Learning-Augmented Streaming Algorithms for Approximating Boolean Max-CSPs*

Yinhao Dong[†] **Pan Peng**[‡] **Ali Vakilian**[§] **Santhoshini Velusamy**[¶] **Xiaoyang Xu**[‖]

Abstract

We study learning-augmented streaming algorithms for estimating the value of Boolean Maximum Constraint Satisfaction Problems (Max-CSPs). Specifically, we consider streaming algorithms equipped with an ε-accurate oracle: for each variable, the oracle outputs a label in $\{-1, 1\}$ that agrees with a fixed optimal assignment with probability $1/2 + \varepsilon$, and disagrees otherwise. Prior work by Dong, Peng, and Vakilian [ITCS 2025] showed that, for the special case of Max-Cut, such an oracle enables a $(1/2 + \Omega(\varepsilon^2))$-approximation using $\mathrm{poly}(1/\varepsilon)$ words of space in insertion-only streams, and $\mathrm{poly}(1/\varepsilon, \log n)$ words of space in dynamic streams. We substantially generalize and strengthen their result, provided that a highly accurate estimate of ε is available. For any Boolean Max-2CSP and any $\eta > 0$ independent of ε, we obtain, with high constant probability, a single-pass $(1 - \eta)$-approximation using $\mathrm{poly}(1/\varepsilon, 1/\eta)$ words of space in insertion-only streams, and $\mathrm{poly}(1/\varepsilon, 1/\eta, \log n)$ words of space in dynamic streams, where n is the number of variables. We further extend our techniques to general Boolean Max-kCSPs for $k \geq 3$, with slightly worse space complexity.

1 Introduction

Constraint Satisfaction Problems (CSPs) form a highly expressive and well-studied class of combinatorial optimization problems. The goal is to assign values from a finite domain to n variables so as to maximize the number of satisfied constraints, where each constraint depends only on a small subset of variables. This framework captures classical problems such as MAX-CUT, MAX-DICUT, and MAX-2AND. Beyond their central role in approximation algorithms and complexity theory, CSP-like finite-domain labeling formulations also appear throughout machine learning, especially in structured prediction, conditional random fields, MAP inference, and energy minimization in graphical models and computer vision [LMP01, TGK03, THJA04, BVZ01]. More recently, CSPs and SAT have also served as benchmarks and targets for learning-based solvers and algorithm-selection methods [XKK18, SLB+19].

Recently, there has been a flurry of work on CSPs in the *streaming* model (e.g. [GVV17, CGS+22b, CGSV24]; see also Section 1.2). In this model, the constraints of the instance arrive sequentially, possibly with deletions in the dynamic setting, and the algorithm is allowed only limited memory, typically sublinear in the input size. Since explicitly outputting an assignment for MAX-CUT or more general MAX-CSPs already requires $\Omega(n)$ space, the standard streaming objective is instead to estimate the optimum value of the instance.

*All authors contributed equally and are listed in alphabetical order.

[†]School of Computer Science and Technology, University of Science and Technology of China. Email: yhdong@mail.ustc.edu.cn.

[‡]School of Computer Science and Technology, University of Science and Technology of China. Email: ppeng@ustc.edu.cn.

[§]Rutgers University. Email: a.vakilian@rutgers.edu.

[¶]University of Waterloo. Email: santhoshini.velusamy@uwaterloo.ca.

[‖]University of California, Santa Barbara. Email: xiaoyangxu@ucsb.edu.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

A central motivation for studying streaming CSPs is that they admit unconditional, information-theoretic lower bounds (typically via communication complexity) capturing the limits of compressing an instance while preserving its optimum value. Streaming algorithms can still achieve meaningful approximations, and a key theme is understanding the trade-off between approximation quality and space complexity (see recent surveys [Sud22, Sin25]). Given these abundant lower bounds, it is natural to ask whether they can be circumvented under reasonable additional assumptions, that is, *beyond the worst-case* setting.

In many applications, the abundance of data, together with powerful prediction models such as deep neural networks and large language models, makes it possible to obtain reliable hints about the underlying optimization instance. This perspective has motivated the rapidly growing area of *learning-augmented algorithms*, also known as *algorithms with predictions*, which studies algorithms that exploit oracle predictions while retaining provable guarantees.[7] A particularly simple and widely used model is that of *ε-accurate predictions*, where the oracle predicts membership in a fixed optimal solution (or an optimal certificate) with probability only slightly better than random guessing, namely $1/2 + \varepsilon$. This model was introduced in [CdG+24, GMM25] for problems such as MAX-CUT and MAX-CSPs, showing that even such mildly biased predictions can be leveraged to obtain provable improvements over prediction-free approximation barriers. For example, for MAX-CUT these results can go beyond the classical 0.878 threshold achieved by the Goemans–Williamson SDP rounding [GW95], which is known to be optimal under the Unique Games Conjecture [KKMO07].

The work most closely related to ours is [DPV25], which studies streaming MAX-CUT with ε-accurate predictions. They obtain a streaming algorithm that breaks the Kapralov–Krachun lower bound [KK19] for beating a $1/2$-approximation in $o(n/2^{\mathrm{poly}(1/\varepsilon)})$ space. This demonstrates that even weak, noisy predictions can fundamentally change the power of streaming algorithms for CSPs.

In this paper, we study *learning-augmented streaming algorithms for Boolean* MAX-CSP*s* and prove that they are significantly more powerful than classical streaming algorithms. In particular, we consider an ε-accurate oracle similar to the one defined in [DPV25], where there is a hidden optimal assignment $\boldsymbol{\sigma}^* \in \{-1,+1\}^n$ to the variables $\{x_1, \ldots, x_n\}$, and the streaming algorithm has (unlimited) oracle access to a prediction vector $\mathbf{Y} \in \{-1,+1\}^n$, where for every variable x_i, $Y_i = \sigma_i^*$ *independently* with probability $1/2 + \varepsilon$.

1.1 Our results

Our main result shows that weakly accurate predictions dramatically increase the power of streaming algorithms for Boolean MAX-kCSPs. For any constant $\varepsilon > 0$, access to an ε-accurate oracle is sufficient to obtain an arbitrarily accurate constant-factor approximation using extremely small space: for insertion-only streams, where constraints are only added, our algorithms use space independent of n up to the word size; for dynamic streams, where constraints may be both inserted and deleted, they use only $\mathrm{poly}(\log n)$ space. We assume that the prediction accuracy parameter ε is known to the algorithm in advance.[8] We now state our main results.

Theorem 1.1. *Let $\varepsilon \in (0, \frac{1}{2}], \delta \in (0,1)$ and $k \in \mathbb{Z}_{>0}$. Let $f : \{-1,1\}^k \to \{0,1\}$ be a Boolean function of arity k. For any $\eta \in (0,1)$, given an ε-accurate oracle $\mathcal{O}$ for a Boolean* MAX-kCSP(f) *instance Ψ, there exists a single-pass algorithm that outputs a $(1-\eta)$-approximation to the maximum value of Ψ with probability at least $1-\delta$, using the following amount of space (in words):*

(1) if $k = 2$, then the algorithm uses $\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$ space in insertion-only streams, and $\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta, \log n)$ space in dynamic streams;

(2) if $k \geq 3$, then the algorithm uses $\exp\big(\mathrm{poly}(k, 1/\varepsilon^k, 1/\eta, \log(1/\delta))\big)$ space in insertion-only streams, and $\exp\big(\mathrm{poly}(k, 1/\varepsilon^k, 1/\eta, \log(1/\delta))\big)\,\mathrm{poly}(\log n)$ space in dynamic streams.

Comparison to previous learning-augmented algorithms. Our results generalize the learning-augmented streaming algorithm of [DPV25] for MAX-CUT to arbitrary Boolean MAX-CSPs, while also improving the approximation–space tradeoff even for the special case of MAX-CUT. Specifically,

[7] See `https://algorithms-with-predictions.github.io/` for an up-to-date repository of papers in this area.

[8] In fact, it suffices to know a highly accurate estimate of ε for our algorithms to work. Specifically, it is enough to have an estimate $\varepsilon' \in (0, 1/2]$ such that $\varepsilon' = (1 \pm O(\eta)) \cdot \varepsilon$ for MAX-2CSPs; for MAX-kCSPs the hidden constant may depend on k. See Appendix E for details. In contrast, the algorithm of [DPV25] for MAX-CUT requires a constant-factor estimate of ε.

[DPV25] obtains a $\frac{1}{2} + \Omega(\varepsilon^2)$ approximation using $\mathrm{poly}(1/\varepsilon)$ space in insertion-only streams and $\mathrm{poly}(1/\varepsilon, \log n)$ space in dynamic streams. In contrast, our algorithm achieves a $(1-\eta)$-approximation for any $\eta > 0$, at the cost of an additional polynomial dependence on $1/\eta$ in space.

Comparison to previous classical algorithms. For one-pass streaming algorithms in the *dynamic* setting without predictions, strong lower bounds are known. Chou et al. [CGSV24] showed that for every MAX-CSP, there exists an explicit constant α such that, for any $\eta > 0$: (i) there is an $(\alpha - \eta)$-approximation streaming algorithm using $\mathrm{poly}(\log n)$ space, whereas (ii) any $(\alpha + \eta)$-approximation requires $\Omega(\sqrt{n})$ space. For Boolean MAX-2CSPs, Chou et al. [CGV20] proved analogous lower bounds even in the insertion-only setting. For example, for MAX-DICUT and MAX-2AND, the threshold is $\alpha = 4/9$, so any $(4/9 + \eta)$-approximation requires $\Omega(\sqrt{n})$ space. For higher arity, the barriers are even stronger: [CGS$^+$22b] showed that, without predictions, no $o(n)$-space streaming algorithm can achieve better than a $2^{-(k-1)}$ approximation for MAX-kAND. Similarly, [CGV20] showed that no $o(\sqrt{n})$-space algorithm for MAX-kSAT can surpass a $1/\sqrt{2}$ approximation. These lower bounds hold for every k and even on instances with bounded maximum degree.

Our learning-augmented algorithms bypass these worst-case barriers. With only weakly accurate independent predictions, they achieve approximations arbitrarily close to 1 using space that is independent of n in insertion-only streams and polylogarithmic in n in dynamic streams, without imposing any bounded-degree assumption. We provide a technical overview in Section 3. Due to space limits, we defer all missing proofs from the main text to the appendix.

1.2 Related work

Streaming CSPs. There has been a long line of work on streaming algorithms for constraint satisfaction problems (CSPs), including: single-pass lower bounds for MAX-CUT [KK15, KKS15, KKSV17, KK19]; multi-pass lower bounds for MAX-CUT [AKSY20, AN21, CKP$^+$23, FMW25] and other CSPs [FMW26]; algorithms and lower bounds for approximating MAX-DICUT [GVV17, CGV20, SSSV23a, SSSV23b, SSSV25, Vel26, ABFS26]; quantum algorithms and lower bounds for MAX-CUT and MAX-DICUT [KP22, KPV24, KPV25]; and results on other specific CSPs, including Unique Games [GT19], monarchy-like predicates [CGS$^+$22a], and MAX-kAND [Sin23]. More broadly, dichotomy theorems and results for streaming and sketching general CSPs have been established in [CGSV21, CGS$^+$22b, CGSV24].

Learning-augmented algorithms. Learning-augmented algorithms have also been actively studied across several areas, including online algorithms [MNS07, LV21, KPS18, LMS25], data structures [KBC$^+$18, Mit18, VKMK21], graph algorithms [DIL$^+$21, GMM25, BDSW24], and sublinear-time algorithms [IVY19, EIN$^+$21, LLL$^+$23]. Our algorithms fall within the framework of learning-augmented streaming algorithms [HIKV19, JLL$^+$20, AIV19, CEI$^+$22, ACN$^+$23, ACG$^+$25b, DJLP25, NPW$^+$26], which focuses on using predictions to improve space–accuracy tradeoffs. More recently, learning-augmented (offline) algorithms have been studied for MAX-CUT and CSPs [CdG$^+$24, GMM25, ACG$^+$25a, AGR25] under variants of the prediction model considered here. In particular, for the MAX-CUT problem where the goal is to output a solution rather than merely estimate its value, [CdG$^+$24] and [ACG$^+$25a] achieved approximation ratios exceeding 0.878 using predictions. Additionally, [GMM25] developed algorithms that yield near-optimal solutions when the average degree exceeds a threshold depending only on the prediction bias, independent of the instance size; they further extended these results to weighted graphs and the MAX-3LIN problem. Finally, [AGR25] studied learning-augmented Boolean satisfiability, obtaining both exponential-time speedups for k-SAT and provable approximation improvements for MAX-SAT variants, including near-optimal guarantees on instances of sufficiently large average degree.

