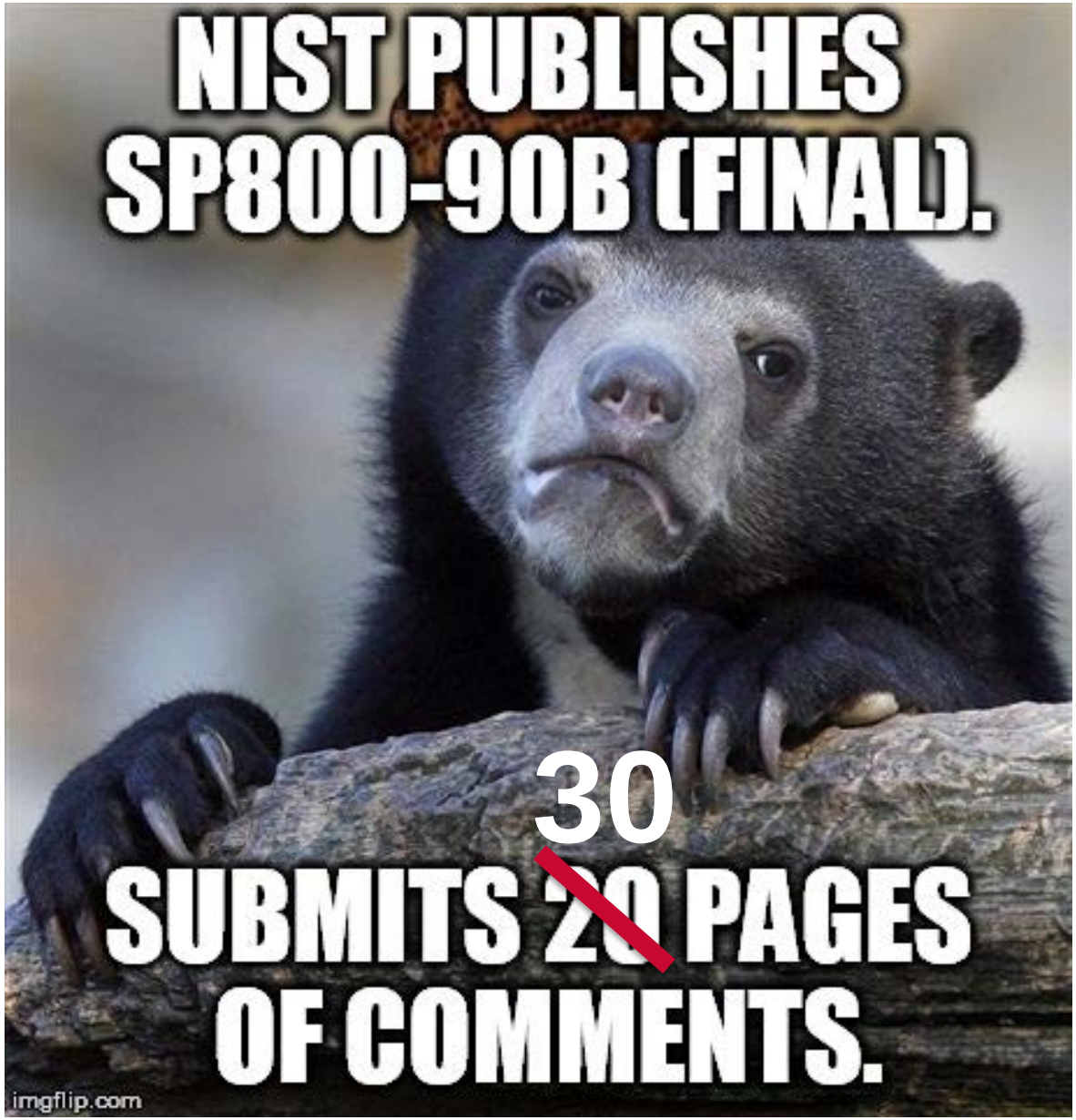
Care and Feeding of the SP800-90B Statistical Tests

UL VS, LLC.
Dr. Joshua Hill

What I Want to Say Today

What I Actually Have to Say Today



Main Comments

1. High-entropy noise sources fail the Restart Sanity Check more than expected. (SP800-90B §3.1.4.3)
2. Entropy and noise sources are required to have the same entropy rate across all process characteristics and all environmental conditions, and are required to be stationary. (SP800-90B §3.2.1 #3 and §3.2.2 #2)
3. The definition of “noise source” makes the described assessment strategy problematic. (SP800-90B §3.1.6)

(+ many other minor technical comments; see our public comments for details.)

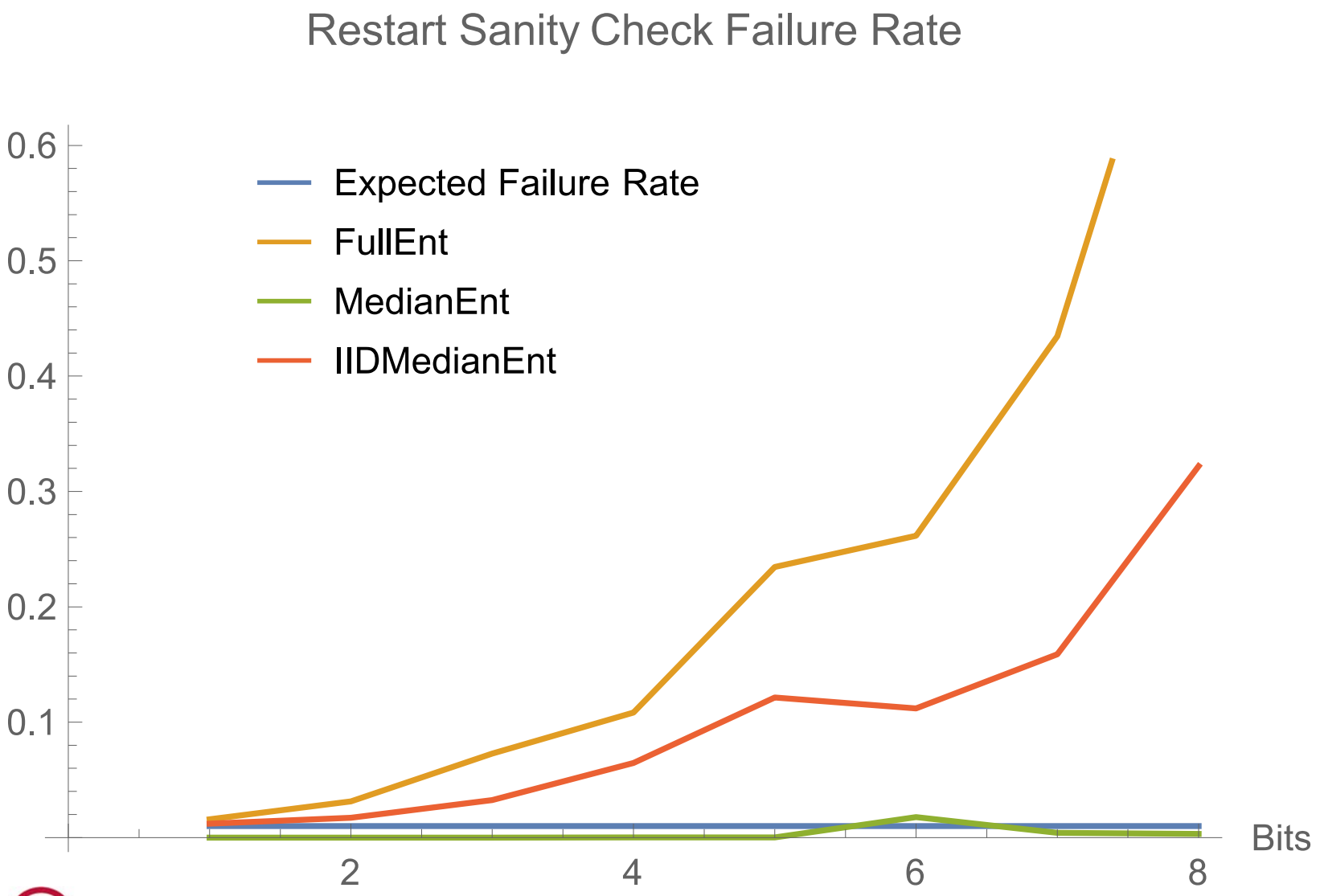
The Restart Sanity Test

- Capture the first 1000 symbols across 1000 separate restarts and store them in a 1000x1000 matrix.
- For each row/column in the matrix, count the number of occurrences of the most common symbol in that row / column. Take the maximum across all rows/columns.
- Calculate a p-value for this maximum based on the assumption that this maximum binomial distribution has a binomial distribution.
- If this test indicates a failure, then the lab/vendor is prohibited from crediting the noise source with any entropy production.
- Probability of false reject is intended to be 0.01 (1%).

