FERMILAB-SLIDES-23-196-STUDENT

Searching Applications of Benford's Law to Investigate Beam Jitter

Presenter : Dominic Fallico, Geneva High School
Mentor: Dr. Arun Saini, APS-TD
TRAC
8/3/2023

Outline

- Introduction to Benford’s Law
  - Definition and Origin
  - Applications/Uses
- Application for particle accelerators
  - PIP2-IT
  - Beam jitters/noise and distribution
- Methodology
  - Generate synthetic and real distributions using Python
  - Extracting First Digits and Analyze
  - Results, Findings, & Shortcomings
- Conclusions

1. Introduction
   - Motivation
   - Benford's Law Definition and Application
2. Methodology
3. Results
4. Conclusion

Motivation and Objective

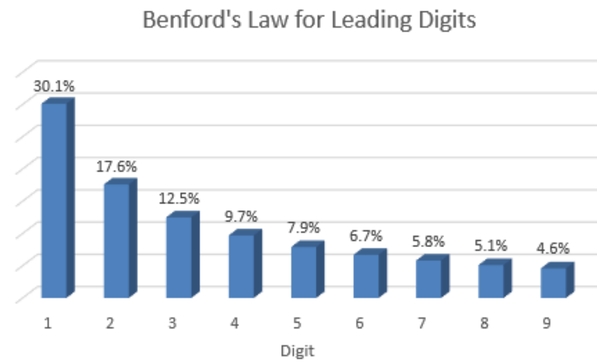
- Explore Applications of Benford's Law to predict jitter and noise in the beam distribution

- Apply Benford's Law on beam measurement data of PIP2IT beamline.

Benford's Law

Origin - Simon Newcomb and worn logarithmic tables → $P_{10} = \log_{10}\left(1 + \frac{1}{d}\right)$

Frank Benford rediscovers in 1938 and extends it to → $P_d = \log_b\left(1 + \frac{1}{d}\right)$



Benford tests it himself

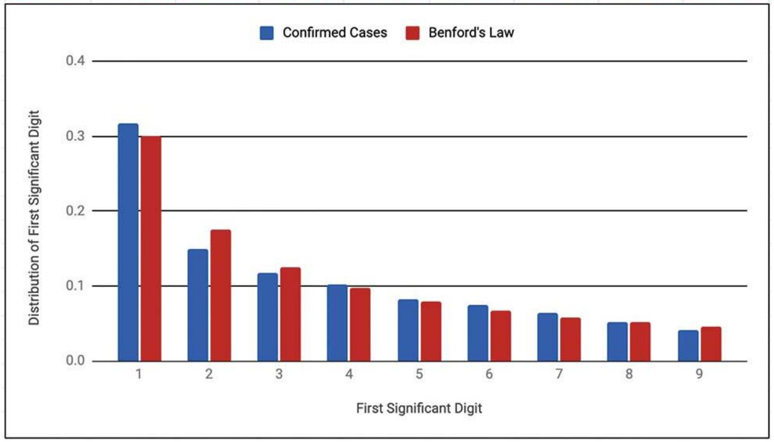
First Digit Distributions Expressed as Percentages for Various Physical Data Sets[a]

	First Digit Frequencies									Number of Values in Each Data Set	Dynamic Range of the Data (max/min)
	1	2	3	4	5	6	7	8	9		
P_D	30.1	17.6	12.49	9.69	7.92	6.69	5.80	5.12	4.58		
Geomagnetic Field	28.9	17.7	13.3	9.4	8.1	6.9	6.1	5.1	4.5	36512	10^{10}
Geomagnetic reversals	32.3	19.4	13.9	11.8	5.3	4.3	3.2	5.4	4.3	93	10^{3}
Seismic wavespeeds below SW-Pacific	30.0	17.6	13.3	9.8	7.9	6.4	5.6	4.89	4.47	423776	10^{6}
Earth's gravity	33.0	16.6	11.2	8.5	7.5	6.7	5.94	5.57	5.03	25917	10^{9}
Exoplanet mass	33.9	15.4	10.7	9.2	6.23	9.47	5.98	4.48	4.48	401	10^{5}
Pulsars rotation freq	33.9	20.7	12.7	7.6	5.3	5.0	4.94	4.67	4.88	1861	10^{4}
Fermi space telescope γ-ray source fluxes	30.3	17.9	13.0	9.9	7.6	6.96	5.23	5.23	2.72	1451	10^{5}
Earthquake depths	31.6	16.9	14.0	8.69	6.98	7.42	5.27	4.58	4.36	248915	10^{2}
S-A seismogram	28.4	15.7	12.5	9.6	8.97	7.37	6.52	6.04	4.93	24000	10^{5}
Green house gas emissions by country	29.9	17.9	11.4	7.6	9.2	8.15	5.97	4.89	4.89	184	10^{4}
Global Temp anomalies in period 1880–2008	27.7	19.4	12.7	12.1	8.9	5.4	6.61	4.32	2.81	1527	10^{2}
Fund. Phys. constants	34.0	18.4	9.2	8.28	8.58	7.36	3.37	5.21	5.52	326	10^{4}
Global Infectious disease cases	33.7	16.7	13.2	10.7	7.3	5.4	4.56	5.07	3.34	987	10^{6}
Geometric series	29.8	17.4	13.0	10.0	7.8	6.6	5.8	5.0	4.6	1000	10^{21}
Fibbonacci sequence	30.0	17.7	12.5	9.6	8.0	6.7	5.7	5.3	4.5	1000	10^{14}
Combined	30.9	17.4	13.2	9.0	7.6	6.4	5.7	4.8	5.0	10000	10^{33}

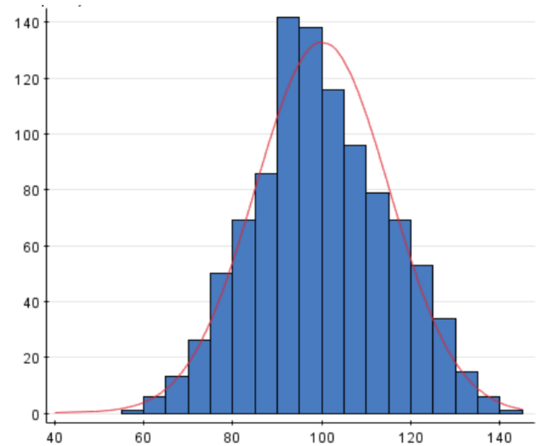
Benford Applications

- Minimal interest except for Knuth in 1968 programming.
- More recent, prominent applications include validity testing:
  - COVID data
  - Fraud Detection & Auditing
  - Election Data
  - Statistical Physics?