2 Preliminaries

Notation. Let $[n] := \{1, 2, \ldots, n\}$ for any $n \in \mathbb{Z}_{>0}$. Throughout the paper we let $G = (V, E)$ be a directed, unweighted graph with n vertices and m edges. We assume that G is loopless and has no repeated directed edges, while both orientations of a vertex pair may appear. For each vertex $v \in V$, let $\deg(v)$ denote the number of edges incident to v, counting both incoming and outgoing edges.

Given a vertex $v \in V$ and disjoint sets $S, T \subseteq V$, let $e(S, v)$ denote the number of edges from S to v; that is, $e(S, v) := |\{(u, v) \in E : u \in S\}|$. Similarly, let $e(v, T)$ denote the number of edges from v

to T; that is, $e(v,T) := |\{(v,u) \in E : u \in T\}|$. We define $e(S,T)$ to be the number of edges from S to T; $e(S,T) := |(u,v) \in E : u \in S \wedge v \in T| = \sum_{v \in T} e(S,v) = \sum_{v \in S} e(v,T)$. A *directed cut* (or *dicut*) (S,T) of G is an ordered bipartition of V into two disjoint subsets S and T. Following the standard convention, we refer to S as the *left* side and T as the *right* side. The *value* of the dicut (S,T) is $e(S,T)$. In the descriptions of our algorithms for MAX-DICUT, for a vertex set $S \subseteq V$, we define S^+ (resp. S^-) as the set of vertices with predicted $+$ (resp. $-$) signs by the given ε-accurate oracle $\mathcal{O}$. Note that S^+ and S^- form a bipartition of S.

2.1 Problem definitions

Problem 2.1 (Boolean MAX-kCSP)**.** Let $k \in \mathbb{Z}_{>0}$. Let $f : \{-1,1\}^k \to \{0,1\}$ be a Boolean function of arity k. Let $x_1, \ldots, x_n$ be Boolean *variables*. Given a Boolean variable x, we refer to x and $-x$ as the *literals* associated with x. In other words, a literal is either a variable itself or its negation. A *constraint* C is a triple $(f, \mathbf{j}, \mathbf{b})$, where $\mathbf{j} = (j_1, \ldots, j_k) \in [n]^k$ has distinct coordinates and $\mathbf{b} = (b_1, \ldots, b_k) \in \{-1,1\}^k$. It represents applying the function f to the literals $b_1 x_{j_1}, \ldots, b_k x_{j_k}$.

An instance Ψ of the Boolean MAX-kCSP(f) problem consists of m constraints $C_1, \ldots, C_m$ over n variables $x_1, x_2, \ldots, x_n$. For an assignment $\boldsymbol{\sigma} \in \{-1,1\}^n$, we say a constraint $C = (f, \mathbf{j}, \mathbf{b})$ is *satisfied* by $\boldsymbol{\sigma}$ if $f(b_1\sigma_{j_1}, \ldots, b_k\sigma_{j_k}) = 1$. The *value* of $\boldsymbol{\sigma}$ on Ψ, denoted $\mathrm{val}_\Psi(\boldsymbol{\sigma})$, is the number of constraints satisfied by $\boldsymbol{\sigma}$. The goal of the problem is to compute the maximum value on Ψ over all assignments, i.e., $\max_{\boldsymbol{\sigma} \in \{-1,1\}^n} \mathrm{val}_\Psi(\boldsymbol{\sigma})$.

Unless stated otherwise, throughout the paper MAX-kCSP refers to Boolean MAX-kCSP. Since the function f will be clear from the context, we will often omit f and write a constraint as $(\mathbf{j}, \mathbf{b})$. We assume that all MAX-kCSP instances are simple and unweighted: no variable appears more than once in a constraint, and no two constraints have the same function f and the same set of literals. For each variable x_i, we further define $\deg(i)$ as the number of constraints in which x_i appears, either positively as x_i or negatively as $-x_i$.

Problem 2.2 (MAX-kAND)**.** MAX-kAND is the special case of MAX-kCSP(f) with $f = \mathrm{AND}$. Specifically, $\mathrm{AND}(b_1\sigma_{j_1}, \ldots, b_k\sigma_{j_k}) = 1$ iff $b_1\sigma_{j_1} = \cdots = b_k\sigma_{j_k} = 1$.

Streaming model. In this work, we consider the streaming model of computation, where the input instance arrives as a sequence that the algorithm scans in one pass, with the goal of outputting a solution for the problem at hand while minimizing space usage. For MAX-CSPs, an instance Ψ is represented as a sequence of constraints, i.e., $\langle C_1, C_2, \ldots \rangle$ where each C_i is a constraint in Ψ, and the goal is to output an estimate of the maximum value of Ψ. When the sequence contains only constraint insertions, we refer to it as an *insertion-only* stream. When it contains both constraint insertions and deletions, we refer to it as a *dynamic* stream and represent it as $\langle (C_1, \Delta_1), (C_2, \Delta_2), \ldots \rangle$, where for each i, $\Delta_i = 1$ (resp. -1) denotes the insertion (resp. deletion) of constraint C_i.

Prediction model. In this work, we assume that the algorithm is equipped with an oracle $\mathcal{O}$ that provides ε-accurate predictions $\mathbf{Y} \in \{\pm 1\}^n$, where $\varepsilon \in (0, \frac{1}{2}]$. We can also assume that each arriving constraint in the stream is annotated with the predicted labels of its associated variables. Specifically, for each variable index $i \in [n]$, $\Pr[Y_i = \sigma_i^*] = \frac{1}{2} + \varepsilon$ and $\Pr[Y_i = -\sigma_i^*] = \frac{1}{2} - \varepsilon$, where $\boldsymbol{\sigma}^* \in \{-1,1\}^n$ is some fixed but unknown optimal assignment to the variables. We also assume that the predictions are independent.[9]

Problem statement. In this paper, we consider the problem of estimating the maximum value of a Boolean MAX-CSP instance, in the streaming model using ε-accurate predictions. We say an algorithm achieves an α-approximation if it outputs a value ALG such that $\alpha \cdot \mathrm{OPT} \leq \mathrm{ALG} \leq \mathrm{OPT}$, where OPT denotes the optimal (maximum) value. We assume that the prediction accuracy parameter ε is known to the algorithm in advance.

3 Technical overview

In this section, we provide a technical overview, using MAX-DICUT to illustrate our ideas.

Problem 3.1 (MAX-DICUT)**.** Given a directed graph $G = (V, E)$, the goal is to output the maximum value of all possible directed cuts, i.e., $\max_{S \subseteq V} e(S, V \setminus S)$.

[9] In fact, our algorithms only require 4-wise independence for MAX-2CSPs.

Note that MAX-DICUT can be viewed as a special case of MAX-2AND: the variables are indexed by the vertices, and each directed edge corresponds to a constraint with $\mathbf{b} = (1, -1)$ (see Problem 2.2).

Specifically, the input corresponds to a graph stream consisting of edge insertions (and deletions) of a directed graph G, arriving one by one in an arbitrary order. The goal is to estimate the value of the maximum directed cut of G, which maximizes the number of edges directed from the first set to the second across all bipartitions of the vertices. This problem naturally generalizes the maximum cut problem MAX-CUT for undirected graphs.

At a high level, we view the oracle predictions, given as a vector $\mathbf{Y} \in \{-1, +1\}^n$, as a noisy encoding of an optimal cut (S^*, T^*), represented by $\mathbf{x}^* \in \{\pm 1\}^n$. We then design a collection of *linear measurements* of the stream that can be *decoded* to recover the optimal value.

Before presenting our approach in detail, we note that a natural strategy, used in recent work on learning-augmented algorithms for streaming MAX-CUT [DPV25], is to follow the predictions, i.e., evaluate the cut induced by predictions. Fix an optimal solution $\mathbf{x}^* \in \{\pm 1\}^n$. Since each endpoint may flip independently with probability $\frac{1}{2} - \varepsilon$, a cut edge under $\mathbf{x}^*$ may survive (i.e., still appears to be cut) under $\mathbf{Y}$ with probability at most $\frac{1}{2} + \mathrm{poly}(\varepsilon)$, and therefore such arguments inherently yield at best a $\frac{1}{2} + \mathrm{poly}(\varepsilon)$ approximation to the optimum value.

For MAX-DICUT, the direct analogue of the survival-based approach is to evaluate the directed cut $(S_{\mathbf{Y}}, T_{\mathbf{Y}})$ induced by predictions, where $S_{\mathbf{Y}} = \{v : Y_v = +1\}$ and $T_{\mathbf{Y}} = \{v : Y_v = -1\}$. However, an optimal edge (u, v) with $u \in S^*$ and $v \in T^*$ contributes to $(S_{\mathbf{Y}}, T_{\mathbf{Y}})$ only if $Y_u = +1$ and $Y_v = -1$, which happens with probability $\frac{1}{4} + O(\varepsilon)$. Hence this paradigm is inherently limited to a $(1/4 + O(\varepsilon))$-approximation for MAX-DICUT. By comparison, without predictions the best known one-pass $O(\log n)$-space algorithms achieve a $(4/9 - \eta)$-approximation [CGV20, GVV17], and with $\widetilde{O}(\sqrt{n})$ space one can reach roughly a 0.485-approximation on general graphs [SSSV23a, SSSV23b]. Very recently, a $(1/2 - \eta)$-approximation in truly sublinear space was obtained for general graphs, matching the $1/2$ barrier [ABFS26, KK19].

Our approach departs from this paradigm: rather than estimating the fraction of optimal cut edges that "survive" the noise, we use *all edge outcomes induced by the noise* and explicitly *invert the noise* to recover the hidden cut value. Fix an optimal bipartition (S^*, T^*) for MAX-DICUT and its corresponding labeling $\mathbf{x}^* \in \{\pm 1\}^n$. The oracle returns $\mathbf{Y} \in \{\pm 1\}^n$ such that for each $v \in V$,

$$\Pr[Y_v = x_v^*] = \tfrac{1}{2} + \varepsilon, \qquad \Pr[Y_v = -x_v^*] = \tfrac{1}{2} - \varepsilon,$$

independently over vertices. Thus, $\mathbf{Y}$ can be seen as the output of a binary symmetric channel applied to the optimal labels.

For a directed edge (u, v), the pair (Y_u, Y_v) falls into one of three categories that are relevant to MAX-DICUT: (1) predicted to go from left to right: $(Y_u, Y_v) = (+1, -1)$, (2) predicted to stay on the same side: $Y_u = Y_v$, (3) predicted to go from right to left: $(Y_u, Y_v) = (-1, +1)$. Let X_0, X_1, X_2 denote the counts of edges in these three categories (restricted to the low-degree part of the graph, as discussed below). Crucially, the expectations $(\mathbb{E}[X_0], \mathbb{E}[X_1], \mathbb{E}[X_2])$ form a *linear transformation* of the unknown triple (k_0, k_1, k_2), where: k_0 is the number of edges going from S^* to T^*, k_1 is the number of edges staying inside S^* or inside T^*, and k_2 is the number of edges going from T^* to S^*. Since the noise model is i.i.d. across vertices, this linear map depends *only* on ε and has full rank for any $\varepsilon > 0$. Therefore, we can explicitly *invert* this system, via Gaussian elimination, to obtain a linear estimator

$$\widetilde{k}_0 = \alpha_0(\varepsilon) \cdot X_0 + \alpha_1(\varepsilon) \cdot X_1 + \alpha_2(\varepsilon) \cdot X_2,$$

satisfying $\mathbb{E}[\widetilde{k}_0] = k_0$. Informally, instead of using the predictions to estimate the number of survived edges of an actual cut, we use the predictions to build an *unbiased decoder* for the *number of optimal cut edges*. This is the core reason we can surpass the $1/2 + \mathrm{poly}(\varepsilon)$-approximation guarantee.

