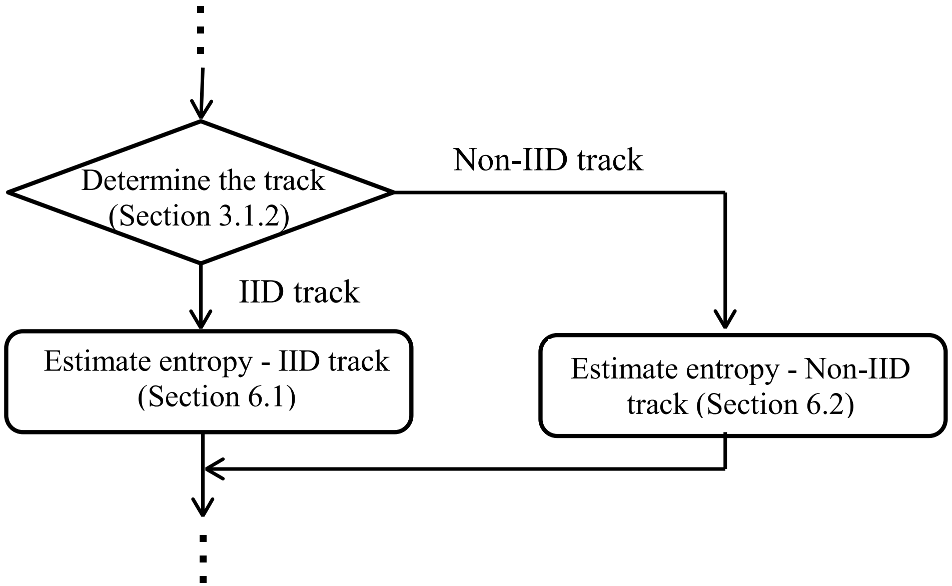
The IID Assumption and YOU!

UL VS, LLC.
Joshua Hill, PhD

Revision: 20190524

Two Roads Diverged in a Wood

Q:



A: ???

In Today's News

- Almost all commercially-available noise source designs are not IID. Some of these noise source designs can be operated using parameters that cause the noise source behavior to be arbitrarily *close* to IID behavior.

- Implementations of IID noise source designs generally have implementation-specific emergent phenomena that induce non-IID behavior.

- SP800-90B does not provide a specification for what constitutes *close enough* to IID. Presently, any deviation from IID-behavior constitutes a failure to be IID.

In summary:

Almost all evaluated noise sources are non-IID.

What is IID?

Definitions

An IID (Independent and Identically Distributed) noise source is:

“A quality of a sequence of random variables for which each element of the sequence has the same probability distribution as the other values, and all values are mutually independent.”

1. Independent: Roughly, events can’t influence each other.

2. Identically Distributed: Roughly, the distribution doesn’t change over time.

IID: Example Designs

Each output is established by:

- the final side displayed after a (fair or unfair) coin is flipped,
- the value shown on a (fair or loaded) die after it is cast,
- the value resulting from a spin of a roulette wheel,
- selection **with replacement.**

Commonality: The sequence is made up of outputs from running multiple rounds of the **exact same** process, where the process's state is **completely reset** between each round.

IID: Notes

IID is not the same thing as *good*.

- Bias is allowed, but variation in the distribution and/or dependence is not.
- This is why it is so easy to estimate entropy for an IID noise source.

Having said that, any full entropy noise source **will be IID**.

IID: Non-Examples

Example of noise sources that are Independent but **not** Identically Distributed:

- Rolling two dice with different numbers of sides.
  - Outputting $Y_{2i} = D4$, $Y_{2i+1} = D6$
- A resetting counter tracking the number of rising edges of a jittery oscillator reset between rounds, when the underlying oscillator's period drifts as conditions change. (e.g., [T 2002])
  - Outputting Y_i = count of rising edges since the last sample

IID: Non-Examples

Example of noise sources that are Identically Distributed but **not** Independent:

- Selection of 4 labeled balls $\{0, \ldots, 15\}$ from an urn, selected **without replacement**. (e.g., Powerball, cards).
  - Outputting Y_{4i} $=$ first draw Y_{4i+1} $=$ second draw, …, Y_{4i+3} = fourth draw

- A (time-homogeneous, irreducible, aperiodic) Markov chain with 4 states, when the transition matrix rows are not all the same.
  - Output Y_i = the current state

IID: Non-Examples

Example of noise sources that are **neither** Independent, **nor** Identically Distributed:

- A random walk with IID Gaussian steps.
  - Output Y_i = the current location

There are many examples of this type of noise source design, including ring oscillators, a frequently sampled noisy circuit (e.g., a reverse-biased diode), event-driven sampling of a free-running timer, etc.

