University of Kent

Great Expectations

A Critique of Current Approaches to Random Number Generation, Testing, and Certification

Darren Hurley-Smith & Julio Hernandez-Castro

University of Kent | Computing

Who are we?

- Prof Julio Hernandez-Castro, University of Kent
- Dr Darren Hurley-Smith, University of Kent

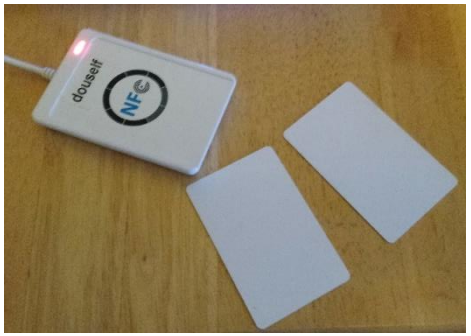
- Research interests:
  - Statistical testing of random number generators
  - Design of new, more robust tests
  - Non-deterministic random number generation
  - Certification and standards

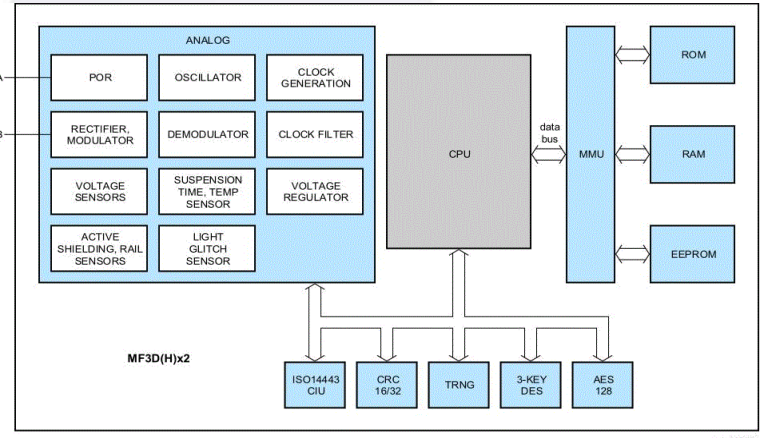
Introduction

- We've been working on this are for a while
- Published a couple results
  - ***Certifiably Biased: An In-Depth Analysis of a Common Criteria EAL4+ Certified TRNG***. D Hurley-Smith, J Hernandez-Castro. IEEE Transactions on Information Forensics and Security 13 (4), 1031-1041, 2018
  - ***Quam Bene Non Quantum: Bias in a Family of Quantum Random Number Generators***. Darren Hurley-Smith and Julio Hernandez-Castro https://eprint.iacr.org/2017/842 and RWC 2018
- And seen many a thing we don't like ~ heavy customer bias
- This presentation is a list of criticisms that reflect all our moaning and whining over the years, hoping to inform better future testing and certification schemes

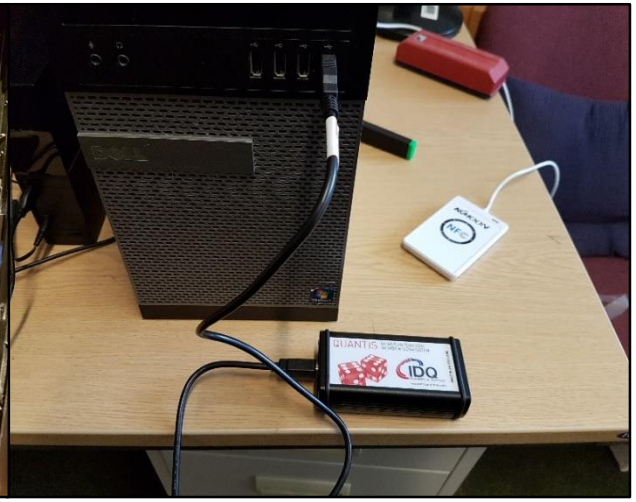
Our Previous Research

- Studying RFID security
- Analysis of small TRNGs
- Identified biases in the EV1 TRNG
  - EV1 is CC EAL4+ certified
  - Responsible disclosure
- Identified bias in Quantis RNGs
  - Presented initial findings at RWC 2018
  - Self-certified, seller shows passes tests
  - Post-processing is essential for QRNGs
  - Responsible disclosure





