

THE SCALING WINDOW OF RANDOM k -SAT

GAIA CARENINI

ABSTRACT. We prove a general upper bound for scaling windows of sparse monotone covering problems. From that, we deduce that for every fixed value $k \geq 3$, the window of random k -SAT is $O(n/\log n)$, improving the Friedgut–Bourgain bound of $O(n/\log \log n)$. We also show that random signed Not-All-Equal- k -SAT and hypergraph non-two-colourability (Property B) have windows of the same order. Our upper bound for scaling windows of sparse monotone covering problems follows from a general strategy combining the strengthening of Bourgain’s sharp threshold theorem by Keevash, Lifshitz, Long, and Minzer with a local-to-random replacement principle. This provides a systematic method for proving scaling-window bounds, addressing a question of Perkins.

1. INTRODUCTION

Random discrete structures typically undergo phase transitions. That is, a relatively small change in a parameter can cause a swift change in the structure of the overall object. This basic phenomenon—first understood in detail for the emergence of the giant component in Erdős–Rényi graphs [14]—has since been found to organize an enormous range of models, from random graphs and hypergraphs to random constraint satisfaction problems, random matrices, and statistical-physics models on random and non-random lattices. Once a transition is known to occur, three further questions present themselves. Where, as a function of the size of the structure, does it occur? Is it sharp, in the sense that the property goes from unlikely to likely over a window that is small compared to the location of the transition? And, quantitatively, how small is that window?

The first two questions now have general theories attached to them. Locating a transition is often approachable by first- and second-moment methods, or by more refined methods from statistical physics. Meanwhile, a substantial body of work, beginning with the influence theory of Kahn, Kalai, and Linial [15] and developed through the sharp-threshold machinery of Bourgain, Friedgut, Hatami, and others [9, 4, 17, 19], gives general sufficient conditions for a monotone property to have a sharp threshold. A major recent development in this direction was the proof by Park and Pham [18] of the expectation-threshold conjecture of Kahn and Kalai [16]. Their work builds on the fractional expectation-threshold theorem of Frankston, Kahn, Narayanan, and Park [21], which in turn developed ideas arising from the breakthrough of Alweiss, Lovett, Wu, and Zhang on the sunflower conjecture [20]. The Park-Pham theorem shows, in wide generality, that the location of a threshold is determined, up to a logarithmic factor, by a simple combinatorial covering parameter.

The third question—the width of the window, once sharpness is known—has no comparable general theory. As Perkins emphasizes in his survey of the area [11], sharp-threshold theorems are typically *qualitative*: they certify that a window is small compared to the threshold location, without producing any usable quantitative bound on how small. The window bounds that do exist in the literature have generally come from arguments tailored to the fine structure of one specific model. The most prominent example is the essentially complete understanding of the random 2-SAT scaling window due to Bollobás, Borgs, Chayes, Kim, and Wilson [2]. Their analysis relies on a detailed combinatorial description of the transition. However, such a description is unavailable once $k \geq 3$.

Random k -SAT is the standard example of the gap between what sharp-threshold theory gives and what one would like to know. Friedgut, with an appendix by Bourgain, showed that random k -SAT has a sharp threshold for every fixed $k \geq 3$ [9]—one of the cornerstone results of the general theory. Sharpness has therefore been known since 1999, but turning the qualitative statement of

Friedgut and Bourgain into an explicit window bound produces only $O(n/\log \log n)$ clauses [8, 11], and this bound has not been improved for more than twenty-five years.

Abbe and Montanari [1] showed that a transition window of order $n/\log^{1+\delta} n$, for any fixed $\delta > 0$, would have major consequences for the satisfiability conjecture, a long-standing conjecture that asserts that, for every fixed k , there is an asymptotically well-defined critical clause density at which a random k -CNF formula transitions from being satisfiable to being unsatisfiable with high probability; see [3] for an overview of the state of the art on the conjecture. Thus, there is real motivation for obtaining a better bound and, more importantly for the present paper, for developing a method that does not depend on the special combinatorics of the fully understood $k = 2$ case.

This paper supplies such a technique. We prove an abstract window theorem, stated in Section 2 and proved in Section 3, which converts a sharp-threshold-type input into an explicit window bound for a monotone event. It has two hypotheses, which can be verified separately in applications: short minimal witnesses for the event must be rare, and the deterministic probability boost supplied by a strong quantitative form of Bourgain’s theorem must be reproducible, with comparable strength, by adding genuinely random coordinates.

The sharp additive input we use is the recent hypercontractivity theorem for global functions of Keevash, Lifshitz, Long, and Minzer [10]. The simulation step—turning a small deterministic set of helpful coordinates into a comparably effective random one—is new; it is what allows the argument to be separated from the specific combinatorics of any one model. This mechanism directly addresses a question posed in Perkins’s survey, which asks for a systematic method for obtaining quantitative window bounds [11].

Applied to random k -SAT, the abstract theorem gives a clause window of order $O(n/\log n)$ for every fixed $k \geq 3$. This improves, by a factor of order $\log n/\log \log n$, the bound that has stood since the original Friedgut–Bourgain argument. We want to be precise about what this does and does not establish: it is an improved *upper* bound on the window, not a matching lower bound, and it falls short of the Abbe–Montanari scale needed to obtain consequences for the satisfiability conjecture. We discuss in Section 10 why the present method, run as far as it can go, produces exactly one power of $\log n$ and no more, and what would be needed to do better.

The k -SAT bound is in fact a special case of a bound for general atomic constraint satisfaction problems (CSPs) over an arbitrary finite alphabet, stated in Section 7. As further applications, we also consider random signed Not-All-Equal- k -SAT, and non-two-colourability of random k -uniform hypergraphs (Property B). In signed NEA- k -SAT, each clause has its own pattern of literal negations and therefore forbids an arbitrary antipodal pair $\{a, \bar{a}\}$ of local assignments. Property B is the unsigned analogue: the variables represent the two vertex colours, and every hyperedge forbids only two monochromatic assignments 0^k and 1^k . Although an individual signed NAE clause can be made unsigned by flipping some variables, these flips need to be consistent across overlapping clauses, so the two models are inherently different. For both problems, we prove an $O(n/\log n)$ window bound.

Random NAE- k -SAT is one of the standard benchmark models in random CSP theory. Its global complement symmetry makes it structurally different from ordinary k -SAT, and it has played an important role both in the study of satisfiability thresholds [22, 6, 23] and in the study of condensation and the geometry of the solution space [24, 25, 26, 27]. Hypergraph two-colorability, traditionally known as Property B, is one of the oldest models in the area—Erdős’s original probabilistic argument for Property B predates the modern threshold literature—and it remains a standard testbed for hypergraph covering thresholds. Treating both models therefore proves that the method applies beyond atomic CSPs while accommodating genuinely different constraint symmetries.

Although these three models are closely related, they have different local symmetries; thus, the local replacement argument takes a different form in each case. Sections 4 and 5 isolate the common part of these arguments and package the model-specific verification into a finite, checkable criterion based on a rigidity certificate and a comparison between the locally induced constraints and the original random constraints.

2. PRELIMINARIES AND OVERVIEW OF THE MAIN THEOREM

Let Ω_n be a finite coordinate set, $N_n = |\Omega_n|$, and let $\mathcal{F}_n \subseteq 2^{\Omega_n}$ be a nontrivial monotone event. We write μ_p for product measure on 2^{Ω_n} , and for $0 < \alpha < 1$ define

$$p_\alpha = p_\alpha(n) := \inf\{p \in [0, 1] : \mu_p(\mathcal{F}_n) \geq \alpha\}.$$

Since $p \mapsto \mu_p(\mathcal{F}_n)$ is continuous, $\mu_{p_\alpha}(\mathcal{F}_n) = \alpha$. For $0 < \eta < 1/2$, define the central window, measured in expected number of coordinates, by

$$W_\eta(n) = N_n(p_{1-\eta} - p_\eta).$$

We also define the central expected-size and variance scales

$$M_\eta(n) = \sup_{p \in [p_\eta, p_{1-\eta}]} pN_n, \quad V_\eta(n) = \inf_{p \in [p_\eta, p_{1-\eta}]} N_n p(1-p).$$

The product-measure formulation is the natural setting for influence inequalities. Random CSPs are more commonly presented with an exact number of independently sampled constraints. In order to keep the differences between these models out of the applications, we outline a generic transfer principle.

For $0 \leq m \leq N_n$, let ν_m be the uniform measure on $\binom{\Omega_n}{m}$, and let $a_m = \nu_m(\mathcal{F}_n)$. The sequence (a_m) is nondecreasing. Define

$$m_\alpha = \min\{m : a_m \geq \alpha\}, \quad \widehat{W}_\eta(n) = m_{1-\eta} - m_\eta.$$

Proposition 2.1 (Transfer from product measure to uniform layers). *For every fixed $0 < \eta < 1/2$, and all sufficiently large n ,*

$$\widehat{W}_\eta(n) \leq W_{\eta/4}(n) + C_\eta \sqrt{M_{\eta/4}(n)} + 1.$$

Consequently, if $M_{\eta/4}(n) = O(n)$ and $W_{\eta/4}(n) = O(n/\log n)$, then $\widehat{W}_\eta(n) = O(n/\log n)$.

Proof. Let $p_- = p_{\eta/4}$, and $p_+ = p_{1-\eta/4}$, and let $B_\pm \sim \text{Bin}(N_n, p_\pm)$. Write $\sigma_\pm^2 = N_n p_\pm(1-p_\pm)$. Choose $t = t(\eta)$ so that $t^{-2} \leq \min\{1/2, \eta/4\}$, and set $r_- = \lfloor N_n p_- - t(\sigma_- + 1) \rfloor$, and $r_+ = \lceil N_n p_+ + t(\sigma_+ + 1) \rceil$. Since $\mu_p(\mathcal{F}_n) = \mathbb{E}_{B \sim \text{Bin}(N_n, p)}$ and (a_m) is nondecreasing, Chebyshev's inequality gives

$$\begin{aligned} \frac{\eta}{4} &= \mathbb{E} a_{B_-} \geq a_{r_-} \mathbb{P}(B_- \geq r_-) \geq \frac{1}{2} a_{r_-}, \\ 1 - \frac{\eta}{4} &= \mathbb{E} a_{B_+} \leq a_{r_+} + \mathbb{P}(B_+ > r_+) \leq a_{r_+} + \frac{\eta}{4}. \end{aligned}$$

Thus $a_{r_-} < \eta$ and $a_{r_+} > 1 - \eta$, whence $m_\eta > r_-$ and $m_{1-\eta} \leq r_+$. Therefore

$$\begin{aligned} \widehat{W}_\eta(n) &\leq N_n(p_+ - p_-) + C_\eta(\sigma_- + \sigma_+ + 1) \\ &\leq W_{\eta/4}(n) + C_\eta \sqrt{M_{\eta/4}(n)} + 1, \end{aligned}$$

after changing the constant. □

Let $\tilde{\nu}_m$ be the law obtained by drawing $Q_1, \dots, Q_m$ independently and uniformly from Ω_n , and retaining the set of distinct sampled coordinates. This is the standard with-replacement model for random constraints. Write $\tilde{a}_m = \tilde{\nu}_m(\mathcal{F}_n)$, define $\tilde{m}_\alpha$ analogously, and let $\widetilde{W}_\eta = \tilde{m}_{1-\eta} - \tilde{m}_\eta$.

Proposition 2.2 (Transfer to independently sampled constraints). *Fix $0 < \eta < 1/2$. Suppose that, throughout the relevant central range, $m = O(n)$, and that $N_n = \Omega(n^2)$. Then*

$$\widetilde{W}_\eta(n) \leq \widehat{W}_{\eta/4}(n) + O_\eta(1).$$

In particular, an $O(n/\log n)$ uniform-layer window gives the same bound in the with-replacement model.

Proof. Let $S_m = \{Q_1, \dots, Q_m\}$, $L_m = |S_m|$, and $D_m = m - L_m$. Conditional on $L_m = \ell$, the set S_m is uniform on $\binom{\Omega_n}{\ell}$. Moreover,

$$\mathbb{E}D_m \leq \sum_{1 \leq i < j \leq m} \mathbb{P}(Q_i = Q_j) = \frac{\binom{m}{2}}{N_n} = O(1)$$

uniformly in the central range. Hence, for a constant $T = T(\eta)$, $\mathbb{P}(D_m > T) \leq \eta/4$.