From Non-IID to IID

It is sometimes possible to use a non-IID noise source design to build an IID noise source design.

This commonly involves discarding data, delaying sampling until some particular state occurs, bundling together distinct outputs into a single output, or resetting the noise source between samples.

From Non-IID to IID: Examples

For our Selection of 4 labeled balls $\{0, \ldots, 15\}$ in an urn, selected **without replacement**

Non-IID:

- Y_{4i} = first draw Y_{4i+1} = second draw, …, Y_{4i+3} = fourth draw

IID Output:

- Package all 4 of these into a single data sample

$$X_i = (Y_{4i}, Y_{4i+1}, \ldots, Y_{4i+3})$$

So long as the system is reset after selecting these 4 values, the X_i samples are IID.

From Non-IID to IID: Examples

- For our (finite, time-homogeneous, irreducible, aperiodic) Markov noise source design, you can use **thinning:**
  - **Thinning** is discarding sufficient states (outputs) so that the k-step transition matrix is *almost* the same as the asymptotic behavior.
- Easy Approach: Use thinning prior to every sample.
- More Efficient Approach: Use thinning prior to taking multiple samples, and combine these samples into one raw data sample.

Q: How close does *almost* have to be to make an IID claim?

A: ¯_(ツ)_/¯

Caution: Stateful Conditioning

If the conditioning function retains state, it is also possible to spoil the IID property in the conditioning stage.

Some examples include:

- outputting a running XOR of (biased) IID raw data.
- output from an LFSR whose state is not reset between inputs of (biased) IID raw data.

Making an IID Claim

To Make an IID Claim (SP800-90B)

1. “The submitter makes an IID claim on the noise source… The submitter shall provide rationale for the IID claim.”

2. “The sequential dataset … is tested using the statistical tests described in Section 5...”

3. “The [restart testing] row and column datasets… are tested using the statistical tests described in Section 5…”

4. “[For non-vetted conditioning components], the conditioned sequential dataset… is tested using the statistical tests described in Section 5.”

To Make an IID Claim (IG 7.18)

1. “Provide a **rigorous proof** in support of this [IID] claim.”

2. “A claim of independence and that of an identical distribution shall be substantiated separately.”

   a) “For an independence claim, a deep understanding of the underlying operation of the noise source is required.”

   b) “A claim of an identical distribution of the samples shall consider a possible deterioration of the source’s entropy generation pattern due to the mechanical or the environmental changes or to the timing variations in human behavior.”