0.31772278	0.14892908	0.11707912	0.10165353	0.08256575	0.07468649	0.06436577	0.05115969	0.04183775
0.30102999	0.17609125	0.12493873	0.09691001	0.07918124	0.06694678	0.05799194	0.05115252	0.04575749



- Non-Examples:
  - Height
  - IQ scores
  - Car prices
  - Phone Numbers



Fermilab

Benford Criteria

Something conforms to Benford's Law if:

- $P_d \approx \log_{10}\left(1 + \frac{1}{d}\right)$
- Orders of magnitude
- Scale invariance
- Data describes the same object
- No artificial max/min imposed on data
- Number cannot be assigned (e.g. phone numbers)

Python

```python
from matplotlib import pyplot as plt
import numpy as np
import math


def benford(d):
    return np.log10(1+1/d)
d_values=np.linspace(.1, 10, 100)
benford_prob=benford(d_values)

numbers = [float(n) for n in range(1, 10)]
benford = [math.log10(1 + 1 / d) for d in numbers]

digits = list(range(1,10))

#Beam Position Data X
beam_data_X = np.loadtxt('ALLBPMX.txt', unpack = False)
plt.style.use('bmh')
plt.hist(beam_data_X, bins = 200,color = 'teal',hatch = '')
plt.xlabel('Horizontal Beam Centroid (\u03BCm)')
plt.ylabel('Pulse Counts')
plt.title('All BPMs')
plt.show()

#Beam Position Data Y
beam_data_Y = np.loadtxt('ALLBPMY.txt', unpack = False)
plt.style.use('bmh')
plt.hist(beam_data_Y, bins = 200,color = 'salmon',hatch = '')
plt.xlabel('Vertical Beam Centroid (\u03BCm)')
plt.ylabel('Pulse Counts')
plt.title('ALL BPMs')
plt.show()
```

```python
#Mean Absolute Deviation
MAD_X = (1/9)*sum(abs(residuals_X))
MAD_Y = (1/9)*sum(abs(residuals_Y))
print('MAD X = ' + str(MAD_X))
print('MAD Y = ' + str(MAD_Y))
```

```python
#First Digit Counts X
digit_counts_X = [0]*9
with open("ALLBPMX 1ST DIGIT.txt","r") as file:
    lines=file.read().splitlines()
    for line in lines:
        if line.isdigit() and int(line) in range (1,10):
            digit_counts_X[int(line)-1]+=1
print(digit_counts_X)

#First Digit Counts Y
digit_counts_Y = [0]*9
with open("ALLBPMY 1ST DIGIT.txt","r") as file:
    lines=file.read().splitlines()
    for line in lines:
        if line.isdigit() and int(line) in range (1,10):
            digit_counts_Y[int(line)-1]+=1
print(digit_counts_Y)

#Digit Proportions X
total_count = sum(digit_counts_X)
digit_proportions_X = [count/total_count for count in digit_counts_X]

#Digit Proportions Y
digit_proportions_Y=[count/total_count for count in digit_counts_Y]
```

```python
#Benford Deviation
actual_X = (digit_proportions_X)
predicted = (benford)

array1 = np.array(actual_X)
array2 = np.array(predicted)
residuals_X = np.subtract(array1,array2)

summand_X = (residuals_X/predicted)**2
summation_X = sum(summand_X)
Benford_Deviation_X = np.sqrt(summation_X)
print("Benford Deviation \u03B4= " + str(Benford_Deviation_X))
```

Calculating Benford Deviation: Methods

Benford Deviation

$$\delta = \sqrt{\sum_{d=1}^{9}\left[\frac{F(d) - P(d)}{P(d)}\right]^2}$$

$\rightarrow P(d) =$ ideal proportion
$\rightarrow F(d) =$ observed proportion

$\delta = 0 - 0.03$ considered "good"

$\delta = 0.6 - 0.7$ shows "preference to small digits"

**Chi-Square

$$\chi^2 = \sum_{1}^{n}\frac{(O_n - E_n)^2}{E_n}$$

$\rightarrow O_n =$ observed frequency
$\rightarrow E_n =$ expected frequency

**Not best metric for Benford's Law

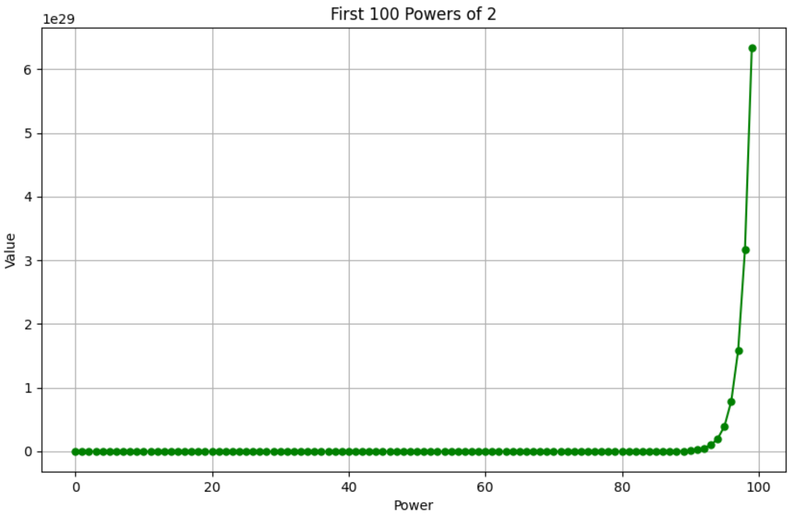
Mean Absolute Deviation

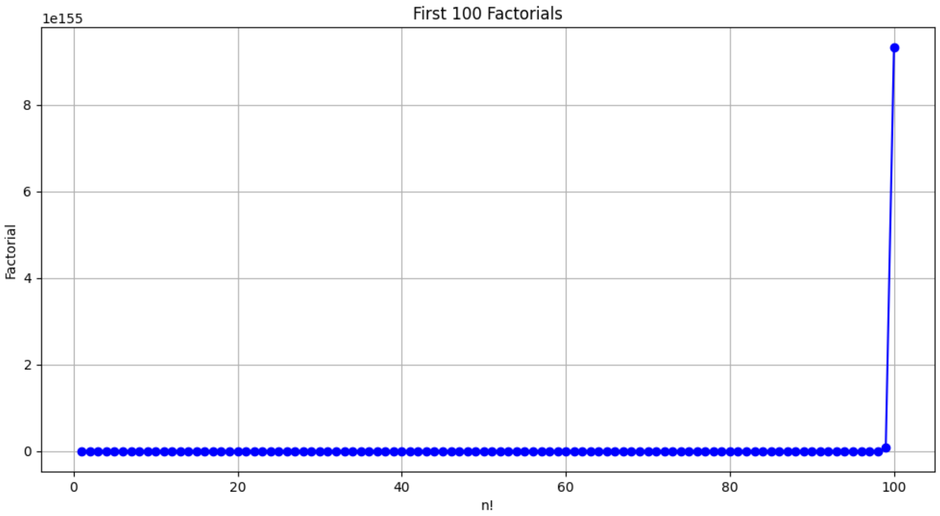
$$MAD = \frac{1}{9}\cdot\sum_{i=1}^{9}|p_i - b_i|$$

