

Formal Verification of Rate–Distortion Bounds for Generalized Gaussian Sources in Lean 4

André M. Carvalho
andremiguel.pt

February 2025

Abstract

We present a Lean 4 formalization of rate–distortion bounds for generalized Gaussian distributions (GGD) under mean-squared error distortion. Our development provides machine-checked proofs of fundamental GGD properties—including normalization, moment integrals, differential entropy, and Fisher information—and establishes a formal derivation of the gap between the rate–distortion function and Shannon’s lower bound via the Gaussian smoothing technique. The formalization cleanly separates measure-theoretic computations (fully proved using Mathlib) from standard information-theoretic identities (recorded as a minimal axiomatic basis). We obtain an explicit closed-form bound on the rate–distortion gap for unit-variance GGD sources parametrized by shape $\beta > 1$. The Lean development is publicly available and provides a foundation for further formalization of lossy source coding theory.

1 Introduction

Rate–distortion theory, introduced by Shannon [1], characterizes the fundamental limits of lossy data compression. For a source with distribution P_X and a distortion measure $d(x, \hat{x})$, the rate–distortion function $R(D)$ specifies the minimum number of bits per symbol required to represent the source such that the expected distortion does not exceed D . While $R(D)$ admits a variational characterization, closed-form solutions exist only for special cases, notably Gaussian sources under squared-error distortion.

For non-Gaussian sources, a common approach is to bound $R(D)$ relative to the *Shannon lower bound* (SLB), which provides a universal lower bound depending only on the source entropy and the distortion level. The gap $\Delta(D) = R(D) - R_{\text{SLB}}(D)$ quantifies how far a source deviates from the “Gaussian ideal” and can be bounded using techniques from estimation theory, particularly through Fisher information and the de Bruijn identity.

1.1 Contributions

This work presents a Lean 4 formalization that:

- (i) Establishes machine-checked proofs of key analytic properties of the generalized Gaussian distribution, including normalization, absolute moment integrals, variance, differential entropy, and Fisher information.
- (ii) Formalizes the Gaussian smoothing framework for bounding the rate–distortion gap, reducing the bound to an integral of Fisher information along the smoothing path.

- (iii) Derives an explicit closed-form upper bound on $\Delta(D)$ for unit-variance GGD sources with shape parameter $\beta > 1$.
- (iv) Identifies a minimal axiomatic basis consisting of standard information-theoretic identities (test-channel achievability, de Bruijn's identity, Stam's inequality), cleanly separating what is proved from what is assumed.

1.2 Related work

Formal verification of information-theoretic results remains relatively unexplored. Prior work in proof assistants has addressed entropy inequalities [6], channel capacity bounds, and basic probability theory, but to our knowledge no prior formalization addresses rate–distortion theory or the Gaussian smoothing technique.

The generalized Gaussian distribution has been studied extensively in signal processing and source coding [7, 3]. The rate–distortion function for GGD sources does not admit a closed form in general, motivating the use of bounds. The Fisher information approach we formalize originates in classical work on entropy power inequalities [4, 5].

2 Preliminaries

2.1 Notation

All rates and entropies are expressed in *bits*. We write $\log_2(\cdot)$ for base-2 logarithms and $\log(\cdot)$ for natural logarithms; conversions use the factor $1/\log 2$. We work on \mathbb{R} and represent probability distributions by densities $f : \mathbb{R} \rightarrow [0, \infty)$ integrating to one.

2.2 Core definitions

Definition 1 (Differential entropy). *For a density f , the differential entropy in bits is*

$$h(f) := - \int_{\mathbb{R}} f(x) \log_2 f(x) dx.$$

Definition 2 (Shannon lower bound). *For mean-squared error distortion with target level $D > 0$, the Shannon lower bound is*

$$R_{\text{SLB}}(D) := h(f) - \frac{1}{2} \log_2(2\pi e D).$$

Definition 3 (Rate–distortion gap). *The rate–distortion function $R(D)$ gives the minimum achievable rate at distortion D . The rate–distortion gap is*

$$\Delta(D) := R(D) - R_{\text{SLB}}(D).$$

For Gaussian sources, $\Delta(D) = 0$ for all $D \leq \sigma^2$. For non-Gaussian sources, the gap is strictly positive but can be bounded using Fisher information.