In Summary

If you try the IID track, then (probably)…

In Summary

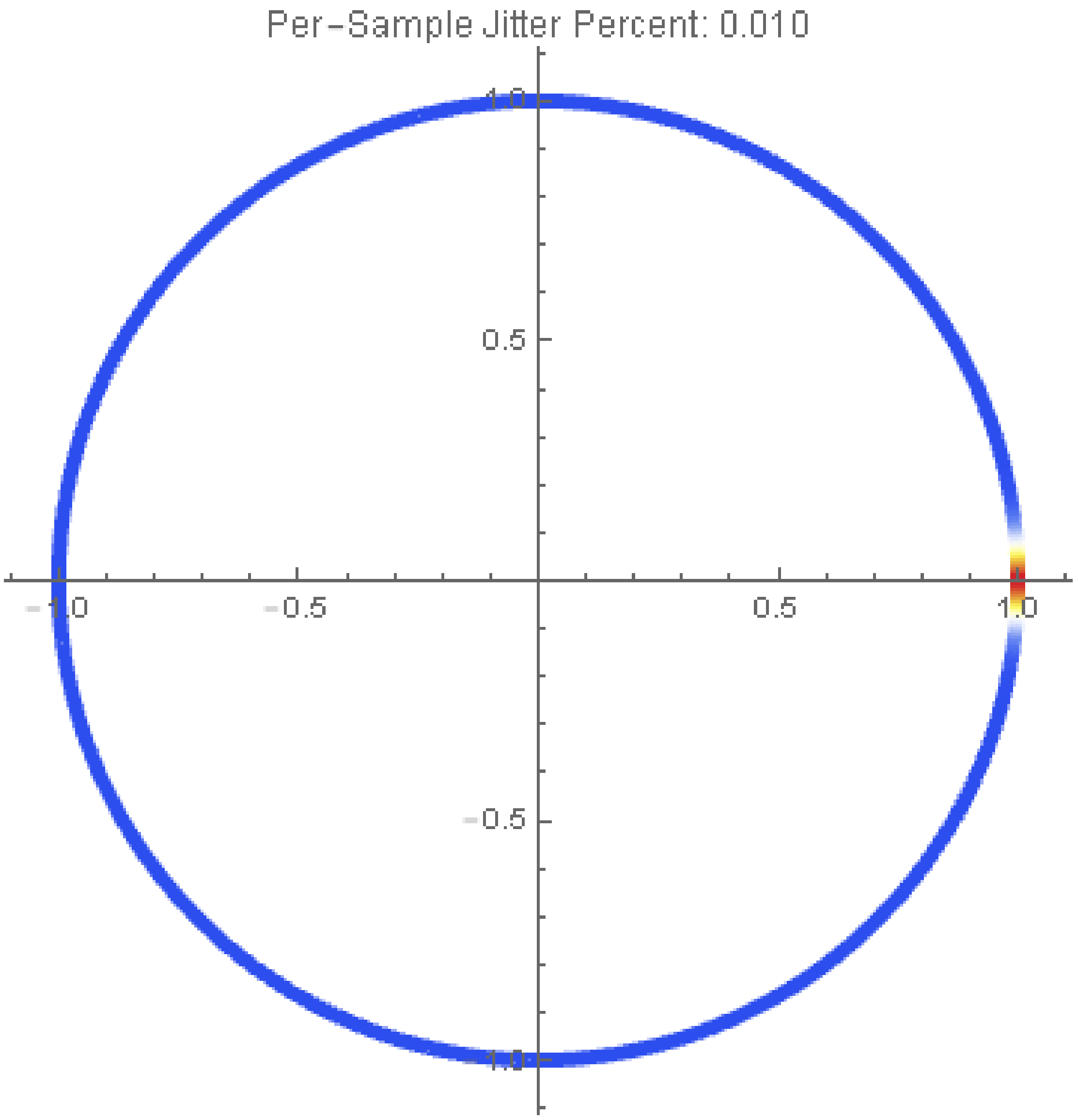
If you try the IID track, then (probably)…