The main obstacle: dependence across edges and variance blow-up. While $\widetilde{k}_0$ is unbiased, it is not immediately concentrated: even though $(Y_v)_{v \in V}$ are independent, the edge indicators are not independent because edges share endpoints. If a vertex participates in many edges, a single prediction bit Y_v influences many summands, and the variance of the edge-type counts can become too large. To control this, we decompose the graph into low- and high-degree parts. We partition V into low-degree and high-degree vertices: $L = \{v \in V : \deg(v) < \theta m\}$ and $H = V \setminus L$, for a ratio $\theta = \Theta(\mathrm{poly}(\varepsilon, \eta) \cdot \delta)$ chosen in the analysis. This decomposition has two key properties:

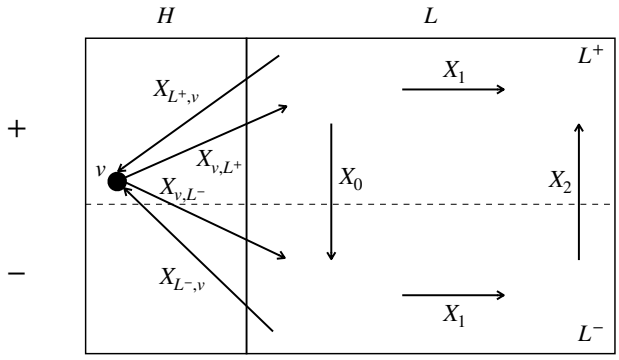

Figure 1: Illustration of the counters maintained in the algorithm for MAX-DICUT. The rectangle represents the vertex set V, partitioned into high-degree vertices H (left) and low-degree vertices L (right). We also partition V by the predicted labels (indicated by the horizontal dashed line): the upper region contains vertices predicted as $+1$ (denoted by "+"), and the lower region contains vertices predicted as -1 (denoted by "−").

- *Low-degree stability:* In the induced subgraph on L, each edge shares endpoints with only $O(\theta m)$ other edges. This bounds the covariance structure and yields $\mathrm{Var}[X_i] \le O(\theta m^2)$ for $i \in \{0, 1, 2\}$, which implies concentration of X_0, X_1, X_2 around their expectations via Chebyshev's inequality. Plugging these bounds into the decoded estimator $\widetilde{k}_L$ gives $|\widetilde{k}_L - k_L| \le \mathrm{poly}(\varepsilon, \eta) \cdot m$ with high probability, where k_L denotes the number of optimal MAX-DICUT edges completely inside L.
- *Few high-degree vertices:* Each $v \in H$ has degree at least θm, hence the number of such vertices is bounded by $|H| \le 2/\theta = \mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$. This allows us to explicitly maintain per-vertex statistics for high-degree vertices.

Decoding the contribution of high-degree vertices using 2-by-2 inversion. For every high-degree vertex $v \in H$ we maintain four counters describing its incident edges to L, separated by the predicted label on the low-degree endpoint: $X_{L^+,v}, X_{L^-,v}, X_{v,L^+}, X_{v,L^-}$. These are the numbers of edges from $L^\pm$ to v and from v to $L^\pm$, respectively, where $L^+ = \{u \in L : Y_u = +1\}$ and $L^- = \{u \in L : Y_u = -1\}$.

Now fix a high-degree vertex v and consider the unknown quantities $k_{S^*,v}, k_{T^*,v}, k_{v,S^*}, k_{v,T^*}$, which count edges between v and the *true* partition $S^* \cap L$ vs. $T^* \cap L$. The expectation of $(X_{L^+,v}, X_{L^-,v})$ is a simple linear system in $(k_{S^*,v}, k_{T^*,v})$ (and similarly for the outgoing direction). Because the noise is independent across low-degree endpoints, this system can be inverted explicitly, yielding unbiased estimators $\widetilde{k}_{S^*,v}$ and $\widetilde{k}_{v,T^*}$. Moreover, each $X_{\cdot,v}$ is a sum of independent Bernoulli variables over low-degree endpoints, so by concentration and a union bound over H, with high probability every $v \in H$ satisfies $|\widetilde{k}_{S^*,v} - k_{S^*,v}| \le \mathrm{poly}(\varepsilon, \eta) \cdot m$ and $|\widetilde{k}_{v,T^*} - k_{v,T^*}| \le \mathrm{poly}(\varepsilon, \eta) \cdot m$.

Figure 1 illustrates the meaning of the maintained counters in the algorithm, including X_0, X_1, X_2 and the incidence counters $X_{L^+,v}, X_{L^-,v}, X_{v,L^+}, X_{v,L^-}$ for each $v \in H$.

Greedy assignment of high-degree vertices. Conditioned on the low-degree part behaving well, we may think of $(S^* \cap L, T^* \cap L)$ as fixed, and decide independently for each $v \in H$ whether it should join the left side or the right side in order to maximize the resulting MAX-DICUT value contributed by edges between v and L. This contribution is exactly $\max\{k_{S^*,v}, k_{v,T^*}\}$ for the true cut, and we approximate it by $\max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\}$. Summing over $v \in H$ and adding the decoded low-degree estimate $\widetilde{k}_L$ yields our final estimator.

"Non-realizable" estimator. A subtle but important point is that the resulting quantity need not correspond to the value of *any explicit cut computed from the predictions* (or any efficiently constructed cut at all); it is best viewed as a *decoded estimate* of the optimal value. Nevertheless, the proof shows that it is a *valid estimator*: it is (i) close to OPT up to $\mathrm{poly}(\varepsilon, \eta) \cdot m$ additive error, and (ii) never exceeds OPT once we account for the small estimation slack.

Finally, we must account for edges whose endpoints both lie in H. Since $|H|$ is bounded by $\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$, the number of such edges becomes negligible compared to OPT under a mild density assumption on m, and thus does not affect the $(1 - o(1))$ approximation factor.

Streaming implementation. Our streaming implementation follows the same high-level techniques as the MAX-CUT algorithm of [DPV25]. We implement the described algorithm in the streaming model by approximating the low-/high-degree decomposition and maintaining just enough statistics to run the same decoding at the end. Since we cannot afford to identify the high-degree set or store all degrees, we first use reservoir sampling (and ℓ_0-sampling in the dynamic setting) to extract a small candidate set $\widetilde{H}$ that contains all truly high-degree vertices with high probability, and let $\widetilde{L} = V \setminus \widetilde{H}$. During the stream we maintain (i) a few exact global counters for the prediction-induced edge types and (ii) compact sketches that, for each $v \in \widetilde{H}$, estimate its incidence to the predicted sides. In post-processing, we subtract the estimated contribution of $\widetilde{H}$ from the global counts to recover the corresponding statistics on $\widetilde{L}$, and then apply the same approach of solving linear systems as in the offline algorithm for the $\widetilde{L}$ subgraph to obtain a near-optimal estimation of the MAX-DICUT value.

Extending to MAX-kAND. For MAX-2AND, our algorithm closely parallels that for MAX-DICUT. A MAX-2AND instance consists of constraints, which generalize directed edges, and each constraint contains two literals, which play a role analogous to edge directions. For $k \geq 3$, the main challenge is choosing the high-degree set itself. It must be small enough to enumerate, but after enumeration the remaining oracle noise must concentrate uniformly for every high assignment. We grow H with the moving threshold $\theta m/(|H| + 1)$ where θ is a parameter constant dependent on η, δ and ε: the harmonic-series bound keeps $|H|$ bounded in terms of $k, \varepsilon, \eta, \delta$, while the same $(|H| + 1)^{-1}$ margin makes every low variable's influence small enough to union-bound over all $2^{|H|}$ decoded values.

Extending to general Boolean MAX-CSPs. There is a standard black-box reduction that transforms any instance of Boolean MAX-kCSP into an instance of MAX-kAND, while preserving the optimum value and increasing the number of constraints by at most a factor of 2^k. Applying this reduction allows us to immediately extend our results to Boolean MAX-CSPs.

4 Our algorithms for MAX-DICUT and MAX-2AND

In this section, we present our algorithms for MAX-DICUT and MAX-2AND with predictions. To illustrate the algorithmic ideas, we first describe the algorithms in the *offline* setting, where the entire input instance is available. We defer the missing details, the streaming implementations, and the extension to MAX-kAND for $k \geq 3$ and, more generally, to MAX-kCSPs to Appendices B, C and D.

4.1 MAX-DICUT

We first describe our algorithm for MAX-DICUT as a warm-up. We defer the details to Appendix B.

We partition the vertices V into low-degree vertices L and high-degree vertices H, where $v \in H$ iff $\deg(v) \geq \theta m$ for $\theta := \Theta(\varepsilon^4 \delta \eta^2)$. We then scan the edges, query the oracle $\mathcal{O}$ on each endpoint, and update a small set of counters from which we then recover an estimate of the MAX-DICUT value.

Specifically, for each edge with both endpoints in L, we maintain three counters: (i) X_0: the number of edges with $(Y_u, Y_v) = (+1, -1)$ (predicted to contribute to the dicut), (ii) X_1: the number of edges with $Y_u = Y_v$ (predicted to stay on the same side), and (iii) X_2: the number of edges with $(Y_u, Y_v) = (-1, +1)$ (predicted to go in the reverse direction). These statistics later let us decode an accurate estimate $\widetilde{k}_L$ of the optimal dicut contribution coming from edges inside L, namely, $\widetilde{k}_L := \alpha_0(\varepsilon) \cdot X_0 + \alpha_1(\varepsilon) \cdot X_1 + \alpha_2(\varepsilon) \cdot X_2$.

Furthermore, for each high-degree vertex $v \in H$, we maintain four incidence counters: (i) $X_{L^+,v}$: the numbers of edges from L^+ to v, (ii) $X_{L^-,v}$: the numbers of edges from L^- to v, (iii) X_{v,L^+}: the numbers of edges from v to L^+, and (iv) X_{v,L^-}: the numbers of edges from v to L^-, where L^+ (resp. L^-) is the set of vertices in L predicted to be $+1$ (resp. -1). Using these counts, we decode accurate estimates of the corresponding edge quantities with respect to the unknown optimal bipartition (S^*, T^*), namely, the number of edges between v and $S^* \cap L$ or $T^* \cap L$ in each direction, denoted by $\widetilde{k}_{S^*,v}$ and $\widetilde{k}_{v,T^*}$, respectively. We then assign each $v \in H$ to the side that maximizes its

estimated contribution to the dicut, and combine this with the estimate from edges inside L to obtain our final MAX-DICUT estimate $\widetilde{k}_L + \sum_{v \in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \Theta(\eta)m$.

4.2 MAX-2AND

Similar to the algorithm for MAX-DICUT, we partition the variables into low-degree variables L and high-degree variables H, where $i \in H$ iff $\deg(i) \geq \theta m$ for $\theta := \Theta(\varepsilon^4 \delta \eta^2)$. We then scan the constraints, query the oracle $\mathcal{O}$ on the variables involved, and update a small set of counters. These counters are analogous to those in the MAX-DICUT case, except that they count the corresponding constraints rather than edges. From these counters, we subsequently recover an estimate of the MAX-2AND value. The algorithm is formally given in Algorithm 1.

Specifically, for each constraint whose associated variables are both in L, we maintain three counters: (i) X_0: the number of such constraints for which both literals are satisfied by the predicted assignment, (ii) X_1: the number for which exactly one literal is satisfied, and (iii) X_2: the number for which neither literal is satisfied. These statistics later let us decode an accurate estimate of the optimal contribution from constraints whose variables both lie in L (Line 20).

Furthermore, for each high-degree variable $i \in H$, we maintain four counters: (i) $X^{\mathrm{pos}}_{i,L(+)}$: the number of constraints in which i appears positively and the low-degree literal is satisfied by the predicted assignment; (ii) $X^{\mathrm{pos}}_{i,L(-)}$: i appears positively and the low-degree literal is not satisfied; (iii) $X^{\mathrm{neg}}_{i,L(+)}$: i appears negatively and the low-degree literal is satisfied; and (iv) $X^{\mathrm{neg}}_{i,L(-)}$: i appears negatively and the low-degree literal is not satisfied. Using these counts, we decode accurate estimates of the corresponding quantities with respect to the unknown optimal assignment $\boldsymbol{\sigma}^*$ (Lines 22 and 23). We then assign each $i \in H$ to maximize its estimated contribution to the value, and combine this with the estimate from constraints whose variables both lie in L to obtain our final MAX-2AND value estimate (Line 24).

