Information Loss in Diffraction-Limited Data Compression

Can we compress our data and still retain the relevant information?

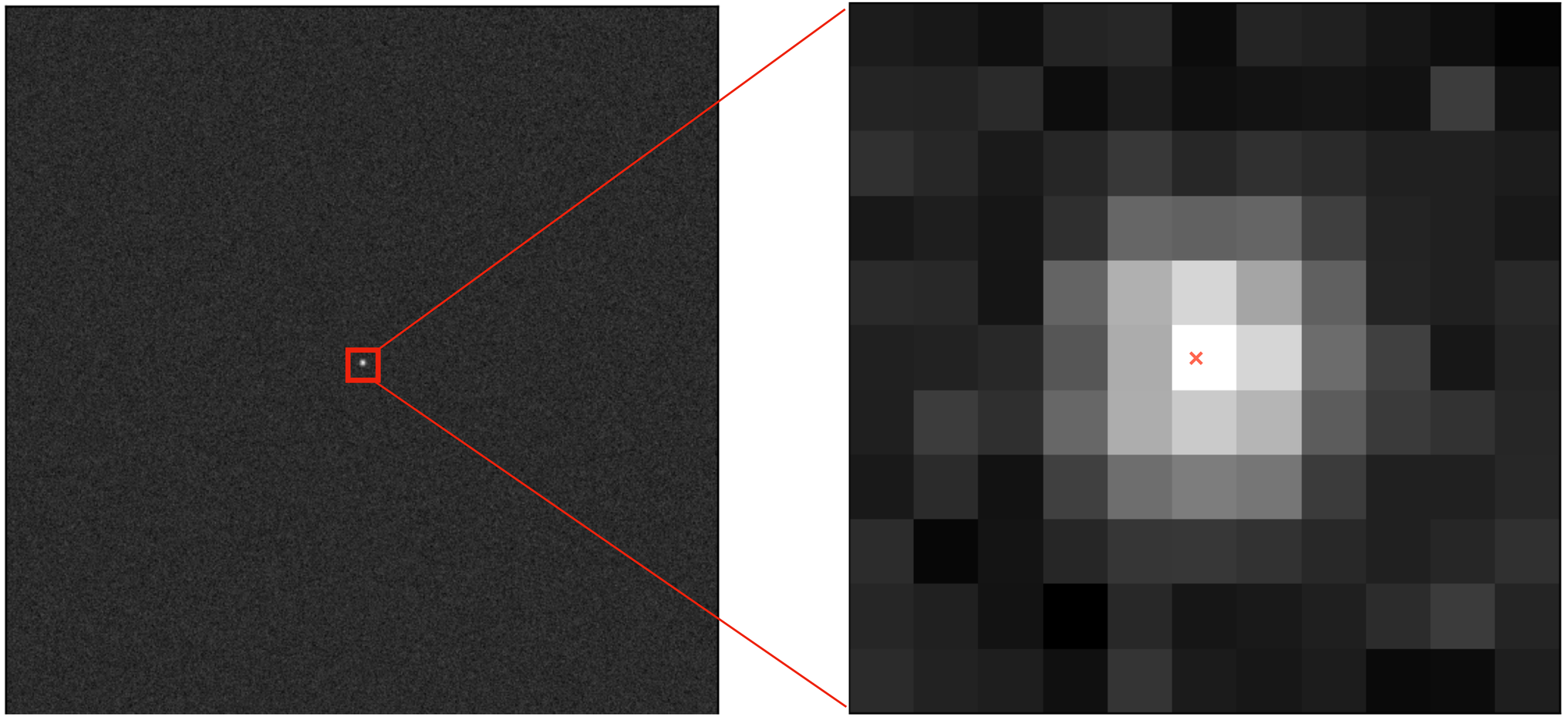
Matthew Hogan and Dr. Michael Vershinin
The University of Utah Department of Physics and Astronomy