If $m < m_{\eta/4}$, then $L_m \leq m$ and monotonicity gives $\tilde{a}_m \leq a_m < \eta$. On the other hand, for $m = m_{1-\eta/4} + T$, the event $D_m \leq T$ implies $L_m \geq m_{1-\eta/4}$, and therefore

$$\tilde{a}_m = \mathbb{E}a_{L_m} \geq \left(1 - \frac{\eta}{4}\right) \mathbb{P}(D_m \leq T) \geq \left(1 - \frac{\eta}{4}\right)^2 > 1 - \eta.$$

It follows that $\tilde{m}_\eta \geq m_{\eta/4}$ and $\tilde{m}_{1-\eta} \leq m_{1-\eta/4} + T$, proving the claim. $\square$

Corollary 2.3 (Linear-scale model transfer). *Suppose $N_n = \Omega(n^2)$, the central expected size is $\Theta(n)$, and the Bernoulli window is $O(n/\log n)$ at every fixed central level. Then the same $O(n/\log n)$ bound holds in the uniform fixed-size model and in the model with independently sampled coordinates.*

Proof. Apply Propositions 2.1 and 2.2. The $O(\sqrt{n})$ and $O_\eta(1)$ transfer errors are $o(n/\log n)$. $\square$

Assume that $\mathcal{F}_n$ is generated by a family $\mathcal{W}_n$ of inclusion-minimal witnesses, meaning $X \in \mathcal{F}_n$ if and only if there exists $Y \in \mathcal{W}_n$ such that $Y \subseteq X$, and each $Y \in \mathcal{W}_n$ is inclusion-minimal with this property.

For $s \geq 1$, define

$$\text{Small}_{n,s} = \{X : \exists Y \in \mathcal{W}_n, |Y| \leq s, Y \subseteq X\},$$

and

$$\mathcal{G}_{n,s} = \{X : \exists Y \in \mathcal{W}_n, |Y| > s, Y \subseteq X\}.$$

Then

$$\mathcal{G}_{n,s} \subseteq \mathcal{F}_n \subseteq \mathcal{G}_{n,s} \cup \text{Small}_{n,s}, \quad (2.1)$$

and $\mathcal{G}_{n,s}$ is monotone.

The event $\text{Small}_{n,s}$ records the presence of a short minimal witness, while $\mathcal{G}_{n,s}$ is the event generated only by witnesses longer than s . The inclusions (2.1) permit us to prune short witnesses before applying a Bourgain-type theorem and then return to the original event with a controlled error.

A set $A \subseteq \Omega_n$ will be called a *clean booster* at density p and scale K if $A \notin \mathcal{F}_n$, $|A| = O(K)$, and

$$\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) - \mu_p(\mathcal{F}_n) \geq \exp(-O(K)).$$

The condition $A \notin \mathcal{F}_n$ is essential: the booster reveals useful local structure but does not already contain a witness.

Definition 2.4 (Simulable clean boosters). We say that $(\mathcal{F}_n)$ has simulable clean boosters in the central interval if the following holds. For every $0 < \eta < 1/2$ and every input constant $C_0 > 0$, there are constants $C_1, c > 0$ such that, whenever $p \in [p_\eta, p_{1-\eta}]$, $1 \leq K \leq c \log V_\eta(n)$, and $A \subseteq \Omega_n$ satisfies $A \notin \mathcal{F}_n$, $|A| \leq C_0 K$, and

$$\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) \geq \mu_p(\mathcal{F}_n) + \exp(-C_0 K),$$

there exist a finite coordinate set Ω' , a monotone event $\mathcal{H} \subseteq 2^{\Omega'}$, and an integer $D \leq \exp(C_1 K)$ such that $|\Omega'|p(1-p) \geq cV_\eta(n)$.

Moreover, if $X' \sim \mu_p$ on $2^{\Omega'}$ and R_D is a multiset of D independent uniformly random coordinates of Ω' , independent of X' , then

$$\mathbb{P}(X' \cup R_D \in \mathcal{H}) \geq \mu_p(\mathcal{H}) + \exp(-C_1 K).$$

The second hypothesis in the theorem below is deliberately stated through an auxiliary event $\mathcal{H}$. In applications, forcing a local booster first exposes a lower-dimensional fibre, and the natural random sprinkling lives on the coordinate set associated with that fibre rather than on the original coordinate set. The variance comparison in Definition 2.4 ensures that this passage does not destroy the scale relevant to the layer estimate.

Theorem 2.5. *Fix $0 < \eta < 1/2$, and suppose $V_\eta(n) \rightarrow \infty$ and $p_{1-\eta} \leq 1/2$ for all sufficiently large n . Assume that $(\mathcal{F}_n)$ satisfies:*

(i) *there are constants $B, c, C > 0$ such that, for every $1 \leq K \leq c \log V_\eta(n)$ and $s = BK$,*

$$\sup_{p \in [p_\eta, p_{1-\eta}]} \mu_p(\text{Small}_{n,s}) \leq \exp(-CK),$$

where, for the constants in Lemma 3.2 with parameter $\eta/2$, $B \geq B_{\eta/2}$, and

$$C \geq \max \left\{ C_{\eta/2}^{\text{wit}} + \log 4, \log(2/\eta) \right\};$$

(ii) *$(\mathcal{F}_n)$ has simulable clean boosters.*

Then

$$W_\eta(n) \leq C_\eta \frac{M_\eta(n)}{\log V_\eta(n)}.$$

The theorem is useful when $M_\eta(n)$ and $V_\eta(n)$ have comparable polynomial size. In all three applications of this paper they are $\Theta(n)$, so the conclusion becomes $W_\eta(n) = O(n/\log n)$.

2.1. Overview of the proof. Suppose that the central window is wide. Choose a parameter K of order $M_\eta(n)/W_\eta(n)$, and delete all minimal witnesses of size at most $s = \Theta(K)$. The small-witness hypothesis makes this pruning negligible throughout the central interval. Russo's formula then provides a density p in that interval at which the pruned event has $pI_p = O(K)$. The sharp additive Bourgain theorem of Keevash, Lifshitz, Long, and Minzer produces a set of $O(K)$ coordinates whose forcing raises the probability by $\exp(-O(K))$.

The first key step is the extraction of a clean booster for the original event. On a configuration newly brought into the pruned event, choose a minimal witness and intersect it with the Bourgain set. Since the witness has size greater than s , whereas the Bourgain set is much smaller than s , this intersection is a proper sub-witness and hence cannot itself force $\mathcal{F}_n$. Pigeonholing over the possible intersections loses only another exponential factor.

The second interface is model-specific. Simulability converts the clean deterministic booster into at most $D = \exp(O(K))$ uniformly random coordinates for an auxiliary monotone event. A universal Boolean-layer estimate shows that adding D random coordinates can increase the probability of an event on N coordinates by at most

$$O\left(\frac{D+1}{\sqrt{Np(1-p)}}\right).$$

Comparing this with the $\exp(-O(K))$ boost gives

$$\exp(-O(K)) \leq \frac{\exp(O(K))}{\sqrt{V_\eta(n)}},$$

and therefore $K = \Omega(\log V_\eta(n))$. Since $K \asymp M_\eta(n)/W_\eta(n)$, this rearranges to the claimed window bound.

For atomic CSPs, a clean booster leaves one exceptional fibres and induces random codimension- $(k-1)$ cylinders there. For signed NAE-SAT, it leaves two antipodal fibres and induces complementary atomic cylinders. For hypergraph two-colourability, it again leaves two antipodal fibres but induces only constant cylinders. The replacement theorems in the later sections are precisely what verify simulability in these three geometries.

3. PROOF OF THE ABSTRACT WINDOW THEOREM

The first delicate step is the passage from the Bourgain-type booster supplied by Keevash, Lifshitz, Long, and Minzer to a *clean* booster, namely a small set which is not itself a witness for $\mathcal{F}_n$. We isolate this step in Lemma 3.2. No structure of the later constraint-satisfaction models is used here; the only inputs are pruning by minimal witnesses and rarity of small witnesses.

For a Boolean function f , let

$$I_p(f) = \sum_{i \in \Omega} \mathbb{P}_p(i \text{ is pivotal for } f)$$

denote its total p -biased influence. With this convention, Russo's formula reads $\frac{d}{dp} \mu_p(f) = I_p(f)$ for monotone f . We use the following theorem of Keevash, Lifshitz, Long, and Minzer [10, Theorem 1.4].

Theorem 3.1 (Sharp additive Bourgain theorem). *For every $0 < \eta < 1/2$ there is $C_\eta < \infty$ such that the following holds. Let $f : 2^\Omega \rightarrow \{0, 1\}$ be monotone, let $0 < p \leq 1/2$, and suppose $\eta \leq \mu_p(f) \leq 1 - \eta$, and $pI_p(f) \leq K$. Then there is a set $J \subseteq \Omega$ such that $|J| \leq C_\eta K$ and*

$$\mu_p(f_{J \rightarrow 1}) \geq \mu_p(f) + \exp(-C_\eta K).$$

Here $f_{J \rightarrow 1}$ is obtained by setting every coordinate of J equal to 1.

Lemma 3.2 (Pruned clean-booster extraction). *Fix $0 < \eta < 1/2$. There are constants $B_\eta > 0$, $C_\eta^{\text{wit}} > 0$, and $C_\eta^{\text{cl}} > 0$ such that the following holds. Let $0 < p \leq 1/2$, let $K \geq 1$, and let $s \geq B_\eta K$. Suppose $\eta \leq \mu_p(\mathcal{G}_{n,s}) \leq 1 - \eta$, $pI_p(\mathcal{G}_{n,s}) \leq K$, and*

$$\mu_p(\text{Small}_{n,s}) \leq \frac{1}{4} \exp(-C_\eta^{\text{wit}} K).$$

Then there is a nonempty set $A \subseteq \Omega_n$ such that $A \notin \mathcal{F}_n$, $|A| \leq C_\eta^{\text{cl}} K$, and

$$\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) \geq \mu_p(\mathcal{F}_n) + \exp(-C_\eta^{\text{cl}} K).$$

Proof. Let $C_K = C_K(\eta)$ be the constant in Theorem 3.1. Set $B_\eta = 2C_K$, $C_\eta^{\text{wit}} = C_K$, and choose

$$C_\eta^{\text{cl}} \geq (1 + \log 2)C_K + \log(4/3).$$

Apply Theorem 3.1 to $\mathcal{G}_{n,s}$. We obtain $J \subseteq \Omega_n$ with $|J| \leq C_K K \leq s/2$ and

$$\mu_p((\mathcal{G}_{n,s})_{J \rightarrow 1}) \geq \mu_p(\mathcal{G}_{n,s}) + \delta, \quad \delta = e^{-C_K K}.$$

Let $X \sim \mu_p$. By monotonicity of $\mathcal{G}_{n,s}$, the last display is equivalent to

$$\mathbb{P}(X \notin \mathcal{G}_{n,s}, X \cup J \in \mathcal{G}_{n,s}) \geq \delta.$$

Since $\mu_p(\text{Small}_{n,s}) \leq \delta/4$,

$$\mathbb{P}(X \notin \mathcal{G}_{n,s}, X \notin \text{Small}_{n,s}, X \cup J \in \mathcal{G}_{n,s}) \geq \frac{3\delta}{4}.$$

On this event, $X \notin \mathcal{F}_n$, by (2.1). Fix an arbitrary ordering of $\mathcal{W}_n$. Since $X \cup J \in \mathcal{G}_{n,s}$, let $Y = Y(X)$ be the first minimal witness in this ordering such that $Y \subseteq X \cup J$, and $|Y| > s$. Set $A_X = Y \cap J$. Then $A_X \neq \emptyset$, since otherwise $Y \subseteq X$, contrary to $X \notin \mathcal{F}_n$. Moreover,

$$|A_X| \leq |J| \leq s/2 < |Y|,$$

so $A_X \subsetneq Y$. If $A_X \in \mathcal{F}_n$, then A_X contains a minimal witness $Y' \in \mathcal{W}_n$, and $Y' \subseteq A_X \subsetneq Y$, contradicting the inclusion-minimality of Y . Thus $A_X \notin \mathcal{F}_n$. Finally,

$$Y = (Y \cap X) \cup (Y \cap J) \subseteq X \cup A_X,$$

so $X \cup A_X \in \mathcal{F}_n$.

There are at most $2^{|J|} \leq \exp(C_K K \log 2)$ possible values of A_X . Hence some fixed nonempty $A \subseteq J$, with $A \notin \mathcal{F}_n$, satisfies

$$\begin{aligned} \mathbb{P}(X \notin \mathcal{F}_n, X \cup A \in \mathcal{F}_n) &\geq \frac{3}{4} \exp(-(1 + \log 2)C_K K) \\ &\geq \exp(-C_\eta^{\text{cl}} K). \end{aligned}$$

Since $|A| \leq C_K K \leq C_\eta^{\text{cl}} K$, and since for a monotone event

$$\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) - \mu_p(\mathcal{F}_n) = \mathbb{P}(X \notin \mathcal{F}_n, X \cup A \in \mathcal{F}_n),$$

the claim follows. $\square$

Remark 3.3. Lemma 3.2 is where Theorem 3.1 is converted into a statement about the original event $\mathcal{F}_n$. The pruning parameter s is chosen much larger than the booster size $|J|$, so that the part of a new large witness lying inside J is a proper sub-witness and hence cannot already force $\mathcal{F}_n$. This is the mechanism that produces the clean condition $A \notin \mathcal{F}_n$.