Case Study:
Ring Oscillators

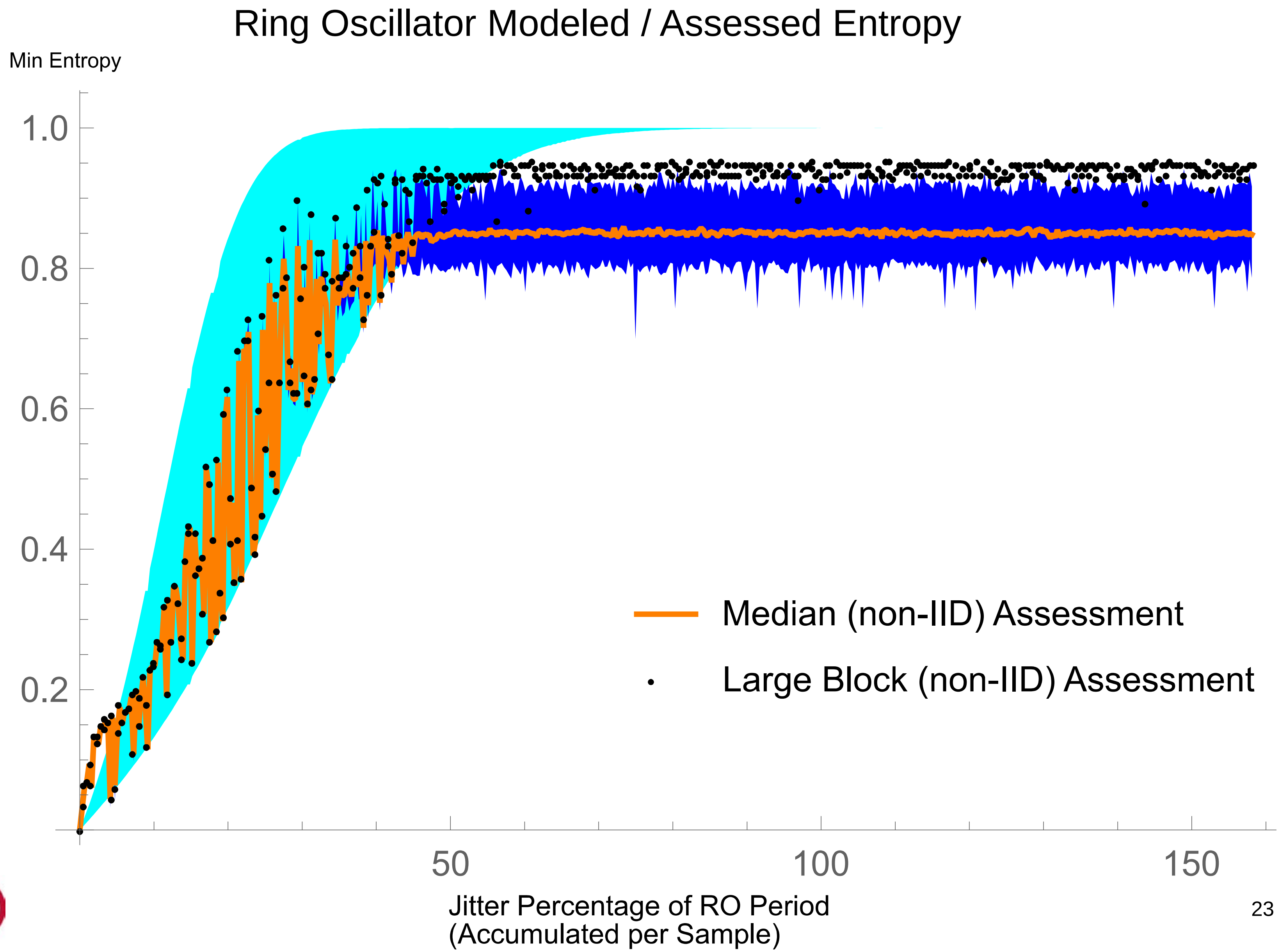
Ring Oscillators

- A free-running ring oscillator has a substantial amount of internal state, namely the current phase of the ring with respect to the sample clock.
- The phase values can be thought of as residing on a unit circle, and the output is established by where on the circle the current phase is when the ring oscillator is sampled.
- The free-running ring oscillator is effectively a random walk on the unit circle.
- When periodically sampled (with most parameter sets) there is substantial autocorrelation between adjacent outputs. Autocorrelation prevents the data from being independent.

Ring Oscillators: PDF as Loop



Ring Oscillators: Entropy



Ring Oscillators: IID?

- The modeled entropy asymptotically approaches full entropy.
- If the output is full entropy, it must be IID.
- Idea: As you let the ring oscillator accumulate uncertainty between samples, the initial condition becomes less important, and the samples become *almost* IID.