Diffraction-Limited Data

Ernst Abbe Limit: $d = \frac{\lambda}{2n \sin\theta} = \frac{\lambda}{2NA}$

In modern optics, d~200 nm, larger than most viruses and proteins of interest

Diffraction results in an Airy disk

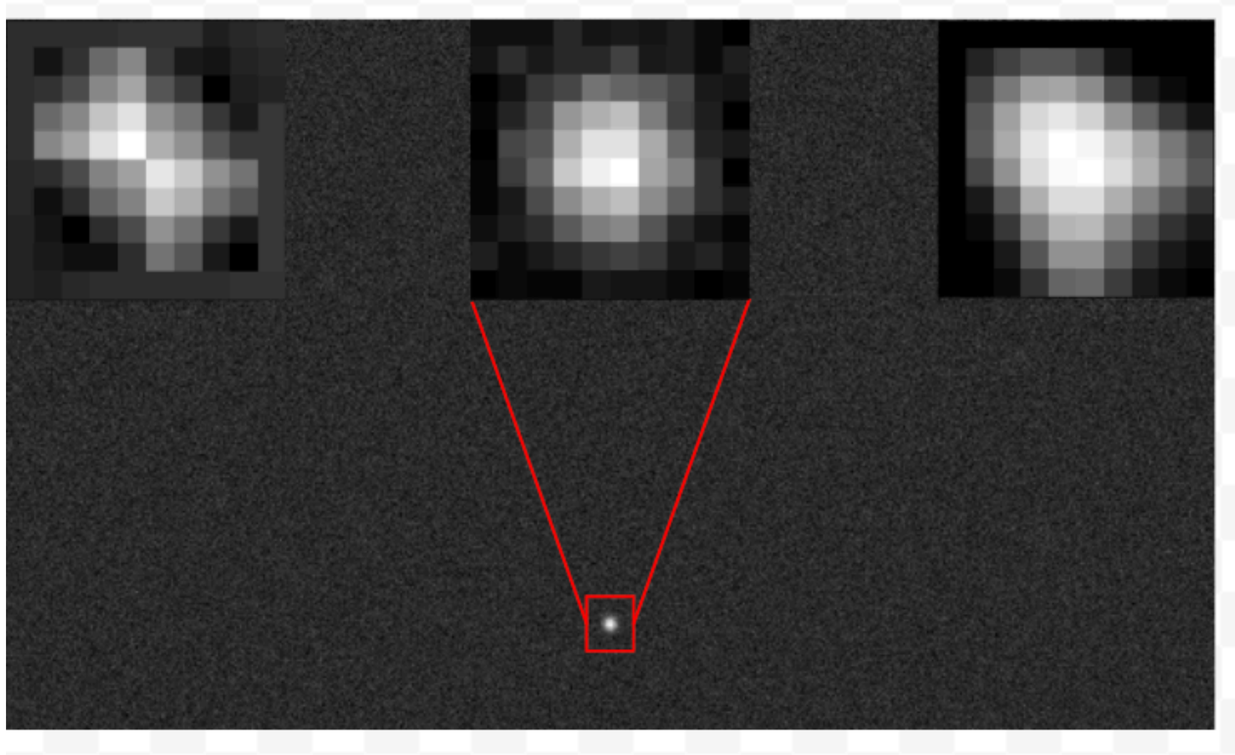


The Cost of Data Storage

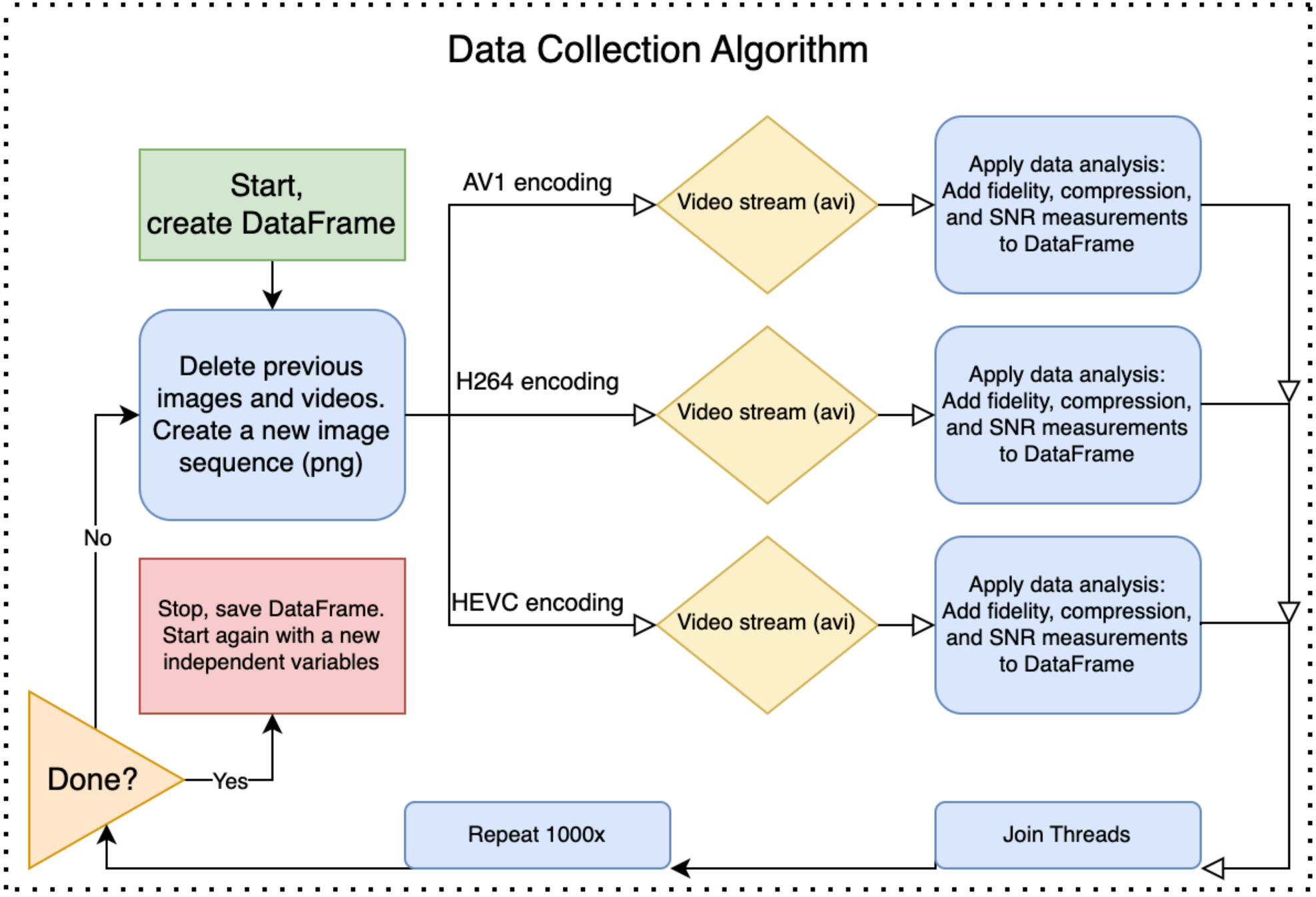
- 1 image = 500 kB
- 1 video (1000 image frames) = 500 MB
- 100 videos (enough for statistical analysis) = 50 GB
- Data retained over multiple years = 50-500 TB
- Monthly payment for storage = **$1,000-$10,000** ($20/TB)
- U.S. Dept. of Energy - "*Data centers are one of the most energy-intensive building types, consuming **10 to 50 times the energy per floor space of a typical commercial office building**. Collectively, these spaces account for approximately 2% of the total U.S. electricity use*"

Compression Could be a Solution

- Open source video processing software is readily available
- Many codecs can easily achieve 95% compression, but are geared toward macroscopic images
- However, compression of diffraction-limited data may have consequences for analysis
- We tested three codecs: AV1, H264, and HEVC

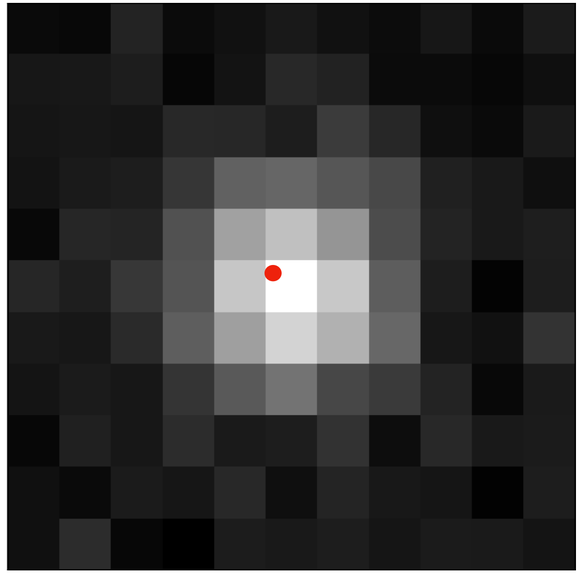


Uncompressed image (center).
Compressed with H264 (left) and HEVC (right)

