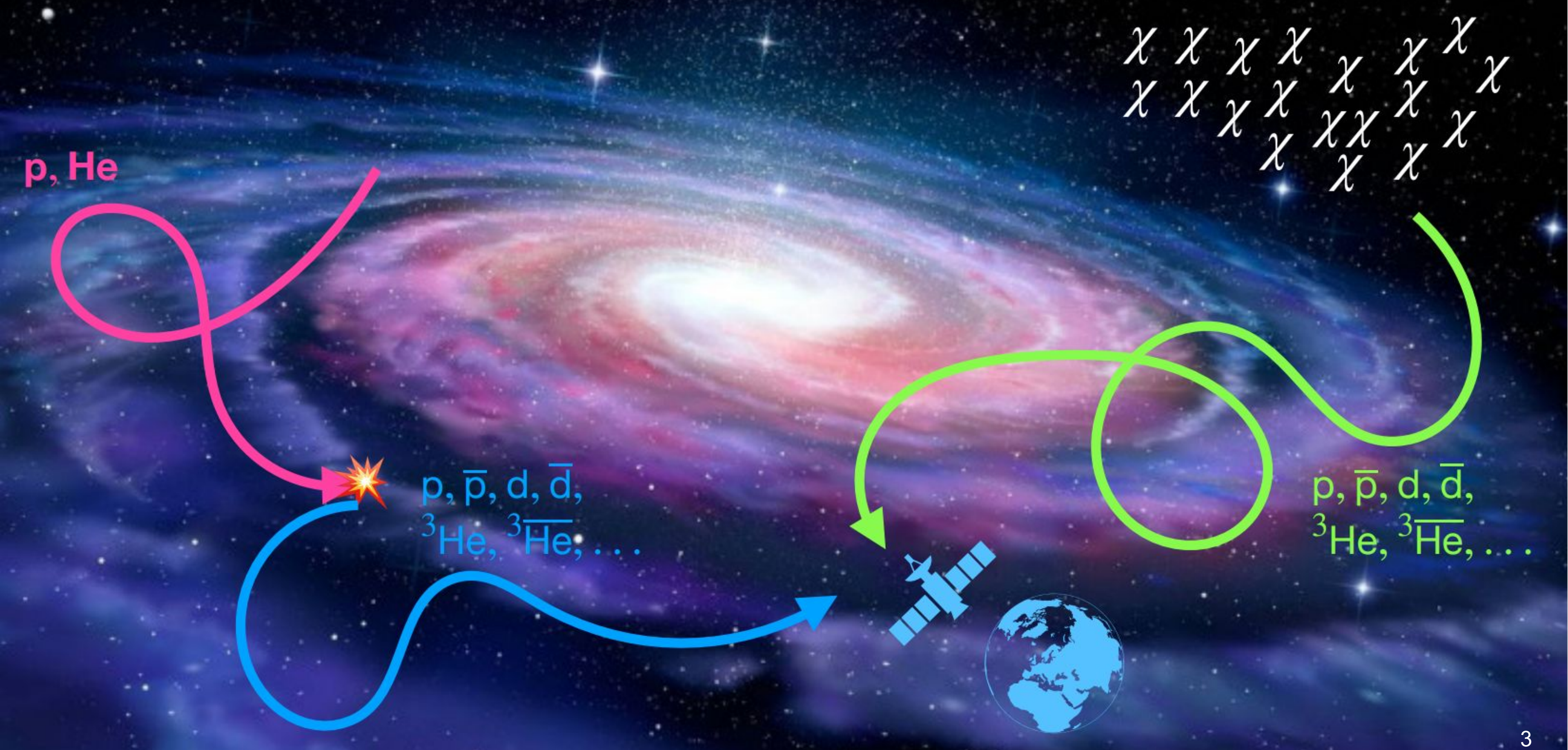


Image from S. Hornung, ICHEP 2020

What are the physics challenges?