Lemma 3.4 (Random-coordinate layer increment). *Let $\mathcal{H} \subseteq 2^\Omega$ be monotone, with $|\Omega| = N$, let $0 < p < 1$, and let $D \geq 0$ be an integer. Let $X \sim \mu_p$, and obtain Y from X by adding D independent uniformly random coordinates of Ω , independent of X , with repetitions allowed and already-present coordinates ignored. Then*

$$\mathbb{P}(Y \in \mathcal{H}) - \mathbb{P}(X \in \mathcal{H}) \leq C \min \left\{ 1, \frac{D+1}{\sqrt{Np(1-p)}} \right\}.$$

Proof. For $0 \leq r \leq N$, define the layer density

$$a_r = \frac{|\mathcal{H} \cap \binom{\Omega}{r}|}{\binom{N}{r}}.$$

Since $\mathcal{H}$ is monotone, the normalized matching property of the Boolean lattice implies that $(a_r)_{r=0}^N$ is nondecreasing.

Let $S = |X| \sim \text{Bin}(N, p)$. Conditional on $S = s$, the set X is uniform on the s -th layer. After adding D coordinates, the resulting set has size at most $s + D$; conditional on its final size it is uniform on the corresponding layer, by symmetry. Thus, by monotonicity and averaging over layers,

$$\mathbb{P}(Y \in \mathcal{H} \mid S = s) - \mathbb{P}(X \in \mathcal{H} \mid S = s) \leq a_{s+D} - a_s,$$

where $a_t = 1$ for $t > N$. Therefore

$$\mathbb{P}(Y \in \mathcal{H}) - \mathbb{P}(X \in \mathcal{H}) \leq \mathbb{E}[a_{S+D} - a_S].$$

For completeness, put $\Delta_j = a_j - a_{j-1} \geq 0$, extending a_j by 0 below 0 and by 1 above N . Then

$$\begin{aligned} \mathbb{E}[a_{S+D} - a_S] &= \sum_j \Delta_j \mathbb{P}(j - D \leq S \leq j - 1) \\ &\leq \sup_t \mathbb{P}(S \in [t - D, t]) \sum_j \Delta_j \\ &\leq \sup_t \mathbb{P}(S \in [t - D, t]). \end{aligned}$$

The standard binomial local bound gives

$$\sup_t \mathbb{P}(S = t) \leq \frac{C}{\sqrt{Np(1-p)}}.$$

Summing over at most $D + 1$ consecutive values gives the claim, together with the trivial bound by 1. $\square$

We now combine pruning, the clean-booster extraction, simulability, and the layer estimate.

Proof of Theorem 2.5. Apply Lemma 3.2 with parameter $\eta/2$, and write its constants as $B_{\eta/2}$, $C_{\eta/2}^{\text{wit}}$, and $C_{\eta/2}^{\text{cl}}$. Apply Definition 2.4 with input constant $C_0 = C_{\eta/2}^{\text{cl}}$, and denote the resulting constants by C_{sim} and c_{sim} . Replacing c by a smaller constant if necessary, we may assume $c \leq c_{\text{sim}}$. By hypothesis (i), $B \geq B_{\eta/2}$, $C \geq C_{\eta/2}^{\text{wit}} + \log 4$, and $e^{-C} \leq \eta/2$.

Assume for contradiction that

$$W_\eta(n) \geq L \frac{M_\eta(n)}{\log V_\eta(n)}$$

for a constant $L = L_\eta$ to be chosen. Put

$$K_0 = \max \left\{ 1, \frac{M_\eta(n)}{W_\eta(n)} \right\}.$$

Then

$$K_0 \leq 1 + \frac{1}{L} \log V_\eta(n).$$

Taking L sufficiently large and then n sufficiently large, we have $K_0 \leq c \log V_\eta(n)$. Let $s = BK_0$.

Hypothesis (i), with $K = K_0$, gives

$$\sup_{p \in [p_\eta, p_{1-\eta}]} \mu_p(\text{Small}_{n,s}) \leq e^{-CK_0} \leq e^{-C} \leq \eta/2.$$

For every $p \in [p_\eta, p_{1-\eta}]$, (2.1) therefore implies

$$\begin{aligned} \mu_p(\mathcal{G}_{n,s}) &\geq \mu_p(\mathcal{F}_n) - \mu_p(\text{Small}_{n,s}) \geq \eta/2, \\ \mu_p(\mathcal{G}_{n,s}) &\leq \mu_p(\mathcal{F}_n) \leq 1 - \eta \leq 1 - \eta/2. \end{aligned}$$

Russo's formula gives

$$\int_{p_\eta}^{p_{1-\eta}} I_p(\mathcal{G}_{n,s}) dp = \mu_{p_{1-\eta}}(\mathcal{G}_{n,s}) - \mu_{p_\eta}(\mathcal{G}_{n,s}) \leq 1.$$

Since $p_{1-\eta} - p_\eta = W_\eta(n)/N_n$, there is $p \in [p_\eta, p_{1-\eta}]$ such that

$$I_p(\mathcal{G}_{n,s}) \leq \frac{N_n}{W_\eta(n)}.$$

Consequently,

$$pI_p(\mathcal{G}_{n,s}) \leq \frac{pN_n}{W_\eta(n)} \leq \frac{M_\eta(n)}{W_\eta(n)} \leq K_0.$$

By Lemma 3.2, there is a nonempty $A \subseteq \Omega_n$ such that $A \notin \mathcal{F}_n$, $|A| \leq C_{\eta/2}^{\text{cl}} K_0$, and

$$\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) \geq \mu_p(\mathcal{F}_n) + e^{-C_{\eta/2}^{\text{cl}} K_0}.$$

The simulability hypothesis now gives a finite set Ω' , a monotone event $\mathcal{H} \subseteq 2^{\Omega'}$, and an integer $D \leq e^{C_{\text{sim}} K_0}$ such that $|\Omega'|p(1-p) \geq c_{\text{sim}} V_\eta(n)$ and

$$\mathbb{P}(X' \cup R_D \in \mathcal{H}) - \mu_p(\mathcal{H}) \geq e^{-C_{\text{sim}} K_0}.$$

By Lemma 3.4,

$$e^{-C_{\text{sim}} K_0} \leq C_{\text{lay}} \frac{D+1}{\sqrt{|\Omega'|p(1-p)}} \leq \frac{C_\eta e^{C_{\text{sim}} K_0}}{\sqrt{V_\eta(n)}}.$$

Taking logarithms, there is a constant $c'_\eta > 0$ such that $K_0 \geq c'_\eta \log V_\eta(n)$ for all sufficiently large n . On the other hand,

$$K_0 \leq 1 + \frac{1}{L} \log V_\eta(n).$$

Choosing $L > 2/c'_\eta$ and then n large yields a contradiction. This proves the claimed bound with $C_\eta = L$. $\square$

Remark 3.5. The hypothesis $p_{1-\eta} \leq 1/2$ is included only to match the form of Theorem 3.1. In the atomic CSP application this condition follows from the linear central scale, since the relevant probabilities satisfy $p = \Theta(n^{1-k})$. Note that no symmetry or CSP structure is used in the abstract theorem beyond the two stated hypotheses.

4. EXCEPTIONAL FIBRES AND TWO LOCAL-TO-RANDOM REDUCTION

The three applications below have different local geometries, but the probabilistic part of their fibre arguments is identical. We isolate it here in two results. The first converts a binomial family of helpful lower-arity objects into a fixed number of independent uniform helpers. The second converts a pointwise power comparison between two classes of local objects into an averaged replacement theorem which preserves a positive excess probability.

Let X be a finite set and let $\mathcal{Q}$ be a finite family of subsets of X . For a subfamily $\mathcal{S} \subseteq \mathcal{Q}$, write $\bigcup \mathcal{S} = \bigcup_{Q \in \mathcal{S}} Q$. A θ -random subfamily of $\mathcal{Q}$ is obtained by retaining each member independently with probability θ .

Lemma 4.1 (Binomial-to-independent helper extraction). *Let B be a random subset of X , and let $\mathcal{H}_\theta$ be a θ -random subfamily of $\mathcal{Q}$, independent of B . Let $\zeta = \mathbb{P}(B = \emptyset)$, and $\lambda = \theta|\mathcal{Q}|$. Suppose that, for some $K \geq 1$, $\Lambda \geq 1$, and $0 < \beta \leq 1/2$, $\lambda \leq \Lambda K$, and $\mathbb{P}(B \subseteq \bigcup \mathcal{H}_\theta) \geq \zeta + \beta$. There is a constant $C = C(\Lambda)$ such that, if*

$$|\mathcal{Q}| \geq \frac{4 \lceil C(K + \log(1/\beta)) \rceil^2}{\beta},$$

then there is an integer $1 \leq d \leq \lceil C(K + \log(1/\beta)) \rceil$ for which independent uniformly random $Q_1, \dots, Q_d \in \mathcal{Q}$, also independent of B , satisfy

$$\mathbb{P}(B \subseteq Q_1 \cup \dots \cup Q_d) \geq \zeta + \frac{\beta}{2}.$$

Proof. Let $L = |\mathcal{H}_\theta| \sim \text{Bin}(|\mathcal{Q}|, \theta)$, and let $T = \lceil C(K + \log(1/\beta)) \rceil$. Choose the constant in the definition of T so that the standard binomial upper-tail estimate gives

$$\mathbb{P}(L > T) \leq \frac{\beta}{8}.$$

Indeed, for $T \geq e^2 \lambda$,

$$\mathbb{P}(L \geq T) \leq \left(\frac{e\lambda}{T} \right)^T \leq e^{-T},$$

and the asserted choice of T ensures both $T \geq e^2 \lambda$ and $e^{-T} \leq \beta/8$.

Conditional on $L = \ell$, the family $\mathcal{H}_\theta$ is a uniformly random ℓ -element subset of $\mathcal{Q}$. Let

$$b_\ell = \mathbb{P}\left(B \subseteq \bigcup \mathcal{S}_\ell\right) - \zeta,$$

where $\mathcal{S}_\ell$ is such a subset, independent of B . Since adding sets can only help, $b_\ell \geq 0$, and $b_0 = 0$. Therefore

$$\begin{aligned} \beta &\leq \sum_{\ell \geq 0} \mathbb{P}(L = \ell) b_\ell \\ &\leq \sum_{1 \leq \ell \leq T} \mathbb{P}(L = \ell) b_\ell + \mathbb{P}(L > T). \end{aligned}$$

Thus some $1 \leq d \leq T$ satisfies $b_d \geq 7\beta/8$.

A uniformly random d -element subset of $\mathcal{Q}$ has the same law as d independent uniform samples conditioned on having no collision. The total variation distance between these two laws is at most

$$\mathbb{P}(\text{a collision}) \leq \frac{\binom{d}{2}}{|\mathcal{Q}|} \leq \frac{\beta}{8}.$$

Consequently the independent sample has excess at least $7\beta/8 - \beta/8 \geq \beta/2$, as required. $\square$

Corollary 4.2 (Logarithmic-scale helper extraction). *Fix positive constants C_1, C_2, a . Suppose $\lambda \leq C_1 K$, $\beta \geq e^{-C_2 K}$, $|\mathcal{Q}| \geq n^a$, and $1 \leq K \leq c \log n$. If $c = c(C_1, C_2, a) > 0$ is sufficiently small, then the conclusion of Lemma 4.1 holds with $1 \leq d \leq CK$ and excess at least e^{-CK} , where $C = C(C_1, C_2, a)$.*

Proof. The parameter T in Lemma 4.1 is $O(K)$. Moreover, $4T^2/\beta \leq CK^2 e^{C_2 K} \leq n^a$ when $K \leq c \log n$ and c is sufficiently small. Apply the lemma and adjust constants. $\square$

Let $\mathcal{Q}$ and $\mathcal{R}$ be finite families of subsets of the same finite set X . For $S \subseteq X$, define $\rho_{\mathcal{Q}}(S) = \mathbb{P}_{Q \in \mathcal{Q}}(S \subseteq Q)$, and $\rho_{\mathcal{R}}(S) = \mathbb{P}_{R \in \mathcal{R}}(S \subseteq R)$, where the objects are chosen uniformly.

Theorem 4.3 (Averaged power replacement). *Let $\gamma \geq 1$, $c_* > 0$, and $\varepsilon_0 \geq 0$. Suppose that*

$$\rho_{\mathcal{Q}}(S) \geq \varepsilon \geq \varepsilon_0 \implies \rho_{\mathcal{R}}(S) \geq c_* \varepsilon^\gamma \quad (4.1)$$

for every $S \subseteq X$. Let B be a random subset of X , independent of all local objects below, and let $\zeta \in [0, 1]$. Suppose that $d \geq 1$ independent uniform members of $\mathcal{Q}$ cover B with probability at least $\zeta + \alpha$, where $\alpha > 0$, and that $\varepsilon := \frac{\alpha}{4d} \geq \varepsilon_0$. Then there is an integer $D \leq C_{c_, \gamma} d^{1+\gamma} \alpha^{-\gamma} \log(2d/\alpha)$ such that D independent uniformly random members of $\mathcal{R}$ cover B with probability at least $\zeta + \alpha/2$.*

Proof. Let $Q_1, \dots, Q_d$ be the original independent members of $\mathcal{Q}$. Replace them one at a time by independent blocks of members of $\mathcal{R}$, using fresh randomness at every step. At a given step, condition on B , on all local objects other than the Q_i currently being replaced, and on every block inserted at earlier steps. Under this conditioning, Q_i remains independent and uniform. Let $S \subseteq B$ be the residual set not covered by the other local objects. The current conditional success probability is $\rho_{\mathcal{Q}}(S)$.