Next, we analyze the approximation guarantee of Algorithm 1.

Theorem 4.1. *Let $\varepsilon \in (0, \frac{1}{2}], \delta, \eta \in (0,1)$, and Ψ be a* MAX-2AND *instance with $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Given an ε-accurate oracle $\mathcal{O}$ for Ψ, Algorithm 1 provides a $(1-\eta)$-approximation for the maximum value of Ψ with probability at least $1-\delta$.*

4.2.1 Key properties

The proof of Theorem 4.1 relies on the following key properties of the counters maintained by Algorithm 1, which are formalized in the lemmas below. We defer the proofs of these lemmas to Appendix C.1.1.

Let X_0, X_1, and X_2 denote the random variables given by the final values of the corresponding counters at the end of the algorithm. We first prove concentration bounds for these random variables around their expectations.

Lemma 4.2. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For each $i \in \{0,1,2\}$, we have $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^2 \eta m}{8}$ with probability at least $1 - \frac{\delta}{6}$.*

Recall that $\boldsymbol{\sigma}^*$ denotes an optimal assignment for the MAX-2AND instance Ψ. Define $k_L := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in L,\ b_1\sigma^*_{j_1} = b_2\sigma^*_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in L (i.e., both low-degree) that are satisfied by $\boldsymbol{\sigma}^*$. Recall that $\widetilde{k}_L$, computed at Line 20 of Algorithm 1, is an estimate of k_L. The following lemma shows that $\widetilde{k}_L$ is close to k_L up to an additive error of $\mathrm{poly}(\eta) \cdot m$ with high probability.

Lemma 4.3. *Let $\delta, \eta \in (0,1)$. We have $|\widetilde{k}_L - k_L| \leq \frac{\eta m}{32}$ with probability at least $1 - \frac{\delta}{2}$.*

For each $i \in H$, define

1. $k^{\mathrm{pos}}_{i,L(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in L, b_i = 1, b_j\sigma^*_j = 1\}|$ to be the number of constraints involving x_i and some low-degree variable x_j such that the literal of x_i appears positively (i.e., $b_i = 1$) and the low-side literal is satisfied by $\boldsymbol{\sigma}^*$ (i.e., $b_j\sigma^*_j = 1$). We analogously define:
2. $k^{\mathrm{pos}}_{i,L(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in L, b_i = 1, b_j\sigma^*_j = -1\}|$.

Algorithm 1 A learning-augmented *offline* algorithm for estimating the MAX-2AND value

Input: MAX-2AND instance $\Psi = (C_1, \ldots, C_m)$, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$
Output: Estimate of the maximum value of Ψ

1: $\theta \leftarrow \frac{\varepsilon^4 \delta \eta^2}{768}$, $L \leftarrow \{i \in [n] : \deg(i) < \theta m\}$, $H \leftarrow [n] \setminus L$
2: **initialize** $X_0 \leftarrow 0, X_1 \leftarrow 0, X_2 \leftarrow 0$.
3: **initialize** $X^{\mathrm{pos}}_{i,L(+)} \leftarrow 0, X^{\mathrm{pos}}_{i,L(-)} \leftarrow 0, X^{\mathrm{neg}}_{i,L(+)} \leftarrow 0, X^{\mathrm{neg}}_{i,L(-)} \leftarrow 0$ for each $i \in H$.
4: **for each** constraint $((j_1, j_2), (b_1, b_2)) \in \Psi$ **do**
5: $\quad Y_{j_1} = \mathcal{O}(j_1), Y_{j_2} = \mathcal{O}(j_2)$
6: $\quad$ **if** $j_1 \in L$ and $j_2 \in L$ **then**
7: $\quad\quad$ **if** $b_1 Y_{j_1} = b_2 Y_{j_2} = 1$ **then** $X_0 \leftarrow X_0 + 1$
8: $\quad\quad$ **if** $b_1 Y_{j_1} \neq b_2 Y_{j_2}$ **then** $X_1 \leftarrow X_1 + 1$
9: $\quad\quad$ **if** $b_1 Y_{j_1} = b_2 Y_{j_2} = -1$ **then** $X_2 \leftarrow X_2 + 1$
10: $\quad$ **if** $j_1 \in H$ and $j_2 \in L$ **then**
11: $\quad\quad$ **if** $b_1 = 1$ **then**
12: $\quad\quad\quad$ **if** $b_2 Y_{j_2} = 1$ **then** $X^{\mathrm{pos}}_{j_1,L(+)} \leftarrow X^{\mathrm{pos}}_{j_1,L(+)} + 1$ **else** $X^{\mathrm{pos}}_{j_1,L(-)} \leftarrow X^{\mathrm{pos}}_{j_1,L(-)} + 1$
13: $\quad\quad$ **else**
14: $\quad\quad\quad$ **if** $b_2 Y_{j_2} = 1$ **then** $X^{\mathrm{neg}}_{j_1,L(+)} \leftarrow X^{\mathrm{neg}}_{j_1,L(+)} + 1$ **else** $X^{\mathrm{neg}}_{j_1,L(-)} \leftarrow X^{\mathrm{neg}}_{j_1,L(-)} + 1$
15: $\quad$ **if** $j_1 \in L$ and $j_2 \in H$ **then**
16: $\quad\quad$ **if** $b_2 = 1$ **then**
17: $\quad\quad\quad$ **if** $b_1 Y_{j_1} = 1$ **then** $X^{\mathrm{pos}}_{j_2,L(+)} \leftarrow X^{\mathrm{pos}}_{j_2,L(+)} + 1$ **else** $X^{\mathrm{pos}}_{j_2,L(-)} \leftarrow X^{\mathrm{pos}}_{j_2,L(-)} + 1$
18: $\quad\quad$ **else**
19: $\quad\quad\quad$ **if** $b_1 Y_{j_1} = 1$ **then** $X^{\mathrm{neg}}_{j_2,L(+)} \leftarrow X^{\mathrm{neg}}_{j_2,L(+)} + 1$ **else** $X^{\mathrm{neg}}_{j_2,L(-)} \leftarrow X^{\mathrm{neg}}_{j_2,L(-)} + 1$
20: $\widetilde{k}_L \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2} + \varepsilon)^2 X_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon) X_1 + (\frac{1}{2} - \varepsilon)^2 X_2]$
21: **for each** $i \in H$ **do**
22: $\quad \widetilde{k}^{\mathrm{pos}}_{i,L(+)} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) X^{\mathrm{pos}}_{i,L(+)} + (\frac{1}{2} - \frac{1}{4\varepsilon}) X^{\mathrm{pos}}_{i,L(-)}$
23: $\quad \widetilde{k}^{\mathrm{neg}}_{i,L(+)} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) X^{\mathrm{neg}}_{i,L(+)} + (\frac{1}{2} - \frac{1}{4\varepsilon}) X^{\mathrm{neg}}_{i,L(-)}$
24: **return** $\widetilde{k}_L + \sum_{i \in H} \max\{\widetilde{k}^{\mathrm{pos}}_{i,L(+)}, \widetilde{k}^{\mathrm{neg}}_{i,L(+)}\} - \frac{\eta m}{16}$

3. $k^{\mathrm{neg}}_{i,L(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in L, b_i = -1, b_j \sigma^*_j = 1\}|$.
4. $k^{\mathrm{neg}}_{i,L(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in L, b_i = -1, b_j \sigma^*_j = -1\}|$.

Recall that $\widetilde{k}^{\mathrm{pos}}_{i,L(+)}$ and $\widetilde{k}^{\mathrm{neg}}_{i,L(+)}$, computed at Lines 22 and 23 of Algorithm 1, are estimates of $k^{\mathrm{pos}}_{i,L(+)}$ and $k^{\mathrm{neg}}_{i,L(+)}$, respectively. The following lemma shows that $\widetilde{k}^{\mathrm{pos}}_{i,L(+)}$ and $\widetilde{k}^{\mathrm{neg}}_{i,L(+)}$ are close to $k^{\mathrm{pos}}_{i,L(+)}$ and $k^{\mathrm{neg}}_{i,L(+)}$, respectively, up to an additive error of $\mathrm{poly}(\varepsilon, \delta, \eta) \cdot m$ with high probability.

Lemma 4.4. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0, 1)$. For each $i \in H$, each of the following holds with probability at least $1 - \frac{\varepsilon^4 \delta^2 \eta^2}{6144}$: (1) $|\widetilde{k}^{\mathrm{pos}}_{i,L(+)} - k^{\mathrm{pos}}_{i,L(+)}| \leq \frac{\varepsilon^4 \delta \eta^3 m}{49152}$; (2) $|\widetilde{k}^{\mathrm{neg}}_{i,L(+)} - k^{\mathrm{neg}}_{i,L(+)}| \leq \frac{\varepsilon^4 \delta \eta^3 m}{49152}$.*

Define $k_H := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in H,\ b_1 \sigma^*_{j_1} = b_2 \sigma^*_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in H (i.e., both high-degree) that are satisfied by $\boldsymbol{\sigma}^*$. Note that $|H| \leq \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{1536}{\varepsilon^4 \delta \eta^2}$. Thus we have the following fact.

Fact 4.5. *It holds that $k_H < |H|^2 \leq \frac{1536^2}{\varepsilon^8 \delta^2 \eta^4}$.*

4.2.2 Proof of Theorem 4.1

Proof. Recall that k_L (resp. k_H) is the number of constraints in Ψ whose associated variables are both in L (resp. H) that are satisfied by $\boldsymbol{\sigma}^*$.

Furthermore, define

$$k_{L,H} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : |\{j_1,j_2\} \cap L| = |\{j_1,j_2\} \cap H| = 1, b_1 \sigma^*_{j_1} = b_2 \sigma^*_{j_2} = 1\}|$$

to be the number of constraints in Ψ whose associated variables are one in L and the other in H that are satisfied by $\boldsymbol{\sigma}^*$.

Therefore, we have $\mathrm{OPT} = k_L + k_H + k_{L,H}$. Let ALG denote the output of Algorithm 1. By union bound, with probability at least $1 - \frac{\delta}{2} - \frac{\varepsilon^4\delta^2\eta^2}{6144} \cdot 2 \cdot |H| \geq 1 - \delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{i\in H} \max\{\widetilde{k}^{\mathrm{pos}}_{i,L(+)}, \widetilde{k}^{\mathrm{neg}}_{i,L(+)}\} - \frac{\eta m}{16} \\
&\leq k_L + \frac{\eta m}{32} + \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} + |H| \cdot \frac{\varepsilon^4\delta\eta^3 m}{49152} - \frac{\eta m}{16} && \triangleright \text{Lemma 4.3 and Lemma 4.4} \\
&\leq k_L + \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} && \triangleright |H| \leq \frac{1536}{\varepsilon^4\delta\eta^2} \\
&\leq \mathrm{OPT},
\end{aligned}$$

where the last step follows since $k_L + \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\}$ is a lower bound on the value of a feasible assignment. Specifically, we fix the restriction of the assignment on L to be $\boldsymbol{\sigma}^*|_L$, and then, for each $i \in H$, we set $\sigma_i \in \{-1, 1\}$ to maximize the number of satisfied constraints involving variable x_i and a low-degree variable.