Some of our other targets



Main issues with current certification schemes

- Only identify egregious failures
- Randomness tests are highly correlated, and research is very limited in the area
- Engineering towards ‘just about’ to pass tests, and ‘just about’ to get the desired certification level
- Closed hardware designs can be certified!
- No analysis of raw entropy, but only sequences after postprocessing
- Certification can be performed over a single device, despite selling millions of them, no manufacturing quality assessed
- Poor understanding of randomness: virginal, binary take instead of an engineering take
- Randomness tests used in certification are a sitting duck
  - Allowing for easy adversarial attacks
- The market is too concerned with speed

Certification, Standards, and Testing

- NIST
  - SP800-90B outlines properties befitting NIST approved entropy sources
  - SP800-22 provides a comprehensive series of statistical tests
  - SP800-22 is still used independently by many manufacturers
- Common Criteria
  - European standard ISO/IEC 15408: a broad set of standards relating to computer security
  - Evaluation Assurance Level (EAL) scheme is a crucial ‘whole device’ evaluation methodology
  - AIS-31 (authored by BSI) provides guidelines and tests for accepted entropy sources
- Some widely used statistical test batteries
  - Federal Information Processing Standard (FIPS) 140-2
  - NIST SP800-22
  - Marsaglia’s Diehard and Tufftests tests
  - Dieharder: Diehard and NIST SP800-22 tests
  - L’Ecuyer’s TestU01
  - BSI’s AIS-31
  - SP800-90B entropy estimation tests (IID and non-IID)

Manufacturer reported testing

Manufacturer	Device	Cost (euros)	Entropy source	Certifications and Tests
NXP	DESFire EV1	0.59	Not disclosed	CC EAL 4+
	DESFire EV2	1.25	Not disclosed	CC EAL 5+
IDQ	Quantis 16M	2,900.00	Beam splitter	NIST SP800-22, METAS, CTL
	Quantis 4M	1,299.00	Beam splitter	NIST SP800-22, METAS, CTL
	Quantis USB	990.00	Beam splitter	NIST SP800-22, METAS, CTL
Comscire	PQ32MU	1211.00	Shot noise	NIST SP800-90B/C, SP800-22, Diehard
Altus Metrum	ChaosKey	45.00	Reverse biased semiconductor junction	FIPS 140-2

Data collection

Device	# samples	Sample size (MB)	Mean data rate (Mbit/s)
DESFire EV1	3	64	-
	100	1	-
DESFire EV2	1	64	-
Quantis 16M	100	2100	15.87
Quantis 4M	100	2100	3.86
Quantis USB	100	2100	3.96
PQ32MU	100	2100	30.99
ChaosKey	10	2100	3.80
urandom	100	2100	-

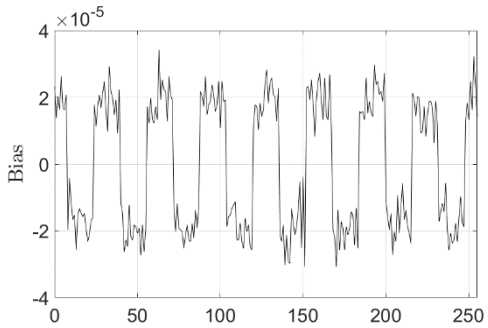
Testing diversity

- Relying on a single battery of tests is not advisable
  - NIST SP800-90B periodically revises their recommended tests
  - IDQ, NXP and Comscire all publish results over multiple batteries (with caveats)