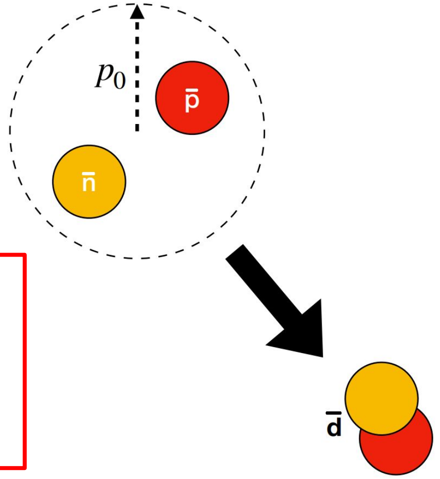


Image from S. Hornung, ICHEP 2020

Detection of cosmic-ray antideuterons is a potential breakthrough approach for the identification of dark matter.

However, there are astrophysical and particle physics uncertainties.

- Mechanism of dark matter annihilation or decay
- Antimatter production
- Galactic propagation/solar modulation
- Geomagnetic deflection
- Atmospheric interactions
- Interactions within the detector

How to simulate antimatter production?

- Monte Carlo (MC) methods are used to simulate the collision of a fast-moving proton with a proton at rest.
- This simulates the interaction of abundant cosmic-ray protons with the very low-density interstellar gas.
- Different physics models have been developed to match experimental data.
- Typical analysis involves hundreds of million of simulated collision “events”.
- Since each event is independent, these simulation are a perfect match for HTC.
- Since production of heavy antimatter is extremely rare, these types of large-scale simulations were not feasible even just a few years ago.

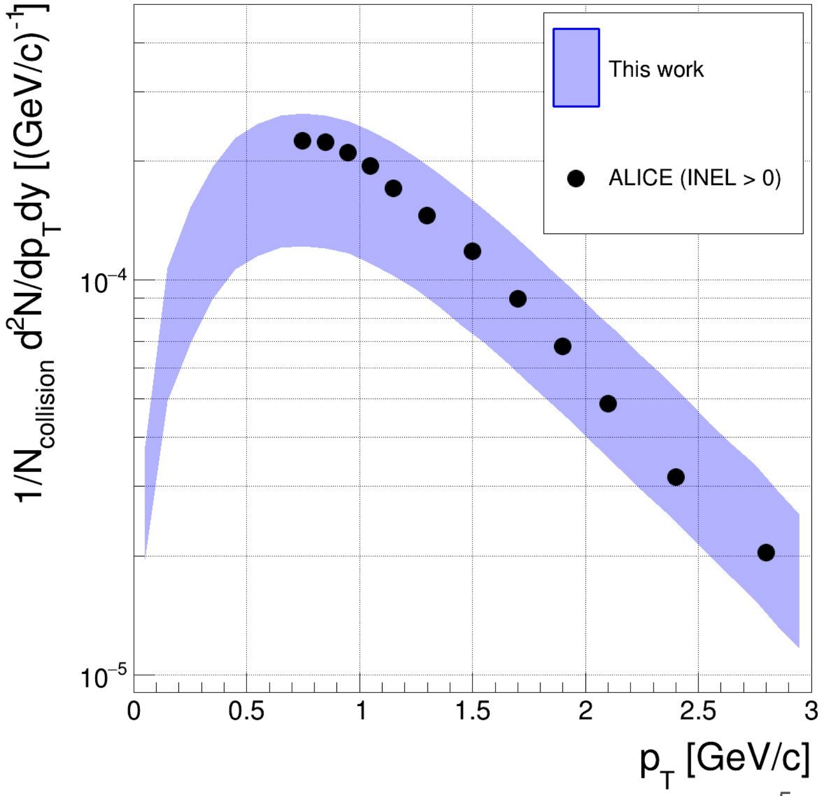


Image from A. Shukla, arXiv:2006.12707

How rare is the production of heavy antimatter?

Why does this need a cluster, and not a regular computer?

- 1 antihelium particle is produced in every 10 billion - 1 trillion events.
- 1 million events need ~1 CPU-hour.
- So to simulate 1 antihelium, need 10,000-1 million CPU hours.
- Since the analysis is statistical in nature, need at least 1000 antihelium to even begin with predictions.
- ~100 trillion events needed!

Data storage challenges

- 10 million events stored in 1-10 GB output file.
- With 100 trillion events, need 10-100 million GB of storage.