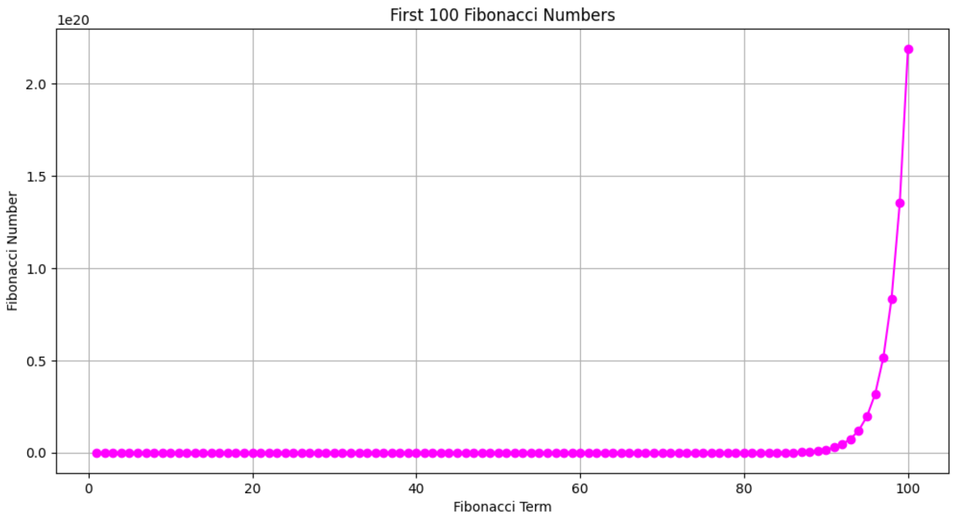
$\rightarrow p_i =$ observed proportions
$\rightarrow b_i =$ expected proportions

MAD	Conformity
$0 - 0.006$	Close
$0.006 - 0.012$	Acceptable
$0.012 - 0.015$	Marginal
> 0.015	Non-conformity