The Restart Sanity Test: Problem

- This test produces fail verdicts much more commonly than anticipated (e.g., a theoretical failure rate of nearly 100% for wide data) due to a test construction issue.
  - The per row/column maximum isn't necessarily the noise source's most likely symbol! (There are as many ways of getting a per row/column maximum as there are distinct symbols.)
  - It might be any of the other symbols.
  - More symbols means more choices, thus a higher chance of failure!
- The underlying distribution for the existing test is really the maximum count of any symbol of a multinomial distribution. That distribution is hard to work with.

The Restart Sanity Test



Restart Sanity Test Proposal: Simulation FTW!

- For the assessed entropy estimate, the tester establishes the appropriate cutoff through simulation.
- The highest cutoff (the “worst case”) occurs when as many symbols as possible have the same probability as the most probable symbol.
- We found that performing 2,000,000 rounds of simulation of the 1000-sample test (analogous to the per-row/column test) provided stable results.
- This is really quick.

Assessment Stability

- SP800-90B requires that all instances of the noise source must behave essentially the same way across all per-part and environmental conditions within its operational range, and be stationary.
- This isn't true for any noise / entropy source that we've ever encountered.
  - Most of the physical sources have: substantial part-to-part variation due to manufacturing variations and substantial temperature and voltage sensitivity, and some depend on the frequency of an external clock.
  - Most non-physical noise sources are dependent on the computer's workload, etc.
  - Some noise sources "seek" high entropy states; this behavior makes the noise source non-stationary.
- Almost no commercially produced noise / entropy sources are capable of passing these requirements.

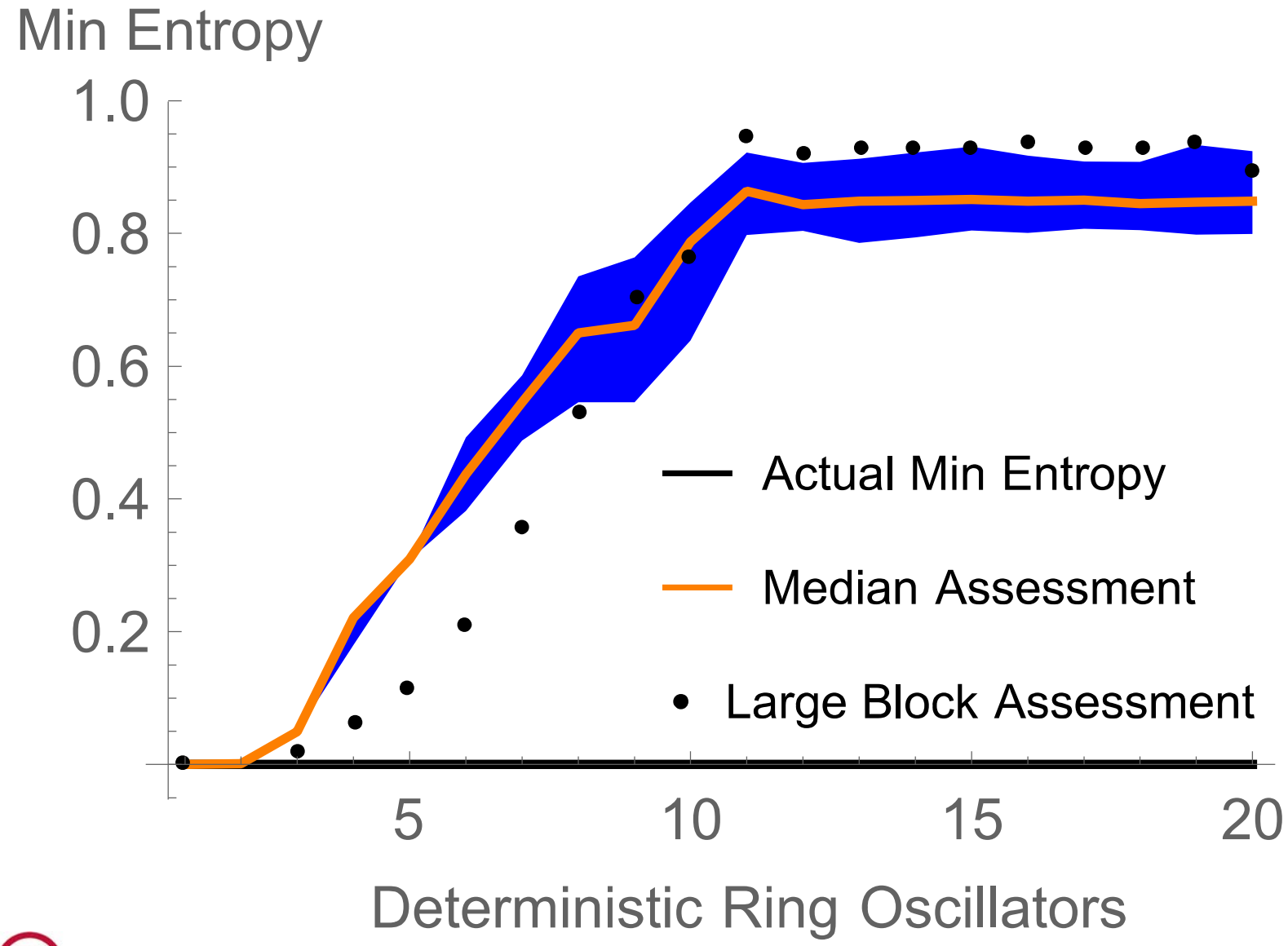
Stability: Suggestions

- The behavior of almost all noise sources is dependent on some set of *entropy-relevant parameters*.
- Require that the vendor produces a list of all such *entropy-relevant parameters*.
- Require assessment across the expected range of entropy-relevant parameters (e.g., across a temperature / voltage / process characteristics envelope).
- The final assessed min entropy value is the smallest assessed value for any entropy-relevant parameters within the expected range.