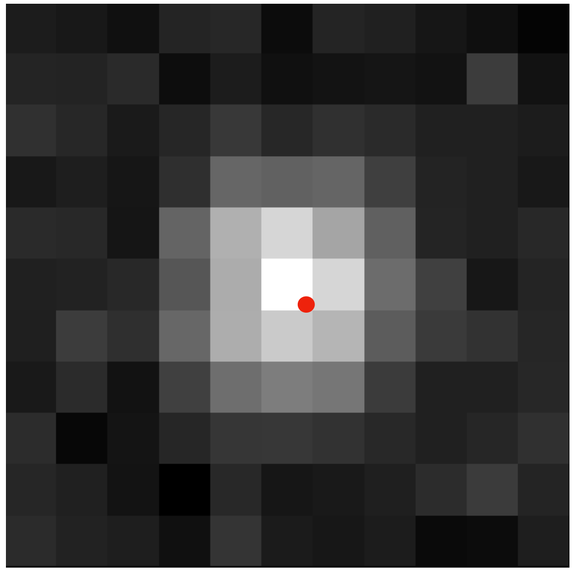

Data Collection Algorithm
Start, create DataFrame
Delete previous images and videos. Create a new image sequence (png)
AV1 encoding
H264 encoding
HEVC encoding
Video stream (avi)
Video stream (avi)
Video stream (avi)
Apply data analysis: Add fidelity, compression, and SNR measurements to DataFrame
Apply data analysis: Add fidelity, compression, and SNR measurements to DataFrame
Apply data analysis: Add fidelity, compression, and SNR measurements to DataFrame
No
Stop, save DataFrame. Start again with a new independent variables
Done?
Yes
Repeat 1000x
Join Threads

Fidelity

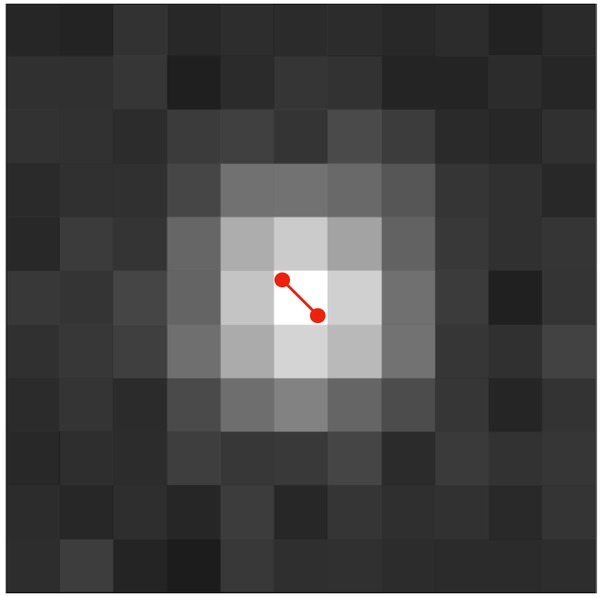
- Fidelity is how well the center position was recovered after compression
- Our measurement of fidelity was the distance between compressed and uncompressed image centers
- Perfect fidelity = 0 pixels