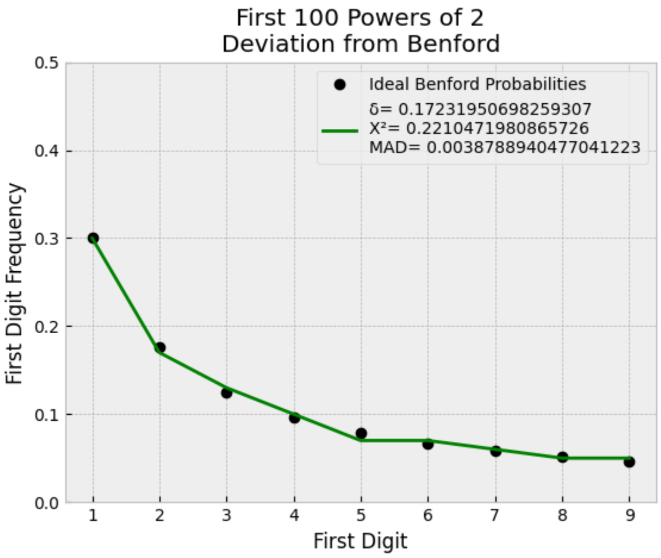
Popular Number Sets Raw Data

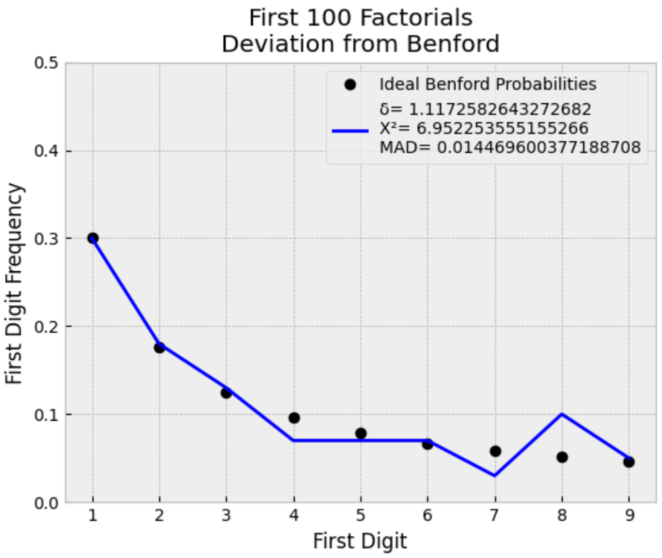


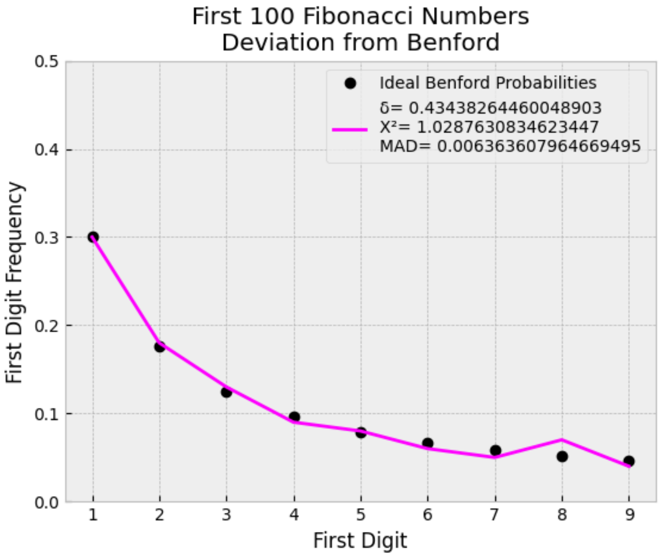




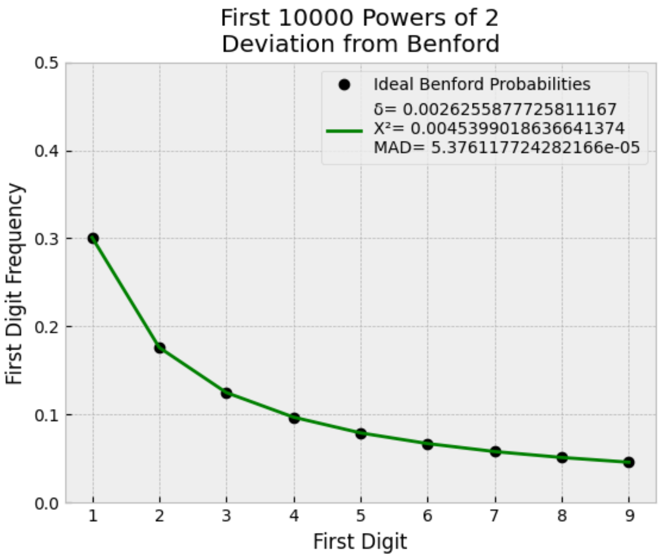
Popular Number Sets Against Benford

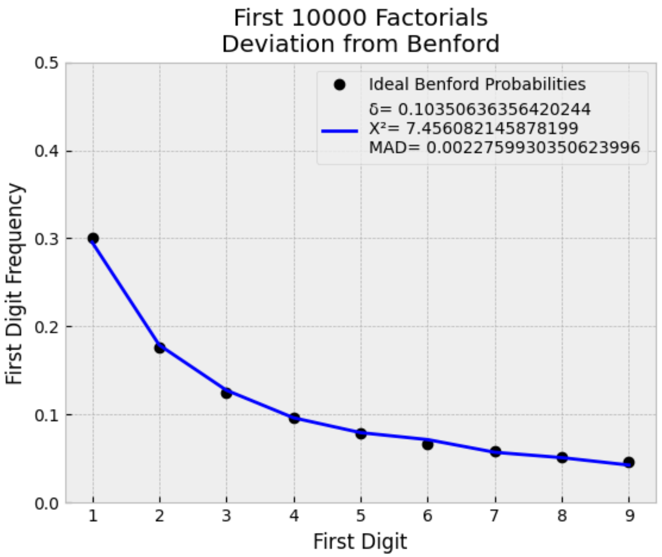


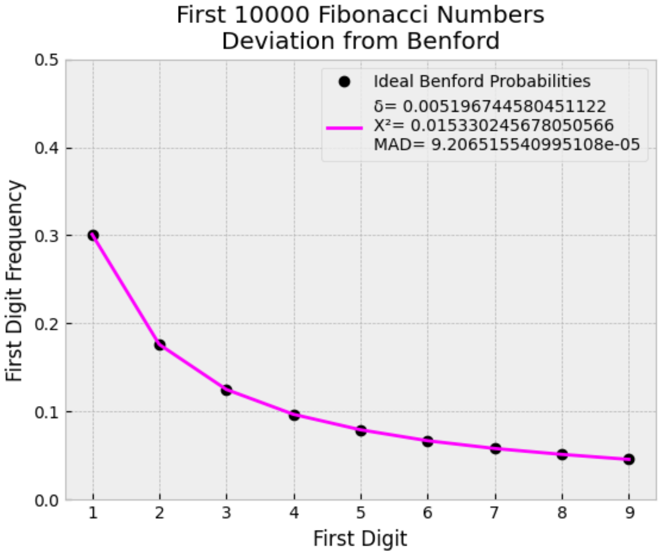




Popular Number Sets Against Benford (again)



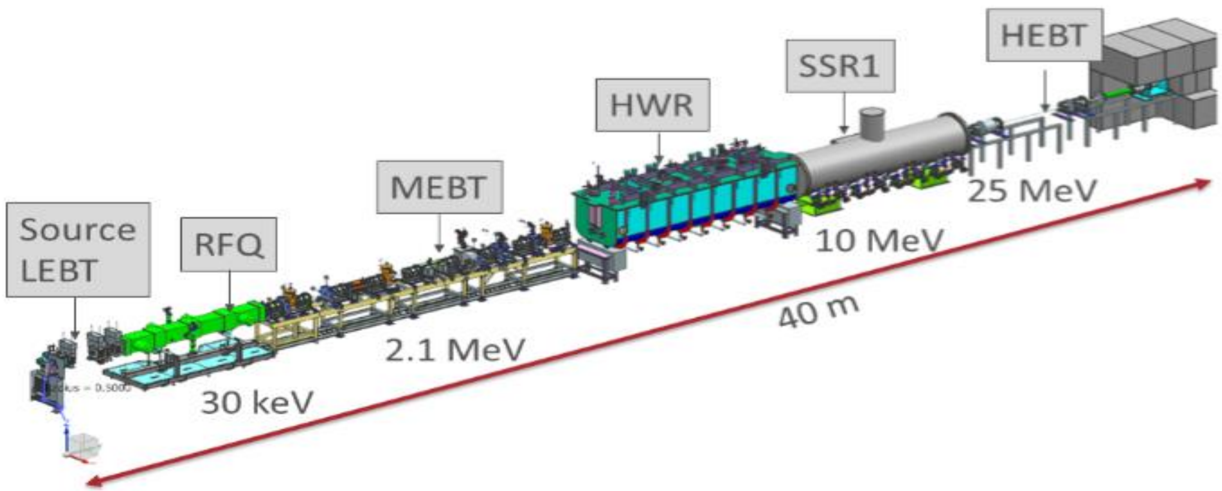




1. Introduction
   - Motivation
   - Benford's Law Definition and Application
2. Methodology
3. Results
   - Beam Measurement Data for PIP2IT beamline
4. Conclusion

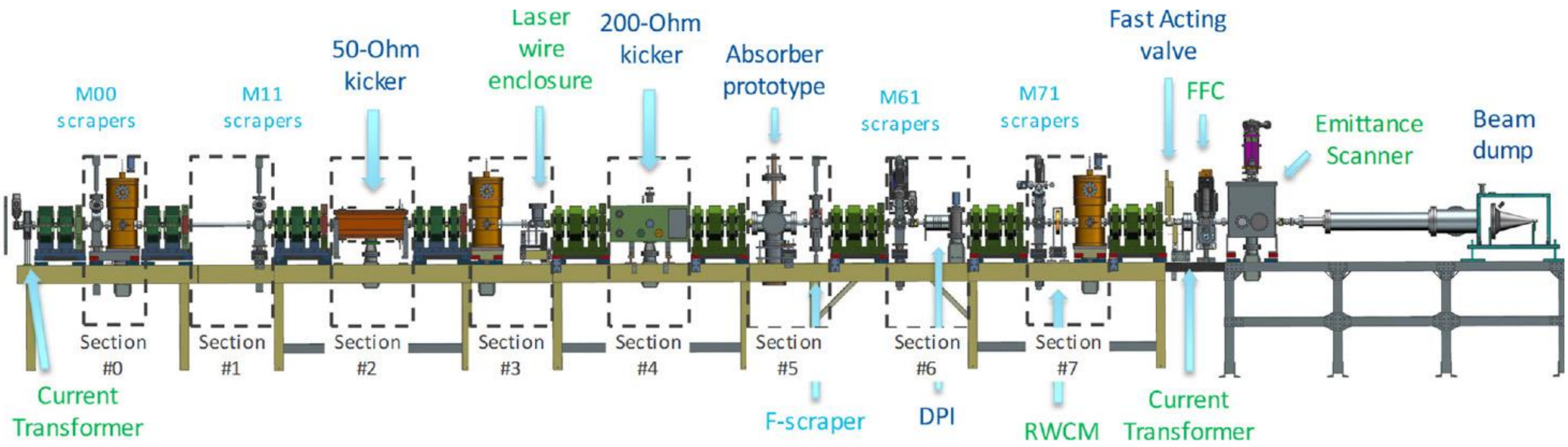
PIP2-IT

- PIP2-IT Goals – address risks for PIP-II linac, test front-end of PIP-II
- Facility was completed & successfully commissioned in 2021



- Starts with ion source, turns Hydrogen gas → beam of H^- ions
- RFQ accelerates beam to MEBT where beam is manipulated.
- Two super conducting cryomodules accelerate beam up to 25 MeV.
- Remaining are beam measurement, diagnostics, and dump station

PIP2-IT Medium Energy Beam Transport (MEBT)



- Beam Horizontal and Vertical positions are measured using Beam Position Monitor (BPMs)
  - BPMs are located at center of each Quadrupole Doublet and triplet.
- BPM data were used to quantify eam position jitters.