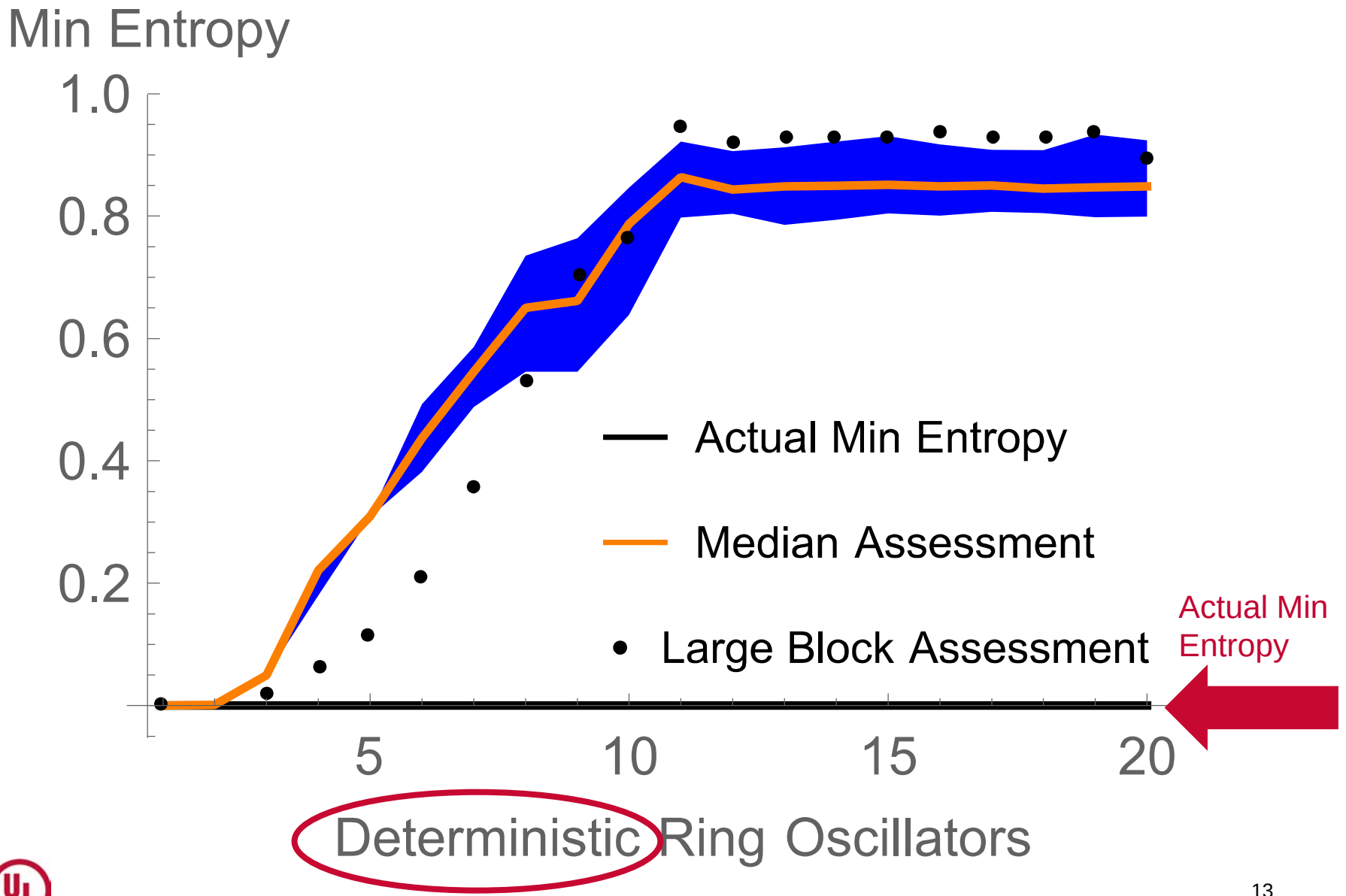
Noise Source Definition

- A “noise source” output can be the XOR of the output of “multiple copies of the same physical noise source”.
- This is problematic for statistical assessment.
  - The XOR of the output of a small number of jitter-free oscillators passes most statistical tests (even though there is absolutely no entropy present). [BBFV 2010]

Noise Source Definition



Noise Source Definition



Noise Source Definition: Proposed Resolution

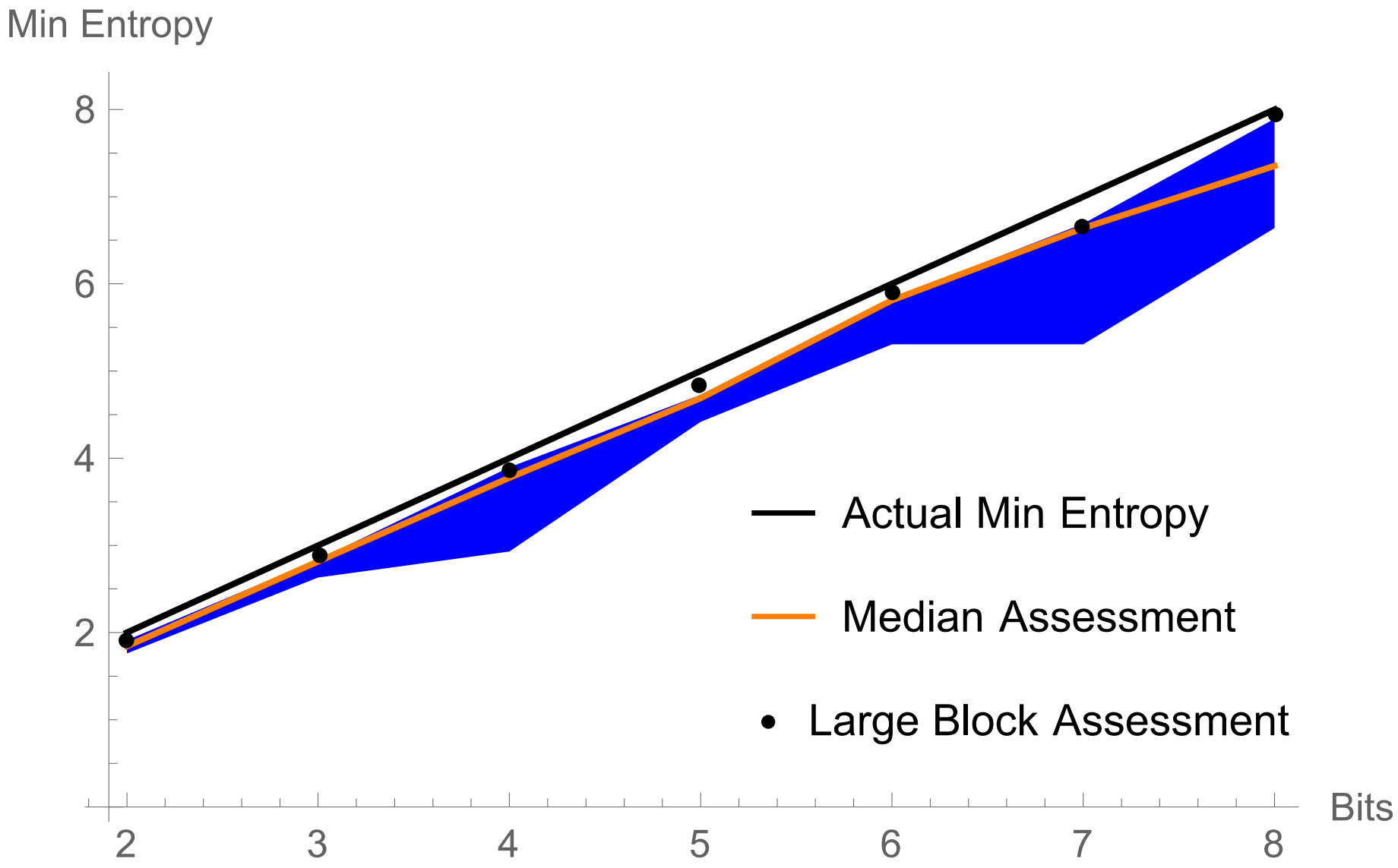
Don’t do that.