2.3 Generalized Gaussian distribution

Definition 4 (GGD density). *For shape parameter $\beta > 0$ and scale parameter $\alpha > 0$, the generalized Gaussian density is*

$$f_{\beta,\alpha}(x) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \exp\left(-\left|\frac{x}{\alpha}\right|^\beta\right).$$

The GGD family includes the Laplace distribution ($\beta = 1$) and the Gaussian distribution ($\beta = 2$) as special cases. For $\beta < 2$, the distribution is leptokurtic (heavier tails than Gaussian); for $\beta > 2$, it is platykurtic.

3 Formalized GGD properties

The following results are proved in Lean using Mathlib's integration and special function libraries.

Theorem 1 (Normalization and moments). *Let $\beta > 0$, $\alpha > 0$, and $p > -1$. Then*

$$\int_{\mathbb{R}} f_{\beta,\alpha}(x) dx = 1$$

and

$$\int_{\mathbb{R}} |x|^p f_{\beta,\alpha}(x) dx = \alpha^p \frac{\Gamma((p+1)/\beta)}{\Gamma(1/\beta)}.$$

In particular, the variance is

$$\text{Var}(X) = \alpha^2 \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}.$$

Definition 5 (Unit-variance scaling). *Define*

$$\alpha_{uv}(\beta) := \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}},$$

so that $X \sim \text{GGD}(\beta, \alpha_{uv}(\beta))$ has unit variance.

Theorem 2 (Differential entropy). *For $\beta > 0$ and $\alpha > 0$, the differential entropy in bits is*

$$h(f_{\beta,\alpha}) = \log_2 \left(\frac{2\alpha \Gamma(1/\beta)}{\beta} \right) + \frac{1}{\beta \log 2}.$$

Theorem 3 (Fisher information). *For $\beta > 1$ and $\alpha > 0$, the density $f_{\beta,\alpha}$ has finite Fisher information given by*

$$J(f_{\beta,\alpha}) = \frac{\beta^2}{\alpha^2} \cdot \frac{\Gamma(2-1/\beta)}{\Gamma(1/\beta)}.$$

For the unit-variance case $\alpha = \alpha_{uv}(\beta)$:

$$J(\beta) := J(f_{\beta,\alpha_{uv}(\beta)}) = \beta^2 \cdot \frac{\Gamma(3/\beta) \Gamma(2-1/\beta)}{\Gamma(1/\beta)^2}.$$

Theorem 4 (Log-concavity). *For $\beta \geq 1$, the GGD density $f_{\beta,\alpha}$ is log-concave.*

4 Rate-distortion gap bound

4.1 The Gaussian smoothing framework

The bound on $\Delta(D)$ follows from the classical Gaussian smoothing technique, which we formalize as a sequence of reductions.

Theorem 5 (RD gap via Fisher information). *Let X be a random variable with density f having finite Fisher information $J(f)$. For any distortion level $D > 0$:*

$$\Delta(D) \leq \frac{D}{2 \log 2} J(f).$$

The proof proceeds through the following steps, formalized in Lean:

Step 1: Gaussian test channel. Consider the reconstruction $\widehat{X} = X + \sqrt{D}Z$ where $Z \sim \mathcal{N}(0, 1)$ is independent of X . By achievability of the test channel:

$$R(D) \leq h(X + \sqrt{D}Z) - h(\sqrt{D}Z).$$

Step 2: Reduction to entropy increment. Since $h(\sqrt{D}Z) = \frac{1}{2} \log_2(2\pi e D)$, subtracting the SLB yields:

$$\Delta(D) \leq h(X + \sqrt{D}Z) - h(X).$$

Step 3: De Bruijn's identity. Let $X_t := X + \sqrt{t}Z$ for $t \in [0, D]$. The integrated de Bruijn identity gives:

$$h(X_D) - h(X_0) = \frac{1}{2 \log 2} \int_0^D J(X_t) dt.$$

Step 4: Fisher information monotonicity. By Stam's inequality, Fisher information is non-increasing under Gaussian convolution:

$$J(X_t) \leq J(X_0) = J(X) \quad \text{for all } t \in [0, D].$$

Step 5: Integration. Combining the above:

$$\Delta(D) \leq \frac{1}{2 \log 2} \int_0^D J(X) dt = \frac{D}{2 \log 2} J(X).$$

4.2 Application to GGD sources

Combining Theorem 5 with the Fisher information formula from Theorem 3 yields our main result.