Furthermore,

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{i\in H} \max\{\widetilde{k}^{\mathrm{pos}}_{i,L(+)}, \widetilde{k}^{\mathrm{neg}}_{i,L(+)}\} - \frac{\eta m}{16} \\
&\geq k_L - \frac{\eta m}{32} + \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} - |H| \cdot \frac{\varepsilon^4\delta\eta^3 m}{49152} - \frac{\eta m}{16} && \triangleright \text{Lemma 4.3 and Lemma 4.4} \\
&\geq k_L - \frac{\eta m}{32} + \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} - \frac{\eta m}{32} - \frac{\eta m}{16} && \triangleright |H| \leq \frac{1536}{\varepsilon^4\delta\eta^2} \\
&\geq k_L + k_{L,H} - \frac{\eta m}{8} && \triangleright \sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} \geq k_{L,H} \\
&= \mathrm{OPT} - k_H - \frac{\eta m}{8} && \triangleright \mathrm{OPT} = k_L + k_H + k_{L,H} \\
&\geq \mathrm{OPT} - \frac{1536^2}{\varepsilon^8\delta^2\eta^4} - \frac{\eta}{2} \cdot \mathrm{OPT} && \triangleright \text{Fact 4.5 and } m \leq 4\mathrm{OPT} \\
&\geq (1 - \eta) \cdot \mathrm{OPT}, && \triangleright \frac{1536^2}{\varepsilon^8\delta^2\eta^4} \leq \frac{\eta m}{8} \leq \frac{\eta}{2} \cdot \mathrm{OPT}
\end{aligned}$$

where the fourth step follows from

$$\begin{aligned}
\sum_{i\in H} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} &= \sum_{i\in H, \sigma^*_i = 1} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} + \sum_{i\in H, \sigma^*_i = -1} \max\{k^{\mathrm{pos}}_{i,L(+)}, k^{\mathrm{neg}}_{i,L(+)}\} \\
&\geq \sum_{i\in H, \sigma^*_i = 1} k^{\mathrm{pos}}_{i,L(+)} + \sum_{i\in H, \sigma^*_i = -1} k^{\mathrm{neg}}_{i,L(+)} = k_{L,H}.
\end{aligned}$$

□

5 Conclusion

We have exhibited that learning-augmented streaming algorithms with oracle access to ε-accurate noisy predictions can obtain highly-accurate estimates for the values of Boolean MAX-CSPs. There are several interesting open questions that arise from our work which we state below.

1. Can we relax the assumption that the streaming algorithm has to know ε *with high accuracy*? For instance, could one obtain similar results assuming only a constant-factor approximation to ε is known to the algorithm? Or can we prove lower bounds showing that this is impossible?
2. Are there suitable generalizations of the prediction models for MAX-CSPs defined on arbitrary q-ary alphabets?

Acknowledgments

This work began as part of a discussion between Pan Peng, Ali Vakilian, and Santhoshini Velusamy when they were attending the Workshop on High-Dimensional and Complex Data Algorithms in Venice, Italy in May 2025. The work was conducted in part while Yinhao Dong and Xiaoyang Xu were visiting the Toyota Technological Institute at Chicago (TTIC) in summer 2025, and while Yinhao Dong, Pan Peng, and Ali Vakilian were attending the Sublinear Algorithms Reunion at the Simons Institute in summer 2025.

Y.D. is supported in part by the National Center for Interdisciplinary Studies and Talents Cultivation, USTC. P.P. is supported in part by NSFC Grant 62272431 and Quantum Science and Technology – National Science and Technology Major Project (Grant No. 2021ZD0302901). A.V. is supported in part by NSF CAREER Award CCF-2542925. S.V. was supported in part by NSF award CCF-2348475 when she was affiliated with TTIC. S.V. is also supported in part by NSERC Discovery grant RGPIN-2026-04689 and ECR supplement DGECR-2026-00421. X.X. was supported by NSF award CCF-2348475 when visiting TTIC.

We thank Yury Makarychev and Daniel Lokshtanov for insightful discussions and valuable feedback.

AI Disclosure

We used AI tools, including GPT-5.4 Pro and GPT-5.5 Pro, throughout this project. Specifically, GPT-5.5 Pro assisted with the generalization of the algorithms for unbounded-degree MAX-kAND with $k \geq 3$ (Appendix C.3) based on the authors' ideas, as well as with the calculations in Appendix E.2. GPT-5.4 Pro assisted with the calculations in Appendix E.1. All other research contributions were made solely by the authors.

AI models were also used to assist with drafting, proofreading, and language editing throughout the writing process. The authors substantially revised and rewrote the AI-generated content and carefully verified all mathematical statements and proofs, making corrections and improvements as needed. The authors verified the AI-assisted results and take full responsibility for the correctness of all claims in this paper.

References

[ABFS26] Amir Azarmehr, Soheil Behnezhad, Shane Ferrante, and Mohammad Saneian. Half-approximating maximum dicut in the streaming setting. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC)*, 2026.

[ACG+25a] Anders Aamand, Justin Y. Chen, Siddharth Gollapudi, Sandeep Silwal, and Hao Wu. Improved approximations for hard graph problems using predictions. In *Forty-second International Conference on Machine Learning (ICML)*, 2025.

[ACG+25b] Anders Aamand, Justin Y. Chen, Siddharth Gollapudi, Sandeep Silwal, and Hao Wu. Learning-augmented frequent directions. In *the Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[ACN+23] Anders Aamand, Justin Chen, Huy Nguyen, Sandeep Silwal, and Ali Vakilian. Improved frequency estimation algorithms with and without predictions. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

[AGR25] Idan Attias, Xing Gao, and Lev Reyzin. Learning-augmented algorithms for boolean satisfiability. *CoRR*, abs/2505.06146, 2025.

[AIV19] Anders Aamand, Piotr Indyk, and Ali Vakilian. (Learned) frequency estimation algorithms under zipfian distribution. *CoRR*, abs/1908.05198, 2019.

[AKSY20] Sepehr Assadi, Gillat Kol, Raghuvansh R. Saxena, and Huacheng Yu. Multi-pass graph streaming lower bounds for cycle counting, max-cut, matching size, and other problems. In *61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 354–364, 2020.

[AN21] Sepehr Assadi and Vishvajeet N. Graph streaming lower bounds for parameter estimation and property testing via a streaming xor lemma. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 612–625, 2021.

[BDSW24] Vladimir Braverman, Prathamesh Dharangutte, Vihan Shah, and Chen Wang. Learning-augmented maximum independent set. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 317 of *LIPIcs*, pages 24:1–24:18, 2024.

[BVZ01] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

[CdG+24] Vincent Cohen-Addad, Tommaso d'Orsi, Anupam Gupta, Euiwoong Lee, and Debmalya Panigrahi. Learning-augmented approximation algorithms for maximum cut and related problems. In *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[CEI+22] Justin Y. Chen, Talya Eden, Piotr Indyk, Honghao Lin, Shyam Narayanan, Ronitt Rubinfeld, Sandeep Silwal, Tal Wagner, David P. Woodruff, and Michael Zhang. Triangle and four cycle counting with predictions in graph streams. In *the Tenth International Conference on Learning Representations (ICLR)*, 2022.

[CGS+22a] Chi-Ning Chou, Alexander Golovnev, Amirbehshad Shahrasbi, Madhu Sudan, and Santhoshini Velusamy. Sketching approximability of (weak) monarchy predicates. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 245 of *LIPIcs*, pages 35:1–35:17, 2022.

[CGS+22b] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, Ameya Velingker, and Santhoshini Velusamy. Linear space streaming lower bounds for approximating csps. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 275–288, 2022.

[CGSV21] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all boolean csps with linear sketches. *CoRR*, abs/2102.12351, 2021.

[CGSV24] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Sketching approximability of all finite csps. *J. ACM*, 71(2):15:1–15:74, 2024.

[CGV20] Chi-Ning Chou, Alexander Golovnev, and Santhoshini Velusamy. Optimal streaming approximations for all boolean max-2csps and max-ksat. In *61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 330–341, 2020.

[CKP+23] Lijie Chen, Gillat Kol, Dmitry Paramonov, Raghuvansh R. Saxena, Zhao Song, and Huacheng Yu. Towards multi-pass streaming lower bounds for optimal approximation of max-cut. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 878–924, 2023.

[CM05] Graham Cormode and Shan Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *J. Algorithms*, 55(1):58–75, 2005.

[DIL+21] Michael Dinitz, Sungjin Im, Thomas Lavastida, Benjamin Moseley, and Sergei Vassilvitskii. Faster matchings via learned duals. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 10393–10406, 2021.

[DJLP25] Yinhao Dong, Shan Jiang, Shi Li, and Pan Peng. Learning-augmented streaming algorithms for correlation clustering. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[DPV25] Yinhao Dong, Pan Peng, and Ali Vakilian. Learning-augmented streaming algorithms for approximating MAX-CUT. In *16th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 325 of *LIPIcs*, pages 44:1–44:24, 2025.

[EIN+21] Talya Eden, Piotr Indyk, Shyam Narayanan, Ronitt Rubinfeld, Sandeep Silwal, and Tal Wagner. Learning-based support estimation in sublinear time. In *9th International Conference on Learning Representations (ICLR)*, 2021.

[FMW25] Yumou Fei, Dor Minzer, and Shuo Wang. Multi-pass streaming lower bounds for approximating max-cut. In *66th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2025.

[FMW26] Yumou Fei, Dor Minzer, and Shuo Wang. A dichotomy theorem for multi-pass streaming csps. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC)*, 2026.

[GMM25] Suprovat Ghoshal, Konstantin Makarychev, and Yury Makarychev. Constraint satisfaction problems with advice. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1202–1221. SIAM, 2025.

[GT19] Venkatesan Guruswami and Runzhou Tao. Streaming hardness of unique games. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 145 of *LIPIcs*, pages 5:1–5:12, 2019.

[GVV17] Venkatesan Guruswami, Ameya Velingker, and Santhoshini Velusamy. Streaming complexity of approximating max 2csp and max acyclic subgraph. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 81 of *LIPIcs*, pages 8:1–8:19, 2017.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[HIKV19] Chen-Yu Hsu, Piotr Indyk, Dina Katabi, and Ali Vakilian. Learning-based frequency estimation algorithms. In *7th International Conference on Learning Representations (ICLR)*, 2019.

[IVY19] Piotr Indyk, Ali Vakilian, and Yang Yuan. Learning-based low-rank approximations. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 7400–7410, 2019.

[JLL+20] Tanqiu Jiang, Yi Li, Honghao Lin, Yisong Ruan, and David P. Woodruff. Learning-augmented data stream algorithms. In *8th International Conference on Learning Representations (ICLR)*, 2020.

[JST11] Hossein Jowhari, Mert Sağlam, and Gábor Tardos. Tight bounds for lp samplers, finding duplicates in streams, and related problems. In *Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS)*, pages 49–58, 2011.

[KBC+18] Tim Kraska, Alex Beutel, Ed H. Chi, Jeffrey Dean, and Neoklis Polyzotis. The case for learned index structures. In *Proceedings of the 2018 International Conference on Management of Data (SIGMOD)*, pages 489–504, 2018.

[KK15] Dmitry Kogan and Robert Krauthgamer. Sketching cuts in graphs and hypergraphs. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 367–376, 2015.

[KK19] Michael Kapralov and Dmitry Krachun. An optimal space lower bound for approximating MAX-CUT. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 277–288, 2019.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37(1):319–357, 2007.

[KKS15] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Streaming lower bounds for approximating max-cut. In *Proceedings of the 2015 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1263–1282, 2015.

[KKSV17] Michael Kapralov, Sanjeev Khanna, Madhu Sudan, and Ameya Velingker. $(1 + \Omega(1))$-approximation to max-cut requires linear space. In *Proceedings of the 2017 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1703–1722, 2017.

[KP22] John Kallaugher and Ojas Parekh. The quantum and classical streaming complexity of quantum and classical max-cut. In *63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 498–506, 2022.

[KPS18] Ravi Kumar, Manish Purohit, and Zoya Svitkina. Improving online algorithms via ml predictions. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 9684–9693, 2018.

[KPV24] John Kallaugher, Ojas Parekh, and Nadezhda Voronova. Exponential quantum space advantage for approximating maximum directed cut in the streaming model. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1805–1815, 2024.

[KPV25] John Kallaugher, Ojas Parekh, and Nadezhda Voronova. How to design a quantum streaming algorithm without knowing anything about quantum computing. In *2025 Symposium on Simplicity in Algorithms (SOSA)*, pages 9–45, 2025.

[LLL+23] Yi Li, Honghao Lin, Simin Liu, Ali Vakilian, and David P. Woodruff. Learning the positions in countsketch. In *11th International Conference on Learning Representations (ICLR)*, 2023.

[LMP01] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML)*, pages 282–289, 2001.

[LMS25] Alexander Lindermayr, Nicole Megow, and Bertrand Simon. Boosting double coverage for k-server via imperfect predictions. *Algorithmica*, 87(11):1477–1517, 2025.

[LV21] Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *J. ACM*, 68(4):24:1–24:25, 2021.

[Mit18] Michael Mitzenmacher. A model for learned bloom filters and optimizing by sandwiching. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 462–471, 2018.

[MNS07] Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. Allocating online advertisement space with unreliable estimates. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC)*, pages 288–294, 2007.