Open Science Grid & User School 2016

- Access to my existing computing resources at CERN was not enough for this project.
- OSG User School made me familiar with the OSG, and the software environment.
- The various pre-installed software modules available on all OSG compute nodes made life much easier.
- The training during the User School, and subsequent prompt online help via email was very useful in deploying my project.

OSG workflow

- Submit file
  - Handle input files to transfer to compute node
  - Launches bash wrapper script
- Bash wrapper script
  - Load software modules
  - Move files, temporary directories etc
  - Launch python script
- Python wrapper script
  - Controls input parameters for C++ programs
  - Much better string handling than bash
  - Launches multiple C++ program in sequence
  - Transfers final histogram file to server in Hawaii
  - Performs clean up, and exits job.
- In hindsight: Use containers (docker/singularity)

Finding solutions to the storage constraints

How to handle the large (100 million GB of output files)?

- Breaking simulations into chunks of 10 million simulated events.
  - Output file of ~10 GB.
- Analyzing the final output file within the job.
  - Extracting all possible data of interest into hundreds of histograms.
  - Lose detailed information at the individual particle level - a compromise.
  - Drastically reducing file size: 10 GB output file -> 100 MB histogram file.
- No local storage!
  - Deleting the large output file, and transferring the 100 MB histogram file from compute node to our server in Hawaii.
- Still end up with 10 million files of 100 MB size = 1000 TB!
  - However, histograms are excellent for “adding” up.
  - Final size of simulation output: ~2 GB!
  - This approach was computationally expensive, but the final storage required was small.
  - Ideal for using multiple clusters.