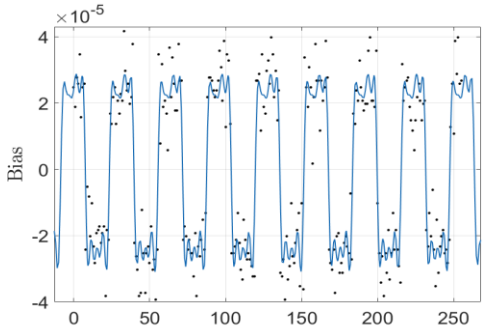
Device	Dieharder (% passed)	NIST SP800-22 (% passed)	TestU01 Alphabits (% passed)	TestU01 Rabbit (% passed)	TestU01 Small Crush (% passed)	TestU01 Crush (% passed)
Q 16M	100	100	54	60	93	47
Q 4M	100	100	3	7	91	3
Q USB	100	100	3	21	89	3
PQ32MU	100	100	91	86	93	84
ChaosKey	100	100	90	90	90	80
urandom	84	100	96	96	92	79

- We present results of Dieharder, NIST SP800-22 and TestU01
  - Dieharder is passed by almost all tested sequences
  - All sequences pass NIST SP800-22
  - TestU01 shows a much greater variance in test results

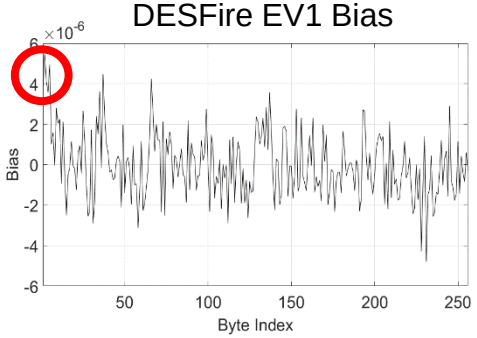
Tests as simple as χ^2 can identify bias



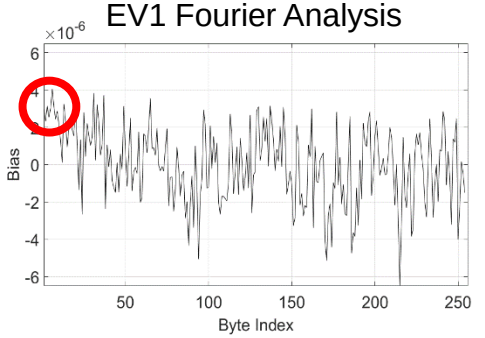
DESFire EV1 Bias



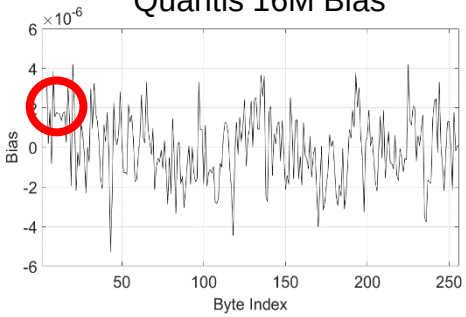
EV1 Fourier Analysis



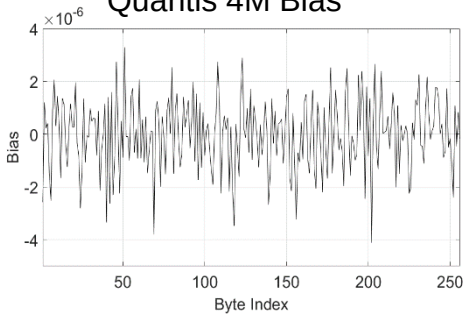
Quantis 16M Bias



Quantis 4M Bias



Quantis USB Bias



urandom Bias

Manufacturer Testing

Device	Diehard	NIST SP800-2	TestU01	"Self-Tested"
Quantis 16M	✓	✓	✗	✓
Quantis 4M	✓	✓	✗	✓
Quantis USB	✓	✓	✗	✓
PQ32MU	✓	✓	✗	✓
ChaosKey	✓	✓	✓	✗

- Diehard and NIST used by all manufacturers for listed devices
- IDQ and Comscire use 'home-brew' tests
  - They claim these tests are more rigorous than NIST/Diehard
- Hardware-RNG test batteries such as TestU01 not used
- PQ32U is 'guaranteed to pass ANY test' ~ "military grade encryption"