[NPW+26] Soham Nagawanshi, Shalini Panthangi, Chen Wang, David P. Woodruff, and Samson Zhou. Learning-augmented moment estimation on time-decay models. In *14th International Conference on Learning Representations (ICLR)*, 2026.

[Sin23] Noah G. Singer. Oblivious algorithms for the Max-kAND problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 275 of *LIPIcs*, pages 15:1–15:19, 2023.

[Sin25] Noah G. Singer. Nine lower bound conjectures on streaming approximation algorithms for csps. *ACM SIGACT News*, 56(4):25–36, 2025.

[SLB+19] Daniel Selsam, Matthew Lamm, Benedikt Bünz, Percy Liang, Leonardo de Moura, and David L. Dill. Learning a SAT solver from single-bit supervision. In *7th International Conference on Learning Representations (ICLR)*, 2019.

[SSSV23a] Raghuvansh R. Saxena, Noah Singer, Madhu Sudan, and Santhoshini Velusamy. Streaming complexity of csps with randomly ordered constraints. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4083–4103, 2023.

[SSSV23b] Raghuvansh R. Saxena, Noah G. Singer, Madhu Sudan, and Santhoshini Velusamy. Improved streaming algorithms for maximum directed cut via smoothed snapshots. In *64th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 855–870, 2023.

[SSSV25] Raghuvansh R. Saxena, Noah G. Singer, Madhu Sudan, and Santhoshini Velusamy. Streaming algorithms via local algorithms for maximum directed cut. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3392–3408, 2025.

[Sud22] Madhu Sudan. Streaming and sketching complexity of csps: A survey (invited talk). In *49th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 229 of *LIPIcs*, pages 5:1–5:20, 2022.

[TGK03] Benjamin Taskar, Carlos Guestrin, and Daphne Koller. Max-margin markov networks. In *Advances in Neural Information Processing Systems 16: Neural Information Processing Systems (NIPS)*, pages 25–32, 2003.

[THJA04] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. Support vector machine learning for interdependent and structured output spaces. In *Machine Learning, Proceedings of the Twenty-first International Conference (ICML)*, ACM International Conference Proceeding Series, 2004.

[Vel26] Santhoshini Velusamy. Near-optimal streaming approximation for max-dicut in sublinear space using two passes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, 2026.

[Vit85] Jeffrey Scott Vitter. Random sampling with a reservoir. *ACM Trans. Math. Softw.*, 11(1):37–57, 1985.

[VKMK21] Kapil Vaidya, Eric Knorr, Michael Mitzenmacher, and Tim Kraska. Partitioned learned bloom filters. In *9th International Conference on Learning Representations (ICLR)*, 2021.

[XKK18] Hong Xu, Sven Koenig, and T. K. Satish Kumar. Towards effective deep learning for constraint satisfaction problems. In *Principles and Practice of Constraint Programming - 24th International Conference (CP)*, Lecture Notes in Computer Science, pages 588–597, 2018.

Organization of the appendix. Appendix A reviews additional technical preliminaries used in the paper. Appendix B presents our algorithm for MAX-DICUT, including the offline analysis deferred from the main text and its streaming implementations; this case serves as a warm-up for our algorithms for MAX-CSPs. Appendix C extends the approach to MAX-kAND, first for $k = 2$, then for bounded- and unbounded-degree instances with $k \geq 3$. Appendix D proves our main theorem by reducing general MAX-kCSPs to MAX-kAND. Finally, Appendix E shows that it suffices for our algorithms to know only a highly accurate estimate of ε. Although the appendix is extensive, the same core decoding framework is used throughout, with additional technical treatments tailored to the different problems.

A Additional technical preliminaries

In this paper, we employ several well-known techniques in the streaming model, as described below.

CountMin sketch. Given a stream of m integers from $[n]$, let f_i denote number of occurrences of i in the stream, for any $i \in [n]$. In CountMin sketch, there are k uniform and independent random hash functions $h_1, h_2, \ldots, h_k : [n] \to [w]$ and an array C of size $k \times w$. The algorithm maintains C, such that $C[\ell, s] = \sum_{j:h_\ell(j)=s} f_j$ at the end of the stream. Upon each query of an arbitrary $i \in [n]$, the algorithm returns $\widetilde{f}_i = \min_{\ell \in [k]} C[\ell, h_\ell(i)]$.

Theorem A.1 (CountMin sketch [CM05])**.** *Let $\mathbf{f} \in \mathbb{R}^n_{\geq 0}$ be a frequency vector with total mass $m = \|\mathbf{f}\|_1$ maintained under streaming updates, and let $\widetilde{f}_i$ denote the CountMin estimate for coordinate $i \in [n]$. For any parameters $\varepsilon, \delta \in (0,1)$, there is a CountMin sketch that uses $O\big(\frac{1}{\varepsilon} \log \frac{1}{\delta}\big)$ words of space and supports point queries such that for every $i \in [n]$,*

$$\widetilde{f}_i \geq f_i \qquad \textit{and} \qquad \Pr\Big[\widetilde{f}_i \leq f_i + \varepsilon m\Big] \geq 1 - \delta.$$

Reservoir sampling. The algorithm samples k elements from a stream, such that at any time $m \geq k$, the sample consists of a uniformly random subset of size k of the elements seen so far. The space complexity of the reservoir sampling algorithm is $O(k(\log n + \log m))$ bits [Vit85].

Definition A.2 (ℓ_0-sampler)**.** Let $\mathbf{x} \in \mathbb{R}^n$ be a non-zero vector and $\delta \in (0,1)$. An ℓ_0-sampler for $\mathbf{x}$ returns FAIL with probability at most δ and otherwise returns some index i such that $x_i \neq 0$ with probability $\frac{1}{|\operatorname{supp}(\mathbf{x})|}$ where $\operatorname{supp}(\mathbf{x}) = \{i \in [n] : x_i \neq 0\}$ is the support of $\mathbf{x}$.

The following theorem states that ℓ_0-samplers can be maintained using a single pass in dynamic streams.

Theorem A.3 (Theorem 2 in [JST11])**.** *Let $\delta \in (0,1)$. There exists a single-pass streaming algorithm for maintaining an ℓ_0-sampler for a non-zero vector $\mathbf{x} \in \mathbb{R}^n$ (with failure probability δ) in the dynamic model using $O(\log^2 n \log \delta^{-1})$ bits of space.*

We also use the following standard concentration inequalities.

Theorem A.4 (Chebyshev's inequality)**.** *Let X be a random variable with expectation $\mathbb{E}[X]$ and variance $\operatorname{Var}[X]$. Then for any $t > 0$,*

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\operatorname{Var}[X]}{t^2}.$$

Theorem A.5 (Bounded differences inequality)**.** *Let $X_1, \ldots, X_n \in \mathcal{X}$ are independent random variables. Let $f : \mathcal{X}^n \to \mathbb{R}$ be a function, and let $Z = f(X_1, \ldots, X_n)$. Suppose that for every $i \in [n]$, changing only the i-th coordinate can change the value of f by at most c_i. Then for every $t > 0$,*

$$\Pr[|Z - \mathbb{E}[Z]| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

B Our algorithms for MAX-DICUT

In this section, we present our learning-augmented streaming algorithm for MAX-DICUT, which serves as a warm-up for our algorithms for MAX-kCSPs.

Algorithm 2 A learning-augmented *offline* algorithm for MAX-DICUT estimation in general graphs

Input: Directed graph $G = (V, E)$, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$
Output: Estimate of the MAX-DICUT value of G

$m \leftarrow |E|, \theta \leftarrow \frac{\varepsilon^4 \delta \eta^2}{768}, L \leftarrow \{v \in V : \deg(v) < \theta m\}, H \leftarrow V \setminus L$
initialize $X_0 \leftarrow 0, X_1 \leftarrow 0, X_2 \leftarrow 0$.
initialize $X_{v,L^+} \leftarrow 0, X_{v,L^-} \leftarrow 0, X_{L^+,v} \leftarrow 0, X_{L^-,v} \leftarrow 0$ for each $v \in H$.
for each directed edge $(u, v) \in E$ **do**
    $Y_u = \mathcal{O}(u), Y_v = \mathcal{O}(v)$
    if $u \in L$ and $v \in L$ **then**
        if $Y_u = 1$ and $Y_v = -1$ **then** $X_0 \leftarrow X_0 + 1$
        if $Y_u = Y_v$ **then** $X_1 \leftarrow X_1 + 1$
        if $Y_u = -1$ and $Y_v = 1$ **then** $X_2 \leftarrow X_2 + 1$
    if $u \in H$ and $v \in L$ **then**
        if $Y_v = 1$ **then** $X_{u,L^+} \leftarrow X_{u,L^+} + 1$ **else** $X_{u,L^-} \leftarrow X_{u,L^-} + 1$
    if $u \in L$ and $v \in H$ **then**
        if $Y_u = 1$ **then** $X_{L^+,v} \leftarrow X_{L^+,v} + 1$ **else** $X_{L^-,v} \leftarrow X_{L^-,v} + 1$
$\widetilde{k}_L \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2} + \varepsilon)^2 X_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon) X_1 + (\frac{1}{2} - \varepsilon)^2 X_2]$
for each $v \in H$ **do**
    $\widetilde{k}_{S^*,v} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) X_{L^+,v} + (\frac{1}{2} - \frac{1}{4\varepsilon}) X_{L^-,v}, \widetilde{k}_{v,T^*} \leftarrow (\frac{1}{2} - \frac{1}{4\varepsilon}) X_{v,L^+} + (\frac{1}{2} + \frac{1}{4\varepsilon}) X_{v,L^-}$
return $\widetilde{k}_L + \sum_{v \in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16}$

B.1 Analysis of the offline algorithm

In this subsection, we analyze the offline algorithm described in Section 4.1. The algorithm is formally given in Algorithm 2. The following theorem states its approximation guarantee.

Theorem B.1. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta, \eta \in (0, 1)$, and G be a directed graph with $m = \Omega(\varepsilon^{-26} \delta^{-10} \eta^{-14})$. Given an ε-accurate oracle $\mathcal{O}$ for the* MAX-DICUT *of G, Algorithm 2 provides a $(1 - \eta)$-approximation for the* MAX-DICUT *value of G with probability at least $1 - \delta$.*

B.1.1 Key properties

We first state several key properties of the maintained counters in Algorithm 2.

Let X_0, X_1, and X_2 denote the random variables given by the final values of the corresponding counters at the end of the algorithm. We first prove concentration bounds for these random variables around their expectations.

Lemma B.2. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0, 1)$. For each $i \in \{0, 1, 2\}$, we have $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^2 \eta m}{8}$ with probability at least $1 - \frac{\delta}{6}$.*

Proof. Let $E_L := \{(u, v) \in E : u, v \in L\}$ denote the set of directed edges with both endpoints in L. Let $m_L := |E_L|$. For each edge $(u, v) \in E_L$, define the following indicator random variables:

$$X_{0,uv} = \begin{cases} 1 & \text{if } Y_u = 1 \wedge Y_v = -1, \\ 0 & \text{otherwise,} \end{cases} \quad X_{1,uv} = \begin{cases} 1 & \text{if } Y_u = Y_v, \\ 0 & \text{otherwise,} \end{cases} \quad X_{2,uv} = \begin{cases} 1 & \text{if } Y_u = -1 \wedge Y_v = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that for $i \in \{0, 1, 2\}$, we have $\mathbb{E}[X_{i,uv}] = \Pr[X_{i,uv} = 1]$ and $X_i = \sum_{(u,v) \in E_L} X_{i,uv}$. Therefore,

$$\begin{aligned} \mathbb{E}\big[X_i^2\big] &= \mathbb{E}\left[\left(\sum_{(u,v) \in E_L} X_{i,uv}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{(u,v) \in E_L} \sum_{(w,z) \in E_L} X_{i,uv} X_{i,wz}\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \mathbb{E}[X_{i,uv}X_{i,wz}] \\
&= \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] \\
&= \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}=\emptyset}} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] + \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}\neq\emptyset}} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] \\
&\leq \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}=\emptyset}} \Pr[X_{i,uv}=1]\Pr[X_{i,wz}=1] + \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}\neq\emptyset}} 1 \\
&\leq \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \mathbb{E}[X_{i,uv}]\mathbb{E}[X_{i,wz}] + m_L \cdot 2 \cdot \theta m \\
&\leq (\mathbb{E}[\sum_{(u,v)\in E_L} X_{i,uv}])^2 + 2\theta m^2 \\
&= (\mathbb{E}[X_i])^2 + 2\theta m^2,
\end{aligned}
$$

where the third-to-last step follows from $\deg(v) < \theta m$ for $v \in L$, and the second-to-last step follows from $m_L \leq m$.