If $\rho_{\mathcal{Q}}(S) < \varepsilon$, deleting Q_i and inserting an arbitrary fresh block loses less than ε . Otherwise, (4.1) gives $\rho_{\mathcal{R}}(S) \geq c_* \varepsilon^\gamma$. Hence a block of $g = \lceil C_{c_*, \gamma} \varepsilon^{-\gamma} \log(1/\varepsilon) \rceil$ independent members of $\mathcal{R}$ fails to contain S with probability at most ε . In either case the replacement loses at most ε in conditional, and therefore unconditional, success probability.

After all d replacements, the total loss is at most $d\varepsilon = \alpha/4$. The final success probability is consequently at least $\zeta + 3\alpha/4$, which is stronger than claimed. Finally, $D = dg$, and substituting $\varepsilon = \alpha/(4d)$ gives the displayed bound after adjusting the constant. $\square$

Proposition 4.4 (Exceptional-fibre criterion for simulability). *Fix $0 < \eta < 1/2$. Suppose that, for every input constant $C_0 > 0$, there are positive constants $C, c, c_{\text{var}}, c_*, C_{\text{th}}, b$, and $\gamma \geq 1$ with the following property. Whenever $p \in [p_\eta, p_{1-\eta}]$, $1 \leq K \leq c \log V_\eta(n)$, and $A \subseteq \Omega_n$ is a clean booster satisfying $A \notin \mathcal{F}_n$, $|A| \leq C_0 K$, and $\mu_p((\mathcal{F}_n)_{A \rightarrow 1}) \geq \mu_p(\mathcal{F}_n) + e^{-C_0 K}$, there exist the following:*

- (i) *a finite state space X , a finite coordinate set Ω' , and subsets $(R_\omega)_{\omega \in \Omega'}$ of X , defining the monotone covering event*

$$\mathcal{H} = \left\{ Y \subseteq \Omega' : X \subseteq \bigcup_{\omega \in Y} R_\omega \right\};$$

- (ii) *a finite helper family $\mathcal{Q}$ of subsets of X , and an integer $1 \leq d \leq CK$, such that, if $X' \sim \mu_p$ on $2^{\Omega'}$, $B = B(X') := X \setminus \bigcup_{\omega \in X'} R_\omega$, and $Q_1, \dots, Q_d$ are independent uniformly random members of $\mathcal{Q}$, independent of X' , then $\mathbb{P}(B \subseteq Q_1 \cup \dots \cup Q_d) \geq \mu_p(\mathcal{H}) + e^{-CK}$;*
- (iii) *the pointwise comparison*

$$\rho_{\mathcal{Q}}(S) \geq \varepsilon \geq C_{\text{th}} V_\eta(n)^{-b} \implies \rho_{\Omega'}(S) \geq c_* \varepsilon^\gamma \quad (4.2)$$

for every $S \subseteq X$, where $\rho_{\Omega'}(S) = \frac{1}{|\Omega'|} |\{\omega \in \Omega' : S \subseteq R_\omega\}|$;

- (iv) *the variance comparison $|\Omega'| p(1-p) \geq c_{\text{var}} V_\eta(n)$.*

Then $(\mathcal{F}_n)$ has simulable clean boosters in the central interval.

Proof. Put $\alpha = e^{-CK}$ and $\varepsilon = \alpha/(4d)$. Since $d \leq CK$, after decreasing c as a function of C and b , the restriction $K \leq c \log V_\eta(n)$ ensures

$$\varepsilon \geq \frac{e^{-CK}}{4CK} \geq C_{\text{th}} V_\eta(n)^{-b}.$$

Apply Theorem 4.3 to the helper family $\mathcal{Q}$ and the indexed native family $(R_\omega)_{\omega \in \Omega'}$, regarded with its Ω' -multiplicity, using (4.2). We obtain an integer

$$D \leq C_{c^*, \gamma} d^{1+\gamma} \alpha^{-\gamma} \log(2d/\alpha) \leq \exp(C'K)$$

for which D independent uniform coordinates $\omega_1, \dots, \omega_D \in \Omega'$ satisfy

$$\mathbb{P}(B \subseteq R_{\omega_1} \cup \dots \cup R_{\omega_D}) \geq \mu_p(\mathcal{H}) + \frac{\alpha}{2}.$$

The event inside the probability is exactly $X' \cup \{\omega_1, \dots, \omega_D\} \in \mathcal{H}$. Since $\alpha/2 \geq e^{-C''K}$ for $K \geq 1$, and condition (iv) gives the required variance comparison, this is precisely Definition 2.4. $\square$

Remark 4.5. In each application, the maximal local quasi-cover and the exceptional fibres are model-specific. Once they have been identified, however, the remainder is formal. The selected constraints meeting the booster variables once form a binomial family of lower-arity helpers, to which Corollary 4.2 applies. A model-specific pointwise estimate of the form (4.1) then invokes Theorem 4.3 and produces the random native constraints required by simulability.

5. A FINITE CRITERION FOR FORBIDDEN-PATTERN SYSTEMS

We now turn the exceptional-fibre mechanism into a predicate-level theorem. The criterion below has a finite combinatorial part, which identifies the exceptional fibres, and an analytic part, which compares one-coordinate derivatives with native constraints.

Fix a finite alphabet Σ , an integer $k \geq 2$, and a nonempty finite family $\mathcal{P}$ of nonempty proper subsets of Σ^k . We assume that $\mathcal{P}$ is closed under permutations of the k coordinates. A $\mathcal{P}$ -constraint on a k -set $S = \{i_1 < \dots < i_k\}$ is a pair (S, P) , with $P \in \mathcal{P}$, and it covers the assignments $x \in \Sigma^n$ for which $(x_{i_1}, \dots, x_{i_k}) \in P$. Let

$$\Omega_n(\mathcal{P}) = \binom{[n]}{k} \times \mathcal{P}, \quad N_n(\mathcal{P}) = |\mathcal{P}| \binom{n}{k},$$

and let $\mathcal{F}_n(\mathcal{P})$ be the event that the selected constraints cover all of Σ^n .

Let $\Gamma \leq \text{Sym}(\Sigma)$ be a finite group such that every $P \in \mathcal{P}$ is invariant under the diagonal action of Γ . Define a preorder on Σ^k by

$$u \preceq_{\mathcal{P}} v \iff \{P \in \mathcal{P} : u \in P\} \subseteq \{P \in \mathcal{P} : v \in P\}. \quad (5.1)$$

For an injection $\phi : [k] \rightarrow R$, write $x_\phi = (x_{\phi(1)}, \dots, x_{\phi(k)})$.

Definition 5.1 (Rigidity). The system $(\mathcal{P}, \Gamma)$ is *rigid* if there is an integer $L \geq k$ such that, whenever R is finite, every symbol of Σ occurs at least L times in $\sigma \in \Sigma^R$, and $\tau \in \Sigma^R$ satisfies

$$\tau_\phi \preceq_{\mathcal{P}} \sigma_\phi \quad \text{for every injection } \phi : [k] \rightarrow R. \quad (5.2)$$

then $\tau = g\sigma$ for some $g \in \Gamma$.

Although Definition 5.1 is stated for arbitrarily large sets, it has a finite certificate. Let $T_L = \Sigma \times [L]$, and $\sigma_L(a, j) = a$. For $a \in \Sigma$, let $T_{L,a} = T_L \cup \{*\}$ and extend σ_L by $\sigma_{L,a}(*) = a$.

Proposition 5.2 (Finite rigidity certificate). *The system $(\mathcal{P}, \Gamma)$ is rigid with parameter L if and only if the following two finite assertions hold.*

- (i) Every $\tau \in \Sigma^{T_L}$ satisfying (5.2) with $\sigma = \sigma_L$ is equal to $g\sigma_L$ for some $g \in \Gamma$.
- (ii) For every $a \in \Sigma$, $g \in \Gamma$, and $\tau \in \Sigma^{T_{L,a}}$ satisfying (5.2) with $\sigma = \sigma_{L,a}$, if $\tau|_{T_L} = g\sigma_L$, then $\tau(*) = g(a)$.

In particular, whether a proposed L is a valid rigidity parameter is decidable by a finite computation depending only on $\Sigma, k, \mathcal{P}, \Gamma, L$.

Proof. Necessity follows by applying rigidity to the two displayed finite templates. Conversely, let $\sigma \in \Sigma^R$ contain every symbol at least L times, and suppose that τ satisfies (5.2). Choose a subset $T \subseteq R$ containing exactly L coordinates of each σ -value. After identifying T with T_L , condition (i) gives $\tau|_T = g\sigma|_T$ for some $g \in \Gamma$. For each $i \in R \setminus T$, apply condition (ii) to $T \cup \{i\}$, with $a = \sigma_i$. It gives $\tau_i = g(\sigma_i)$. Thus $\tau = g\sigma$ on all of R . $\square$

The connection with exceptional fibres is immediate. For $\sigma \in \Sigma^R$, let

$$M(R, \sigma) = \{(S, P) \in \Omega_R(\mathcal{P}) : \sigma_S \notin P\}. \quad (5.3)$$

Lemma 5.3 (Rigidity gives a finite exceptional orbit). *Suppose $(\mathcal{P}, \Gamma)$ is rigid and every symbol occurs at least L times in σ . Then the assignments not covered by $M(R, \sigma)$ are exactly $\Gamma\sigma = \{g\sigma : g \in \Gamma\}$.*

Proof. Every $g\sigma$ is uncovered because each forbidden set P is Γ -invariant. If τ is uncovered, then there is no constraint (S, P) for which $\tau_S \in P$ but $\sigma_S \notin P$. Equivalently, $\tau_\phi \preceq_{\mathcal{P}} \sigma_\phi$ for every injection $\phi : [k] \rightarrow R$, where closure of $\mathcal{P}$ under coordinate permutations removes the choice of ordering. Rigidity now gives $\tau = g\sigma$. $\square$

For $a \in \Sigma$ and $P \in \mathcal{P}$, define the first-coordinate derivative $\partial_a P = \{z \in \Sigma^{k-1} : (a, z) \in P\}$. Let $\mathcal{D}(\mathcal{P}) = \{(a, P) : a \in \Sigma, P \in \mathcal{P}, \partial_a P \neq \emptyset\}$ be an indexed family, so equal derivative sets retain their multiplicities. For $S \subseteq \Sigma^m$, let $\rho_{\mathcal{P}, m}(S)$ be the probability that a uniformly random native constraint $(U, P) \in \binom{[m]}{k} \times \mathcal{P}$ contains S , and let $\rho_{\partial\mathcal{P}, m}(S)$ be the probability that a uniformly random indexed derivative constraint

$$(U, a, P) \in \binom{[m]}{k-1} \times \mathcal{D}(\mathcal{P})$$

contains S , where its covered set is $\{x \in \Sigma^m : x_U \in \partial_a P\}$.

Definition 5.4 (Power-replaceable derivatives). The system $\mathcal{P}$ has *power-replaceable one-coordinate derivatives* if there are constants $c_*, C_*, b > 0$ and $\gamma \geq 1$ such that, for every m and every $S \subseteq \Sigma^m$,

$$\rho_{\partial\mathcal{P}, m}(S) \geq \varepsilon \geq C_* m^{-b} \implies \rho_{\mathcal{P}, m}(S) \geq c_* \varepsilon^\gamma. \quad (5.4)$$

The indexing convention in Definition 5.4 is important. After balancing a local assignment, every indexed derivative has the same number of one-vertex extensions to a native constraint. Thus the helpers induced by random crossing constraints form an exact binomial subfamily of the indexed derivative family.

For $s \geq 1$, let $\text{Small}_{n,s}(\mathcal{P})$ be the event that the random $\mathcal{P}$ -instance contains an inclusion-minimal cover with at most s constraints.

Theorem 5.5 (Finite-predicate logarithmic-window criterion). *Fix $0 < \eta < 1/2$ and a symmetric forbidden-pattern system $(\Sigma, k, \mathcal{P}, \Gamma)$. Suppose:*

- (i) $(\mathcal{P}, \Gamma)$ is rigid;
- (ii) $\mathcal{P}$ has power-replaceable one-coordinate derivatives;
- (iii) uniformly for $p \in [p_\eta, p_{1-\eta}]$,

$$pN_n(\mathcal{P}) = \Theta(n). \quad (5.5)$$

- (iv) for every $A_0, B_0 > 0$, there is $c = c(A_0, B_0) > 0$ such that, whenever $1 \leq K \leq c \log n$,

$$\sup_{p \in [p_\eta, p_{1-\eta}]} \mu_p(\text{Small}_{n, B_0 K}(\mathcal{P})) \leq e^{-A_0 K}. \quad (5.6)$$

Then

$$W_\eta(n) \leq C_{\eta, \mathcal{P}} \frac{n}{\log n}.$$

If conditions (iii)–(iv) also hold with η replaced by $\eta/4$, then the same bound holds at central level η in the uniform fixed-size model and in the model of independently sampled constraints.