Jitter and Noise

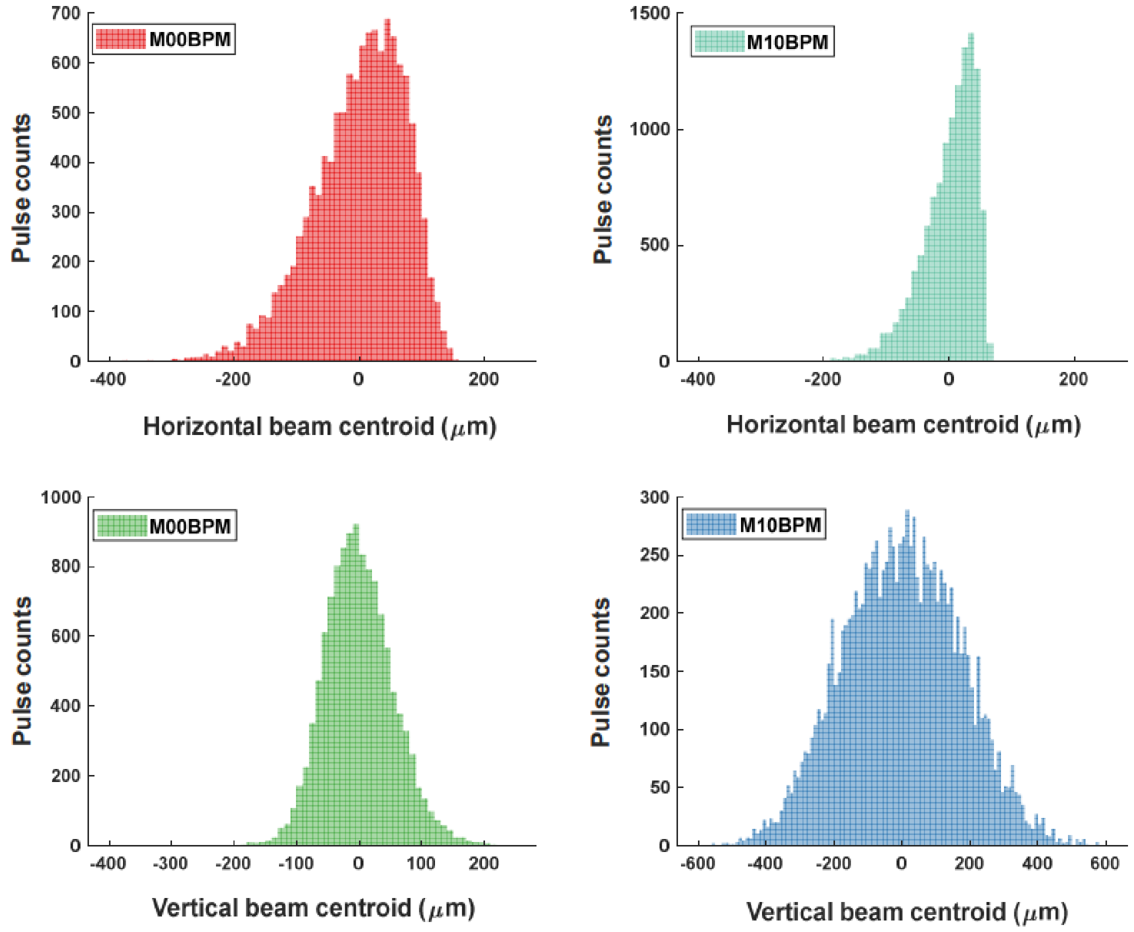
The variation in timing or arrival of particles or beams at specific points within the accelerator system. Arises from accelerator defects, external disturbances, fluctuations in the power supply.

Affect efficiency of accelerator performance, thus affecting measurement accuracy and experimental outcomes.

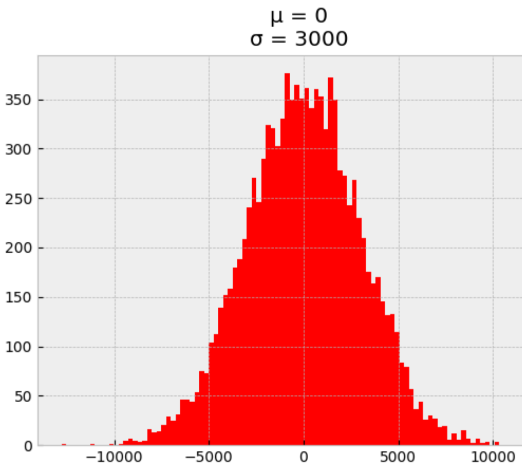
If we want to take more data, it has to be accurate!

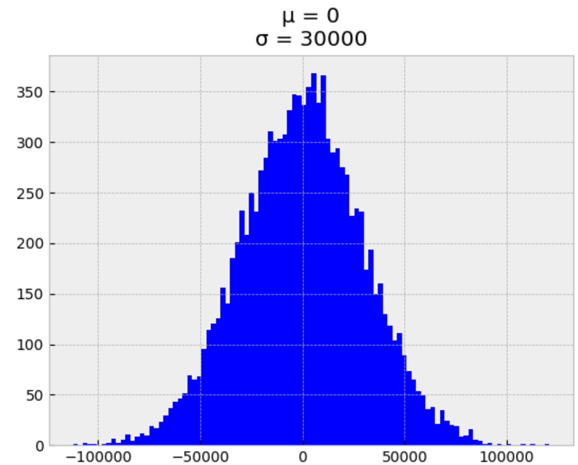
Beam Position Data

- Beam Position measured almost 12,000 times in all nine BPMs.
- Position varies in both planes
- Origin of jitters unknown, methods to resolve unsuccessful.
- Ultimately impacts were negligible.