Uncompressed image center



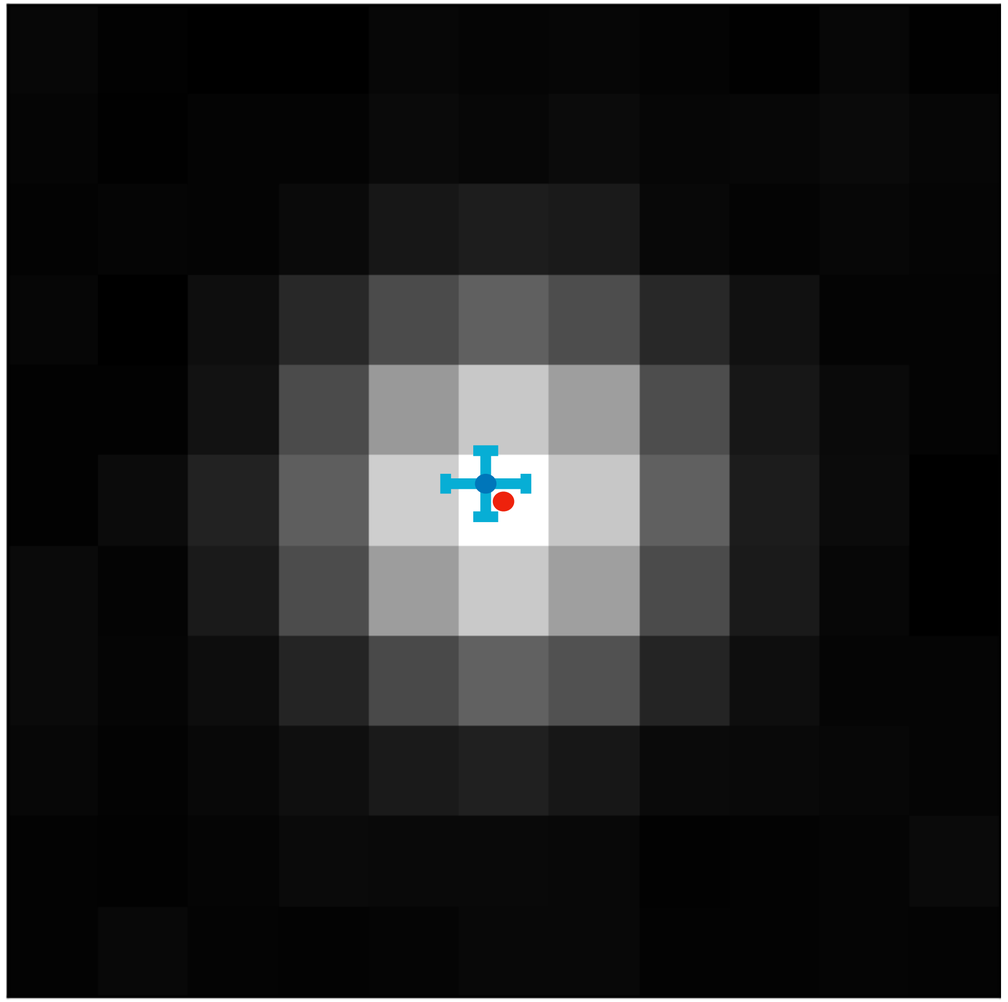
Compressed image center



Fidelity is the distance between the two

Success

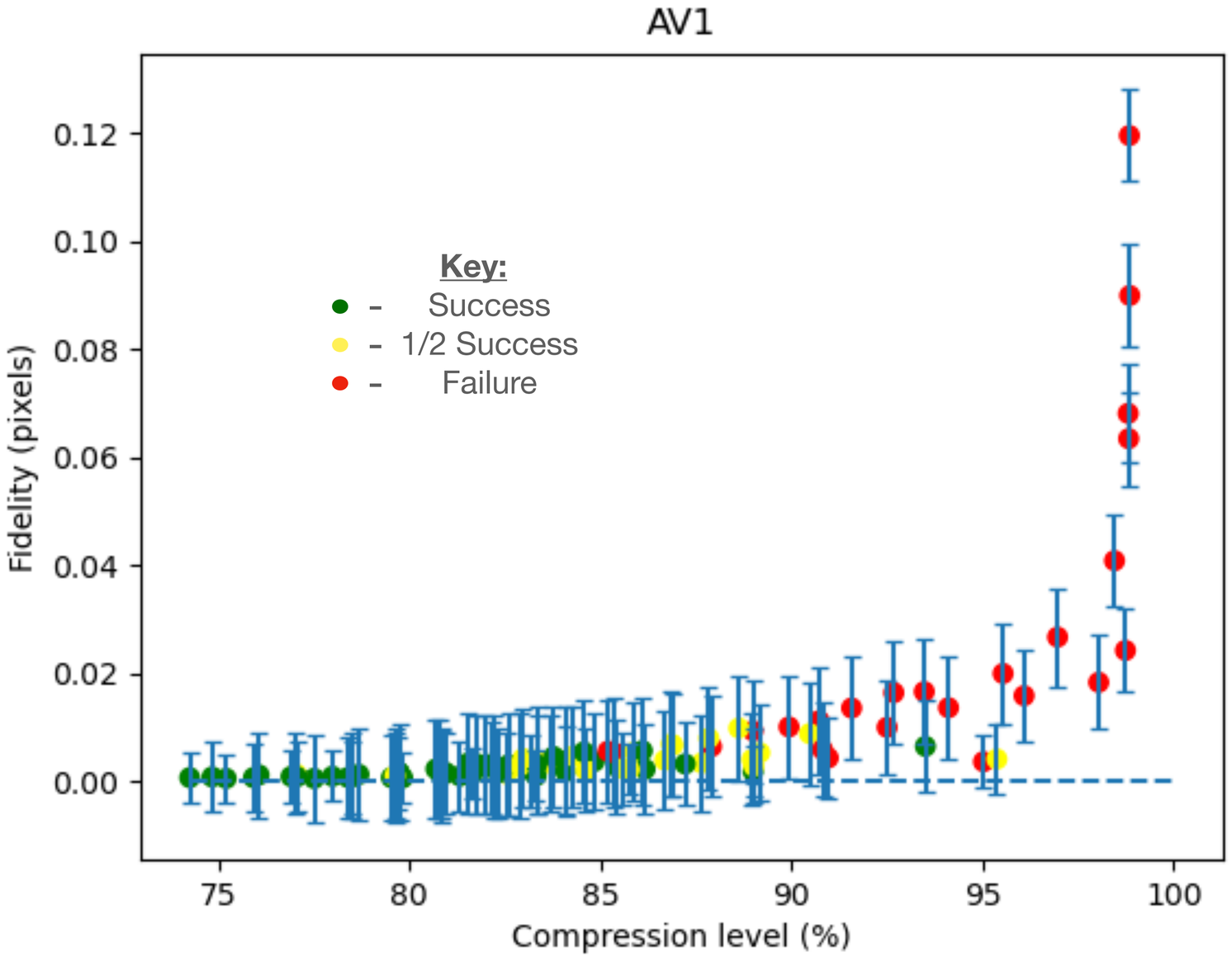
- Since the images were synthetic, the actual ground truth center location was known
- If the true x and y coordinates were captured in the 95% CIs, then it was a successful compression
- Half-success was possible if the x *or* y coordinate was captured, but not both
- Uncompressed images had a measured success of 95 ± 1%