Proof. Condition (iv) implies the short-witness hypothesis of Theorem 2.5, after choosing A_0, B_0 to dominate the constants in the clean-booster lemma. By (5.5), throughout the central interval

$$p = \Theta(n^{1-k}), \quad M_\eta(n) = \Theta(n), \quad V_\eta(n) = \Theta(n). \quad (5.7)$$

and in particular $p_{1-\eta} \leq 1/2$ for large n . It remains to verify simulability.

Fix an arbitrary input constant $C_0 > 0$, let $1 \leq K \leq c \log n$, and let $A \subseteq \Omega_n(\mathcal{P})$ be a clean booster with $|A| \leq C_0 K$, and $\mu_p((\mathcal{F}_n(\mathcal{P}))_{A \rightarrow 1}) \geq \mu_p(\mathcal{F}_n(\mathcal{P})) + e^{-C_0 K}$. Let $\beta = e^{-C_0 K}$. Let R_0 be the variables used by A . Since A is not a cover, choose $\sigma_0 \in \Sigma^{R_0}$ uncovered by A . Add fresh variables and extend σ_0 so that, on the resulting set R , every symbol occurs the same number $r_* \geq L$ of times. Then, $|R| = O_{\Sigma, k, C_0}(K)$. After relabelling variables, assume R is an initial segment of $[n]$. This does not change the law because $\mathcal{P}$ is closed under coordinate permutations.

Force the maximal family $M(R, \sigma)$ from (5.3). It contains A , so the probability increment remains at least β . By Lemma 5.3, the only uncovered fibres over R are $F_{g\sigma}$, $g \in \Gamma$. Put $m = n - |R|$. Constraints disjoint from R form an independent random $\mathcal{P}$ -instance on m variables. Let $B \subseteq \Sigma^m$ be its uncovered set. Because every $P \in \mathcal{P}$ is Γ -invariant, B is Γ -invariant.

Consider selected constraints meeting R in exactly one vertex. Since R is an initial segment, that vertex occupies the first coordinate of the local pattern. On the fibre F_σ , such a constraint induces a member of the indexed derivative family. Conversely, every indexed derivative has exactly r_* extensions, one for each local vertex carrying its specified symbol. Hence the induced helpers form a θ -random subfamily of

$$\binom{[m]}{k-1} \times \mathcal{D}(\mathcal{P}), \quad \theta = 1 - (1-p)^{r_*},$$

independently of B , and their expected number satisfies

$$\lambda \leq r_* p |\mathcal{D}(\mathcal{P})| \binom{m}{k-1} = O_{\eta, \mathcal{P}, C_0}(K). \quad (5.8)$$

On the fibre $F_{g\sigma}$, the same crossing constraint induces the g -image of its helper on F_σ . Since $B = gB$, coverage of B on the distinguished fibre automatically covers all exceptional fibres.

The probability that any selected non-forced constraint meets R in at least two vertices is at most

$$Cp \sum_{j=2}^k |R|^j n^{k-j} = O_{\eta, \mathcal{P}, C_0}(K^2/n) = o(\beta). \quad (5.9)$$

after decreasing the constant in $K \leq c \log n$. Let $E_{\geq 2}$ denote this exceptional event. Off $E_{\geq 2}$, the instance with $M(R, \sigma)$ forced is a cover if and only if the induced derivative family covers B on the distinguished fibre, and hence on every exceptional fibre. Therefore

$$\begin{aligned} \mathbb{P}(B \text{ is covered by the binomial derivative family}) &\geq \mu_p((\mathcal{F}_n(\mathcal{P}))_{M(R, \sigma) \rightarrow 1}) - \mathbb{P}(E_{\geq 2}) \\ &\geq \mu_p(\mathcal{F}_n(\mathcal{P})) + \beta - o(\beta). \end{aligned}$$

A cover on the m outside variables covers the original instance after the remaining variables are added, so $\mu_p(\mathcal{F}_n(\mathcal{P})) \geq \mu_p(\mathcal{F}_m(\mathcal{P}))$. Consequently, for large n , the binomial derivative family covers B with probability at least

$$\mu_p(\mathcal{F}_m(\mathcal{P})) + \frac{\beta}{2}. \quad (5.10)$$

Since the indexed derivative family has $\Theta(m^{k-1})$ members, Corollary 4.2 gives $d = O_{\eta, \mathcal{P}, C_0}(K)$ independent uniform derivative helpers with excess $\alpha = e^{-O_{\eta, \mathcal{P}, C_0}(K)}$.

Apply Proposition 4.4 with $X = \Sigma^m$, $\Omega' = \Omega_m(\mathcal{P})$, $\mathcal{H} = \mathcal{F}_m(\mathcal{P})$, and with the indexed derivative family as $\mathcal{Q}$. The pointwise comparison is precisely (5.4). Since $m = n - O(K)$ and (5.7) holds,

$$|\Omega'| p(1-p) = \Theta(n) \geq cV_\eta(n).$$

Thus clean boosters are simulable. The abstract theorem yields the Bernoulli bound. If the central-scale and witness hypotheses hold at level $\eta/4$, apply the Bernoulli conclusion at that level. Since $N_n(\mathcal{P}) = \Theta(n^k) = \Omega(n^2)$, Corollary 2.3 then gives the two fixed-size formulations at level η . $\square$

Remark 5.6 (Finite rigidity certificates). For fixed $(\Sigma, k, \mathcal{P}, \Gamma)$, the rigidity part of the theorem can be checked by finite enumeration using Proposition 5.2. The remaining predicate-specific question is the power comparison (5.4). In the examples below this reduces to comparing the number of coordinates frozen by a set of assignments. Thus the theorem separates the problem into a finite local classification and a quantitative extremal inequality for the derivative family.

6. POINTWISE CUBICAL REPLACEMENT OVER A FINITE ALPHABET

Fix integers $q \geq 2$ and $k \geq 2$. A codimension- j cylinder in $[q]^m$ is obtained by fixing j distinct coordinates to specified values. Let $\mathcal{U}_{m,q}^{(j)}$ be the family of codimension- j cylinders, so $|\mathcal{U}_{m,q}^{(j)}| = q^j \binom{m}{j}$. For $B \subseteq [q]^m$, let $\rho_j(B)$ be the probability that a uniformly random codimension- j cylinder contains B . If $B \neq \emptyset$, let $\ell(B)$ be the number of coordinates on which all points of B have the same value. Then

$$\rho_j(B) = \frac{\binom{\ell(B)}{j}}{q^j \binom{m}{j}},$$

with the convention that this is zero if $\ell(B) < j$. If $B = \emptyset$, then $\rho_j(B) = 1$.

Lemma 6.1 (Pointwise cubical replacement). *Fix $q \geq 2$ and $k \geq 2$. There are constants $c_{q,k}, C_{q,k} > 0$ such that if $B \subseteq [q]^m$ and $\rho_{k-1}(B) \geq \varepsilon \geq C_{q,k} m^{-(k-1)}$, then $\rho_k(B) \geq c_{q,k} \varepsilon^{k/(k-1)}$.*

Proof. The case $B = \emptyset$ is trivial. Otherwise write $\ell = \ell(B)$. If $C_{q,k}$ is sufficiently large, then $\rho_{k-1}(B) \geq C_{q,k} m^{-(k-1)}$ implies $\ell \geq 2k$. Hence

$$\rho_{k-1}(B) \asymp_{q,k} \left(\frac{\ell}{m}\right)^{k-1}, \quad \rho_k(B) \asymp_{q,k} \left(\frac{\ell}{m}\right)^k.$$

The claimed inequality follows. $\square$

Corollary 6.2 (Averaged cubical replacement with preserved excess). *Fix $q \geq 2$ and $k \geq 2$. Let B be a random subset of $[q]^m$, independent of all auxiliary cylinders below, and let $\zeta \in [0, 1]$. Suppose that $d \geq 1$ independent uniformly random codimension- $(k-1)$ cylinders cover B with probability at least $\zeta + \alpha$, where $\alpha > 0$, and suppose that $\frac{\alpha}{4d} \geq C_{q,k} m^{-(k-1)}$, where $C_{q,k}$ is the constant from Lemma 6.1. Then there is an integer*

$$D \leq C'_{q,k} d^{1+k/(k-1)} \alpha^{-k/(k-1)} \log(2d/\alpha)$$

such that D independent uniformly random codimension- k cylinders cover B with probability at least $\zeta + \alpha/2$.

Proof. Apply Theorem 4.3 with $\mathcal{Q} = \mathcal{U}_{m,q}^{(k-1)}$, $\mathcal{R} = \mathcal{U}_{m,q}^{(k)}$, $\gamma = k/(k-1)$, and the pointwise estimate from Lemma 6.1. $\square$

7. RANDOM ATOMIC CSPs

Fix $q \geq 2$ and $k \geq 2$. The variables take values in $[q] = \{0, \dots, q-1\}$. An atomic constraint is specified by a k -set $S \subset [n]$ and a pattern $a \in [q]^S$; it forbids assignments satisfying $x_S = a$. In the dual covering language, this constraint is the codimension- k cylinder $\{x \in [q]^n : x_S = a\}$. Let $\mathcal{C}_{n,q,k}$ be the set of all atomic constraints, so

$$N_{n,q,k} = |\mathcal{C}_{n,q,k}| = q^k \binom{n}{k}.$$

The random atomic CSP includes each coordinate of $\mathcal{C}_{n,q,k}$ independently with probability p . Let $\text{Cover}_{n,q,k} \subseteq 2^{\mathcal{C}_{n,q,k}}$ be the event that the chosen cylinders cover all of $[q]^n$, equivalently that the CSP is unsatisfiable.

For $0 < \eta < 1/2$, define

$$W_{\eta,q,k}(n) = N_{n,q,k}(p_{1-\eta} - p_\eta),$$

where p_α is the generalized inverse defined above for the event $\text{Cover}_{n,q,k}$.

The second delicate step is the passage from a minimally unsatisfiable atomic instance to the q -ary Tarsi deficiency theorem. We spell this out before using the deficiency bound in the linear-location and small-witness estimates.

An atomic constraint on a variable set S with forbidden pattern $a \in [q]^S$ is identified with the codimension- k subcube $\{x \in [q]^R : x_S = a\}$ inside the cube on any larger variable set $R \supseteq S$. Thus an unsatisfiable atomic instance on variables R is the same thing as a subcube cover of $[q]^R$.

Lemma 7.1 (Minimal unsatisfiability gives a tight q -ary cover). *Let Y be an inclusion-minimal unsatisfiable atomic q -ary instance, and let R be the set of variables used by Y . Then the subcubes corresponding to the constraints of Y , viewed inside $[q]^R$, form a minimal cover of $[q]^R$ using all coordinates of R . In particular, Y is an instance to which the q -ary Tarsi deficiency theorem applies.*

Proof. Every assignment $\sigma \in [q]^R$ extends to an assignment of all ambient variables. Since Y is unsatisfiable, this extension violates at least one constraint of Y . That constraint depends only on variables in R , so its corresponding subcube contains σ . Hence the subcubes cover $[q]^R$.

The cover is minimal. Indeed, if a constraint $Q \in Y$ were redundant in the cover of $[q]^R$, then $Y \setminus \{Q\}$ would still cover $[q]^R$, and therefore would still be unsatisfiable, contradicting the inclusion-minimality of Y . Finally, every coordinate of R is used by some constraint by the definition of R . Thus the cover is tight in the sense needed for the Tarsi bound. $\square$

We use the following q -ary generalization of Tarsi's lemma for subcube covers.

Theorem 7.2 (q -ary Tarsi lemma for subcube covers). *Every tight minimal subcube cover of $[q]^v$ has size at least $(q-1)v + 1$.*

This is [7, Theorem 3.11], proved there as a special case of a matroid-cover version of Tarsi's lemma. There is also a direct and equivalent clause-set formulation in [13, Corollary 1.9.9]. Explicitly, an atomic constraint $Q = (S, a)$ corresponds to the generalised clause

$$C_Q = \bigvee_{i \in S} (x_i \neq a_i).$$

The assignments falsifying C_Q are precisely the points of the subcube covered by Q . Consequently, an atomic instance is unsatisfiable exactly when the associated generalised clause-set is unsatisfiable, and inclusion-minimality agrees in the two languages. For a uniform alphabet of size q , Kullmann's weighted number of variables is

$$\text{wn}(F) = \sum_{i \in \text{var}(F)} (|D_i| - 1) = (q-1)v,$$

so his inequality $c(F) - \text{wn}(F) \geq 1$ gives the same bound. Combining either formulation with Lemma 7.1 gives the exact atomic statement used below. For the Boolean historical form, see Aharoni-Linial [12], who explicitly attribute the result to an unpublished observation of Tarsi.

