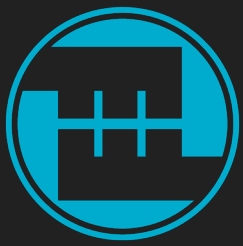
The SP 800-90B Approved Health Tests and Their Cutoffs

Joshua E. Hill, PhD

ICMC N23A
20240919

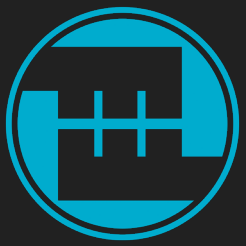
In Summary

- The SP 800-90B cutoff procedures yields APT and RCT with low statistical power for most noise source designs.
  - Designers are allowed to choose cutoffs that yields a more powerful APT and RCT (which also have false positive rates closer to the targeted values).
- The APT window size should be tailored for the noise source distribution.
  - This is also allowed.
- [Theseus] can help designers to perform the empirical cutoff and window size selection.



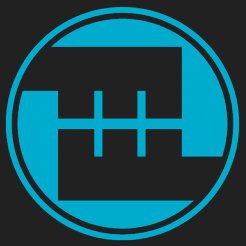
Part 0: Statistical Tests

- Statistical tests are traditionally analyzed using their *false positive rate* (α, here the probability that the test incorrectly detects a failure when the entropy source is working normally).
- From a security perspective, we're more interested in the *statistical power* $(1 - \beta)$ of the test (here the probability that the test correctly detects a failure when the entropy source is in a particular failure mode).
- This is done because there are many ways for a source to fail (so the power is failure-mode specific), but there is only one α.
- Commonly (but not always), increasing α results in an increase in statistical power and decreasing α results in a decrease in statistical power.



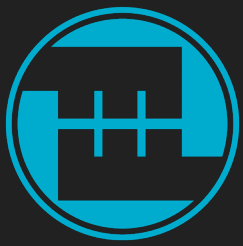
Important APT/RCT Characteristics

- The APT and RCT are categorical, so data encoding (e.g., the raw data width) isn't relevant.
- For many noise sources the distribution parameter that establishes the false positive rate for a selected cutoff is the probability of the most likely symbol, $p_{\max}$ ($p_{\max} \approx 2^{-H}$ for a stationary/IID noise source).
- Here, we discuss this probability in terms of the "apparent entropy":
  - $H_{\text{apparent}} = -\log_2 p_{\max}$
  - $H \leq H_{\text{apparent}}$ (in all sources because of the MCV estimator).
  - If the source is stationary and IID, these values are **theoretically** equal.



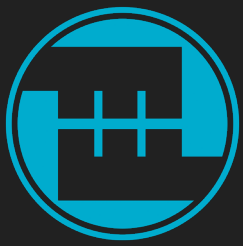
Part 1: The APT

1. A = next()
2. B = 1
3. For i = 1 to W–1
    a) If (A == next()) B=B+1
    b) If (B ≥ C) signal a failure
4. Go to Step 1



The APT Failure Mode

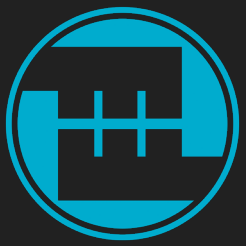
- The APT is intended to (eventually) detect the most likely symbol becoming dramatically more common than expected.
  - This situation naturally arises, even for **some** non-IID/non-stationary sources. [Hill 2023]
- The APT is unlikely to detect any failure mode that does not cause a single symbol to become dramatically more common.
- Many conventional designs have failure modes that are detectable using the APT.
- A related test based on testing the concatenation of several outputs can also yield useful results.