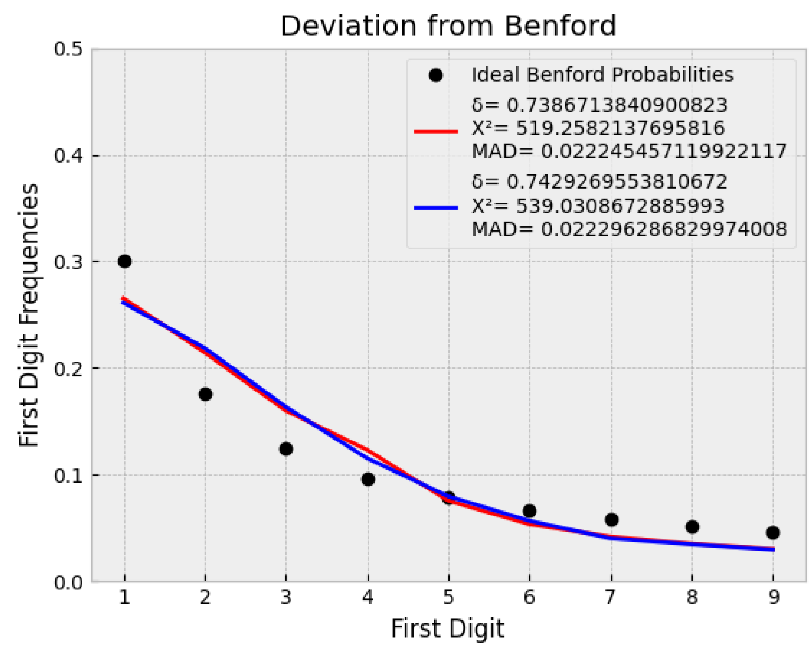


Testing Benford on Normal Distributions



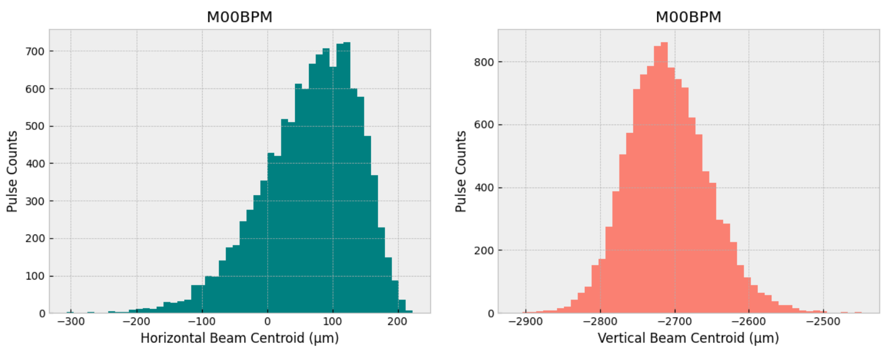


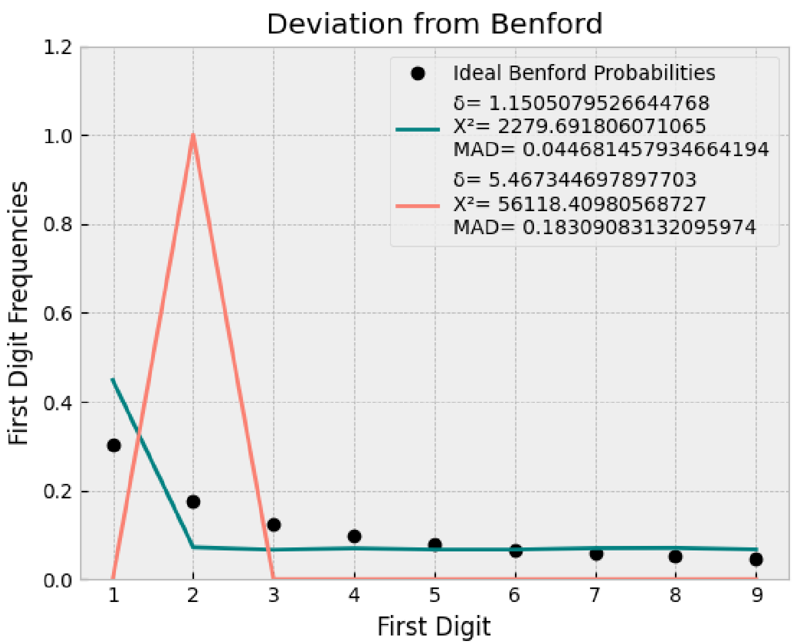
	Red	Blue
μ	0	0
σ	3000	30000
N	11000	11000
δ	≈ 0.739	≈ 0.743
χ^2	≈ 519.258	≈ 539.031
MAD	0.0222	0.0222

