The impact of digitization on the entropy generation rates of physical sources of randomness

Joseph D. Hart$^{1,2,*}$, Thomas E. Murphy$^{2,3}$, Rajarshi Roy$^{2,3,4}$, Gerry Baumgartner$^{5}$

$^1$ Dept. of Physics
$^2$ Institute for Research in Electronics & Applied Physics
$^3$ Dept. of Electrical & Computer Engineering
$^4$ Institute for Physical Science and Technology
$^5$ Laboratory for Telecommunication Science

*jhart12@umd.edu
Why Physical RNG?

Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin.

--John von Neumann
Physical RNG

• Algorithms can only produce pseudo-random numbers
• For true random numbers, we turn to physical systems
• Can be FASTER because not limited by CPU clock
• Need to be post-processed to remove bias, etc.
• Important differences between pseudo-RNG and physical RNG should be reflected in evaluation metrics
Electronic Physical RNG Today (Intel Ivy Bridge Processors)

- 3 Gb/s raw RNG rate
- Raw bits are not directly used (nor accessible)
- Continuously re-seeds a pseudo-random generator
- Instruction: RDRND
Chaotic Semiconductor Laser

- Fluctuations are fast and chaotic
  (sensitive dependence on initial conditions)

Amplified Spontaneous Emission

Comparison of Optical RNG Methods – Recent Research

- PicoQuant
- ID Quantique
- Whitewood

Random Bit Rate
Sampling Rate
Instrumentation Limit
NIST SP 800-22rev1a

• Easy to implement
• Publicly accessible standard
• Even non-cryptographically secure Pseudo-RNG methods (e.g., Mersenne Twister) will pass all tests
• Only works on binary data (1s and 0s), not analog data or waveforms
• Most physical RNG methods require post-processing to pass tests
Post-Processing of Digitized Waveforms

• Least Significant Bit Extraction:

What is the source of entropy? (waveform, digitizer, thermal noise?)
Entropy estimates

• Try to quantify the number of random bits allowed to be harvested from a physical system
• Works on raw data, not post-processed data
• Can help reveal *where* the entropy is coming from
• Can be slower, require more data than NIST SP 800-22rev1a
Dynamical systems approach to entropy generation

• Kolmogorov-Sinai (or metric) entropy

\[ H = - \frac{1}{d\tau} \sum p(i_1, \ldots, i_d) \log_2 p(i_1, \ldots, i_d) \]

• Analog of Shannon entropy for dynamical system

• Allows for direct comparison of dynamical processes, stochastic processes, and mixed processes
Discretization of analog signals
Time-delay embedding

• Reconstruct phase-space of dynamical system from measurement of one variable

\[ \mathbf{x}(t) = (x(t), x(t - T), \ldots, x(t - (d - 1)T)) \]

Lorenz attractor

Numerically Estimating Entropy

Box Counting Method

Entropy of chaotic systems

For small $\varepsilon$

$$h(\varepsilon) = h_{KS} = \frac{1}{\ln(2)} \sum_{\lambda_i > 0} \lambda_i$$

$$X_{t+1} = 4X_t(1 - X_t)$$
$(\varepsilon-\tau)$ entropy of noise

$h(\varepsilon) \sim -\log_2(\varepsilon)$

Gaussian random variable

P. Gaspard and X. Wang,
*Physics Reports*,
Volume 235, 1993
Noisy chaotic systems

\[ Z_{t+1} = X_t + aR_t \]
\[ X_{t+1} = 4X_t(1 - X_t) \]

R is random Gaussian variable

Case study:
Amplified Spontaneous Emission (ASE)

• Least-significant bits contribute considerable entropy (not optical!)
• Entropy rolls off with sample rate
Entropy Rate - ASE

- Entropy rolls off with sample rate
NIST SP 800-90B
Entropy Estimates

• Most Common Value Estimate
• Collision Estimate—based on mean time until first repeated value
• Markov Estimate—measures dependencies between consecutive values
• Compression Estimate—estimates how much the dataset can be compressed
• Other more complicated tests...
Entropy rate as a function of measurement resolution ($\varepsilon$)

- Instrumentation Limit
- IID Entropy Rate

![Graph showing entropy rate as a function of measurement resolution.](image)

50 GSamples/s

Entropy [Gbits/s]

$\varepsilon$ [bits]

$P(V)$

Voltage [V]
Entropy rate as a function of measurement resolution ($\varepsilon$)

- Instrumentation Limit
- IID Entropy Rate

![Graph showing entropy rate as a function of measurement resolution.](image)
Entropy rate as a function of measurement resolution ($\varepsilon$)

- Instrumentation Limit
- IID Entropy Rate

![Graph showing entropy rate vs. measurement resolution with different types of data points and a shaded area indicating the instrumentation limit.](image-url)
Entropy rate as a function of measurement resolution ($\epsilon$)

- Instrumentation Limit
- IID Entropy Rate

Entropy [Gbits/s]

$\epsilon$ [bits]

50 GSamples/s
Entropy rate as a function of measurement resolution ($\varepsilon$)

- Instrumentation Limit
- IID Entropy Rate

### 50 GSamples/s

<table>
<thead>
<tr>
<th>$\varepsilon$ [bits]</th>
<th>Entropy [Gbits/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

### 1 GSample/s

<table>
<thead>
<tr>
<th>$\varepsilon$ [bits]</th>
<th>Entropy [Gbits/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Capturing Temporal Correlations

- Instrumentation Limit
- IID Entropy Rate

Graphs showing the relationship between sampling frequency and entropy rate with different values of $\epsilon$: 9 bits and 5 bits. The graphs compare various compression techniques (MCV, Collision, Markov) against the theoretical limits (Instrumentation Limit and IID Entropy Rate).
Chaotic Laser

- Fluctuations are fast and chaotic (sensitive dependence on initial conditions)
Entropy rate—laser chaos

- Significant portion of entropy comes from background noise (especially at high resolution)
Conclusions:

• Important to look at entropy as a function of measurement resolution and sampling frequency

• Different physical processes can generate entropy, even within the same experiment

• Measurement determines which physical entropy generation processes you observe

• Entropy estimates should consider analog data, not post-processed bit stream
To learn more about:

**Entropy generation in noisy chaotic systems:**

**Amplified spontaneous emission:**