APT Cutoffs

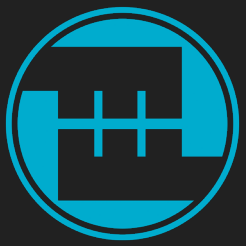
In the SP 800-90B Section 4.4.2 APT cutoff procedure:

- There are relevant corrections ([HJ 2019, Comment 10b]), but they do not change the essential results.
- The SP 800-90B Section 4.4.2 analysis approach:
  - Makes an underlying IID/stationary distribution assumption.
  - Bounds the false positive rate (α) using the assumption that the APT reference symbol (A) is the most likely symbol.
- The actual false positive rate and failure-mode-specific statistical power associated with a particular APT cutoff selection are distribution-dependent and can be estimated via large-scale simulation.



APT Cutoffs

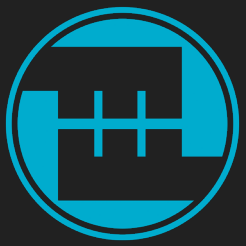
- Most noise sources aren't IID or stationary so if the APT cutoff is established using the assessed entropy, then the targeted false positive rate isn't likely attained.
  - In this case the actual false positive rate is likely to be much smaller than intended.
  - The statistical power may be compromised for some failure modes.
- Very large cutoffs (near the window size, associated with very high entropy assessments) and very small cutoffs (near 0, associated with very low entropy assessments) are likely to result in a lower than desired statistical power. [Hill 2024]
- Reducing the symbol space prior to testing ("mapping") is likely to make the APT worse (both a higher false positive rate and a lower statistical power). [Hill 2024]



APT Window

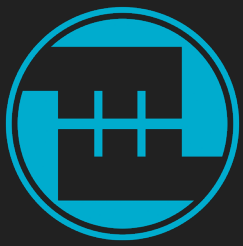
- SP 800-90B directs a window choice (W) of 1024 for binary noise sources and 512 for non-binary sources.
- [90B Shall ID #79] states:

  “The window size must be specified, but other values than 1024 for binary samples and 512 for non-binary samples may be allowed with justification”.
- A larger window size can make the APT more powerful but can increase the lag between failure and detection.
- The window size should be increased for noise sources with a very high or very low apparent entropy.
  - Choosing W as 1024, 2048 and 4096 often yield more powerful tests.