SP800-90B: Overview of Performance

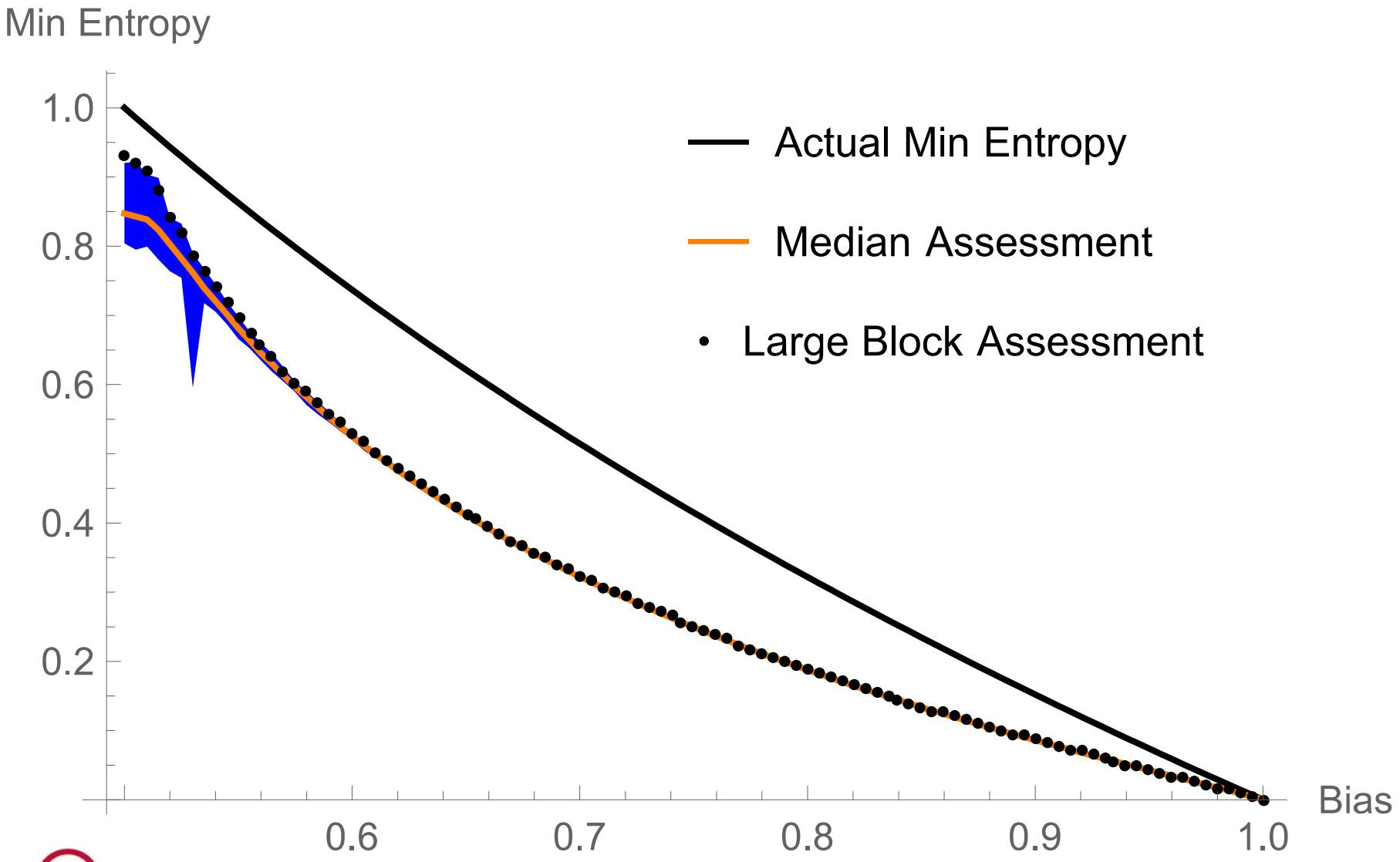
- We've constructed various simulated noise sources.
- We've constructed various models.
- Something something…

- This work is a larger scale version of DJ Johnston's 2017 work using NIST's reference python implementation (which is based on the draft 2016 document)
- This testing occurred using only the full set of **non-IID** tests.
- For each parameter setting, the results represent 100 tests of 1 million samples each, and a single test of 100 million samples (the "large block assessment")
- Blue regions show the range of assessments. Green regions reflect modeling range.