Composite of 1000 images. Red is true center, blue is best guess

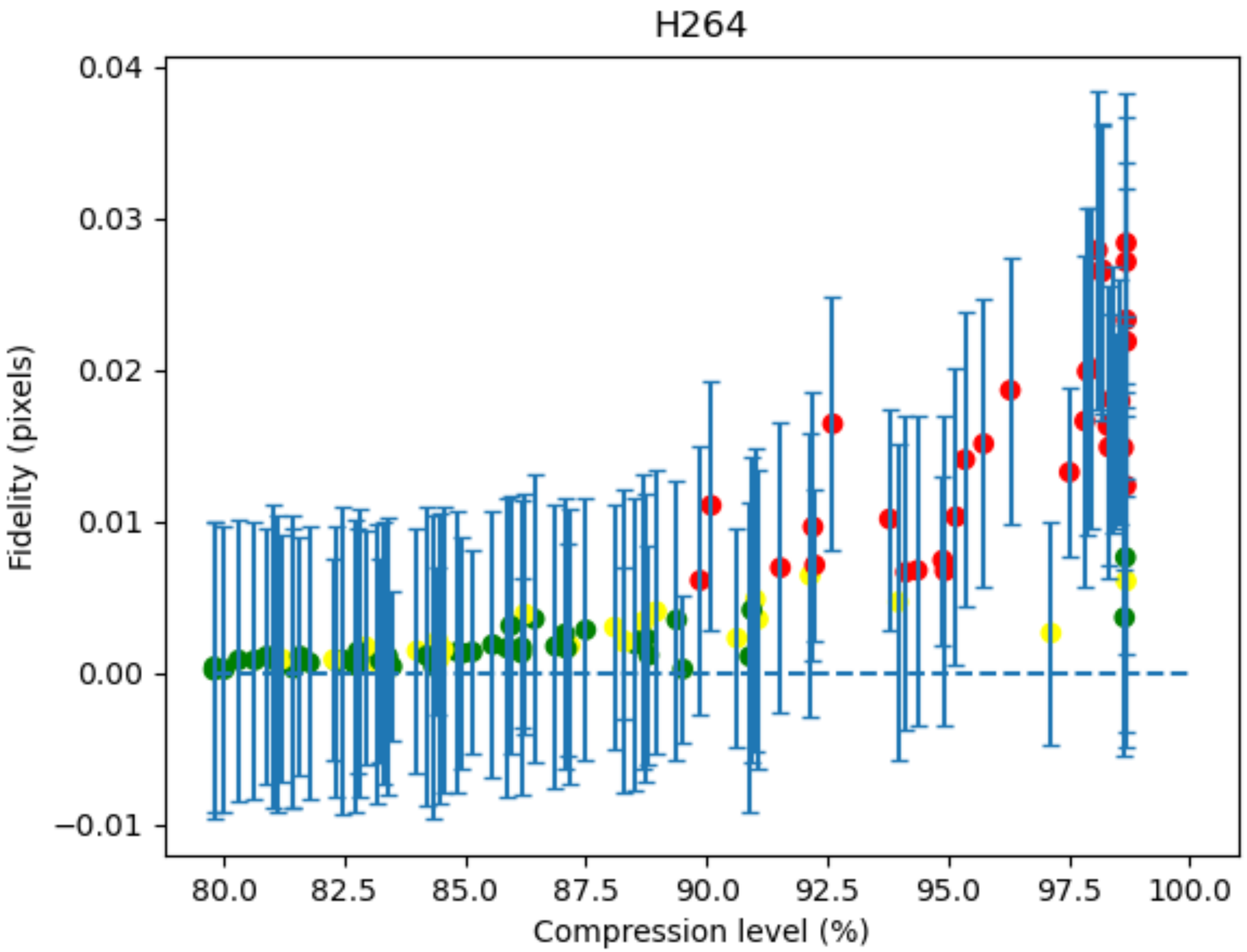
AV1 Results

- Below 80% compression, very strong fidelity and success
- Fidelity diverges asymptotically as compression level approaches 100%
- Average success of 64 ± 4%



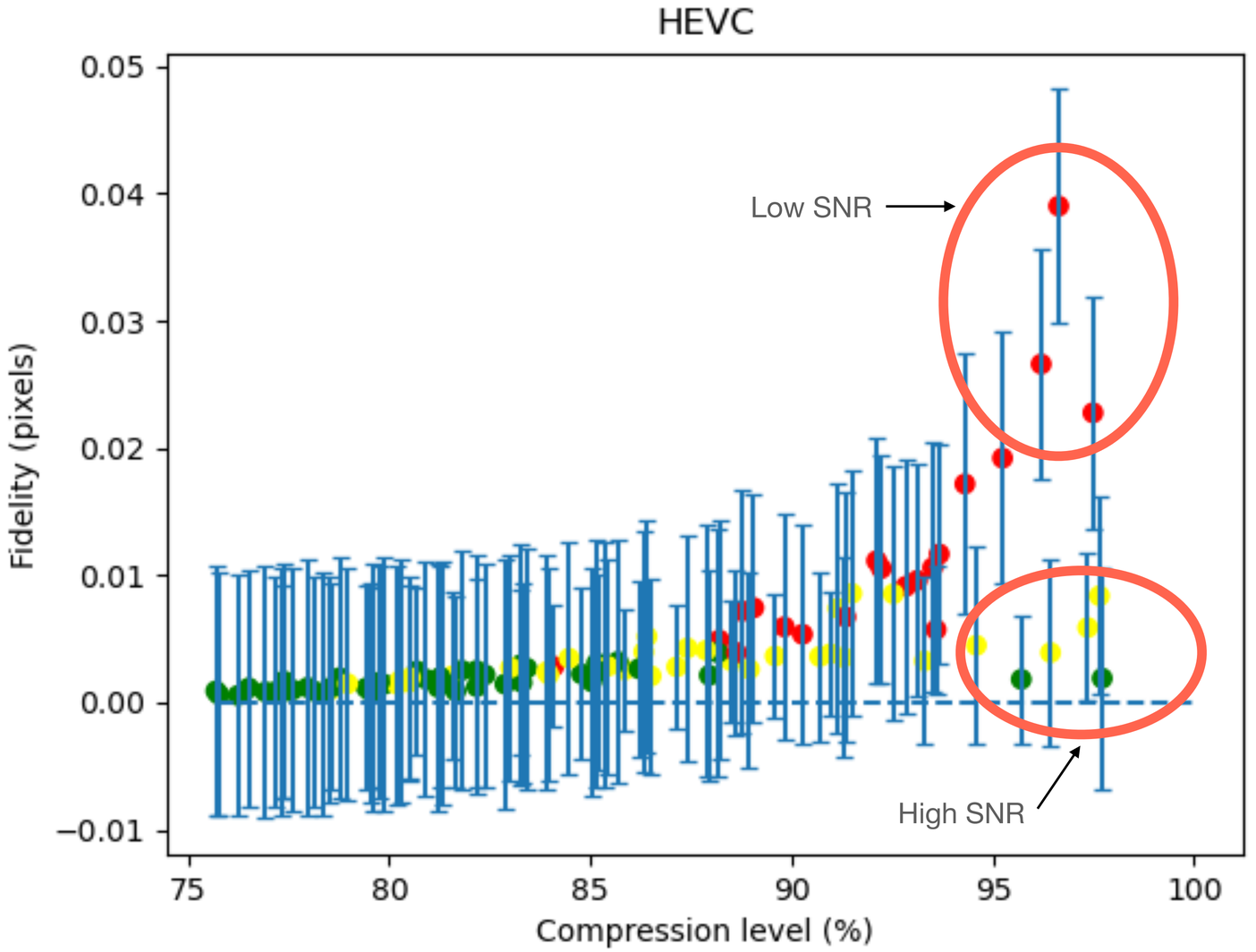
H264 Results

- Good results up to about 90% compression
- Similar trend to AV1
- Best fidelity overall, all within 1/20th of a pixel
- Most consistent
- Average success of $54 \pm 4\%$