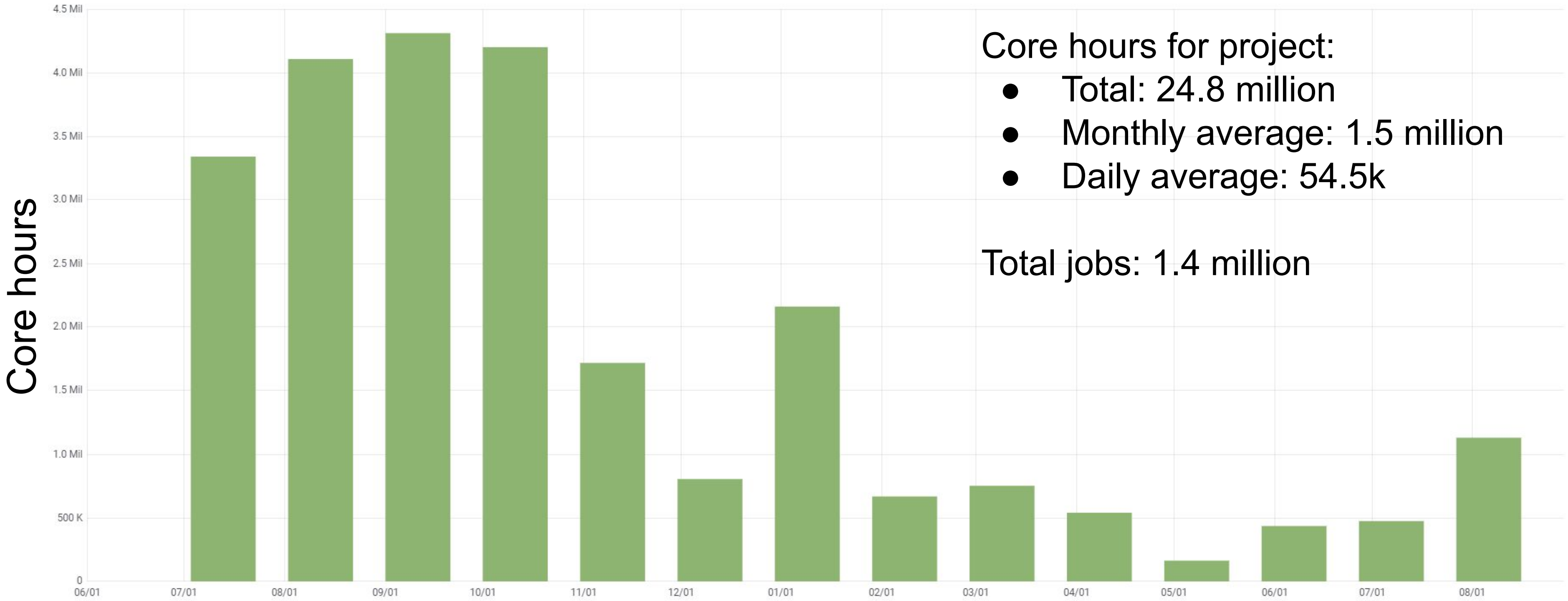
OSG job statistics

- After performing tests, I switched off saving the log files.
  - With ~ 1 millions jobs, logs will take up all my user storage.
  - Potential to slow down the submit node.
- Typical resourced requested
  - CPU: 1
  - Memory: 1-2 GB
  - Disk space (10-15 GB).
- Job failure rate was low, about ~1%-5%.
- The compute nodes were widely distributed across the US. The table on the right shows the top 20 facilities for my jobs.

Total core hours by facility

SU ITS	10.93 Mil
IRISHEP-SSL-UCHICAGO	6.28 Mil
IIT - Illinois Institute of Technology	2.439 Mil
UColorado_HEP	719 K
SLATE-K8S-UCHICAGO	410 K
BNL ATLAS Tier1	300 K
Utah-SLATE-Notchpeak	293 K
UCSD CMS Tier2	265 K
NWICG_NDCMS	259 K
ASU Research Computing	234 K
MWT2 ATLAS UC	234 K
Utah-SLATE-Lonepeak	233 K
UConn-OSG	208 K
ICC-SLATE-HTC	170 K
New Mexico State AggieGrid	165 K
Utah-SLATE-Kingspeak	123.1 K
New Mexico State Discovery	116.6 K
OU ATLAS	106.0 K
FSU_HNPGRID	97.7 K
GLOW	95.5 K

OSG overall statistics



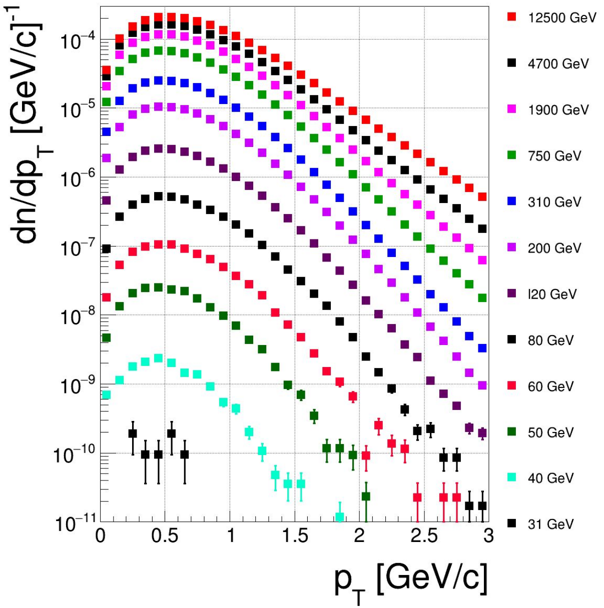
Core hours for project:

- Total: 24.8 million
- Monthly average: 1.5 million
- Daily average: 54.5k

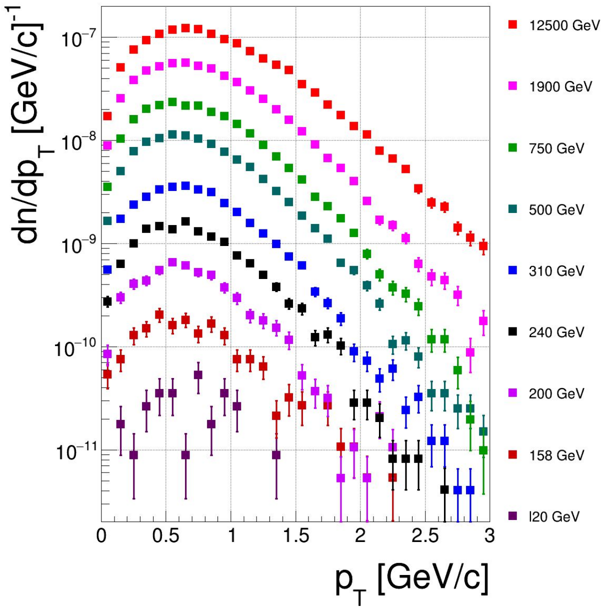
Total jobs: 1.4 million

Images from A. Shukla, arXiv:2006.12707

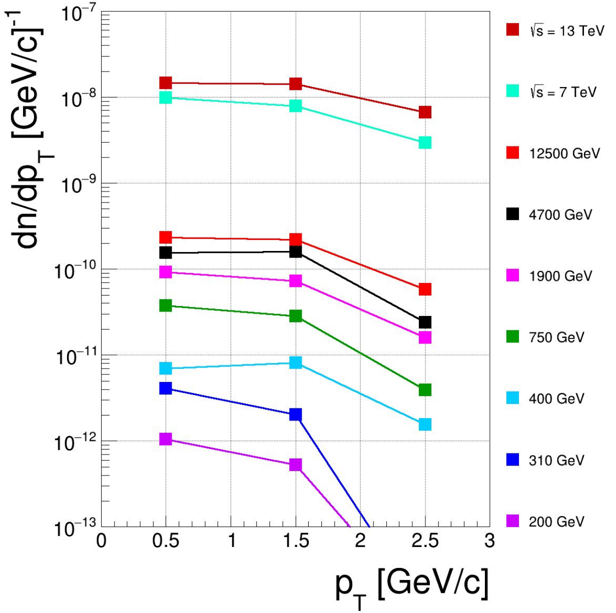
A look at some results



~1 billion antideuterons simulated @ ~40 per million events.



~500k antihelium-3 simulated @ ~17 per billion events.



1,698 antihelium-4 simulated @ ~58 per trillion events.

- First time these spectra could be calculated. Already well received by the large ALICE community.
- Accepted to be published in Phys. Rev. D. Available online at https://arxiv.org/abs/2006.12707.

Next Steps

- Computational resources are never enough!
- There is always demand for a larger cluster, more storage, more CPUs etc.
- Hopefully, the OSG will continue to be a critical resource in our future projects.

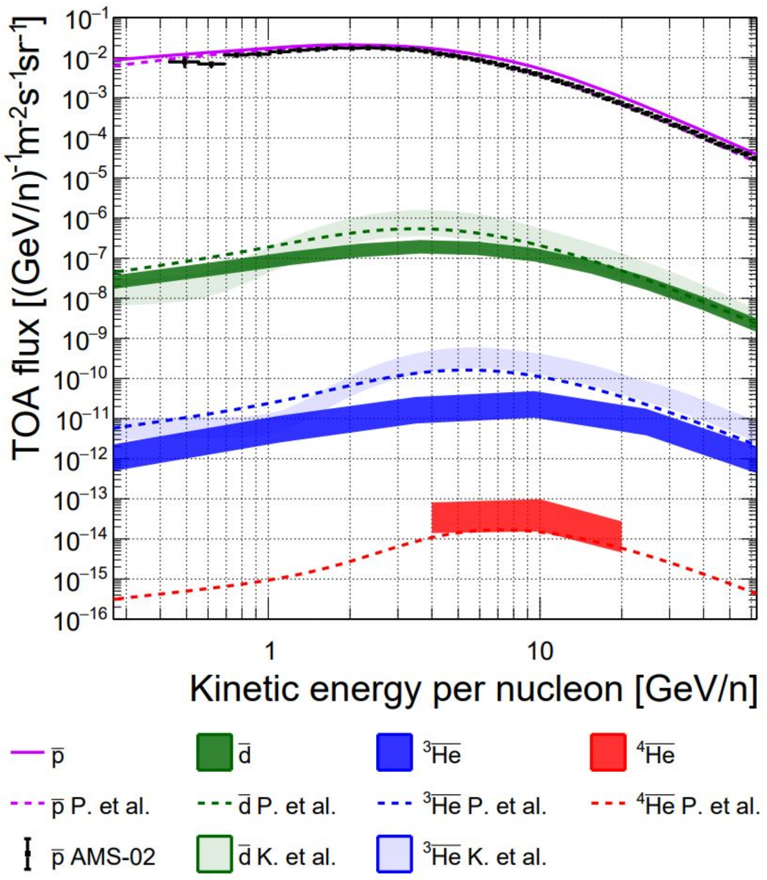


Image from A. Shukla, arXiv:2006.12707