Number of samples and their size

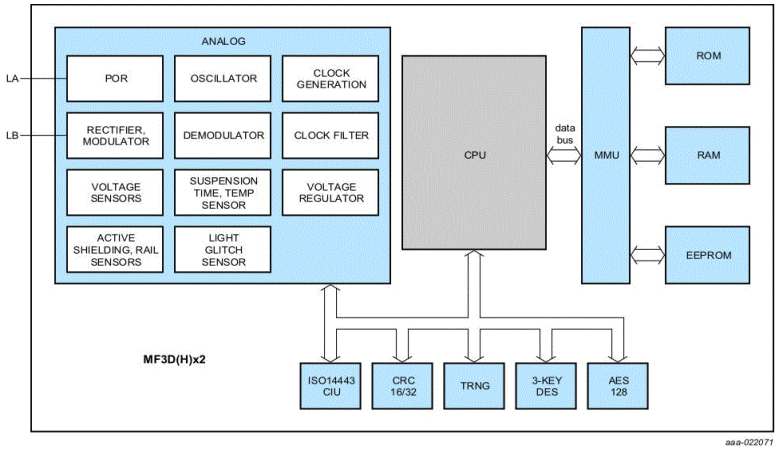
- Quantis devices
  - 1 billion bits tested using NIST SP800-22 (recommended value)
  - Diehard uses the same sample size
  - ‘Large files’ mentioned in official documentation but no how large
  - METAS and IDQ Randomness Test Report v2.0 2010 reports only mention 4M

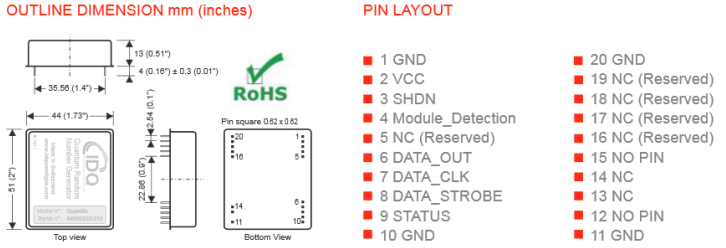
- Comscire PQ32MU
  - Two sample sizes mentioned: 80 million and 1 million bits
  - Test selection & parameters modified to suit small sample size: not standard
  - SP800-22 reports 188 tests statistics, Comscire only reports 148 of them
  - No explicit mention of whether results are from a single sample or multiple ones

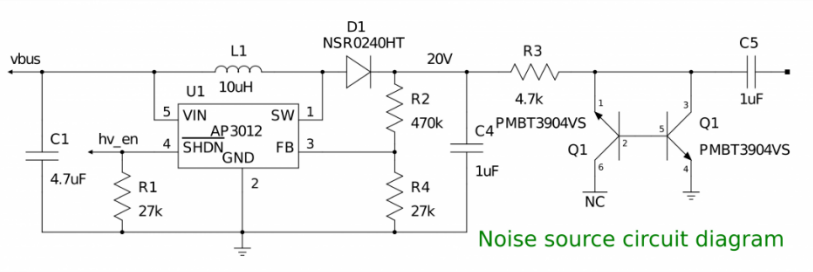
- Neither manufacturer states how many devices were tested
  - Selection criteria not disclosed
  - It is strongly implied that single-device testing was used for self-testing
  - It is also strongly implied that 3rd party testing also tolerates single-device testing
  - Both companies definitely perform QA on finished devices, why not in these tests?