Then we have $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \leq 2\theta m^2$. By Chebyshev's inequality,

$$
\begin{aligned}
\Pr\left[|X_i - \mathbb{E}[X_i]| > \frac{\varepsilon^2\eta m}{8}\right] &\leq \frac{\mathrm{Var}[X_i]}{(\varepsilon^2\eta m/8)^2} \leq \frac{128\theta m^2}{\varepsilon^4\eta^2 m^2} && \triangleright\ \mathrm{Var}[X_i] \leq 2\theta m^2 \\
&= \frac{128\theta}{\varepsilon^4\eta^2} < \frac{\delta}{6}. && \triangleright\ \theta < \frac{\varepsilon^4\delta\eta^2}{768}
\end{aligned}
$$

□

Recall that (S^*, T^*) denotes the optimal bipartition corresponding to some fixed MAX-DICUT of G. Define $k_L := |\{(u,v) \in E : u \in S^* \cap L, v \in T^* \cap L\}|$ to be the number of directed edges with both endpoints in L (i.e., both low-degree) that contribute to the MAX-DICUT value.

Recall that $\widetilde{k}_L$, computed at Line 14 of Algorithm 2, is an estimate of k_L. The following lemma shows that $\widetilde{k}_L$ is close to k_L up to an additive error of $\mathrm{poly}(\eta) \cdot m$ with high probability.

Lemma B.3. *Let $\delta, \eta \in (0,1)$. We have $|\widetilde{k}_L - k_L| \leq \frac{\eta m}{32}$ with probability at least $1 - \frac{\delta}{2}$.*

Proof. Let k_0 denote the number of edges from $S^* \cap L$ to $T^* \cap L$. Let k_1 denote the number of edges whose endpoints both lie in $S^* \cap L$ or both lie in $T^* \cap L$. Let k_2 denote the number of edges from $T^* \cap L$ to $S^* \cap L$. Note that $k_0 = k_L$. According to Algorithm 2, we have

$$
\begin{aligned}
\mathbb{E}[X_0] &= \left(\frac{1}{2}+\varepsilon\right)^2 \cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot k_1 + \left(\frac{1}{2}-\varepsilon\right)^2 \cdot k_2, \\
\mathbb{E}[X_1] &= 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot k_0 + \left[\left(\frac{1}{2}+\varepsilon\right)^2 + \left(\frac{1}{2}-\varepsilon\right)^2\right] \cdot k_1 + 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot k_2, \\
\mathbb{E}[X_2] &= \left(\frac{1}{2}-\varepsilon\right)^2 \cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot k_1 + \left(\frac{1}{2}+\varepsilon\right)^2 \cdot k_2.
\end{aligned}
$$

Therefore, by Gaussian elimination, we have

$$
k_L = k_0 = \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2 \cdot \mathbb{E}[X_0] - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot \mathbb{E}[X_1] + \left(\frac{1}{2}-\varepsilon\right)^2 \cdot \mathbb{E}[X_2]\right] \quad (1)
$$

By applying union bound on Lemma B.2, we have that with probability at least $1 - \frac{\delta}{6} \cdot 3 = 1 - \frac{\delta}{2}$, $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^2\eta m}{8}$ for all $i \in \{0,1,2\}$. Therefore,

$$
\widetilde{k}_L = \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2 X_0 - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) X_1 + \left(\frac{1}{2}-\varepsilon\right)^2 X_2\right]
$$

$$
\begin{aligned}
&\leq \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\mathbb{E}[X_0]+\frac{\varepsilon^2\eta m}{8}\right)-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\mathbb{E}[X_1]-\frac{\varepsilon^2\eta m}{8}\right)\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\mathbb{E}[X_2]+\frac{\varepsilon^2\eta m}{8}\right)\right]\\
&= k_L+\frac{\eta m}{32}\left[\left(\frac{1}{2}+\varepsilon\right)^2+\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)+\left(\frac{1}{2}-\varepsilon\right)^2\right]\\
&= k_L+\frac{\eta m}{32}\left(\frac{3}{4}+\varepsilon^2\right)\\
&\leq k_L+\frac{\eta m}{32},
\end{aligned}
$$

where the third-to-last step follows from Equation (1). Similarly, $\widetilde{k}_L \geq k_L - \frac{\eta m}{32}$.

□

For each high-degree vertex $v \in H$, define

1. $k_{S^*,v} := |\{(u,v) \in E : u \in S^* \cap L\}|$ to be the number of directed edges from $S^* \cap L$ to v.
2. $k_{T^*,v} := |\{(u,v) \in E : u \in T^* \cap L\}|$ to be the number of directed edges from $T^* \cap L$ to v.
3. $k_{v,S^*} := |\{(v,u) \in E : u \in S^* \cap L\}|$ to be the number of directed edges from v to $S^* \cap L$.
4. $k_{v,T^*} := |\{(v,u) \in E : u \in T^* \cap L\}|$ to be the number of directed edges from v to $T^* \cap L$.

Recall that $\widetilde{k}_{S^*,v}$ and $\widetilde{k}_{v,T^*}$, computed at Line 16 of Algorithm 2, are estimates of $k_{S^*,v}$ and k_{v,T^*}, respectively. The following lemma shows that $\widetilde{k}_{S^*,v}$ and $\widetilde{k}_{v,T^*}$ are close to $k_{S^*,v}$ and k_{v,T^*}, respectively, up to an additive error of $\mathrm{poly}(\varepsilon,\delta,\eta)\cdot m$ with high probability.

Lemma B.4. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For each vertex $v \in H$, each of the following inequalities holds with probability at least $1 - \frac{\varepsilon^4\delta^2\eta^2}{6144}$:*

(1) $|\widetilde{k}_{S^*,v} - k_{S^*,v}| \leq \frac{\varepsilon^4\delta\eta^3 m}{49152}$.

(2) $|\widetilde{k}_{v,T^*} - k_{v,T^*}| \leq \frac{\varepsilon^4\delta\eta^3 m}{49152}$.

Proof. We prove the first inequality below; the second inequality follows by the same argument.

For each vertex $v \in H$, let $X_{L^+,v}, X_{L^-,v}, X_{v,L^+}$, and X_{v,L^-} denote the random variables given by the final values of the corresponding counters at the end of the algorithm. We have

$$
\begin{aligned}
\mathbb{E}[X_{L^+,v}] &= \left(\frac{1}{2}+\varepsilon\right)\cdot k_{S^*,v} + \left(\frac{1}{2}-\varepsilon\right)\cdot k_{T^*,v},\\
\mathbb{E}[X_{L^-,v}] &= \left(\frac{1}{2}-\varepsilon\right)\cdot k_{S^*,v} + \left(\frac{1}{2}+\varepsilon\right)\cdot k_{T^*,v},\\
\mathbb{E}[X_{v,L^+}] &= \left(\frac{1}{2}+\varepsilon\right)\cdot k_{v,S^*} + \left(\frac{1}{2}-\varepsilon\right)\cdot k_{v,T^*},\\
\mathbb{E}[X_{v,L^-}] &= \left(\frac{1}{2}-\varepsilon\right)\cdot k_{v,S^*} + \left(\frac{1}{2}+\varepsilon\right)\cdot k_{v,T^*}.
\end{aligned}
$$

By solving the above system of linear equations, we obtain

$$
\begin{aligned}
k_{S^*,v} &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\mathbb{E}[X_{L^+,v}] + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\mathbb{E}[X_{L^-,v}],\\
k_{v,T^*} &= \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\mathbb{E}[X_{v,L^+}] + \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\mathbb{E}[X_{v,L^-}].
\end{aligned}
$$

According to Line 16 of Algorithm 2 and by linearity of expectation, we have

$$\begin{aligned}\mathbb{E}[\widetilde{k}_{S^*,v}] &= \mathbb{E}\left[\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)X_{L^+,v}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)X_{L^-,v}\right]\\ &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\mathbb{E}[X_{L^+,v}]+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\mathbb{E}[X_{L^-,v}]\\ &= k_{S^*,v}.\end{aligned}$$

Next, we calculate $\mathrm{Var}[\widetilde{k}_{S^*,v}]$. Note that

$$\begin{aligned}\widetilde{k}_{S^*,v} &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)X_{L^+,v}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)X_{L^-,v}\\ &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)X_{L^+,v}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)(e(L,v)-X_{L^+,v})\\ &= \frac{1}{2\varepsilon}X_{L^+,v}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)e(L,v),\end{aligned}$$

where $e(L,v)$ is the number of edges from L to v. Therefore, $\mathrm{Var}[\widetilde{k}_{S^*,v}] = \frac{1}{4\varepsilon^2}\mathrm{Var}[X_{L^+,v}]$.

Given a high-degree vertex $v \in H$, for each incoming edge $(u,v) \in E$ with $u \in L$, define an indicator random variable X_{uv} as follows:

$$X_{uv} = \begin{cases}1 & \text{if } Y_u = 1,\\ 0 & \text{otherwise.}\end{cases}$$

Note that $X_{L^+,v} = \sum_{(u,v)\in E: u\in L} X_{uv}$ and $\mathbb{E}[X_{uv}] = \Pr[X_{uv}=1]$. Then $\mathrm{Var}[X_{uv}] = \mathbb{E}[X_{uv}^2] - (\mathbb{E}[X_{uv}])^2 = \mathbb{E}[X_{uv}]\cdot(1-\mathbb{E}[X_{uv}]) = (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon) = \frac{1}{4}-\varepsilon^2 \le \frac{1}{4}$. Since X_{uv}'s are independent, we have

$$\mathrm{Var}[X_{L^+,v}] = \sum_{(u,v)\in E: u\in L}\mathrm{Var}[X_{uv}] \le \frac{m}{4}.$$

So, $\mathrm{Var}[\widetilde{k}_{S^*,v}] = \frac{1}{4\varepsilon^2}\mathrm{Var}[X_{L^+,v}] \le \frac{m}{16\varepsilon^2}$. By Chebyshev's inequality,

$$\begin{aligned}\Pr\left[|\widetilde{k}_{S^*,v}-k_{S^*,v}| \ge \frac{\varepsilon^4\delta\eta^3 m}{49152}\right] &= \Pr\left[|\widetilde{k}_{S^*,v}-\mathbb{E}[\widetilde{k}_{S^*,v}]| \ge \frac{\varepsilon^4\delta\eta^3 m}{49152}\right]\\ &\le \frac{\mathrm{Var}[\widetilde{k}_{S^*,v}]}{(\varepsilon^4\delta\eta^3 m/49152)^2} \le \frac{49152^2/16}{\varepsilon^{10}\delta^2\eta^6 m} \le \frac{\varepsilon^4\delta^2\eta^2}{6144},\end{aligned}$$

where the last step follows from our assumption that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. □

Define $k_H := |\{(u,v)\in E : u\in S^*\cap H, v\in T^*\cap H\}|$ to be the number of directed edges with both endpoints in H (i.e., both high-degree) that contribute to the MAX-DICUT value. Note that $|H| \le \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{1536}{\varepsilon^4\delta\eta^2}$. Thus, we have the following observation.

Fact B.5. *It holds that* $k_H < |H|^2 \le \frac{1536^2}{\varepsilon^8\delta^2\eta^4}$.

B.1.2 Proof of Theorem B.1

Proof. Recall that k_L (resp. k_H) is the number of directed edges with both endpoints in L (resp. H) that contribute to the MAX-DICUT value.

Furthermore, define

$$k_{L,H} := |\{(u,v)\in E : u\in S^*\cap L, v\in T^*\cap H\}|$$

to be the number of directed edges from L to H that contribute to the MAX-DICUT value, and

$$k_{H,L} := |\{(u,v)\in E : u\in S^*\cap H, v\in T^*\cap L\}|$$

to be the number of directed edges from H to L that contribute to the MAX-DICUT value.