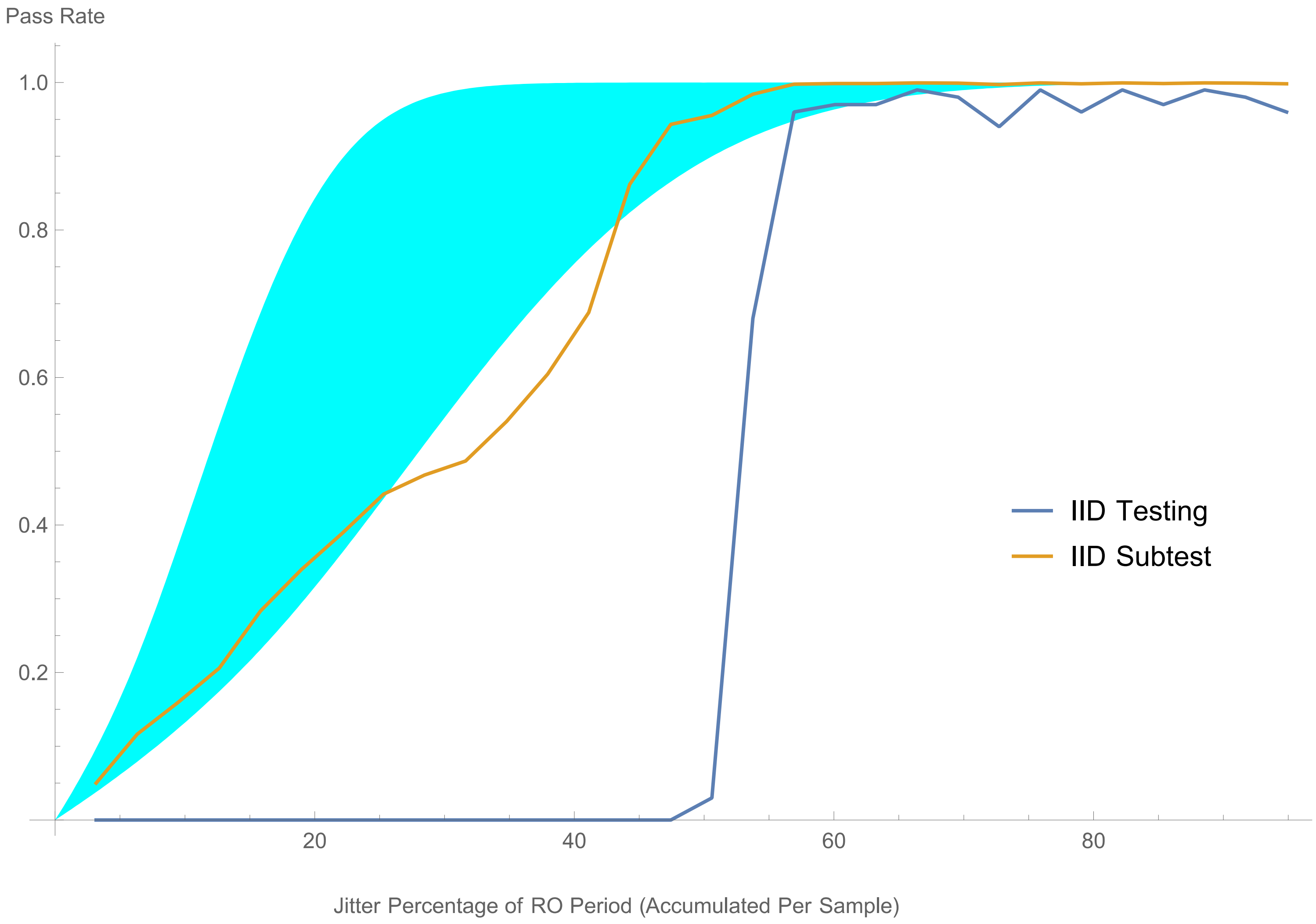
Q: How close does *almost* have to be to make an IID claim?