Corollary 7.3 (Tarsi deficiency for minimal atomic CSPs). *Every inclusion-minimal unsatisfiable atomic q -ary instance using v variables has at least $(q-1)v + 1$ constraints.*

We next show that the coarse linear location of the atomic threshold is automatic in exactly the range needed below. Let $a = q-1$, and $\delta = a(k-1) - 1$. Thus $\delta > 0$ is equivalent to

$$\gamma_{q,k} = (k-1) - \frac{1}{q-1} > 0.$$

Proposition 7.4 (Linear lower bound for atomic unsatisfiability). *Fix $q \geq 2$ and $k \geq 2$, and suppose $(q-1)(k-1) > 1$. There is $c = c(q,k) > 0$ such that, if $pN_{n,q,k} \leq cn$, and $N_{n,q,k} = q^k \binom{n}{k}$, then $\mu_p(\text{Cover}_{n,q,k}) = o(1)$. Equivalently, a random atomic q -ary k -CSP with expected number of constraints at most cn is satisfiable with high probability.*

Proof. Write $N = N_{n,q,k}$. Put $m_0 = pN$, and assume $m_0 \leq cn$, where $c > 0$ will be chosen sufficiently small. For a set $S \subseteq [n]$, let X_S be the number of selected atomic constraints whose support is contained in S . If $|S| = v$, then

$$X_S \sim \text{Bin}\left(q^k \binom{v}{k}, p\right), \quad \text{and} \quad \mathbb{E}X_S = pq^k \binom{v}{k} = m_0 \frac{\binom{v}{k}}{\binom{n}{k}} \leq C_{q,k} cn \left(\frac{v}{n}\right)^k.$$

Set $a = q - 1$. We claim that, with high probability, there is no set $S \subseteq [n]$ with $X_S \geq a|S| + 1$. It is enough to bound the event $X_S \geq av$. By the standard binomial upper-tail estimate,

$$\mathbb{P}(X_S \geq av) \leq \left(\frac{e\mathbb{E}X_S}{av}\right)^{av} \leq \left(C_{q,k}c \left(\frac{v}{n}\right)^{k-1}\right)^{av}.$$

Therefore the expected number of sets S with $|S| = v$ and $X_S \geq av$ is at most

$$\begin{aligned} \binom{n}{v} \left(C_{q,k}c \left(\frac{v}{n}\right)^{k-1}\right)^{av} &\leq \left[\frac{en}{v} \left(C_{q,k}c \left(\frac{v}{n}\right)^{k-1}\right)^a\right]^v \\ &= \left[C'_{q,k}c^a \left(\frac{v}{n}\right)^{a(k-1)-1}\right]^v. \end{aligned}$$

Since $a(k-1) - 1 = \delta > 0$, choose $c = c(q, k) > 0$ so small that $C'_{q,k}c^a < 1/4$. Then

$$\sum_{v=k}^n \left[C'_{q,k}c^a \left(\frac{v}{n}\right)^\delta\right]^v = o(1).$$

Indeed, for fixed L , the contribution of $k \leq v < L$ tends to zero, while the contribution of $L \leq v \leq n$ is at most $\sum_{v \geq L} 4^{-v}$, which can be made arbitrarily small by taking L large.

Now suppose the selected instance were unsatisfiable. Choose an inclusion-minimal unsatisfiable subinstance Y , and let S be the set of variables used by Y . Write $v = |S|$ and $t = |Y|$. By Corollary 7.3, $t \geq (q-1)v + 1 = av + 1$. But all constraints of Y are supported inside S , so $X_S \geq t \geq av + 1$, contradicting the high-probability event proved above. Hence the whole instance is satisfiable with high probability. $\square$

Proposition 7.5 (Linear upper bound for atomic unsatisfiability). *Fix $q \geq 2$ and $k \geq 2$. There is $C = C(q, k) > 0$ such that, if $pN_{n,q,k} \geq Cn$, then $\mu_p(\text{Cover}_{n,q,k}) = 1 - o(1)$.*

Proof. Let Z be the number of satisfying assignments. A fixed assignment $x \in [q]^n$ is ruled out by exactly one atomic constraint on each k -set of variables. Hence

$$\mathbb{P}(x \text{ satisfies all selected constraints}) = (1-p)^{\binom{n}{k}}.$$

Therefore

$$\mathbb{E}Z = q^n (1-p)^{\binom{n}{k}} \leq \exp\left(n \log q - p \binom{n}{k}\right).$$

Since

$$p \binom{n}{k} = \frac{pN_{n,q,k}}{q^k},$$

we have $\mathbb{E}Z = o(1)$ as soon as $pN_{n,q,k} \geq Cn$ with $C > q^k \log q$. Markov's inequality then gives $Z = 0$ with high probability. $\square$

Corollary 7.6 (Automatic linear central scale). *Fix $q \geq 2$, $k \geq 2$, and $0 < \eta < 1/2$. Suppose $\gamma_{q,k} = (k-1) - 1/(q-1) > 0$. Then there are constants $0 < c_{\eta,q,k} < C_{\eta,q,k} < \infty$ such that every central point $p \in [p_\eta, p_{1-\eta}]$ satisfies $c_{\eta,q,k}n \leq pN_{n,q,k} \leq C_{\eta,q,k}n$. Consequently, $M_\eta(n) = \Theta_{\eta,q,k}(n)$, and $V_\eta(n) = \Theta_{\eta,q,k}(n)$.*

Proof. The lower bound on $pN_{n,q,k}$ follows from Proposition 7.4: if $pN_{n,q,k} \leq cn$ with c sufficiently small, then $\mu_p(\text{Cover}_{n,q,k}) = o(1)$, so p cannot lie in the central interval for large n . The upper bound follows from Proposition 7.5: if $pN_{n,q,k} \geq Cn$ with C sufficiently large, then $\mu_p(\text{Cover}_{n,q,k}) = 1 - o(1)$,

so p cannot lie in the central interval. Since $pN_{n,q,k} = \Theta_{\eta,q,k}(n)$, and $p = O_{\eta,q,k}(n^{1-k}) = o(1)$, we also have $N_{n,q,k}p(1-p) = \Theta_{\eta,q,k}(n)$. Thus $M_\eta(n) = \Theta_{\eta,q,k}(n)$ and $V_\eta(n) = \Theta_{\eta,q,k}(n)$. $\square$

Let $\text{Small}_s^{(q,k)}$ be the event that the random instance contains a subfamily of at most s atomic constraints covering all assignments on the variables it uses. Equivalently, it contains a minimal unsatisfiable atomic sub-CSP with at most s constraints.

Lemma 7.7 (Small atomic covers are rare). *Assume $\gamma_{q,k} = (k-1) - 1/(q-1) > 0$. There are constants $c_{q,k}, c'_{q,k} > 0$ such that, if $s \leq c_{q,k} \log n$, then uniformly for all $p = O_{q,k}(n^{1-k})$,*

$$\mu_p(\text{Small}_s^{(q,k)}) \leq n^{-c'_{q,k}}.$$

Proof. It is enough to count inclusion-minimal unsatisfiable atomic subinstances with at most s constraints. Let such an instance have t constraints and use v variables. By Corollary 7.3, $t \geq (q-1)v + 1$, so $v \leq \frac{t-1}{q-1}$. For fixed t and v , the number of possible labelled subinstances using a prescribed v -set of variables is at most $\binom{q^k(v)}{t}$. Therefore, by the union bound, the probability that there is such a subinstance with $t \leq s$ constraints is at most

$$\sum_{1 \leq t \leq s} \sum_{v \leq (t-1)/(q-1)} \binom{n}{v} \binom{q^k(v)}{t} p^t.$$

Using $p = O_{q,k}(n^{1-k})$, $v \leq t$, and the standard estimates for binomial coefficients, the summand is at most

$$\exp(C_{q,k}t \log t) n^{v-(k-1)t} \leq \exp(C_{q,k}t \log t) n^{(t-1)/(q-1)-(k-1)t}.$$

Since $(k-1) - 1/(q-1) = \gamma_{q,k} > 0$, this is at most

$$\exp(C_{q,k}t \log t) n^{-\gamma_{q,k}t-1/(q-1)}.$$

If $s \leq c_{q,k} \log n$ and $c_{q,k}$ is sufficiently small, then $\exp(C_{q,k}t \log t) \leq n^{\gamma_{q,k}t/2}$ for all $t \leq s$. Summing over $t \leq s$ and the allowed values of v gives $\mu_p(\text{Small}_s^{(q,k)}) \leq n^{-c'_{q,k}}$, for some constant $c'_{q,k} > 0$. $\square$

For the atomic model, take $\Sigma = [q]$, let $\mathcal{P}_{q,k}^{\text{at}} = \{\{a\} : a \in [q]^k\}$, and $\Gamma = \{\text{id}\}$.

Proposition 7.8 (Atomic predicates are rigid and derivative-replaceable). *The system $(\mathcal{P}_{q,k}^{\text{at}}, \Gamma)$ is rigid. Its nonempty one-coordinate derivatives are precisely the singleton patterns in $[q]^{k-1}$, with constant multiplicity, and they satisfy (5.4) with $\gamma = \frac{k}{k-1}$, and $b = k-1$.*

Proof. For the singleton pattern system, $u \preceq_{\mathcal{P}} v$ holds exactly when $u = v$. Thus (5.2) forces $\tau = \sigma$, so rigidity holds, for example with $L = k$. Fixing the first coordinate of a singleton either gives the empty set or a singleton in $[q]^{k-1}$; every nonempty derivative occurs with the same multiplicity. Hence random derivative constraints are exactly random codimension- $(k-1)$ cylinders, while native constraints are codimension- k cylinders. The required power comparison is Lemma 6.1. $\square$

Theorem 7.9 (Atomic CSP window bound). *Fix $q \geq 2$, $k \geq 2$, and $0 < \eta < 1/2$. Suppose $\gamma_{q,k} = (k-1) - 1/(q-1) > 0$. Then*

$$W_{\eta,q,k}(n) \leq C_{\eta,q,k} \frac{n}{\log n}.$$

Proof. The linear central scale is Corollary 7.6. By Lemma 7.7, for every fixed $A_0, B_0 > 0$, the estimate (5.6) holds after decreasing the constant in $K \leq c \log n$. Proposition 7.8 verifies rigidity and power-replaceability. Apply Theorem 5.5. $\square$

Corollary 7.10 (Fixed- k SAT). *For every fixed $k \geq 3$ and every $0 < \eta < 1/2$, the central clause window of random k -SAT satisfies*

$$W_{\eta,k}(n) \leq C_{\eta,k} \frac{n}{\log n}.$$

Proof. Random k -SAT is the Boolean atomic model $q = 2$. The condition $\gamma_{2,k} > 0$ is exactly $k \geq 3$. Apply Theorem 7.9. $\square$

Corollary 7.11 (Standard fixed-size atomic CSPs). *The conclusions of Theorem 7.9 and Corollary 7.10 hold with the same $O(n/\log n)$ window in the uniform fixed-size model and in the standard model with independently sampled constraints.*

Proof. This is already included in Theorem 5.5; alternatively apply Corollary 2.3 using $N_{n,q,k} = q^k \binom{n}{k} = \Omega(n^2)$ and Corollary 7.6. $\square$

8. RANDOM SIGNED NAE- k -SAT

We now give a second application of the abstract theorem. The variables are Boolean. A signed NAE clause is specified by a k -set $S \subseteq [n]$ and an antipodal class $[a] = \{a, \bar{a}\} \subseteq \{0, 1\}^S$. It is violated exactly by assignments whose restriction to S belongs to $[a]$. Thus, in the covering language, a signed NAE clause is the union of two antipodal codimension- k cylinders,

$$\{x \in \{0, 1\}^n : x_S = a\} \cup \{x \in \{0, 1\}^n : x_S = \bar{a}\}.$$

Let $\mathcal{N}_{n,k}$ be the set of all signed NAE clauses. Then

$$N_{n,k}^{\text{NAE}} = |\mathcal{N}_{n,k}| = 2^{k-1} \binom{n}{k}.$$

We include each element of $\mathcal{N}_{n,k}$ independently with probability p , and write $\text{Cover}_{n,k}^{\text{NAE}}$ for the event that the selected violated sets cover $\{0, 1\}^n$, equivalently that the formula is NAE-unsatisfiable. For $0 < \eta < 1/2$, define

$$W_{\eta,k}^{\text{NAE}}(n) = N_{n,k}^{\text{NAE}}(p_{1-\eta} - p_\eta), \quad \mu_{p_\alpha}(\text{Cover}_{n,k}^{\text{NAE}}) = \alpha.$$

The analogue of the deficiency input is the following elementary observation.