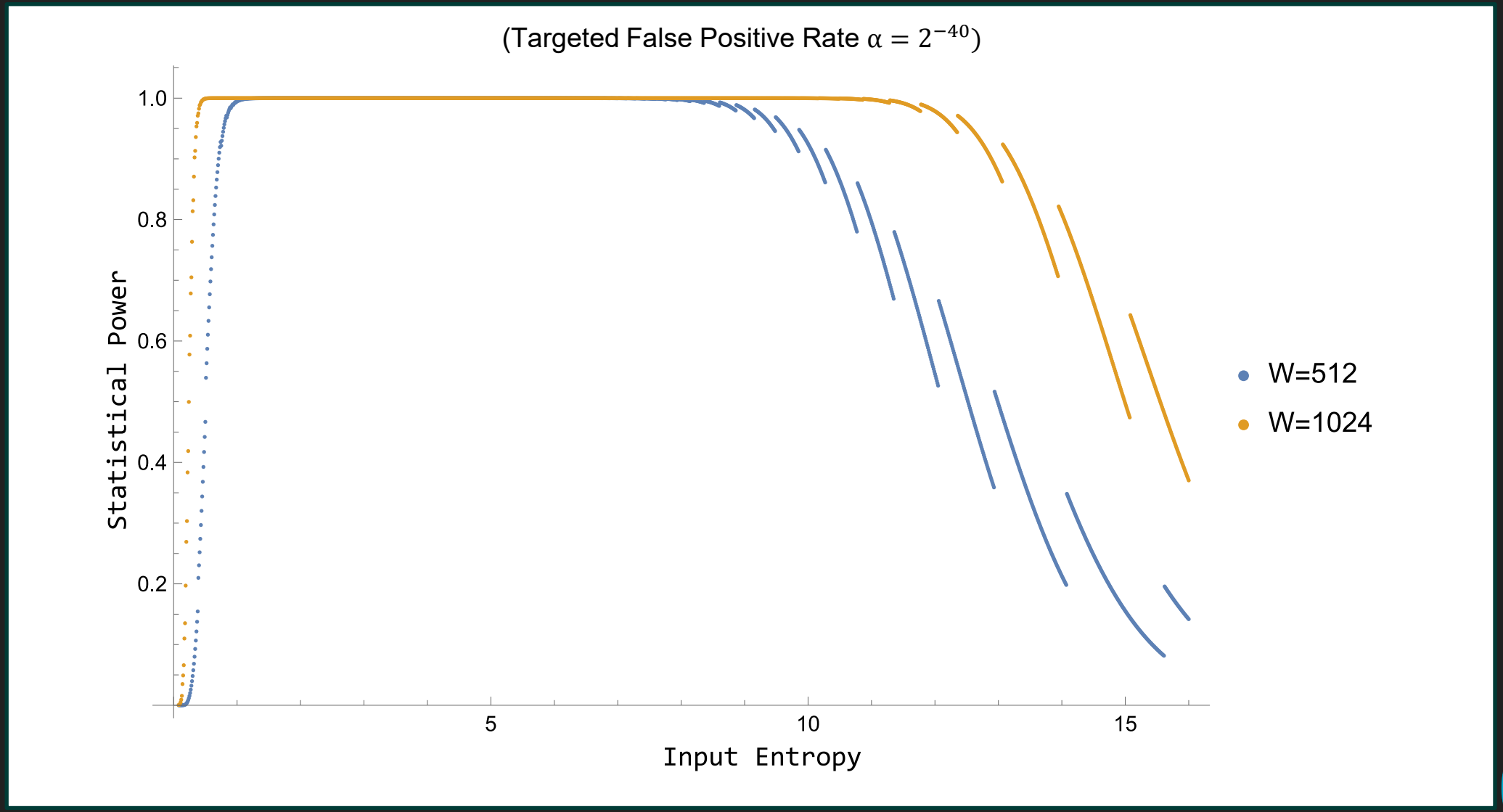


The Impact of the APT Window Size

- To capture the “best case” view of the APT, we’ll examine a modeled stationary IID source.
- This noise source’s output follows the near-uniform distribution from [HD 2012] which produces all possible 16-bit output symbols.
- The cutoff for each test is established using the SP 800-90B Section 4.4.2 procedure.
- We’ll examine the statistical power of the test while in the SP 800-90B Section 4.5 Criterion (b) half-entropy failure mode.



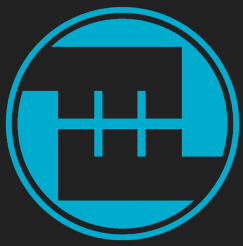
The Impact of the APT Window Size



APT: Does Exactly What It Says on the Tin

The APT does well when detecting the failure mode it was **configured** to detect.

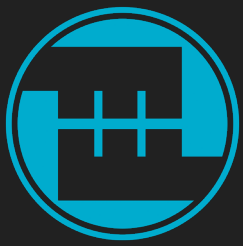
- Setting the cutoffs using the apparent entropy is different than setting them using the assessed entropy.
- If you configure the APT based on the assessed entropy (commonly less than half the apparent entropy), then this APT isn't likely to detect a half-apparent-entropy failure mode.
- If a failure mode doesn't increase $p_{\max}$ then the APT isn't likely to detect it.



Part 2: The RCT

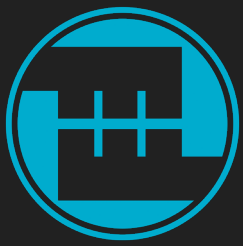
1. A = next()
2. B = 1
3. X = next()
4. If (A == X)
    a) B = B+1
    b) If (B ≥ C), signal a failure

   Else
    a) A = X
    b) B = 1
5. Go to Step 3



The RCT Failure Mode

- The RCT is intended to detect long runs (i.e., a single symbol serially repeating).
  - For most noise source designs, this is a total failure test.
  - This situation naturally arises for most noise sources (even for non-IID/non-stationary sources). [Hill 2023]
- The RCT is unlikely to detect any failure mode that does not result in long runs of values.