A: ¯_(ツ)_/¯

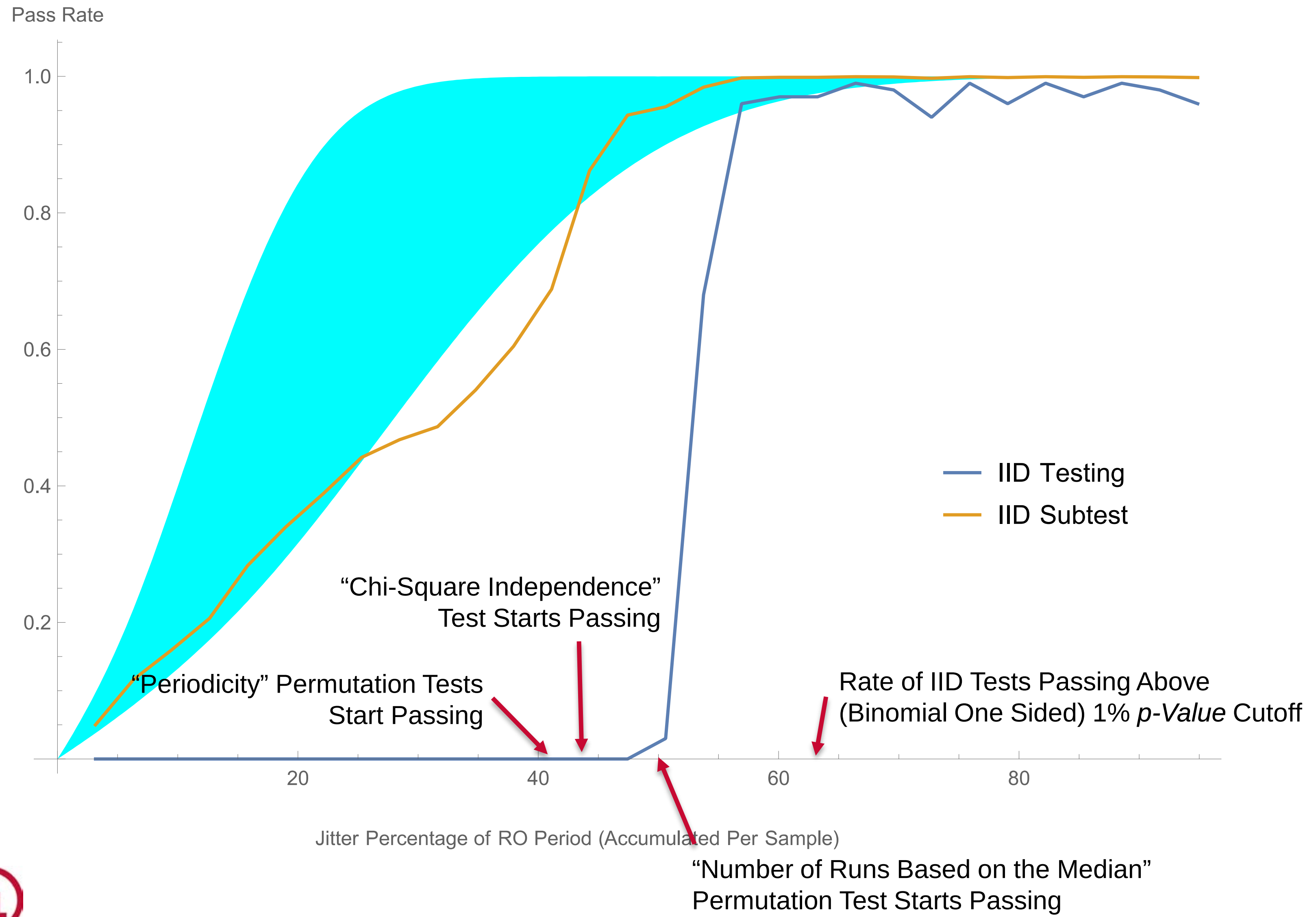
Ring Oscillators: IID Testing

- We don’t know how close to IID you need to be in order to support an IID claim under SP800-90B, but we can tell what actually passes the IID testing.
- We simulated a ring oscillator and generated 100 1-million sample sets for each parameter setting.
- We varied the jitter percentage of the ring oscillator period and performed the testing specified in SP800-90B Section 5 on each set.

Ring Oscillators: IID Testing Results



Ring Oscillators: IID Testing Results



Ring Oscillators: IID Testing Summary

- Idealized ring oscillators start passing all the subtests within IID testing at the expected rates when the jitter percentage is about 64% of the ring oscillator period.
- Ring oscillator designs typically have a jitter percentage between 0.01% and 5%.
- This cutoff is only valid for ideal ring oscillators. The actual cutoff depends on the noise proportion that is due to local Gaussian noise.

Ring Oscillators: Worked Example

1GHz ring oscillator, sampled at 1MHz, with a per-sample 5% jitter percentage, 30% of which is due to local Gaussian noise.

Option 1: Non-IID Source

- Each sample contains more than 0.0175851 bits of entropy, produced at one-million samples per second.
- ≈ 22 **ms** to seed a DRBG to a 256-bit security strength.

Option 2: IID(-ish) Source

- Requires decimating at a rate of 1:1821.
- Each sample contains more than 0.977408 bits of entropy, produced at about 549 samples per second.
- ≈716 **ms** to seed a DRBG to a 256-bit security strength.

It is much more efficient to use this as a non-IID noise source.

In Summary

Vendors:

- You probably shouldn't attempt to make an IID claim.

Labs:

- You probably aren't encountering IID sources, even in the instance where the vendor claims IID.

NIST, **CMVP**, **NIAP** and **Other Regulators**:

- Consider removing the IID track.
- If the IID track remains, consider establishing a maximum amount of *mutual information* that can be tolerated between random variables that can qualify as *independent*, and a maximum *statistical distance* that can be tolerated between random variables' distributions that can qualify as *identically distributed*.

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.
- [HJ 2018] Hill and Jackson. *NIST Special Publication 800-90B Comments*. http://bit.ly/UL90BCOM
- [T 2002] Thomas E. Tkacik. *A Hardware Random Number Generator*. CHES 2002.
- [SP800-90B] Turan, Barker, Kelsey, McKay, Baish, and Boyle. *Special Publication 800-90B: Recommendation for Entropy Sources Used for Random Bit Generation*. January 2018.
- [FIPS IG] *Implementation Guidance for FIPS 140-2 and the Cryptographic Module Validation Program*. May 7, 2019.

THANK YOU.