The dreaded Blackbox

- Public disclosure is rare
  - Intellectual property a priority
  - NXP (upper left) and IDQ (middle) provide only general diagrams
  - This makes independent hardware evaluation much harder
- Required for CC EAL certification
  - NDA protected disclosure
  - Provides a degree of ‘independent’ evaluation
  - Still only 1 additional assessor
  - Open-standards benefit from crowd-testing
- A manipulated RNG can pass tests
  - A simple counter can pass FIPS 140-2 as long as >34% of values are randomly generated
- ChaosKey is open hardware design







Conclusion and Recommendations

- We should only certify open hardware designs
  - At the very least, make the reasoning for the design and the entropy gathering public
- Analyse the raw data, not just the postprocessed/whitened/unbiased/cleaned one after hashing
- Don't base certification on a single device, take into account expected market to check also the manufacturing quality into account.
  - My proposal is to check at least sqrt(sqrt(n)) when n is the number of sold devices until next certification, so ~177 for 1 billion, or ~32 for 1 million, 10 for 10.000
  - Bernstein, D. J., Chang, Y. A., Cheng, C. M., Chou, L. P., Heninger, N., Lange, T., & Van Someren, N. (2013, December). **Factoring RSA keys from certified smart cards: Coppersmith in the wild**. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 341-360). Springer, Berlin, Heidelberg.
- See randomness certification as an engineering problem, not a pure mathematical one, where binary answers are possible
  - Certify bias bounds, no perfect randomness
- Independence/correlation of randomness test is a pressing issue, particularly in embedded devices (selecting sets of tests that require only a small footprint, etc.)

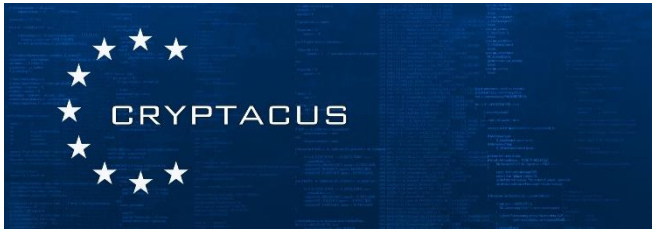
Conclusion and Recommendations

- Don't believe a Quantum Random Number generator at face value, or a TRNG for that matter
- Self-certification is a joke that should invalidate the product claiming such analysis is the only that has been carried out
- Speed is an interesting marketing target that kind of works, but is frequently inversely correlated with security
- Keep a moving target in the tests, so that the target of evaluation is not a sitting duck and designers don't simply design with minimal security to pass these tests in mind
  - We want to catch these by using either some private tests or a sufficiently large set of thousands of tests so that optimising for them is almost impossible
  - Mrazek, Vojtech, et al. **Evolving boolean functions for fast and efficient randomness testing** Proceedings of the Genetic and Evolutionary Computation Conference. ACM, 2018.

Acknowledgements

This work received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No.700326 (RAMSES project)

We would like to thank ECOST – CRYPTACUS action for their valuable and insightful discussion of this work



We would like to convey our thanks to NXP and ID Quantique (IDQ) for their timely and professional responses to our responsible disclosure, particularly, not suing us

Thank you for listening

Questions?

THE UK'S EUROPEAN UNIVERSITY

www.kent.ac.uk

University of Kent

Introducing difficult-to-detect artificial bias

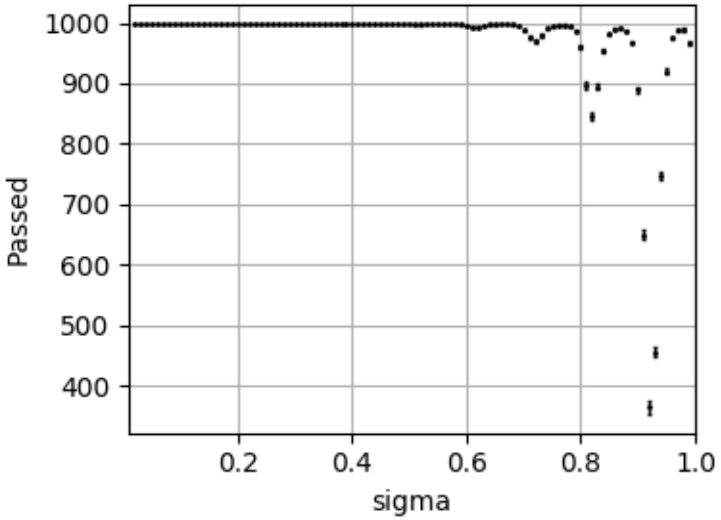
- FIPS 140-2 tests
  - 5 tests
  - Available in the rng-tools suite
  - 4 of these tests are used in the AIS-31 test suite
- Sigma counter
  - A simple counter that occurs with probability σ