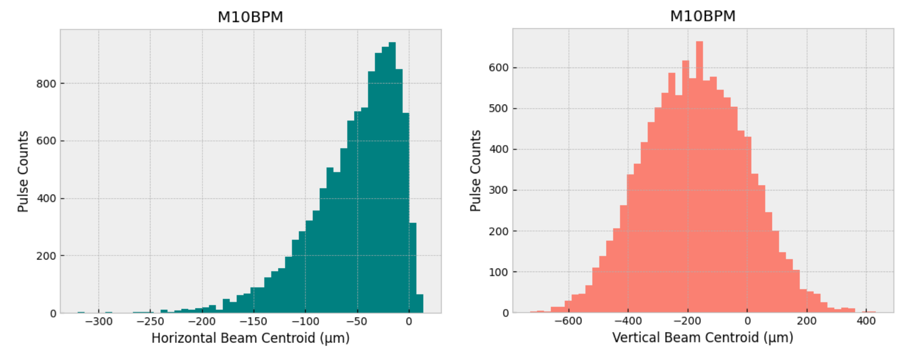


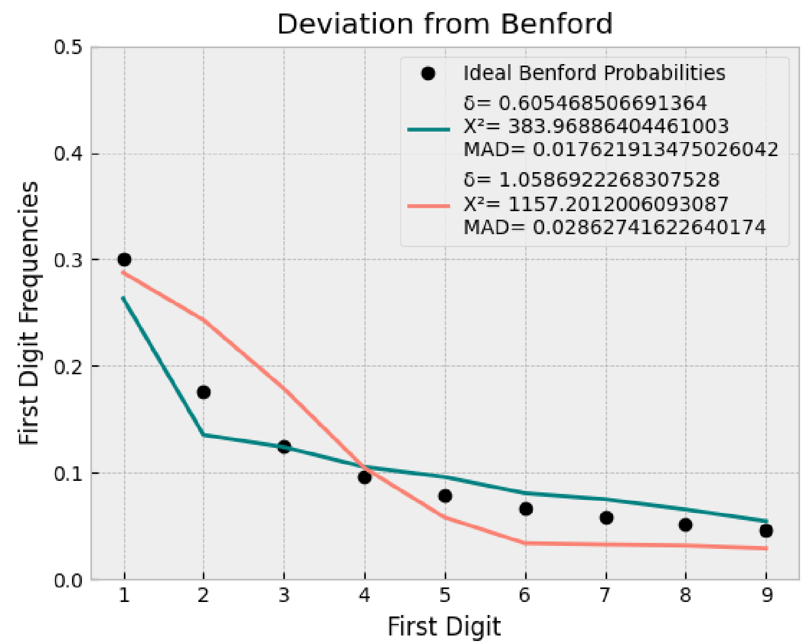
Fermilab

PIP2-IT BPM Data: First two BPMs

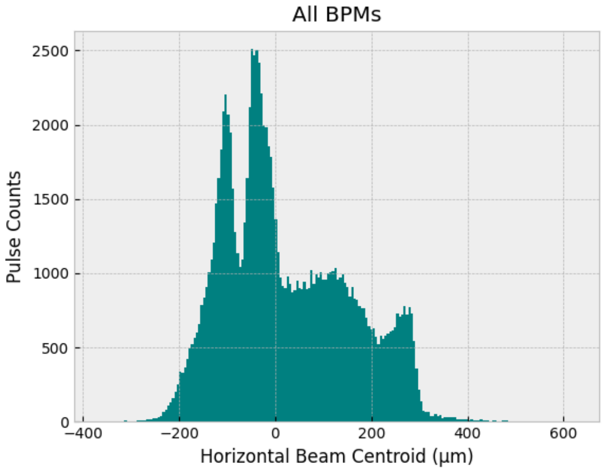


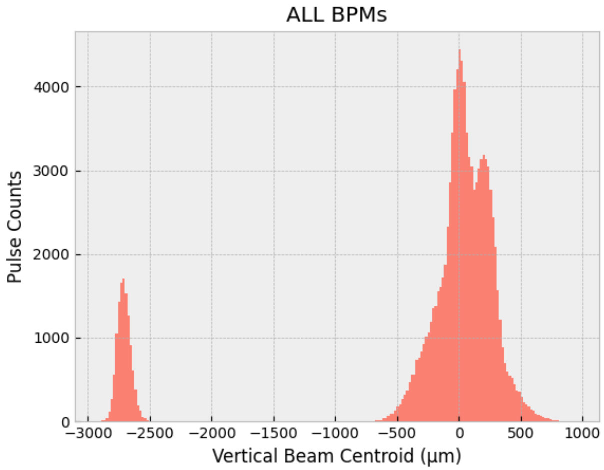




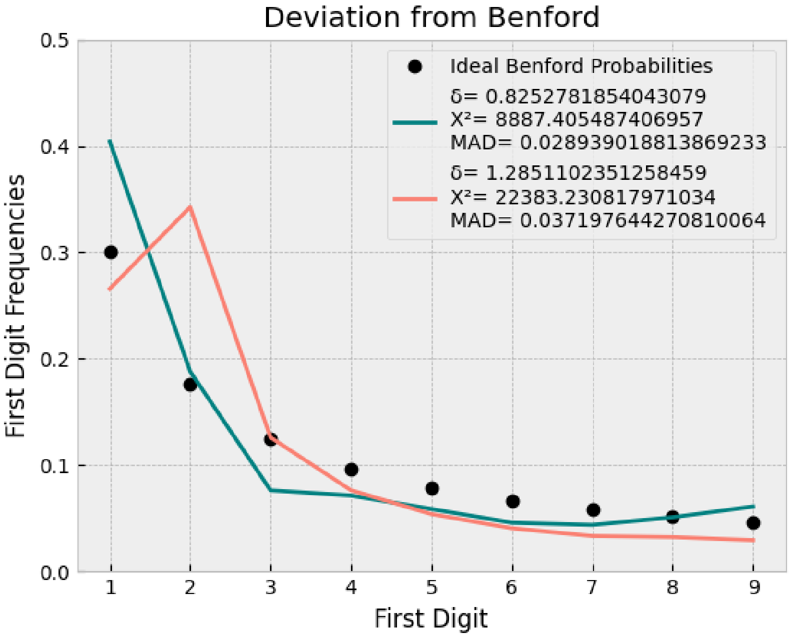


PIP2-IT BPM Data: All BPMs





	Horizontal	Vertical
μ	22.5	-252.52
σ	130.92	890.35
N	107946	107946
δ	≈ 0.825	≈ 1.285
χ^2	≈ 8887.41	≈ 22383.23
MAD	0.0289	0.0372



Findings/Results

Using MAD, for all nine BPMs, 0/18 distributions exhibited even marginal Benford conformity.

Using δ, about 4/18 showed “preference to leading digits” but no conformity.

Poor conformity – why?

Conclusions

Noise → Randomness → Benford Deviation

Experimental noise is persistent, even at low temps

Ambiguity in metrics

Benford’s law is somewhat trivial and postdictive, not predictive

Where to go from here?

Other test methods

- Kolmogorov-Smirnov Test yielded similar results for

Examine other details:

- Outliers
- Skewness
- Orders of magnitude

Analyze additional beam data

Take Home

Immersed in the Fermilab culture

- Mentored by and interact with top scientists

Cross-disciplinary

Math, Stats, Python, Particle Accelerators, & Beam Dynamics

Every teacher's favorite question

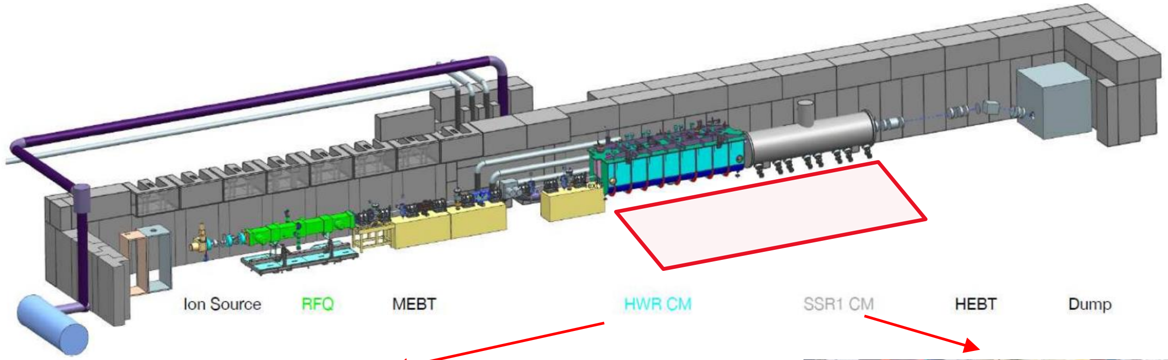
Acknowledgements

This work was done under the Teacher Research Associate (TRAC) Program at Fermilab.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Thank you to Harry Cheung, Don Lincoln, and the other TRAC teachers for a great experience, and a special thank you to my mentor Arun Saini for his guidance this summer.

PIP2-IT SRF Phase



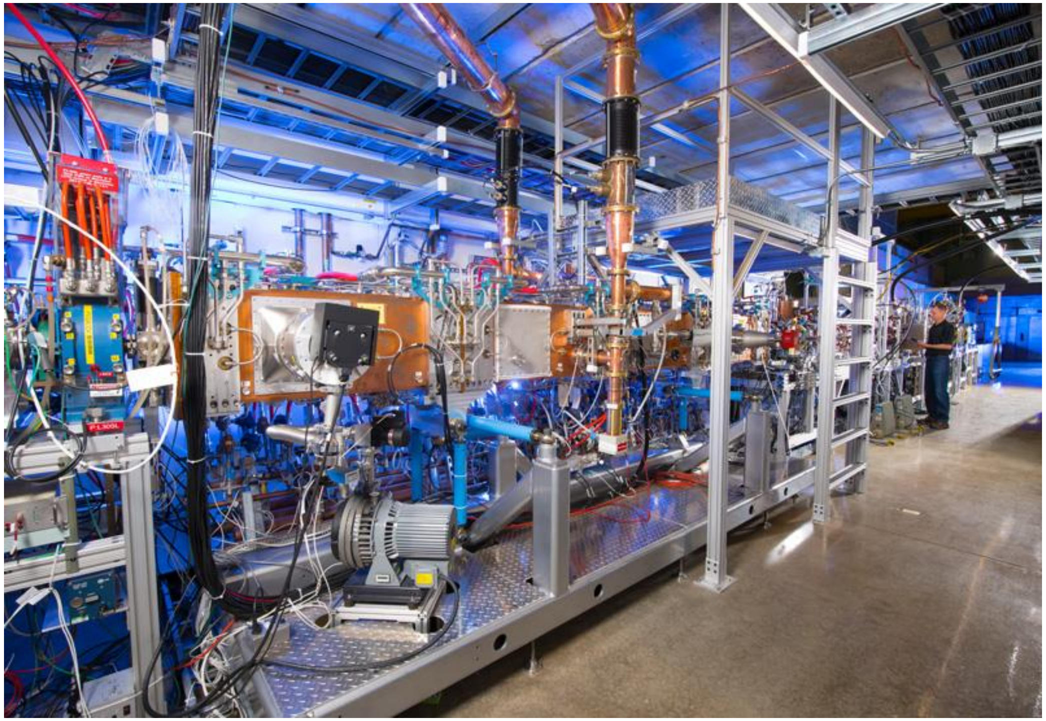
HWR CM & SSR1 CM are first two in PIP-II