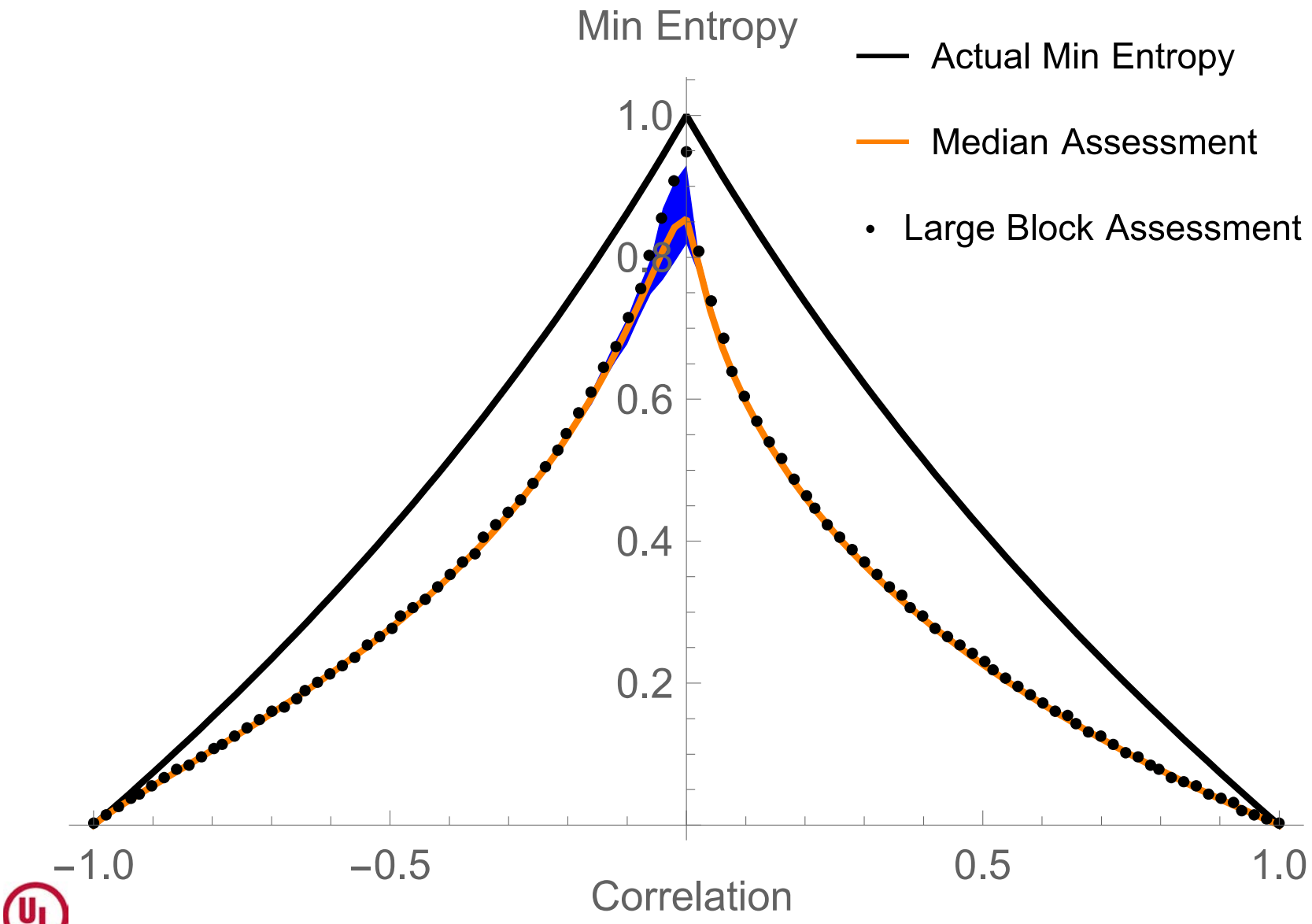
Simulated Ideal Source



Biased Bit Source

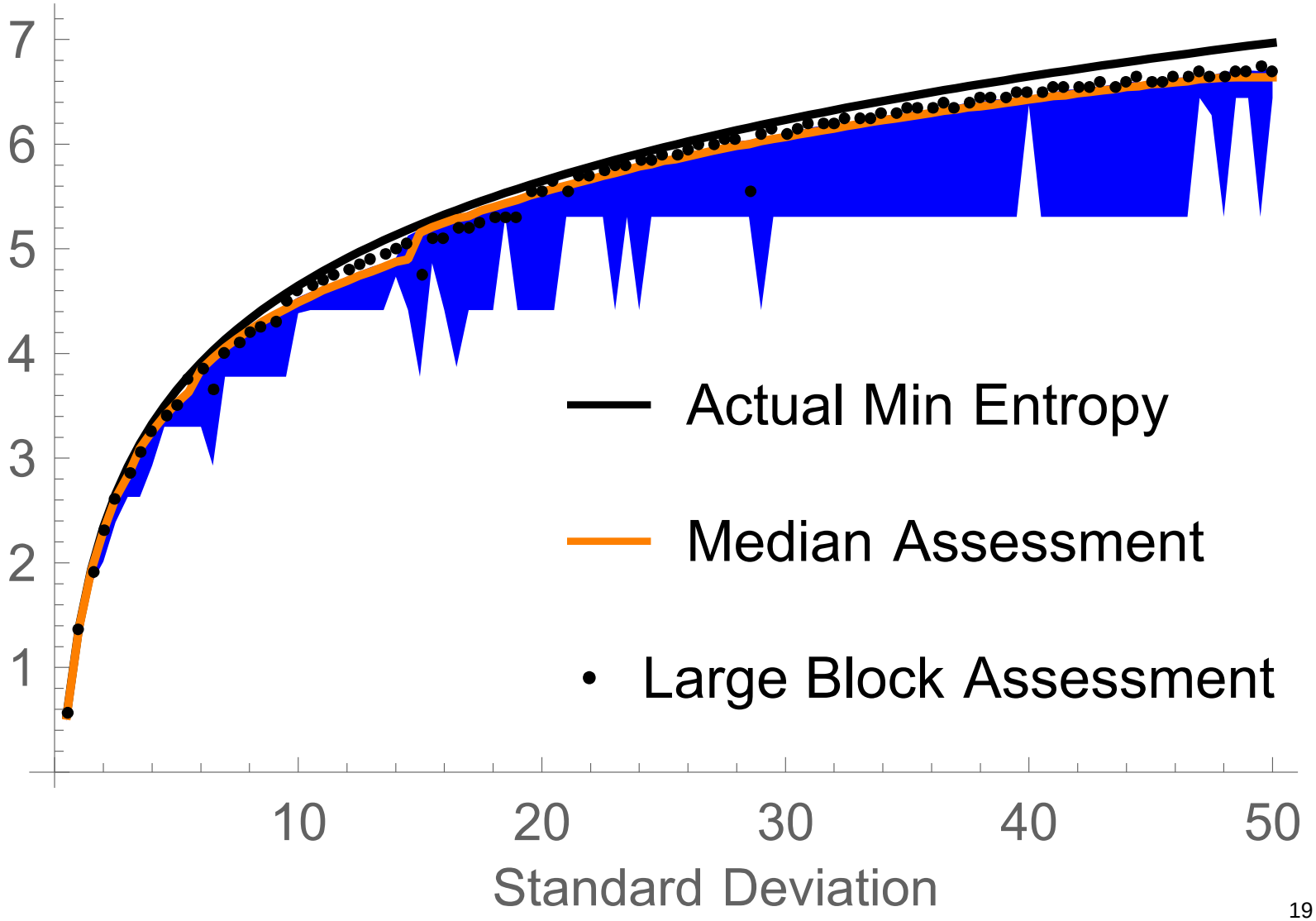


Correlated Bits



Gaussian Noise Source, 8-bit ADC

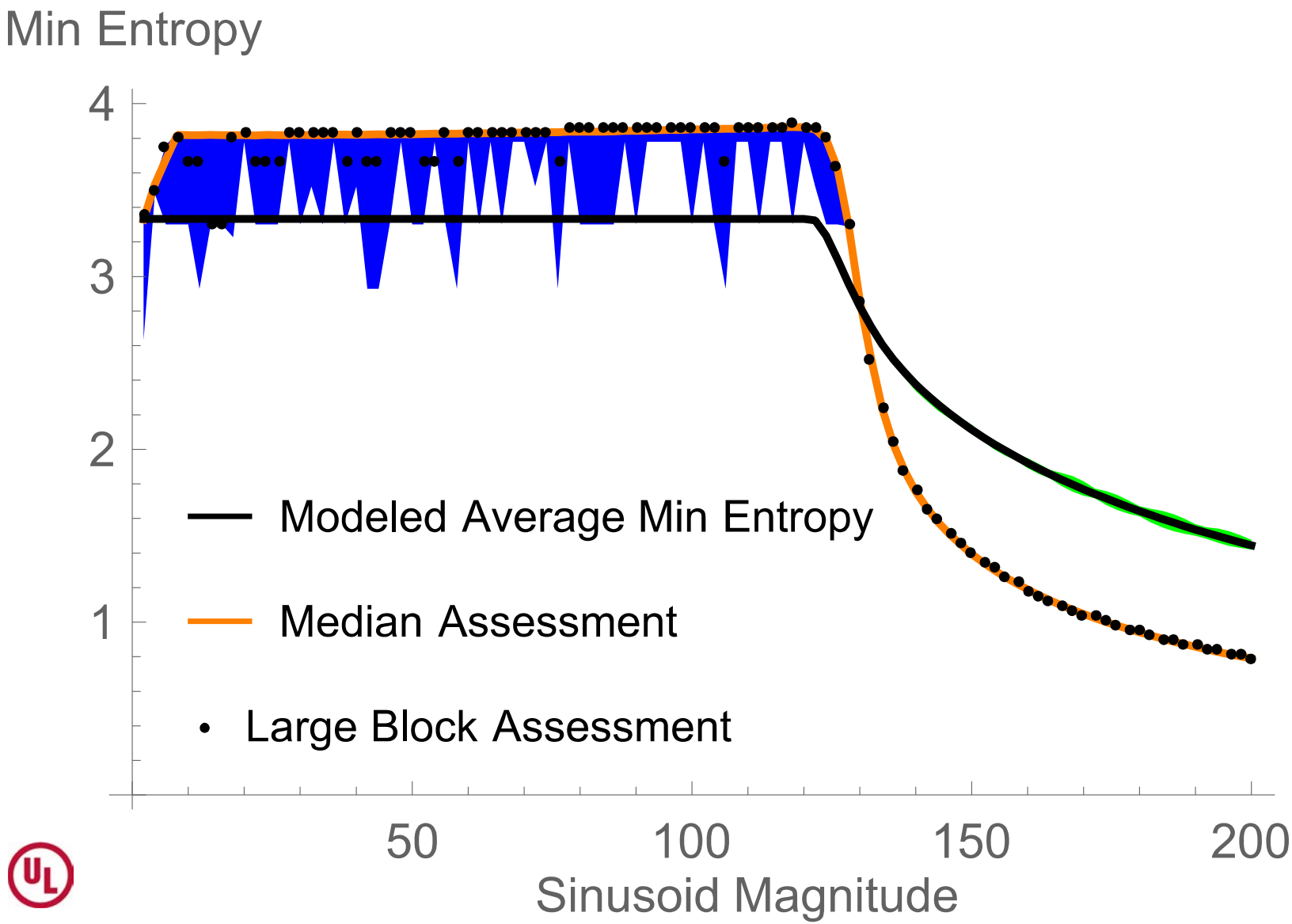
Min Entropy



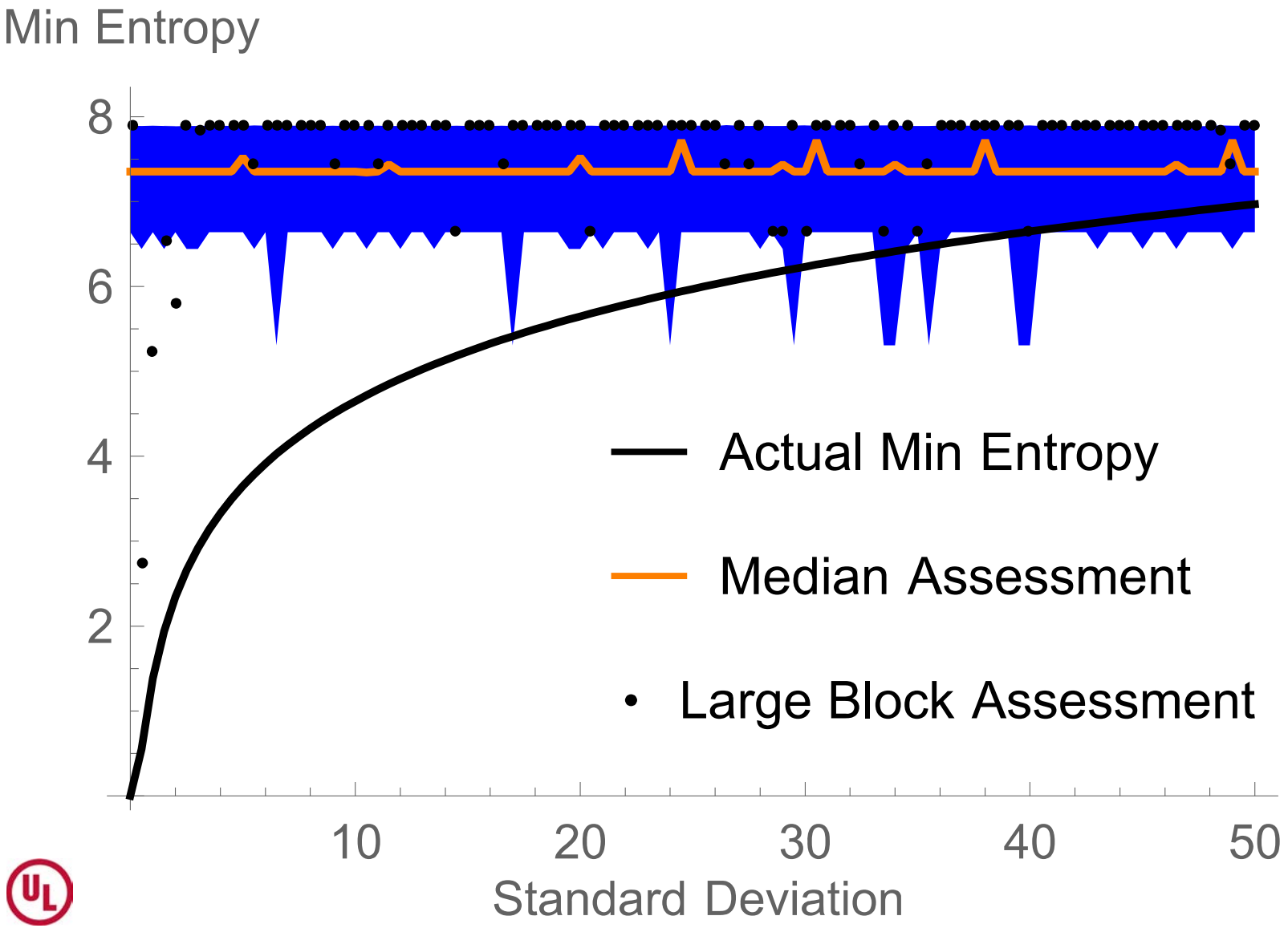
Comments Thus Far

- These are all actually either IID sources, or sources where the dependency can be easily teased out by the 90B statistical tests.
