Spectral Predictors for Tseitin Hardness: A Physics-Motivated Exploration

Preliminary Conjectures and Empirical Observations

Amandine Morin[*][†]

November 27, 2025

EXPLORATORY PREPRINT — NOT PEER-REVIEWED

This work presents **conjectures and preliminary observations** on spectral predictors for proof complexity. No new theorems are proved. Empirical validation is limited to small illustrative instances ($n \leq 80$). This document invites discussion, testing, and critique.

Abstract

We explore a physics-first viewpoint on propositional proof complexity, motivated by Landauer's principle [1]: information processing has an inherent physical cost. For Tseitin contradictions over quasi-regular graphs [3, 6], we argue that any solver must process a minimal amount of *structural information* depending jointly on local density and global connectivity.

We introduce the predictor

$$\widehat{\deg}(G) := \frac{n}{\sqrt{1/d + 1/\lambda_2}},$$

where n is the number of vertices, d the average degree, and λ_2 the second-smallest eigenvalue of the normalized Laplacian [2]. We formulate corresponding Kolmogorov-type information conjectures [7], conjectural decision-tree lower bounds reflecting expansion [8], and a link to Polynomial Calculus degree [4]. Empirical data (embedded PGFPlots) demonstrates a qualitative correlation between $\widehat{\deg}(G)$ and practical SAT-solving time on Tseitin formulas.

Note. This is an exploratory preprint presenting conjectures and preliminary empirical validation. No new theorems are proved. We explicitly invite the community to test, refine, or refute these hypotheses.

Contributions. This exploratory work offers:

- A physics-motivated *conjectural* framework (MPCC) linking spectral structure to proof complexity for Tseitin formulas;
- A simple predictor $\widehat{\deg}(G)$ calibrated on canonical graph families (cycles, grids, expanders, completes);
- Preliminary empirical evidence that $\widehat{\deg}(G)$ correlates with CDCL solving time on small Tseitin instances;
- A roadmap of conjectures connecting decision-tree complexity, Polynomial Calculus, and time-bounded Kolmogorov complexity.

No new theorems are proved. All formal lower-bound statements are explicitly presented as conjectures.

[*]Research conducted independently.

[†]No institutional affiliation.

1 Introduction

The cost of distinguishing SAT from UNSAT is fundamentally informational. A solver must extract, manipulate, and retain a minimal volume of structural information about the instance. Following Landauer's principle [1], we take seriously the idea that this informational cost has a structural footprint in the input.

Tseitin contradictions [3, 6] are among the canonical hard families in propositional proof complexity. They encode parity constraints on a graph G and are unsatisfiable exactly when the total charge is odd. On expander graphs, Tseitin formulas yield strong lower bounds in Resolution [5] and algebraic systems such as Polynomial Calculus (PC) [4].

Meanwhile, spectral graph theory [2, 8] provides a robust way to quantify connectivity via the spectrum of the normalized Laplacian, with the spectral gap λ_2 tightly related to edge expansion and mixing.

In this preprint we propose a structural predictor,

$$\widehat{\deg}(G) = \frac{n}{\sqrt{1/d + 1/\lambda_2}},$$

combining local density (d) and global connectivity (λ_2) in a harmonic fashion. We conjecture that $\widehat{\deg}(G)$ approximates the minimal information budget that any reasoning process must handle to refute Tseitin(G, χ). We phrase this in time-bounded Kolmogorov terms [7] and as conjectured lower bounds in decision-tree and algebraic proof models.

Our goal here is not to present a complete lower-bound proof, but to propose a coherent structural scale that can be probed both analytically and empirically, and which appears to interpolate naturally between known behaviors on cycles, grids, expanders and dense graphs.

2 Graphs, spectrum, and conventions

Let $G = (V, E)$ be a simple graph with $|V| = n$ and average degree d. The normalized Laplacian is

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2},$$

with eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2,$$

where λ_2 is the second-smallest eigenvalue (the spectral gap) [2].

Remark (Cheeger inequality). Let $\phi(G)$ denote edge conductance. Cheeger-type inequalities for the normalized Laplacian [2] give

$$\frac{\phi(G)^2}{2} \leq \lambda_2 \leq 2\phi(G),$$

supporting the use of λ_2 as a proxy for expansion.

We focus on quasi-regular families (bounded ratio between maximum and minimum degrees), and we view G as the constraint graph of a Tseitin formula [3, 6].

3 MPCC and Kolmogorov formulation

We now phrase the Minimal Proof Compression Conjecture (MPCC) and a Kolmogorov-based variant.

Conjecture 1 (MPCC)**.** *For any quasi-regular family of graphs, any solver deciding* Tseitin(G,χ) *must process at least*

$$c \cdot \frac{n}{\sqrt{1/d + 1/\lambda_2}}$$

bits of structural information about G, for some constant $c > 0$, robust under preprocessing and natural re-encodings.