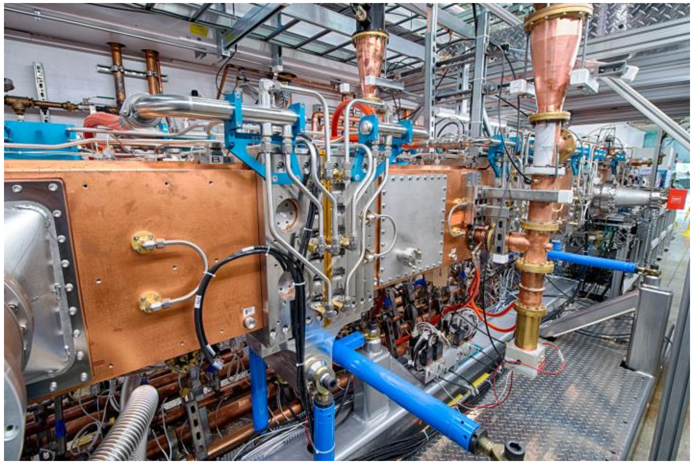
Niobium insulated cavities, cooled to 2° K

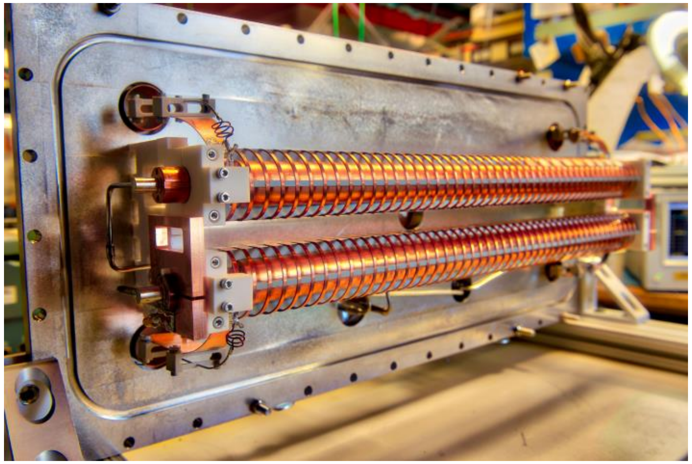
RF doubles from HWR to SSR1

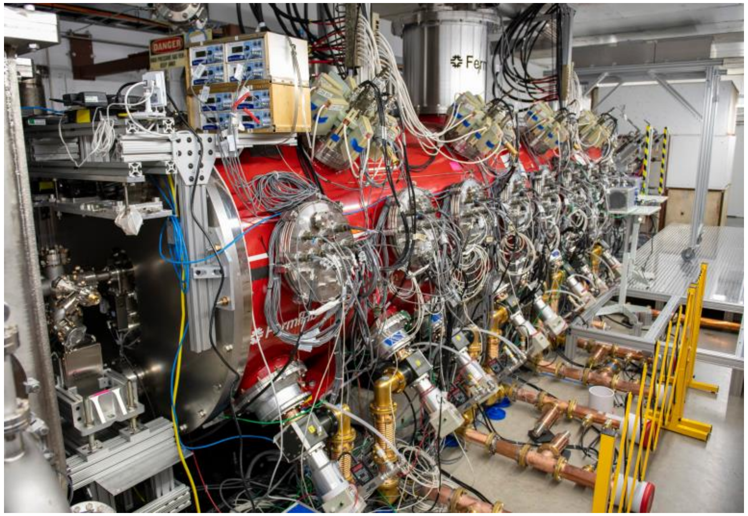
More Pictures

Top Right – PIP2IT Facility
Top Left – RFQ at PIP2IT
Bottom Left – PIP2IT Chopper
Bottom Right – SSR1 at PIP2IT









References

A. Saini, B. Hanna, J. Huber, L. Prost, V. Scarpine, J. Steimel, S.V.L.S. Rao, C. Richard, Beam tests of the PIP-II Injector Test Radio Frequency Quadrupole, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 978, 2020, 164368, ISSN 0168-9002, https://doi.org/10.1016/j.nima.2020.164368.

Azevedo CDS, Gonçalves RF, Gava VL, Spinola MM. A Benford's law based method for fraud detection using R Library. MethodsX. 2021 Nov 11;8:101575. doi: 10.1016/j.mex.2021.101575. PMID: 35004209; PMCID: PMC8720889.

Cheesinglee. (2022, March 24). *Detecting numeric irregularities with Benford's law*. The Official Blog of BigML.com. https://blog.bigml.com/2015/05/15/detecting-numeric-irregularities-with-benfords-law/

C. Richard, M. Alvarez, J. P. Carneiro, B. Hanna, L. Prost, A. Saini, V. Scarpine, A. Shemyakin; Measurements of a 2.1 MeV H^- beam with an Allison scanner. *Rev Sci Instrum* 1 July 2020; 91 (7): 073301. https://doi.org/10.1063/5.0004502

Deckert, J., Myagkov, M., & Ordeshook, P. (2011). Benford's Law and the Detection of Election Fraud. *Political Analysis, 19*(3), 245-268. doi:10.1093/pan/mpr014

Durtschi, Cindy & Hillison, William & Pacini, Carl. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. J. Forensic Account. 5.

Dustin (2021) *Using Benford's law to detect fraud*, *Dustin K MacDonald*. Available at: http://dustinkmacdonald.com/using-benfords-law-to-detect-fraud/ (Accessed: 25 July 2023).

Fewster, R.. (2009). A Simple Explanation of Benford's Law. The American Statistician. 63. 26-32. 10.1198/tast.2009.0005.

Knuth, D. E. (1997). *The art of computer programming* (3rd ed.). Addison Wesley.

Leah. (2020). Accelerating science globally: Pip-II engineers continue designs for particle-propelling machine from Home. Retrieved from https://news.fnal.gov/2020/06/accelerating-science-globally-pip-ii-engineers-continue-designs-for-particle-propelling-machine-from-home/

Lijing Shao, Bo-Qiang Ma,The significant digit law in statistical physics,Physica A: Statistical Mechanics and its Applications,Volume 389, Issue 16,2010, Pages 3109-3116, ISSN 0378-4371, https://doi.org/10.1016/j.physa.2010.04.021

Maahs, C. M. and D. (n.d.). *Benford's law and covid-19 data*. CHANCE. https://chance.amstat.org/2021/04/benfords-law/#:~:text=For%20example%2C%20phone%20numbers%20in,would%20start%20with%20a%201.

Nigrini, M. J. (2012). Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. United States: Wiley.

Sambridge, M., Tkalčić, H., and Jackson, A. (2010), Benford's law in the natural sciences, *Geophys. Res. Lett.*, 37, L22301, doi:10.1029/2010GL044830.

Superconducting radio-frequency technology. Proton Improvement Plan-II. (n.d.). https://pip2.fnal.gov/how-it-works/superconducting-radio-frequency-technology/

Yang, Dikun. (2016). Characterization of noise in airborne-transient electromagnetic data using Benford's law. 1043-1047. 10.1190/segam2016-13972000.1.