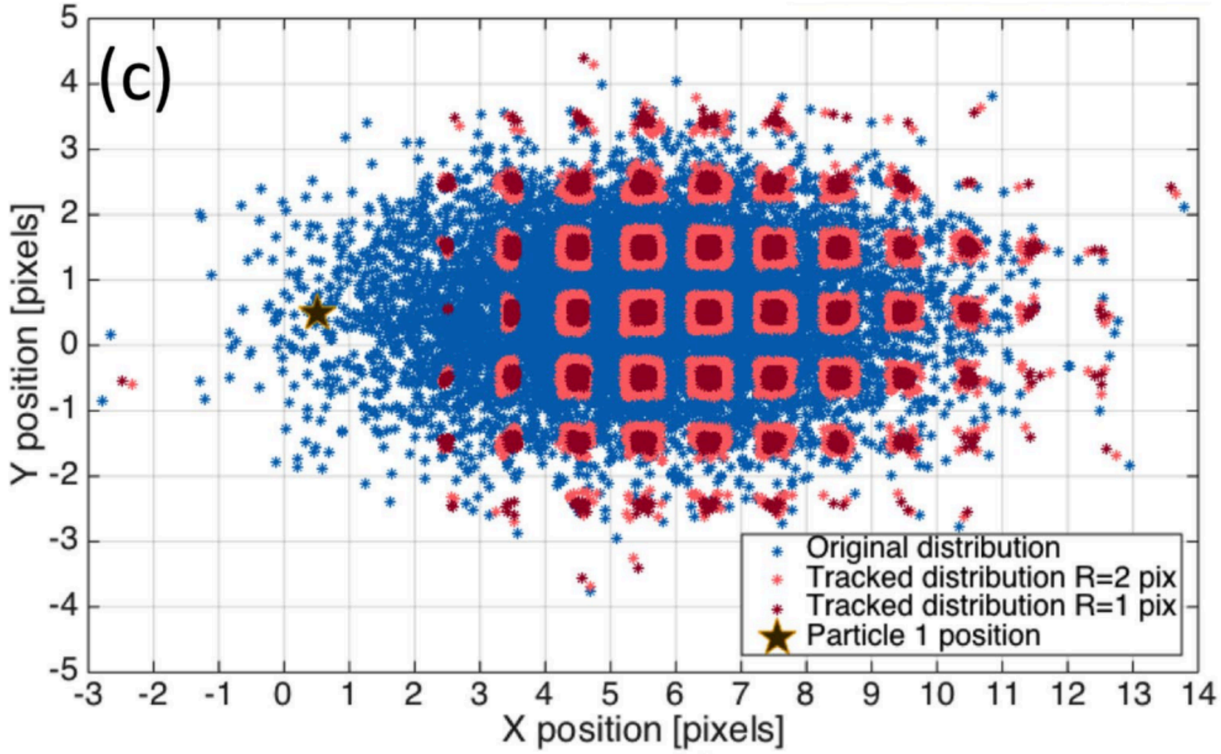
HEVC Results

- Again, good results below 80% compression
- Bifurcation at high compression levels between variable SNRs
- Average success of $64 \pm 4\%$



Summary

- Energy-efficient and cost-effective archival of TB-scale data sets necessitates their compression
- We generated synthetic images, compressed them, and measured centroid differences
- Our results suggest that up to 80% compression is potentially feasible for single images
- Pixel-locking effects may prove challenging to overcome in time-dependent measurements



Pixel-locking effect (Yifat et. al. 2017)

Thank You!

References

- Cheezum, M. K., et al. Quantitative Comparison of Algorithms for Tracking Single Fluorescent Particles. Biophysical Journal 81, 4 (2001). https://doi.org/10.1016/s0006-3495(01)75884-5.

- Carter, B., et al. Tracking single particles: a user-friendly quantitative evaluation. Physical Biology 2, 60 (2005). http://doi.org/10.1088/1478-3967/2/1/008

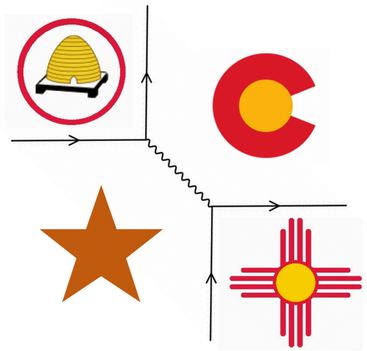
- Svoboda, K., Mitra, P. P., & Block, S. M. (1994). Fluctuation analysis of motor protein movement and single enzyme kinetics. Proceedings of the National Academy of Sciences, 91(25), 11782–11786. https://doi.org/10.1073/pnas.91.25.11782

- Yifat, Yuval, et al. “Analysis and Correction of Errors in Nanoscale Particle Tracking Using the Single-Pixel Interior Filling Function (Spiff) Algorithm.” *Scientific Reports*, vol. 7, no. 1, 2017, https://doi.org/10.1038/s41598-017-14166-6.

Acknowledgments

- Thank you to the J. Irvin and Norma K. Swigart endowment for funding this project.