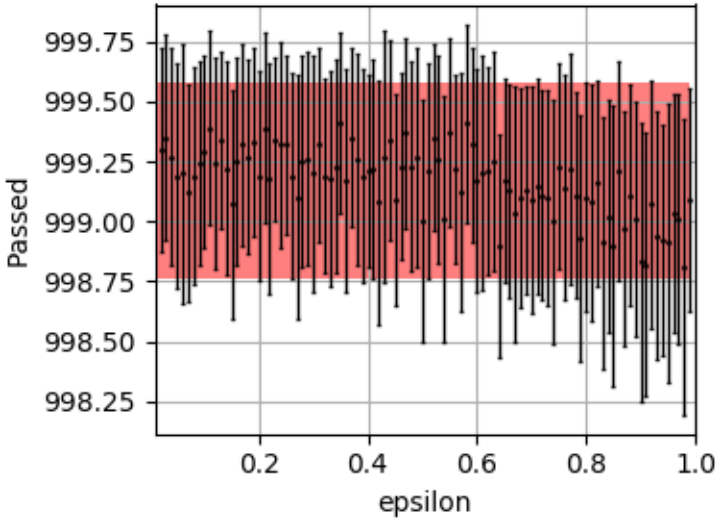
$$X_i = \begin{cases} c \mod 256, & \text{with probability } \sigma \\ R_i, & \text{with probability } 1-\sigma \end{cases}$$

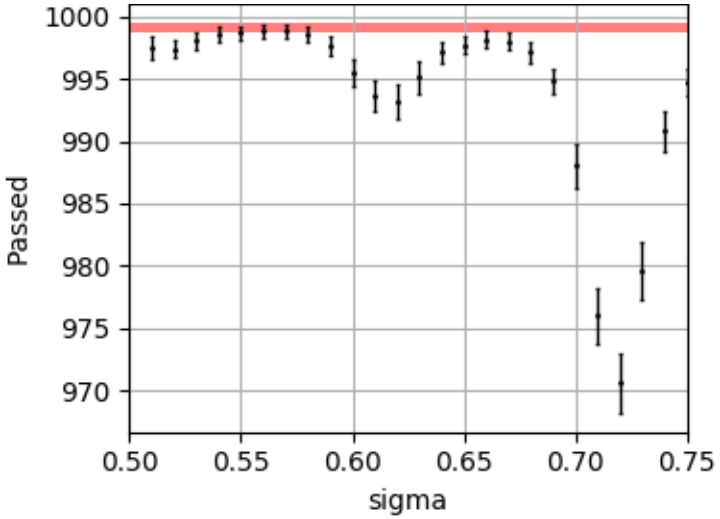
- Epsilon hole
  - With probability ϵ, a byte of value int(255) is discarded and a new byte generated

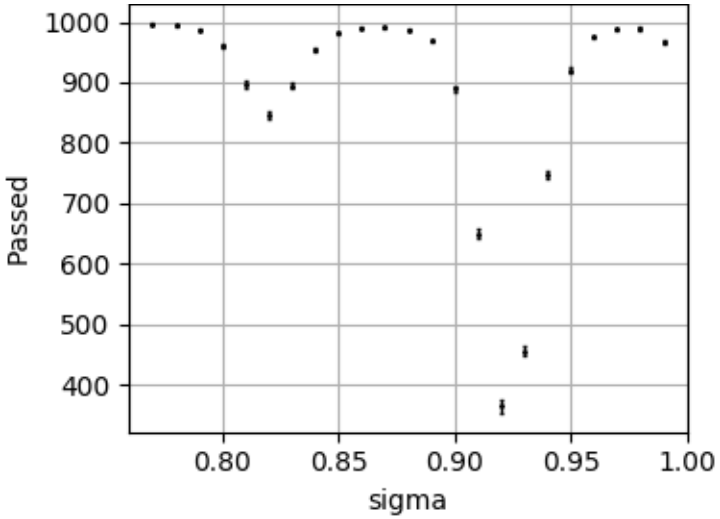
$$X_i = \begin{cases} R_i - \{0xff\}, & \text{with probability } \epsilon \\ R_i, & \text{with probability } 1-\epsilon \end{cases}$$