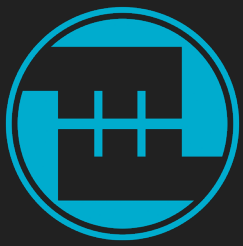


RCT Cutoffs

- The SP 800-90B Section 4.4.1 analysis approach:
  - Makes an underlying stationary/IID assumption.
- The actual false positive rate and failure-mode-specific statistical power associated with a particular RCT cutoff selection are distribution-dependent and can be estimated via large-scale simulation.
- Most noise sources aren't stationary or IID, so if the RCT cutoff is established using the assessed entropy, then the targeted false positive rate isn't likely attained using this procedure.
  - In this case the actual false positive rate is likely to be much smaller than intended.
  - The statistical power is likely to be compromised for some failure modes.

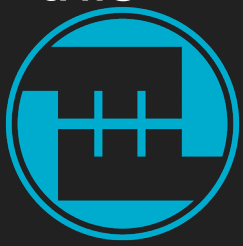
RCT: Does Exactly What It Says on the Tin

- The RCT does well when detecting the failure mode it was **configured** to detect.
  - Setting the cutoffs using the apparent entropy is different than setting them using the assessed entropy.



Part 3: Choosing Alternative Cutoffs

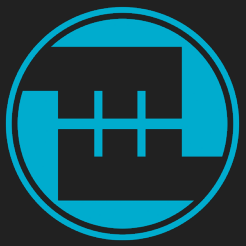
- NIST allows choosing cutoffs based on entropy estimates other than the assessed entropy (H).
  - Such choices need to be justified in the EAR.
  - Basing these cutoffs on the apparent entropy or empirically generally results in strictly lower cutoffs for the same false positive rate.
  - A lower cutoff makes the tests fail in a strict superset of outputs, resulting in more sensitive tests.
- The vendor/lab can also classify these alternative cutoff selections as developer-defined health tests.
  - The SP 800-90B Section 4.5 requirements are trivial to argue in this case.



Part 4: Finding Empirical Cutoffs

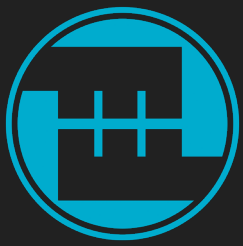
- The [Theseus] package has tools (*apt* and *rct*) for establishing cutoffs for a selected false positive rate.
- The data requirement can be substantial for small values of α.
- For the APT, one would ideally have a data set with approximately

$$\frac{10W}{\alpha} \text{ symbols.}$$



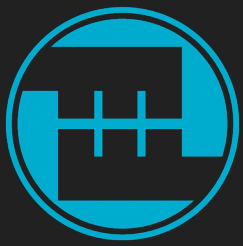
In Summary

- The SP 800-90B cutoff procedures yields APT and RCT with low statistical power for most noise source designs.
  - Designers are allowed to choose cutoffs that yields a more powerful APT and RCT (which also have false positive rates closer to the targeted values).
- The APT window size should be tailored for the noise source distribution.
  - This is also allowed.
- [Theseus] can help designers to perform the empirical cutoff and window size selection.



Designer Meta Summary

- It is fairly easy to do better with the APT and RCT.
- You are allowed to do better.
- Let’s do better!



References

- [HD 2012] Patrick Hagerty and Tom Draper. *Entropy Bounds and Statistical Tests*. 2012.
- [HJ 2019] Joshua E. Hill and Benjamin Jackson. *NIST Special Publication 800-90B Comments*. Version 1.9, December 2019.
- [Hill 2023] Joshua E. Hill. *Health Testing for Periodically Sampled Ring Oscillators*. Version 20230602.
- [Hill 2024] Joshua E. Hill. ~~*Everything*~~*Some of What You Always Wanted to know About the APT (But Were Afraid to Ask)*. Version 20240319-2. March 2024.
- [Theseus] https://github.com/KeyPair-Consulting/Theseus.
- [90B Shall] NIST. 90B Shall Statements.