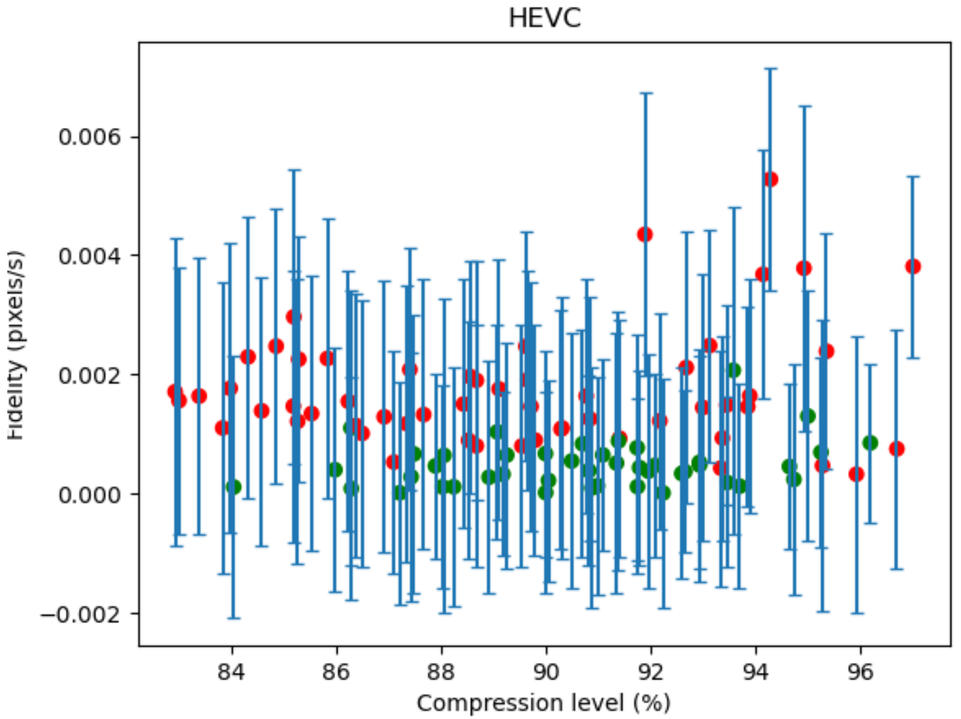
THE UNIVERSITY OF UTAH®

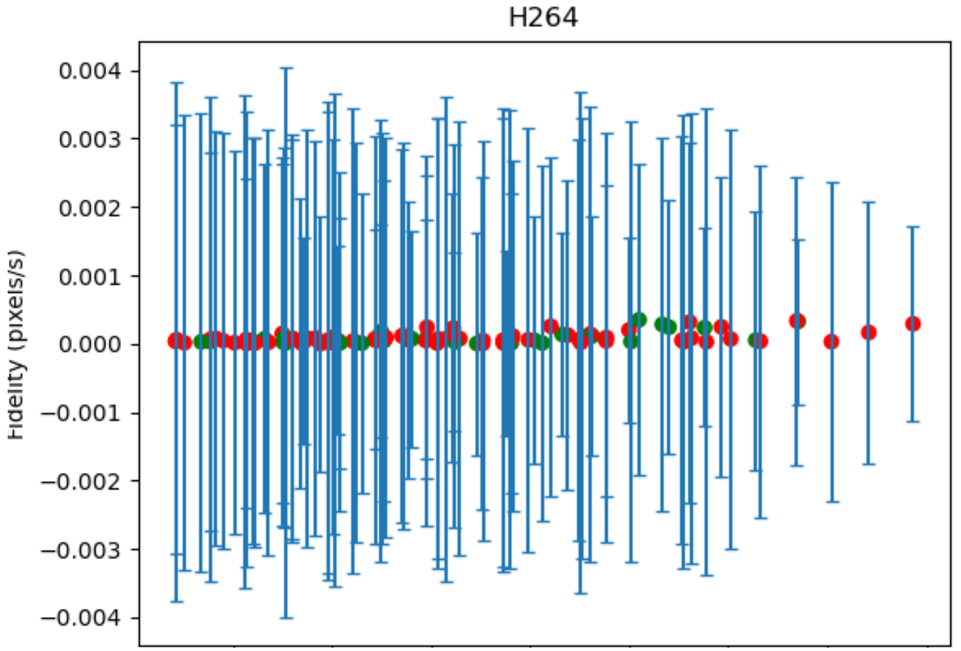


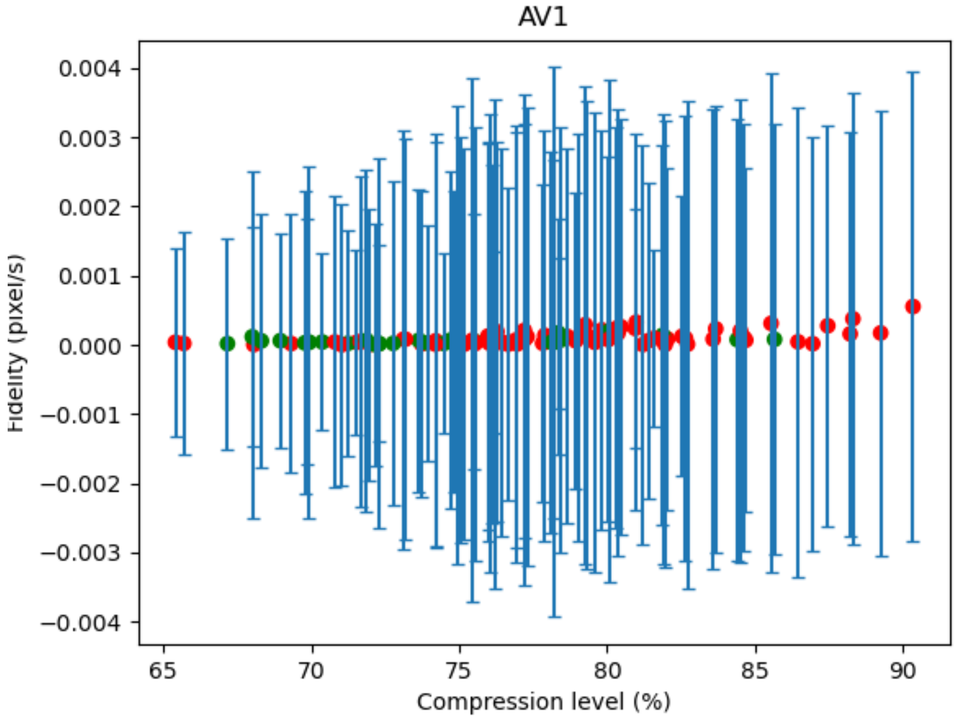
APS physics

Velocity

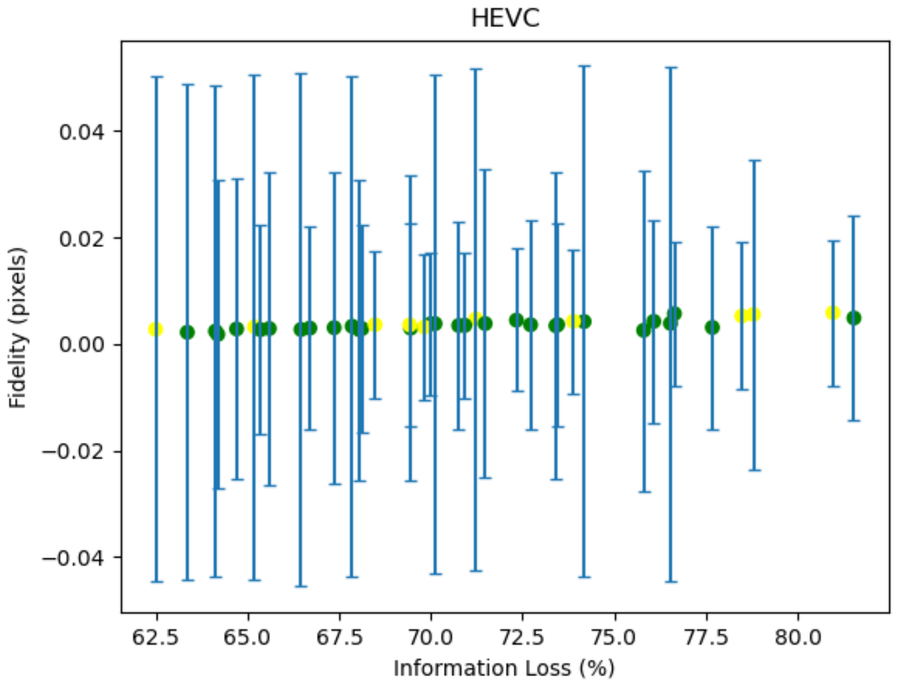