Theorem 6 (GGD rate-distortion gap bound). *Let $X \sim \text{GGD}(\beta, \alpha_{uv}(\beta))$ with $\beta > 1$ (unit variance). For any $D > 0$:*

$$\Delta(D) \leq \frac{D}{2 \log 2} \cdot \beta^2 \cdot \frac{\Gamma(3/\beta) \Gamma(2 - 1/\beta)}{\Gamma(1/\beta)^2}.$$

Remark 1 (Tighter logarithmic bound). *A sharper bound follows from exact integration of the Stam inequality:*

$$\Delta(D) \leq \frac{1}{2 \log 2} \log(1 + D \cdot J(X)).$$

This refinement is stated but not yet proved in the Lean development.

5 Axiomatic basis

The formalization treats the following standard results as axioms, cleanly separating analytic computations (proved) from information-theoretic foundations (assumed):

- **Rate-distortion function:** Existence of $R(D)$ satisfying standard properties (non-negativity, monotonicity, convexity).
- **Test-channel achievability:** The inequality $R(D) \leq h(X + \sqrt{D}Z) - h(\sqrt{D}Z)$ for the Gaussian test channel.
- **Integrated de Bruijn identity:** The entropy increment formula in terms of Fisher information.
- **Stam's inequality:** Monotonicity of Fisher information under Gaussian convolution.

These results are well-established in the literature [3, 4, 5]. Formalizing them would require substantial measure-theoretic infrastructure for mutual information and conditional distributions, which we leave to future work.

6 Formalization details

The Lean 4 development is organized as follows:

- `RateDistortion/Basic.lean`: Core definitions (differential entropy, SLB, log-concavity, finite Fisher information predicate).
- `RateDistortion/Axioms/`: Axiomatic information-theoretic facts.
- `RateDistortion/GaussianSmoothing.lean`: Reduction of RD gap to Fisher information bound.
- `RateDistortion/GGD/`: GGD-specific results (density, moments, entropy, Fisher information, log-concavity).
- `RateDistortion/GGDRDBound.lean`: Main theorem combining the framework with GGD formulas.

Table 1 provides an index of key Lean declarations.

7 Discussion

The complete Lean 4 development is available at <https://github.com/andremiguelc/transformer-compression>

This formalization demonstrates that substantive rate-distortion bounds can be machine-checked with a modest axiomatic basis. The separation between analytic GGD computations and information-theoretic identities provides a clear interface: the former are fully verified using Mathlib, while the latter constitute a well-defined trusted base.

Several directions for future work present themselves:

- **Completing the logarithmic bound:** The tighter $\log(1 + DJ)$ form requires careful integration of the exact Stam inequality.

Lean declaration	Location
gfd_integral_eq_one	GGD/Moments.lean
gfd_abs_moment_integral	GGD/Moments.lean
gfd_second_moment	GGD/Moments.lean
gfd_entropy_bits	GGD/Entropy.lean
gfdFisherInfo_eq_fisherInfo	GGD/FisherInfo.lean
gfd_fisher_info_unitVar	GGD/FisherInfo.lean
gfd_logconcave	GGD/LogConcave.lean
rdGap_bits_via_fisherBound	GaussianSmoothing.lean
gfd_rd_gap_bound_bits_unitVar	GGDRDBound.lean
gfd_rd_gap_bound_fisher	GGDRDBound.lean

Table 1: Index of principal Lean declarations.

- **Formalizing the axioms:** De Bruijn’s identity and Stam’s inequality could be proved from measure-theoretic first principles, though this requires infrastructure for score functions and relative entropy.
- **Extension to other source families:** The Gaussian smoothing framework applies broadly; formalizing bounds for other distributions (e.g., exponential, uniform) would be natural extensions.
- **Achievability proofs:** Formalizing the converse—that the RD function is achievable—would complete the picture.

Acknowledgments

This work benefited from AI assistance during its development. Claude Opus 4.5 contributed to strategic planning, problem formulation, and proof review. GPT-5.2 Codex assisted with formalization, Lean proof development, and cross-checking between the mathematical exposition and formal statements.

The Lean development builds on the Mathlib library; we thank its contributors for the integration and special function infrastructure that made this work possible.

References

- [1] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. [doi:10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x)
- [2] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.
- [4] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.
- [5] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991. [doi:10.1109/18.104312](https://doi.org/10.1109/18.104312)

- [6] R. Affeldt, M. Hagiwara, and J. Sénizergues. Formalization of Shannon's theorems. *Journal of Automated Reasoning*, 53(1):63–103, 2014. doi:[10.1007/s10817-013-9298-1](https://doi.org/10.1007/s10817-013-9298-1)
- [7] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai. Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*, 5(1):1–40, 2018. doi:[10.1186/s40488-018-0088-5](https://doi.org/10.1186/s40488-018-0088-5)