Lemma 8.1 (Minimum degree in a minimal NAE-unsatisfiable formula). *Let Y be an inclusion-minimal NAE-unsatisfiable signed NAE- k -SAT instance. If Y has t clauses and uses v variables, then every used variable has clause-degree at least two. Consequently, $v \leq \frac{k}{2}t$.*

Proof. Suppose that a used variable x occurs in exactly one clause Q . By minimality, $Y \setminus \{Q\}$ has a NAE-satisfying assignment. Keep this assignment on all variables other than x . If the other $k-1$ signed literals in Q already take both Boolean values, either choice of x satisfies Q . Otherwise they all take the same value, and we choose x so that its signed literal takes the opposite value. Since x occurs nowhere else, this extends the assignment to a satisfying assignment of Y , a contradiction. Hence every used variable has degree at least two, and counting variable-clause incidences gives $2v \leq kt$. $\square$

Proposition 8.2 (Linear lower bound for signed NAE-unsatisfiability). *Fix $k \geq 3$. There is $c = c(k) > 0$ such that, if $pN_{n,k}^{\text{NAE}} \leq cn$, then $\mu_p(\text{Cover}_{n,k}^{\text{NAE}}) = o(1)$.*

Proof. Put $N = N_{n,k}^{\text{NAE}}$ and $m_0 = pN \leq cn$. For $S \subseteq [n]$, let X_S be the number of selected signed NAE clauses whose support is contained in S . If $|S| = v$, then

$$X_S \sim \text{Bin}\left(2^{k-1} \binom{v}{k}, p\right) \quad \text{and} \quad \mathbb{E}X_S = m_0 \frac{\binom{v}{k}}{\binom{n}{k}} \leq C_k cn \left(\frac{v}{n}\right)^k.$$

Set $b = 2/k$. Choosing c small enough that the relevant upper-tail threshold exceeds the mean, the standard binomial estimate gives

$$\mathbb{P}(X_S \geq bv) \leq \left(C_k c \left(\frac{v}{n}\right)^{k-1}\right)^{bv}.$$

Therefore the expected number of v -sets S with $X_S \geq bv$ is at most

$$\binom{n}{v} \left(C_k c \left(\frac{v}{n} \right)^{k-1} \right)^{bv} \leq \left[C'_k c^b \left(\frac{v}{n} \right)^{b(k-1)-1} \right]^v.$$

Here

$$b(k-1) - 1 = \frac{k-2}{k} > 0.$$

Choose $c = c(k)$ so that $C'_k c^b < 1/4$. For each fixed L , the contribution from $k \leq v < L$ tends to zero, while the contribution from $v \geq L$ is at most $\sum_{v \geq L} 4^{-v}$. Thus, with high probability, no S satisfies $X_S \geq 2|S|/k$.

If the selected formula were NAE-unsatisfiable, take an inclusion-minimal NAE-unsatisfiable subformula Y , and let S be its set of used variables. Writing $v = |S|$ and $t = |Y|$, Lemma 8.1 gives $t \geq 2v/k$. All clauses of Y are supported inside S , so $X_S \geq 2v/k$, a contradiction. $\square$

Proposition 8.3 (Linear upper bound for signed NAE-unsatisfiability). *Fix $k \geq 2$. There is $C = C(k) > 0$ such that, if $pN_{n,k}^{\text{NAE}} \geq Cn$, then $\mu_p(\text{Cover}_{n,k}^{\text{NAE}}) = 1 - o(1)$.*

Proof. Let Z be the number of NAE-satisfying assignments. A fixed assignment violates exactly one signed NAE clause on each k -set. Hence

$$\mathbb{E}Z = 2^n (1-p)^{\binom{n}{k}} \leq \exp \left(n \log 2 - p \binom{n}{k} \right).$$

Since $p \binom{n}{k} = pN_{n,k}^{\text{NAE}}/2^{k-1}$, this tends to zero once $pN_{n,k}^{\text{NAE}} \geq Cn$ with $C > 2^{k-1} \log 2$. Markov's inequality finishes the proof. $\square$

Corollary 8.4 (Automatic linear central scale for signed NAE-SAT). *Fix $k \geq 3$ and $0 < \eta < 1/2$. Uniformly for $p \in [p_\eta, p_{1-\eta}]$, $pN_{n,k}^{\text{NAE}} = \Theta_{\eta,k}(n)$, and $N_{n,k}^{\text{NAE}} p(1-p) = \Theta_{\eta,k}(n)$.*

Proof. The lower and upper bounds follow from propositions 8.2 and 8.3. They also give $p = O_{\eta,k}(n^{1-k}) = o(1)$, so multiplication by $1-p$ does not change the order. $\square$

Let $\text{Small}_{n,s}^{\text{NAE}}$ denote the event that the random formula contains an inclusion-minimal NAE-unsatisfiable subformula with at most s clauses.

Lemma 8.5 (Small signed NAE witnesses are rare). *Fix $k \geq 3$. There are constants $c_k, c'_k > 0$ such that, if $s \leq c_k \log n$, then uniformly for all $p = O_k(n^{1-k})$,*

$$\mu_p(\text{Small}_{n,s}^{\text{NAE}}) \leq n^{-c'_k}.$$

Proof. By Lemma 8.1, a minimal NAE-unsatisfiable formula with t clauses uses at most $kt/2$ variables. A union bound gives

$$\mu_p(\text{Small}_{n,s}^{\text{NAE}}) \leq \sum_{1 \leq t \leq s} \sum_{v \leq kt/2} \binom{n}{v} \binom{2^{k-1} \binom{v}{k}}{t} p^t.$$

Using $p = O_k(n^{1-k})$, $v \leq kt/2$, and the standard binomial-coefficient estimates, each summand is at most

$$\exp(C_k t \log(t+1)) n^{v-(k-1)t} \leq \exp(C_k t \log(t+1)) n^{-(k-2)t/2}.$$

For $t \leq c_k \log n$, the exponential factor is at most $n^{(k-2)t/4}$ for all sufficiently large n . Summing over t and v yields a fixed polynomial saving. $\square$

Let

$$\mathcal{P}_k^{\text{NAE}} = \{ \{a, \bar{a}\} : a \in \{0, 1\}^k \}, \quad \Gamma = \{\text{id}, \text{comp}\}.$$

Repeated antipodal pairs are identified in $\mathcal{P}_k^{\text{NAE}}$.

Lemma 8.6 (Pointwise atomic-to-NAE replacement). *Fix $k \geq 2$. There are constants $c_k, C_k > 0$ such that, if $\rho_{k-1}^{\text{at}}(B) \geq \varepsilon \geq C_k m^{-(k-1)}$, then $\rho_k^{\text{NAE}}(B) \geq c_k \varepsilon^{k/(k-1)}$, where ρ_{k-1}^{at} is the probability that a random atomic codimension- $(k-1)$ cylinder contains B , and ρ_k^{NAE} is the probability that a random signed NAE clause contains B .*

Proof. The case $B = \emptyset$ is immediate. Otherwise, let $\ell(B)$ be the number of coordinates on which every point of B takes the same value. Then

$$\rho_{k-1}^{\text{at}}(B) = \frac{\binom{\ell(B)}{k-1}}{2^{k-1} \binom{m}{k-1}}.$$

For C_k sufficiently large, the assumed lower bound implies $\ell(B) \geq 2k$. Whenever a signed NAE clause chooses its k coordinates among these frozen coordinates and takes the antipodal class of the frozen pattern, it covers B . Therefore

$$\rho_k^{\text{NAE}}(B) \geq \frac{\binom{\ell(B)}{k}}{2^{k-1} \binom{m}{k}}.$$

The two displays give the claimed power bound. $\square$

Proposition 8.7 (Signed NAE predicates are rigid and derivative-replaceable). *For every $k \geq 3$, the system $(\mathcal{P}_k^{\text{NAE}}, \Gamma)$ is rigid. Its one-coordinate derivatives, with their indexed multiplicities, are uniformly distributed atomic singleton patterns in $\{0, 1\}^{k-1}$, and they satisfy (5.4) with $\gamma = \frac{k}{k-1}$, and $b = k - 1$.*

Proof. For signed NAE patterns, $u \preceq_{\mathcal{P}} v$ if and only if $u \in \{v, \bar{v}\}$. Suppose (5.2) holds but $\tau \notin \{\sigma, \bar{\sigma}\}$. Then there are coordinates i, j for which $\tau_i = \sigma_i$ and $\tau_j \neq \sigma_j$, after globally complementing τ if necessary. Extend $\{i, j\}$ to a k -set. On that set, τ is neither σ nor its complement, contradicting the local preorder condition. Thus the system is rigid, for example with $L = k$. Fixing one coordinate of an antipodal pair gives a singleton pattern; every singleton derivative occurs with the same indexed multiplicity. The power comparison is Lemma 8.6. $\square$

Theorem 8.8 (Signed NAE- k -SAT window bound). *For every fixed $k \geq 3$ and every $0 < \eta < 1/2$,*

$$W_{\eta, k}^{\text{NAE}}(n) \leq C_{\eta, k} \frac{n}{\log n}.$$

Proof. The central linear scale is Corollary 8.4. Lemma 8.5 gives (5.6) after decreasing the logarithmic range as a function of A_0, B_0 . Proposition 8.7 supplies the finite predicate certificate. Apply Theorem 5.5. $\square$

Corollary 8.9 (Standard fixed-size signed NAE-SAT). *The conclusion of Theorem 8.8 holds in the uniform fixed-size model and in the standard model with independently sampled signed NAE clauses.*

Proof. This follows from the final assertion of Theorem 5.5. $\square$

9. RANDOM HYPERGRAPH TWO-COLOURABILITY

We now treat the unsigned model. Let $\mathcal{E}_{n, k} = \binom{[n]}{k}$, and include every k -set independently with probability p . The resulting random k -uniform hypergraph is written $H_k(n, p)$. A two-colouring $x \in \{0, 1\}^n$ is proper if no selected edge is monochromatic. In the covering language, an edge $S \in \mathcal{E}_{n, k}$ covers $Q_S = \{x : x_S = 0^S\} \cup \{x : x_S = 1^S\}$. Let $\text{N2Col}_{n, k} \subseteq 2^{\mathcal{E}_{n, k}}$ be the monotone event that the selected hypergraph is not two-colourable. Put $N_{n, k}^{\text{hyp}} = \binom{n}{k}$, and define

$$W_{\eta, k}^{\text{hyp}}(n) = N_{n, k}^{\text{hyp}}(p_{1-\eta} - p_{\eta}), \quad \mu_{p_{\alpha}}(\text{N2Col}_{n, k}) = \alpha.$$

The elementary minimum-degree argument from Lemma 8.1 has the following unsigned form.

Lemma 9.1 (Minimum degree in a minimal non-two-colourable hypergraph). *Let Y be an inclusion-minimal non-two-colourable k -uniform hypergraph. If Y has t edges and uses v vertices, then every used vertex has degree at least two. Consequently, $v \leq \frac{k}{2}t$.*

Proof. Suppose that a used vertex u lies in a unique edge e . By minimality, $Y \setminus \{e\}$ has a proper two-colouring. Keep this colouring away from u . If $e \setminus \{u\}$ already contains both colours, either colour for u works; otherwise colour u with the colour opposite to the common colour on $e \setminus \{u\}$. This properly colours Y , a contradiction. Counting incidences now gives $2v \leq kt$. $\square$

Proposition 9.2 (Linear lower bound for non-two-colourability). *Fix $k \geq 3$. There is $c = c(k) > 0$ such that, if $pN_{n,k}^{\text{hyp}} \leq cn$, then $\mu_p(\text{N2Col}_{n,k}) = o(1)$.*

Proof. Put $N = N_{n,k}^{\text{hyp}}$ and $m_0 = pN \leq cn$. For $S \subseteq [n]$, let X_S be the number of selected edges contained in S . If $|S| = v$, then

$$\mathbb{E}X_S = m_0 \frac{\binom{v}{k}}{\binom{n}{k}} \leq C_k cn \left(\frac{v}{n}\right)^k.$$

Set $b = 2/k$. For c sufficiently small, the standard binomial upper-tail estimate gives

$$\mathbb{P}(X_S \geq bv) \leq \left(C_k c \left(\frac{v}{n}\right)^{k-1}\right)^{bv}.$$

Hence the expected number of v -sets with $X_S \geq bv$ is at most

$$\left[C'_k c^b \left(\frac{v}{n}\right)^{b(k-1)-1}\right]^v.$$

Since $b(k-1) - 1 = (k-2)/k > 0$, choosing c small and summing over $v \geq k$ shows that, with high probability, no S satisfies $X_S \geq 2|S|/k$. If the selected hypergraph were not two-colourable, an inclusion-minimal non-two-colourable subhypergraph Y , on a vertex set S , would satisfy $|Y| \geq 2|S|/k$ by Lemma 9.1, a contradiction. $\square$

Proposition 9.3 (Linear upper bound for non-two-colourability). *Fix $k \geq 3$. There is $C = C(k) > 0$ such that, if $pN_{n,k}^{\text{hyp}} \geq Cn$, then $\mu_p(\text{N2Col}_{n,k}) = 1 - o(1)$.*

Proof. Let Z be the number of proper two-colourings. A colouring with colour-class sizes a and $n-a$ has

$$\binom{a}{k} + \binom{n-a}{k}$$

monochromatic potential edges. By convexity, this quantity is minimized for a balanced colouring, and for all sufficiently large n it is at least $c_k \binom{n}{k}$, for a constant $c_k > 0$. Consequently,

$$\mathbb{E}Z \leq 2^n (1-p)^{c_k \binom{n}{k}} \leq \exp\left(n \log 2 - c_k p N_{n,k}^{\text{hyp}}\right).$$

This tends to zero once $pN_{n,k}^{\text{hyp}} \geq Cn$ with $C > \log 2/c_k$. Markov's inequality finishes the proof. $\square$

Corollary 9.4 (Automatic linear central scale for hypergraph two-colourability). *Fix $k \geq 3$ and $0 < \eta < 1/2$. Uniformly for $p \in [p_\eta, p_{1-\eta}]$, $pN_{n,k}^{\text{hyp}} = \Theta_{\eta,k}(n)$, and $N_{n,k}^{\text{hyp}} p(1-p) = \Theta_{\eta,k}(n)$.*

Proof. Apply Propositions 9.2 and 9.3. They also imply $p = O_{\eta,k}(n^{1-k}) = o(1)$. $\square$

Let $\text{Small}_{n,s}^{\text{hyp}}$ be the event that the random hypergraph contains an inclusion-minimal non-two-colourable subhypergraph with at most s edges.