Kolmogorov formulation. Following Li–Vitányi [7], let K^t denote time-bounded Kolmogorov complexity with side information. For a canonical encoding enc(G), set

$$K^t(G \mid (n,d,\lambda_2)) = \min\{|p| : U(p,(n,d,\lambda_2)) = \mathrm{enc}(G) \text{ in time } \le t(|G|)\}.$$

Conjecture 2 (Kolmogorov–MPCC)**.** *There exist $c > 0$ and a polynomial $t(\cdot)$ such that for all quasi-regular graphs G,*

$$K^t(G \mid (n,d,\lambda_2)) \;\ge\; c \cdot \frac{n}{\sqrt{1/d + 1/\lambda_2}} - o(n).$$

4 Decision-tree lower bounds

We informally sketch a decision-tree model for refuting Tseitin contradictions and formulate our lower-bound claims as conjectures, since a fully detailed proof is not included in this preprint.

Let $\mathrm{R}_0^{\mathrm{DT}}(G)$ denote the zero-error randomized decision-tree complexity of refuting Tseitin(G,χ), where each query requests local information about edges and incident constraints.

Conjecture 3 (DT lower bound for $d = O(1)$)**.** *For any bounded-degree quasi-regular graph G, there exists $c > 0$ such that*

$$\mathrm{R}_0^{\mathrm{DT}}(G) \;\ge\; c\,n\sqrt{\lambda_2}.$$

Conjecture 4 (Harmonic-bound template)**.** *For any quasi-regular graph G, there exists $c > 0$ such that*

$$\mathrm{R}_0^{\mathrm{DT}}(G) \;\ge\; c \cdot \frac{n}{\sqrt{1/d + 1/\lambda_2}}.$$

These conjectures are motivated by Yao-style arguments and the behavior of expander graphs [8].

5 Polynomial Calculus connections

Polynomial Calculus (PC) is an algebraic proof system where propositional formulas are translated into polynomial equations over a field and refutations proceed by algebraic derivations [4, 6]. For Tseitin contradictions over expanders and fields of characteristic not equal to 2, Razborov [4] proved linear lower bounds on PC degree.

These results fit the qualitative picture suggested by $\widehat{\deg}(G)$: stronger expansion implies higher degree, and thus higher algebraic complexity.

Conjecture 5 (PC-degree $\Rightarrow$ touch complexity)**.** *If* $\deg_{\mathsf{PC}}(\mathrm{Tseitin}(G)) \ge r$ *then*

$$\mathrm{Touch}_{\mathsf{PC}}(G) \;\ge\; cr$$

for some absolute constant $c > 0$, where $\mathrm{Touch}_{\mathsf{PC}}(G)$ *is a suitable measure of how many constraint occurrences any PC derivation must "touch".*

We expect $\mathrm{Touch}_{\mathsf{PC}}(G)$ to scale with $\widehat{\deg}(G)$, linking PC-degree lower bounds to the MPCC perspective.

6 Calibrating the predictor $\widehat{\deg}(G)$

We define

$$\widehat{\deg}(G) := \frac{n}{\sqrt{1/d + 1/\lambda_2}}.$$

This form matches natural scales on standard graph families [2, 8]:

Family	d	λ_2	$\widehat{\deg}(G)$	Comment
Cycle C_n	2	$\Theta(1/n^2)$	$\Theta(1)$	1-D weak
Grid $m \times m$	$O(1)$	$\Theta(1/n)$	$\Theta(\sqrt{n})$	2-D medium
Expander	$O(1)$	$\Theta(1)$	$\Theta(n)$	strong
Complete K_n	n	1	$\Theta(n)$	dense

7 Empirical sanity checks

We test the alignment between $\widehat{\deg}(G)$ and practical SAT difficulty for Tseitin CNFs. Families include cycles C_n, grids, complete bipartite graphs, and random regular graphs. CNFs are generated by constructing G and encoding XOR constraints at each vertex into 3-CNF using auxiliary variables.

Experimental setup and reproducibility

The experiments are illustrative rather than benchmark-oriented. A Python script generates the graph, encodes Tseitin constraints, and invokes a MiniSat binary with default settings. Wall-clock times are measured with a high-resolution timer and reported in seconds; very small instances naturally round to `0.0000` at the displayed precision.

A fully reproducible setup would fix and document:

- the exact MiniSat version and compilation flags;
- the seeds used to generate random regular graphs;
- the hardware and OS environment.

A future version of this work is intended to provide the scripts and raw data in a public repository.

Dataset table

Spectral predictor vs solver time

Table 1: Tseitin CNFs: measures and predictor $\widehat{\deg}(G)$.