- We would be surprised if the tools overestimate the entropy in the prior cases.
- The assessments seem to generally track in a pleasing way.
- The tests seem well behaved in each of these cases.
- Large block assessments don’t appear to be a significant advantage here.

Narrow Gaussian Noise Source, 8-bit ADC, Sinusoidal Bias



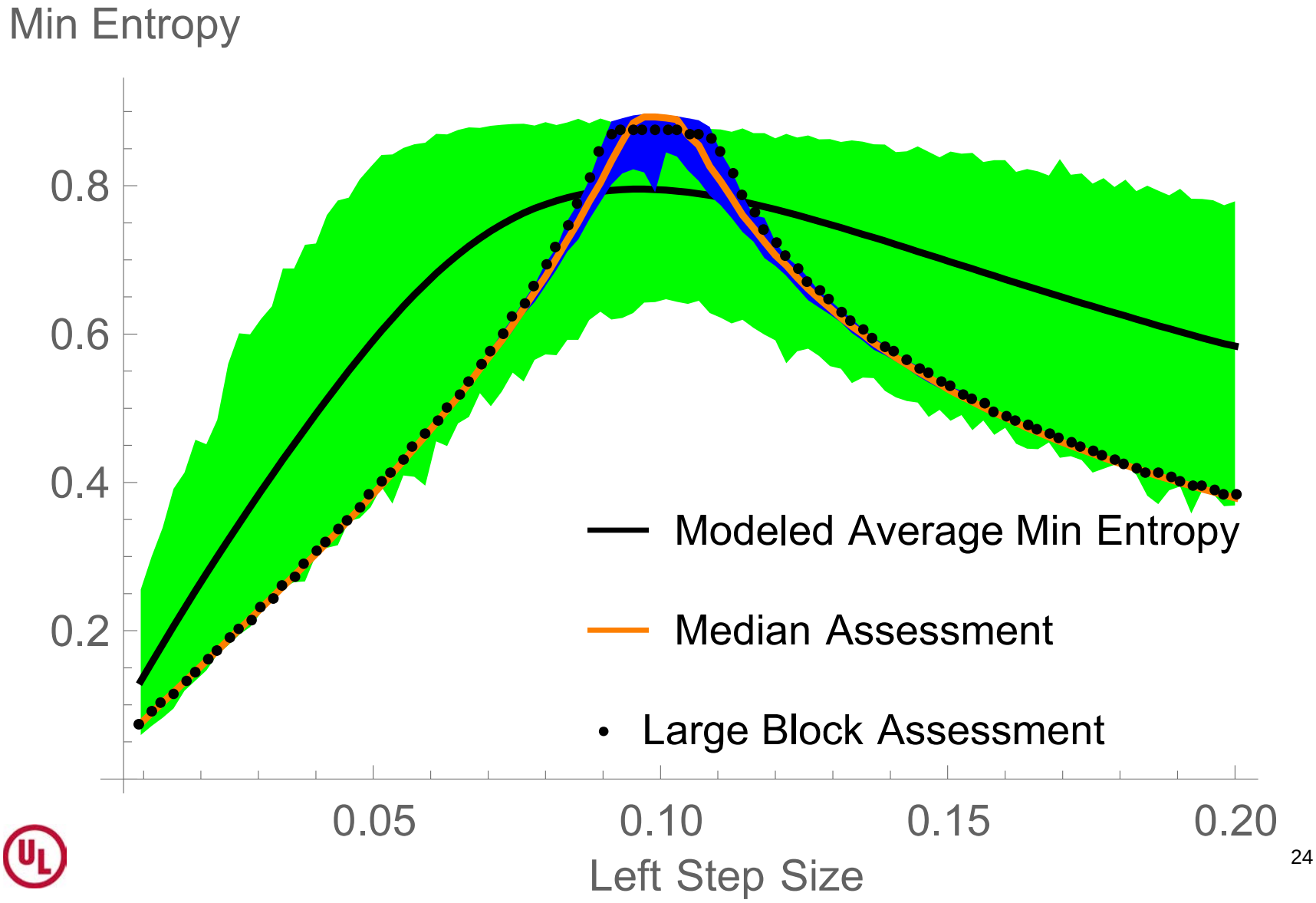
Gaussian Noise Source, 8-bit ADC, LFSR Processed