- Uncompressed success of only 32 ± 5%
- Pixel-locking is an issue

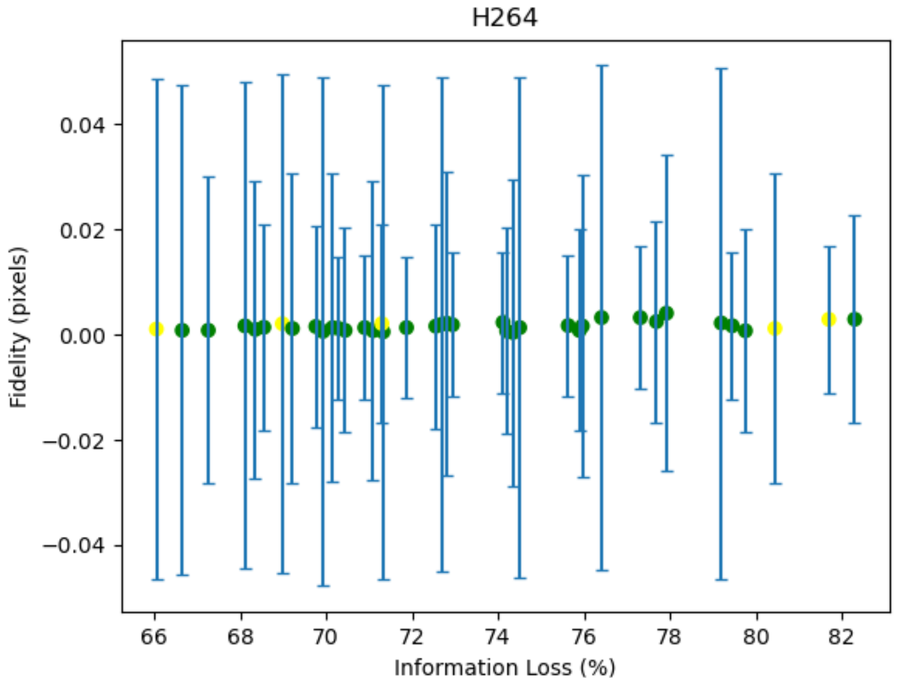


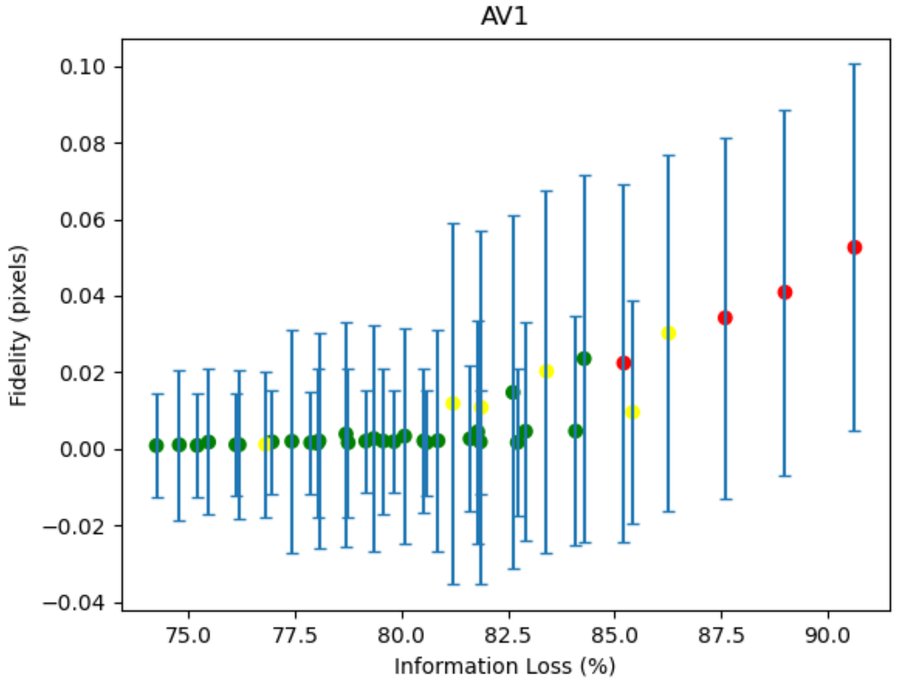




2D Gaussian Fit







Shot Noise

- Very little information to lose