Lemma 9.5 (Small non-two-colourable witnesses are rare). *Fix $k \geq 3$. There are constants $c_k, c'_k > 0$ such that, if $s \leq c_k \log n$, then uniformly for $p = O_k(n^{1-k})$,*

$$\mu_p(\text{Small}_{n,s}^{\text{hyp}}) \leq n^{-c'_k}.$$

Proof. By Lemma 9.1, a minimal witness with t edges uses at most $kt/2$ vertices. Therefore

$$\mu_p(\text{Small}_{n,s}^{\text{hyp}}) \leq \sum_{1 \leq t \leq s} \sum_{v \leq kt/2} \binom{n}{v} \binom{\binom{v}{k}}{t} p^t.$$

Each summand is at most

$$\exp(C_k t \log(t+1)) n^{v-(k-1)t} \leq \exp(C_k t \log(t+1)) n^{-(k-2)t/2}.$$

For a sufficiently small constant in $s \leq c_k \log n$, summing over t, v gives a fixed polynomial saving. $\square$

Let

$$\mathcal{P}_k^{\text{hyp}} = \{\{0^k, 1^k\}\}, \quad \Gamma = \{\text{id}, \text{comp}\}.$$

For $B \subseteq \{0, 1\}^m$, let $\rho_{k-1}^{\text{const}}(B)$ be the probability that a random constant codimension- $(k-1)$ cylinder contains B , and let $\rho_k^{\text{hyp}}(B)$ be the probability that the bad set of a random k -edge contains B .

Lemma 9.6 (Pointwise constant-cylinder-to-edge replacement). *Fix $k \geq 2$. There are constants $c_k, C_k > 0$ such that, if $\rho_{k-1}^{\text{const}}(B) \geq \varepsilon \geq C_k m^{-(k-1)}$, then $\rho_k^{\text{hyp}}(B) \geq c_k \varepsilon^{k/(k-1)}$.*

Proof. The case $B = \emptyset$ is immediate. For $c \in \{0, 1\}$, let $L_c(B) = \{i \in [m] : x_i = c \text{ for every } x \in B\}$, and put $\ell_c = |L_c(B)|$. Then

$$\rho_{k-1}^{\text{const}}(B) = \frac{\binom{\ell_0}{k-1} + \binom{\ell_1}{k-1}}{2 \binom{m}{k-1}}.$$

Hence, for some $c \in \{0, 1\}$,

$$\frac{\binom{\ell_c}{k-1}}{\binom{m}{k-1}} \geq \varepsilon.$$

For C_k sufficiently large, this implies $\ell_c \geq 2k$. Every k -set contained in $L_c(B)$ has a bad set containing B , and therefore

$$\rho_k^{\text{hyp}}(B) \geq \frac{\binom{\ell_c}{k}}{\binom{m}{k}}.$$

The usual binomial-coefficient comparisons give the claim. $\square$

Proposition 9.7 (The Property B predicate is rigid and derivative-replaceable). *For every $k \geq 3$, the system $(\mathcal{P}_k^{\text{hyp}}, \Gamma)$ is rigid. Its one-coordinate derivatives are the two constant patterns 0^{k-1} and 1^{k-1} , and they satisfy (5.4) with $\gamma = \frac{k}{k-1}$, and $b = k-1$.*

Proof. For this pattern system, $u \preceq_{\mathcal{P}} v$ says that if u is constant, then v is constant. Let σ have both colour classes of size at least k , and suppose (5.2) holds. Then every τ -monochromatic k -set is σ -monochromatic. If a colour class of τ has size at least k , it must lie inside one colour class of σ ; otherwise it contains a k -set meeting both σ -classes. Neither τ -class can have size at most $k-1$, because the other class would still meet both σ -classes and have size at least k . Thus the two τ -classes are exactly the two σ -classes, and $\tau \in \{\sigma, \bar{\sigma}\}$. Hence rigidity holds with $L = k$.

Fixing the first coordinate of $\{0^k, 1^k\}$ gives one of the two constant $(k-1)$ -patterns. The derivative comparison is Lemma 9.6. $\square$

Theorem 9.8 (Random hypergraph two-colourability window). *For every fixed $k \geq 3$ and every $0 < \eta < 1/2$,*

$$W_{\eta,k}^{\text{hyp}}(n) \leq C_{\eta,k} \frac{n}{\log n}.$$

Proof. The central linear scale is Corollary 9.4. Lemma 9.5 gives the arbitrarily strong logarithmic-scale witness estimate (5.6). Proposition 9.7 verifies rigidity and derivative replacement. Apply Theorem 5.5. $\square$

Corollary 9.9 (Standard fixed-size hypergraph model). *The conclusion of Theorem 9.8 holds in the uniform $H_k(n, m)$ model and in the model obtained by independently sampling m edges.*

Proof. This follows from the final assertion of Theorem 5.5. $\square$

10. CONCLUDING REMARKS

10.1. The Abbe–Montanari reduction. The scaling-window problem is connected to the still-open problem of proving convergence of the critical density for random fixed- k SAT. Abbe and Montanari showed that a clause-window bound of the form

$$W_{\eta,k}(n) = O\left(\frac{n}{\log^{1+\delta} n}\right)$$

for some fixed $\delta > 0$ would imply the satisfiability conjecture [1]; see also [11]. At a high level, their reduction uses a window narrow enough that the finite-size uncertainty in the critical density becomes summable along an appropriate comparison of instance sizes. One can then promote concentration around the n -dependent critical points to convergence of those points. Thus, a gain of any fixed positive power beyond one logarithm would have consequences substantially stronger than a finite-size scaling estimate.

Our result reaches the boundary scale $n/\log n$. It improves the previous $n/\log \log n$ upper bound by a factor of order $\log n/\log \log n$, but it does not meet the strictly stronger hypothesis required by the Abbe–Montanari reduction. It is therefore tempting to seek an iteration of the present argument, or a more efficient replacement step, that extracts a better power of $\log n$.

The proof itself explains why a straightforward iteration cannot supply that extra power. At scale K , Theorem 3.1 gives a deterministic boost of size $\exp(-O(K))$. The simulation lemmas replace the clean booster by $D \leq \exp(O(K))$ random coordinates, while the universal layer estimate bounds the resulting increase by $O((D+1)/\sqrt{V_\eta(n)})$. Consequently the strongest comparison available from these three ingredients has the form

$$\exp(-O(K)) \leq \frac{\exp(O(K))}{\sqrt{V_\eta(n)}},$$

which forces only $K = \Omega(\log V_\eta(n))$. Reapplying the same argument does not improve the exponent: each new booster again has exponential strength and exponential simulation cost, and the layer estimate depends only on the total number of sprinkled coordinates.

A superlogarithmic denominator would therefore require a genuinely stronger input at one of three places: a Bourgain-type theorem producing a larger boost, a subexponential simulation of clean boosters, or a model-specific replacement for the Boolean-layer estimate that exploits more geometry than monotonicity alone.

Acknowledgments. The author is grateful to Imre Leader for his guidance and support. This work was supported by the CB European PhD Studentship funded by Trinity College, Cambridge.

REFERENCES

- [1] E. Abbe and A. Montanari, On the concentration of the number of solutions of random satisfiability formulas, *Random Structures & Algorithms* 45 (2014), no. 3, 362–382.
- [2] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim and D. B. Wilson, The scaling window of the 2-SAT transition, *Random Structures & Algorithms* 18 (2001), no. 3, 201–256.
- [3] J. Ding, A. Sly and N. Sun, Proof of the satisfiability conjecture for large k , *Annals of Mathematics. Second Series* 196 (2022), no. 1, 1–388.
- [4] D. Achlioptas and Y. Peres, The threshold for random k -SAT is $2^k \log 2 - O(k)$, *Journal of the American Mathematical Society* 17 (2004), no. 4, 947–973.
- [5] D. Achlioptas and C. Moore, Random k -SAT: two moments suffice to cross a sharp threshold, *SIAM Journal on Computing* 36 (2006), no. 3, 740–762.
- [6] A. Coja-Oghlan and K. Panagiotou, Catching the k -NAESAT threshold, in *Proceedings of the 2012 ACM Symposium on Theory of Computing (STOC, 2012)*, ACM, New York, 2012, 899–907.

- [7] Y. Filmus, E. A. Hirsch, S. Kurz, F. Ihringer, A. Riazanov, A. V. Smal and M. Vinyals, Irreducible subcube partitions, *The Electronic Journal of Combinatorics* 30 (2023), no. 3, Paper No. P3.29, 39 pp.
- [8] D. Bal, On sharp thresholds of monotone properties: Bourgain’s proof revisited, arXiv:1302.1162, 2013.
- [9] E. Friedgut, with an appendix by J. Bourgain, Sharp thresholds of graph properties, and the k -SAT problem, *Journal of the American Mathematical Society* 12 (1999), no. 4, 1017–1054.
- [10] P. Keevash, N. Lifshitz, E. Long and D. Minzer, Hypercontractivity for global functions and sharp thresholds, *Journal of the American Mathematical Society* 37 (2024), no. 1, 245–279.
- [11] W. Perkins, Searching for (sharp) thresholds in random structures: where are we now?, *Bulletin of the American Mathematical Society* 62 (2025), no. 1, 113–143.
- [12] R. Aharoni and N. Linial, Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas, *Journal of Combinatorial Theory, Series A* 43 (1986), no. 2, 196–204.
- [13] O. Kullmann, Constraint satisfaction problems in clausal form I: Autarkies and deficiency, *Fundamenta Informaticae* 109 (2011), no. 1, 27–81.
- [14] P. Erdős and A. Rényi, On the evolution of random graphs, *Publicationes Mathematicae Instituti Hungarici Academiae Scientiarum* 5 (1960), 17–61.
- [15] J. Kahn, G. Kalai and N. Linial, The influence of variables on Boolean functions, in *Proceedings of the 29th Annual Symposium on Foundations of Computer Science (FOCS, 1988)*, IEEE Computer Society, 1988, 68–80.
- [16] J. Kahn and G. Kalai, Thresholds and expectation thresholds, *Combinatorics, Probability and Computing* 16 (2007), no. 3, 495–502.
- [17] H. Hatami, A structure theorem for Boolean functions with small total influences, *Annals of Mathematics. Second Series* 176 (2012), no. 1, 509–533.
- [18] J. Park and H. T. Pham, A proof of the Kahn–Kalai conjecture, *Journal of the American Mathematical Society* 37 (2024), no. 1, 235–243.
- [19] R. O’Donnell, *Analysis of Boolean Functions*, Cambridge University Press, Cambridge, 2014.
- [20] R. Alweiss, S. Lovett, K. Wu and J. Zhang, Improved bounds for the sunflower lemma, *Annals of Mathematics. Second Series* 194 (2021), no. 3, 795–815.
- [21] K. Frankston, J. Kahn, B. Narayanan and J. Park, Thresholds versus fractional expectation-thresholds, *Annals of Mathematics. Second Series* 194 (2021), no. 2, 475–495.
- [22] D. Achlioptas and C. Moore, The asymptotic order of the random k -SAT threshold, in *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS, 2002)*, IEEE Computer Society, 2002, 779–788.
- [23] J. Ding, A. Sly and N. Sun, Satisfiability threshold for random regular NAE-SAT, *Communications in Mathematical Physics* 341 (2016), no. 2, 435–489.
- [24] A. Sly, N. Sun and Y. Zhang, The number of solutions for random regular NAE-SAT, *Probability Theory and Related Fields* 182 (2022), nos. 1–2, 1–109.
- [25] D. Nam, A. Sly and Y. Sohn, One-step replica symmetry breaking of random regular NAE-SAT, in *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS, 2021)*, IEEE Computer Society, 2022, 310–318.
- [26] D. Nam, A. Sly and Y. Sohn, One-step replica symmetry breaking of random regular NAE-SAT II, *Communications in Mathematical Physics* 405 (2024), no. 3, article 61.
- [27] A. Sly and Y. Sohn, Local geometry of NAE-SAT solutions in the condensation regime, *Probability Theory and Related Fields* 191 (2025), no. 3, 979–1033.

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, CAMBRIDGE, UNITED KINGDOM

Email address: gc645@cam.ac.uk