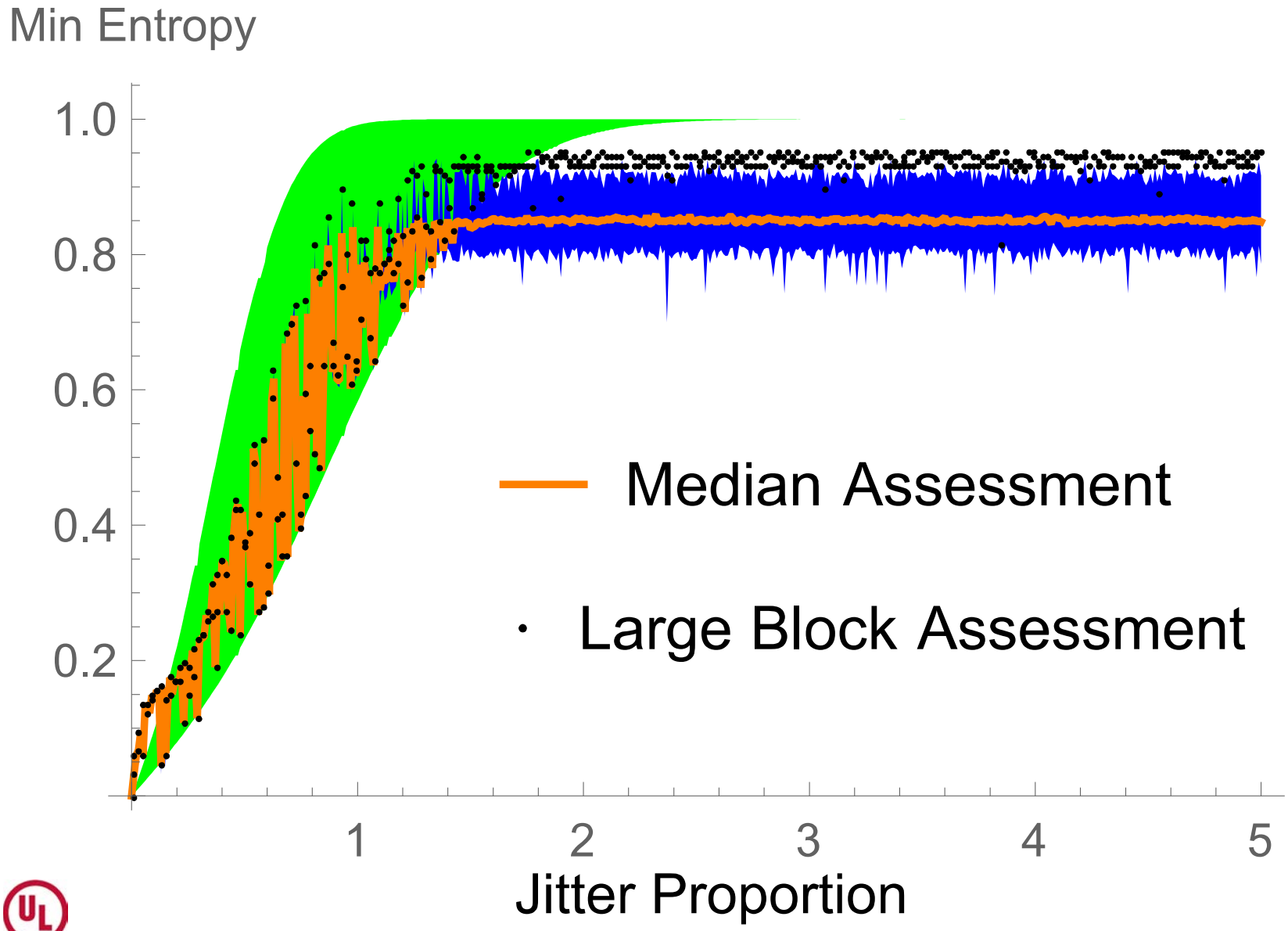
Some Complications

- Even small additions of wholly deterministic variation induce substantial overestimates of entropy.
- It is **vital** to test only raw data, and to filter out extraneous signals.
- Don’t perform statistical testing on conditioned data!
- Large block assessments don’t seem to offer a major advantage here.