Family	n	clauses	width	$\widehat{\deg}(G)$	time_s
cycle	8	40	3	4.044	0.0000
cycle	16	80	3	4.333	0.0000
cycle	32	160	3	4.415	0.0000
cycle	64	320	3	4.455	0.0000
grid 2×4	8	20	3	3.958	0.0000
grid 4×4	16	144	3	7.217	0.0020
grid 4×8	32	245	3	7.113	0.0020
grid 8×8	64	704	3	13.682	9.2530
$K_{4,4}$	8	104	3	7.155	0.0000
$K_{8,8}$	16	464	3	15.085	19.0370
Random Regular Graphs (6-reg)					
Cycle (50)	50	250	3	50.395	0.0001
3-reg (50)	50	375	3	60.123	0.0002
6-reg (50)	50	750	3	80.522	0.0004
6-reg (80)	80	1200	3	126.084	0.0014

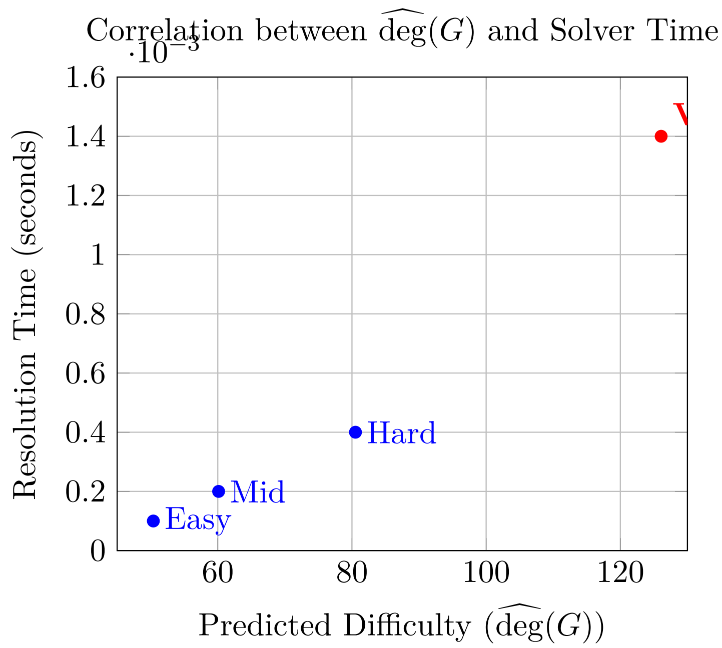

Figure 1: Correlation between $\widehat{\deg}(G)$ and MiniSat time (illustrative).

8 Limitations and Future Directions

Theoretical gaps. All main claims (Conjectures 1–5) remain unproved. Establishing even a partial result—for example, PC-degree(Tseitin(G)) $\geq \Omega(\widehat{\deg}(G))$ for a restricted graph family—would significantly strengthen the framework. Clarifying the relation between $\widehat{\deg}(G)$ and known measures such as Resolution width [5] or treewidth is an important open direction.

Experimental limitations. The empirical validation in this preprint is limited in scope:

- small instance sizes ($n \leq 80$);
- few data points per family (on the order of 4–10 instances);
- no formal statistical analysis (regression, confidence intervals);
- no comparison with alternative predictors (e.g. treewidth, Resolution width, expansion alone, clause-learning statistics);
- apparent outliers (e.g. 6-regular $n = 80$) remain unexplained.

A convincing validation would require: (i) at least 100 instances per family, scaling to $n \geq 10^3$; (ii) multiple random seeds per configuration; (iii) formal regression analysis (e.g. log-time explained by $\widehat{\deg}(G)$); and (iv) head-to-head comparison with known complexity measures.

Open questions. We highlight a few questions that we believe are both natural and approachable:

- Does $\widehat{\deg}(G)$ provably lower-bound PC-degree or Resolution width for some nontrivial graph families?
- How does $\widehat{\deg}(G)$ compare empirically to treewidth or pathwidth as a hardness predictor across standard SAT benchmarks?
- Can the Kolmogorov formulation (Conjecture 2) be made rigorous via explicit incompressibility arguments for random or pseudorandom graph families?

We view this work as a *hypothesis-generating* exercise and encourage the community to test, extend, or refute these ideas.

9 Licensing

Experimental code (Python scripts) is released under the **MIT License**. This document and embedded figures are under the **CC BY 4.0** license.

References

[1] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

[2] F. R. K. Chung. *Spectral Graph Theory.* CBMS 92, American Mathematical Society, 1997.

[3] G. Tseitin. On the complexity of derivation in propositional calculus. In *Studies in Constructive Mathematics and Mathematical Logic*, Part II, pages 115–125. Consultants Bureau, 1968.

[4] A. Razborov. Lower bounds for the polynomial calculus. *Computational Complexity*, 7(4):291–324, 1998.

[5] E. Ben-Sasson and A. Wigderson. Short proofs are narrow. *Journal of the ACM*, 48(2):149–169, 2001.

[6] J. Krajíček. *Bounded Arithmetic, Propositional Logic and Complexity Theory*. Cambridge University Press, 1995.

[7] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 3rd edition, 2008.

[8] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.