FIPS 140-2 Results
Sigma and Epsilon FIPS 140-2 raw pass rates



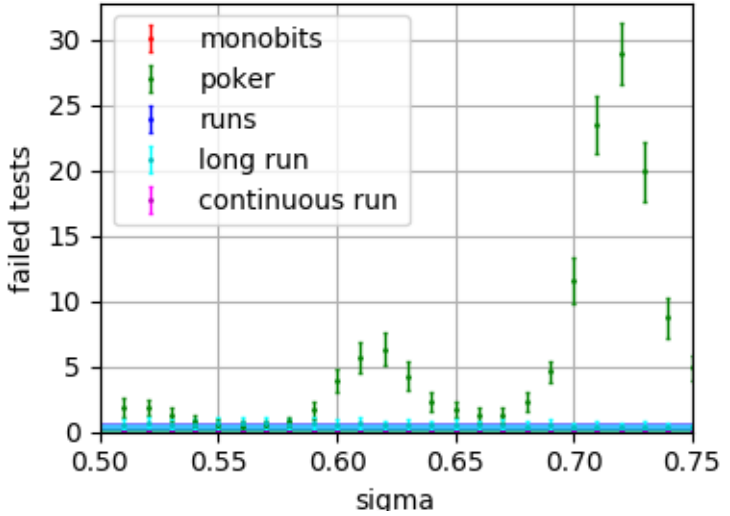




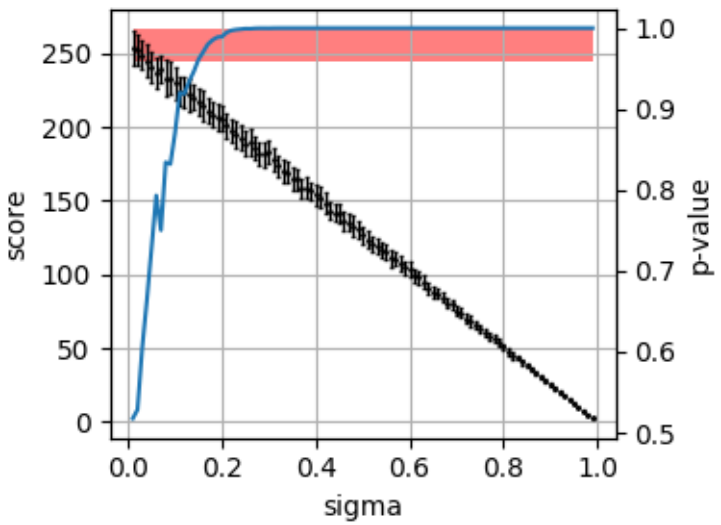


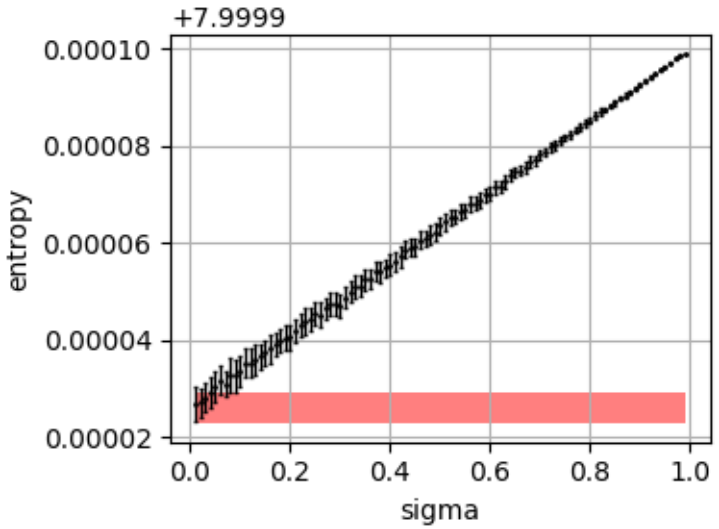
Sigma counter: Points of Interest

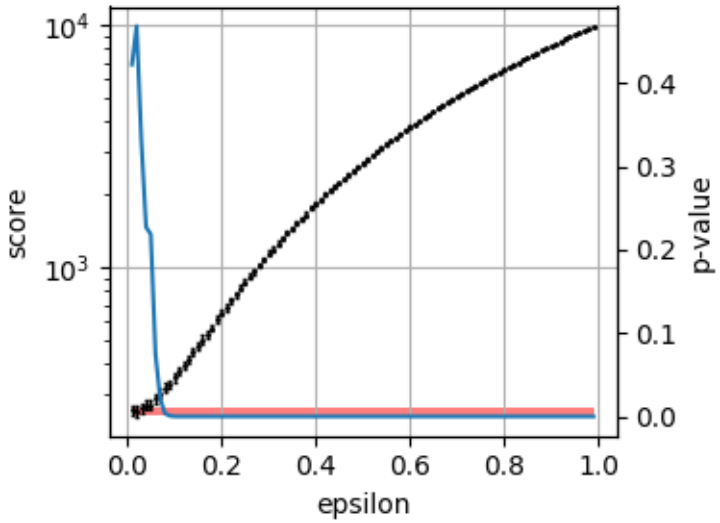
- Sigma counters were effective
  - 1000 iterations of FIPS 140-2
  - Further validated for 2GB of data
- Pass FIPS 140-2 up to 70% bias
- Only Poker test fails!
  - Some false positives at low bias
  - As bias increases, all other tests consistently pass
  - Poker consistently fails
- Poker is below battery failure threshold!