SUMS Model, 384-bit Block Size



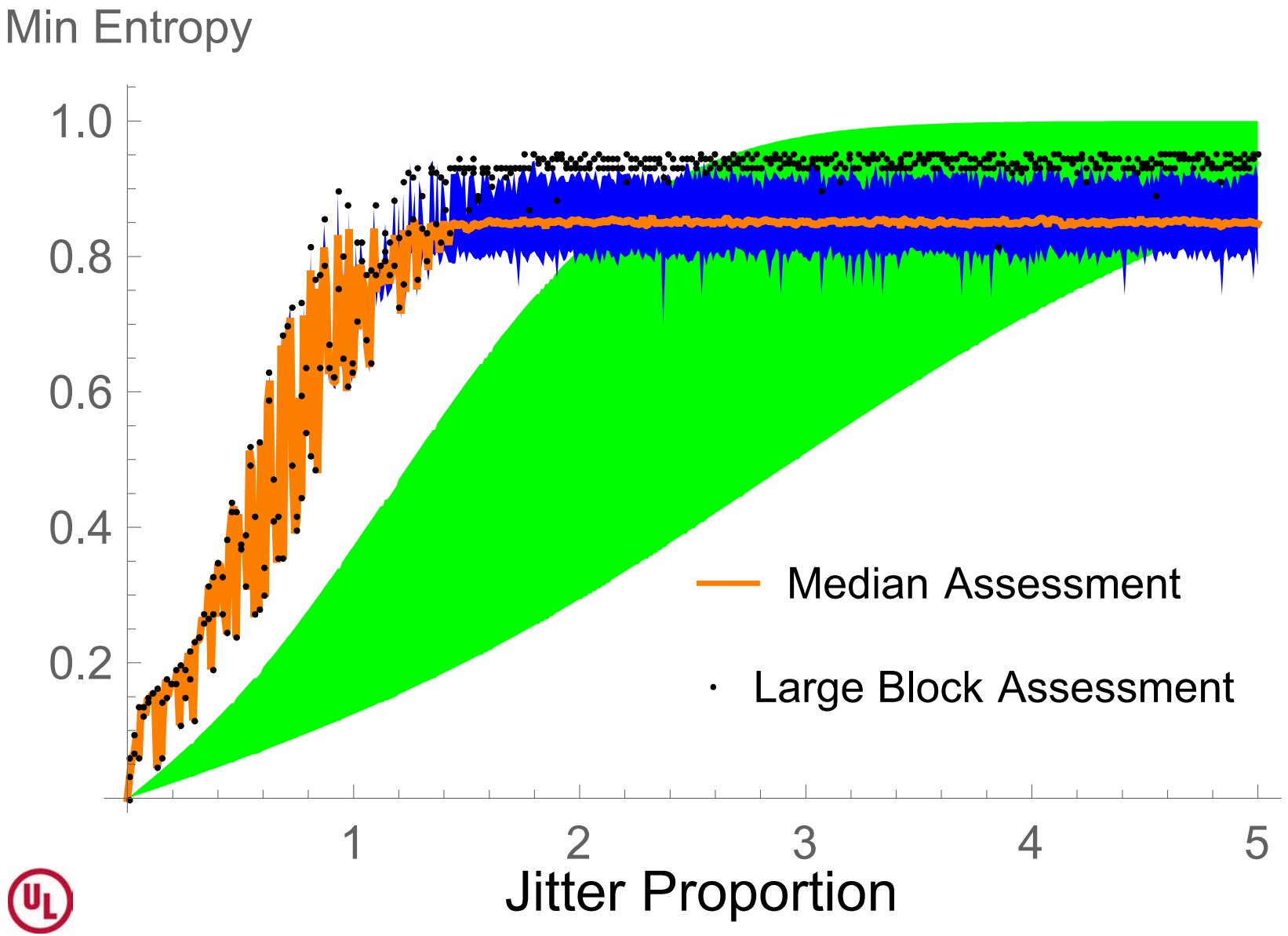
Idealized Ring Oscillator



Curiouser Still…

- These models are now somewhat complicated, and can return a range of entropy values for each parameter.
- The assessed values generally lie within the expected ranges, but **the lower end of the modeled range is the value that ought to be used**; this is lower than the value produced by the SP800-90B tests.
- These model ranges can themselves vary with data block size being accounted for.
- Large block assessments don’t seem to offer a major advantage here.

Practical Ring Oscillator [BLMT 2011]



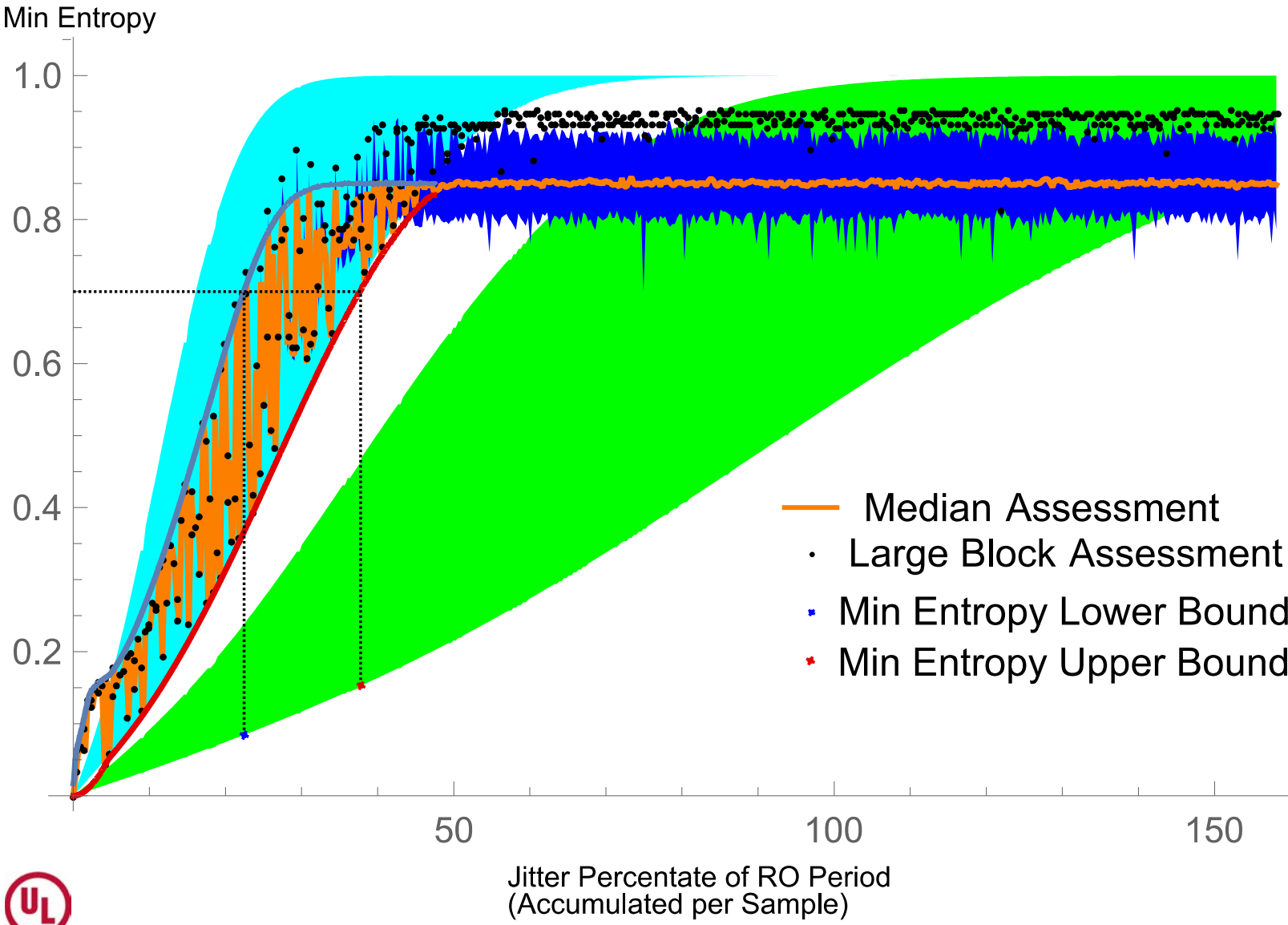
Well, “Leap” does RHYME with “Weep”…

- The results of testing for a particular source forms a distribution. Single results aren’t very meaningful.
- When we account for predictable and/or “worst case” behavior, things get worse.
- The statistical assessment doesn’t consistently underestimate the modeled min entropy in more complicated systems.
- These suggests that assessment of non-trivial non-IID sources should commonly be further reduced.
- In the scenarios we tested, performing statistical analysis on large blocks didn’t seem to offer any significant advantage over taking the median over many smaller assessments.

More Ring Oscillator!

- A particular statistical assessment corresponds to a range of possible jitter percentages.
- We can deduce a lower bound for the per-sample jitter percentage from the statistical testing results.
- If the vendor can model the local Gaussian jitter %, g, use that, otherwise, bound at 30%.
- Run statistical testing on a large sample of output from the ring oscillator, and use these results to establish a lower bound for the overall per-sample jitter percentage, σ.
- Use this $g\sigma$ within a ring oscillator model. This model produces a lower min entropy bound appropriate for use as $H_{\text{submitter}}$

Practical Ring Oscillator [BLMT 2011]



References

- [BLMT 2011] Baudet, Lubicz, Micolod, and Tassiaux. *On the security of oscillator-based random number generators*. Journal of Cryptology, April 2011, Volume 24, Issue 2.
- [BBFV 2010] Bochard, Bernard, Fischer, and Valtchanov. *True-Randomness and Pseudo-Randomness in Ring Oscillator-Based True Random Number Generators*. International Journal of Reconfigurable Computing, Vol. 2010.
- [HKM 2012] Hamburg, Kocher, and Marson. *Analysis of Intel's Ivy Bridge Digital Random Number Generator*.
- [HJ 2018] Hill and Jackson. *NIST Special Publication 800-90B Comments*. http://bit.ly/2jwKN9R
- [J 2017] Johnston, David. *STS-2.1.2 and SP800-90B Assessment Suite Anomalous results*. https://github.com/dj-on-github/90B_check
- [SP800-90B] Turan, Barker, Kelsey, McKay, Baish, and Boyle. *Special Publication 800-90B: Recommendation for Entropy Sources Used for Random Bit Generation*. January 2018.

THANK YOU.