  - Maybe we need to consider distribution of failures?

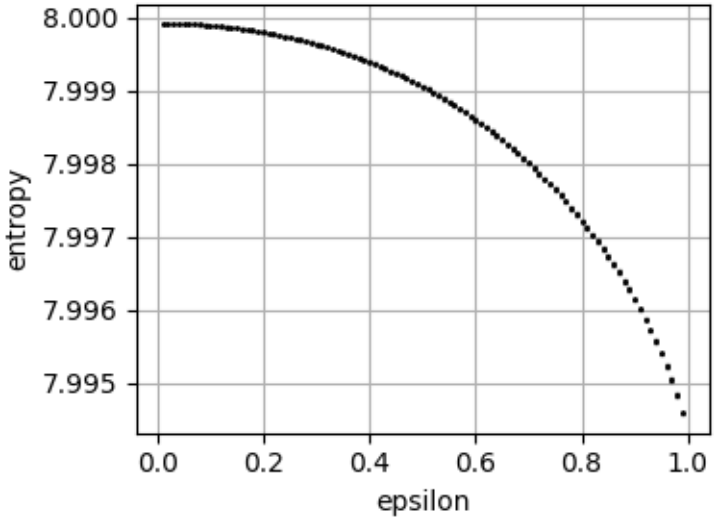


Results of Ent









Entropy Estimation

- A more realistic entropy estimate can be gained
  - Processing chunks larger than the single bits and bytes that Ent processes
- For σ of 1:
  - A 32-bit sequence possesses 8 bits of true entropy
- For σ of 0.9:
  - A 32-bit sequence possesses 8.57 bits, of 0.2678 bits of entropy per bit
- For σ of 0.5:
  - σ counters pass at this level of bias
  - 24.15 bits, or 0.7546 bits per bit, of entropy in a 32-bit sequence