Therefore, we have $\mathrm{OPT} = k_L + k_H + k_{L,H} + k_{H,L}$. Let ALG denote the output of Algorithm 2. By union bound, with probability at least $1 - \frac{\delta}{2} - \frac{\varepsilon^4\delta^2\eta^2}{6144} \cdot 2 \cdot |H| \geq 1 - \delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{v\in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\leq k_L + \frac{\eta m}{32} + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} + |H| \cdot \frac{\varepsilon^4\delta\eta^3 m}{49152} - \frac{\eta m}{16} && \triangleright \text{ Lemma B.3 and Lemma B.4} \\
&\leq k_L + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} && \triangleright\ |H| \leq \frac{1536}{\varepsilon^4\delta\eta^2} \\
&\leq \mathrm{OPT},
\end{aligned}$$

where the last step follows since $k_L + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\}$ is a lower bound on the value of a feasible dicut of G. Specifically, we fix the restriction on L to be $(S^* \cap L, T^* \cap L)$, and then assign each vertex $v \in H$ to the side that maximizes the resulting dicut value.

Furthermore,

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{v\in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\geq k_L - \frac{\eta m}{32} + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} - |H| \cdot \frac{\varepsilon^4\delta\eta^3 m}{49152} - \frac{\eta m}{16} && \triangleright \text{ Lemma B.3 and Lemma B.4} \\
&\geq k_L - \frac{\eta m}{32} + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} - \frac{\eta m}{32} - \frac{\eta m}{16} && \triangleright\ |H| \leq \frac{1536}{\varepsilon^4\delta\eta^2} \\
&\geq k_L + k_{L,H} + k_{H,L} - \frac{\eta m}{8} && \triangleright \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} \geq k_{L,H} + k_{H,L} \\
&= \mathrm{OPT} - k_H - \frac{\eta m}{8} && \triangleright\ \mathrm{OPT} = k_L + k_H + k_{L,H} + k_{H,L} \\
&\geq \mathrm{OPT} - \frac{1536^2}{\varepsilon^8\delta^2\eta^4} - \frac{\eta}{2} \cdot \mathrm{OPT} && \triangleright \text{ Fact B.5 and } m \leq 4\mathrm{OPT} \\
&\geq (1-\eta) \cdot \mathrm{OPT}, && \triangleright\ \frac{1536^2}{\varepsilon^8\delta^2\eta^4} \leq \frac{\eta m}{8} \leq \frac{\eta}{2} \cdot \mathrm{OPT}
\end{aligned}$$

where the fourth step follows from

$$\begin{aligned}
\sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} &= \sum_{v\in H\cap S^*} \max\{k_{S^*,v}, k_{v,T^*}\} + \sum_{v\in H\cap T^*} \max\{k_{S^*,v}, k_{v,T^*}\} \\
&\geq \sum_{v\in H\cap S^*} k_{v,T^*} + \sum_{v\in H\cap T^*} k_{S^*,v} \\
&= k_{H,L} + k_{L,H}.
\end{aligned}$$

□

Remark B.6. In fact, it suffices to know a highly accurate estimate of ε for Algorithm 2 to work. Specifically, it is enough to have an estimate $\varepsilon' \in (0, 1/2]$ such that $\varepsilon' = (1 \pm O(\eta)) \cdot \varepsilon$. We first update the degree threshold in Line 1 of Algorithm 2 to θm, where $\theta := \frac{\varepsilon'^4\delta\eta^2}{3072}$. Then we replace ε by ε' in the estimators. Namely, in Line 14, we set $\widetilde{k}_L := \frac{1}{4\varepsilon'^2}[(\frac{1}{2}+\varepsilon')^2 X_0 - (\frac{1}{2}+\varepsilon')(\frac{1}{2}-\varepsilon')X_1 + (\frac{1}{2}-\varepsilon')^2 X_2]$; in Line 16, for each $v \in H$, we set $\widetilde{k}_{S^*,v} := (\frac{1}{2} + \frac{1}{4\varepsilon'})X_{L^+,v} + (\frac{1}{2} - \frac{1}{4\varepsilon'})X_{L^-,v}$ and $\widetilde{k}_{v,T^*} := (\frac{1}{2} - \frac{1}{4\varepsilon'})X_{v,L^+} + (\frac{1}{2} + \frac{1}{4\varepsilon'})X_{v,L^-}$. The full modified algorithm and its analysis are given in Appendix E.

B.2 Our algorithm in insertion-only streams

In this section we show how to implement the offline estimator (Algorithm 2) in the streaming model. Suppose the edges of G arrive one by one in an insertion-only stream, in an arbitrary order.

Recall that Algorithm 2 (i) partitions the vertices into low-degree L and high-degree H, (ii) maintains edge-type counts within $G[L]$ induced by the predictions and approximately computes (via a small linear system) an estimate of the optimal dicut contribution from L, and (iii) for each $v \in H$ maintains incidence counts between v and L^+, L^- (in both directions) to optimize v's contribution in the dicut estimation. The main challenge in the streaming setting is that we do *not* know H and L in advance, and we cannot afford to maintain degrees for all vertices explicitly.

Our streaming implementation is built around three ingredients that together let us "simulate" the offline estimator:

1. We maintain a uniform sample F of $\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$ edges using reservoir sampling. At the end of the stream we set $\widetilde{H}$ to be the set of endpoints of edges in F. With high probability, every truly high-degree vertex appears in F (hence $H \subseteq \widetilde{H}$, where H is the set of high degree vertices defined in Algorithm 2), while $|\widetilde{H}| = \mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$ remains small. We then define $\widetilde{L} = V \setminus \widetilde{H}$.
2. Since we cannot store per-vertex incident counts, we maintain four CountMin sketches $\mathbf{CM}^{\mathrm{in}}[V^+]$, $\mathbf{CM}^{\mathrm{out}}[V^+]$, $\mathbf{CM}^{\mathrm{in}}[V^-]$, and $\mathbf{CM}^{\mathrm{out}}[V^-]$. Intuitively, they estimate (for each vertex v) the number of incoming/outgoing edges between v and the predicted sides $V^+ = \{u : \mathcal{O}(u) = +1\}$ and $V^- = \{u : \mathcal{O}(u) = -1\}$. When an edge (u, v) arrives, we query both endpoints and update the appropriate sketches: we increment the estimate of edges from $V^{\mathcal{O}(u)}$ into v, and the estimate of edges from u into $V^{\mathcal{O}(v)}$. This suffices to recover (up to a small additive error) the quantities $e(\widetilde{L}^+, v), e(v, \widetilde{L}^+), e(\widetilde{L}^-, v), e(v, \widetilde{L}^-)$ for each $v \in \widetilde{H}$ at the end of the stream.
3. In parallel, we keep exact counters for $e(V^+, V^-), e(V^-, V^+), e(V^+, V^+), e(V^-, V^-)$ over the entire stream. These provide the global edge-type statistics induced by the predictions.

The algorithm is formally given in Algorithm 3. We have the following theoretical guarantee stating its approximation ratio and space complexity.

Theorem B.7. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta \in (0, 1)$. For any $\eta \in (0, 1)$, given an ε-accurate oracle $\mathcal{O}$ for the* MAX-DICUT *of G, there exists a single-pass $(1 - \eta)$-approximation algorithm for estimating the* MAX-DICUT *value of G in insertion-only streams. The algorithm uses $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6} \log \frac{1}{\varepsilon\delta\eta})$ words of space. The approximation holds with probability at least $1 - \delta$.*

We prove Theorem B.7 by establishing several lemmas that justify the correctness of our streaming simulation of Algorithm 2 using the high-degree threshold θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$, i.e., $H := \{v \in V : \deg(v) \geq \theta m\}$. As we will see, this choice of threshold guarantees that our streaming-constructed partition $(\widetilde{L}, \widetilde{H})$ satisfies the key properties required for the offline analysis, and yields the desired success probability. We begin by showing that reservoir sampling identifies a superset $\widetilde{H}$ of the high-degree vertices H with high probability. Recall that $\widetilde{H}$ is the set of endpoints of all edges in the uniformly sampled edge set F.

Lemma B.8. *Let $\delta \in (0, 1)$. Then $\widetilde{H}$ contains all high-degree vertices H in G (i.e., $H \subseteq \widetilde{H}$) with probability at least $1 - \frac{\delta}{8}$.*

Proof. Let $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$. Consider a vertex $v \in H$, so $\deg(v) \geq \theta m$. For each directed edge e_i incident to v, define an indicator random variable

$$X_i = \begin{cases} 1 & \text{if } e_i \in F, \\ 0 & \text{otherwise.} \end{cases}$$

So, $\mathbb{E}[X_i] = \Pr[X_i = 1] = \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. We define $X := \sum_{i \in [\deg(v)]} X_i$ to denote the number of edges incident to v that are sampled by the end of the stream. Then, $\mathbb{E}[X] = \sum_{i \in [\deg(v)]} \mathbb{E}[X_i] = \frac{\beta \deg(v)}{\varepsilon^8\delta^3\eta^4 m}$.

Next, we bound $\mathrm{Var}[X]$. For any $i \in [\deg(v)]$, $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \mathbb{E}[X_i] - (\mathbb{E}[X_i])^2 \leq \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. Since $(X_i)_{i \in [\deg(v)]}$ are negatively correlated, $\mathrm{Var}[X] \leq \sum_{i \in [\deg(v)]} \mathrm{Var}[X_i] \leq \deg(v) \cdot \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} = \frac{\beta \deg(v)}{\varepsilon^8\delta^3\eta^4 m}$. Therefore, by Chebyshev's inequality,

$$\Pr\left[v \notin \widetilde{H}\right] = \Pr[X \leq 0] = \Pr\left[X \leq \mathbb{E}[X] - \frac{\beta \deg(v)}{\varepsilon^8\delta^3\eta^4 m}\right]$$

Algorithm 3 Estimating the MAX-DICUT value in insertion-only streams

Input: Directed graph G as an insertion-only stream of edges, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$

Output: Estimate of the MAX-DICUT value of G

▷ **Pre-processing phase**

initialize $e(V^+, V^+) \leftarrow 0, e(V^-, V^-) \leftarrow 0, e(V^+, V^-) \leftarrow 0, e(V^-, V^+) \leftarrow 0, m \leftarrow 0$

initialize four CountMin sketches $\mathbf{CM}^{\text{in}}[V^+], \mathbf{CM}^{\text{out}}[V^+], \mathbf{CM}^{\text{in}}[V^-]$ and $\mathbf{CM}^{\text{out}}[V^-]$ with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{28}$.

▷ **Streaming phase**

for each edge (u, v) in the stream **do**

  $m \leftarrow m + 1$

  $Y_u = \mathcal{O}(u), Y_v = \mathcal{O}(v)$

  if $Y_u = 1$ and $Y_v = -1$ **then** $e(V^+, V^-) \leftarrow e(V^+, V^-) + 1$

  if $Y_u = Y_v = 1$ **then** $e(V^+, V^+) \leftarrow e(V^+, V^+) + 1$

  if $Y_u = Y_v = -1$ **then** $e(V^-, V^-) \leftarrow e(V^-, V^-) + 1$

  if $Y_u = -1$ and $Y_v = 1$ **then** $e(V^-, V^+) \leftarrow e(V^-, V^+) + 1$

  UPDATE$(\mathbf{CM}^{\text{in}}[V^{Y_u}][v])$, UPDATE$(\mathbf{CM}^{\text{out}}[V^{Y_v}][u])$

In parallel, uniformly sample $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ edges F in the stream via reservoir sampling.

▷ **Post-processing phase**

for each $v \in V$ **do**

  $\widetilde{e}(V^+, v) \leftarrow$ QUERY$(\mathbf{CM}^{\text{in}}[V^+][v])$, $\widetilde{e}(v, V^+) \leftarrow$ QUERY$(\mathbf{CM}^{\text{out}}[V^+][v])$

  $\widetilde{e}(V^-, v) \leftarrow$ QUERY$(\mathbf{CM}^{\text{in}}[V^-][v])$, $\widetilde{e}(v, V^-) \leftarrow$ QUERY$(\mathbf{CM}^{\text{out}}[V^-][v])$

$\widetilde{H} \leftarrow \bigcup_{(u,v)\in F}\{u, v\}, \widetilde{L} \leftarrow V \setminus \widetilde{H}$

$\widetilde{H}^+ \leftarrow \{v \in \widetilde{H} : Y_v = 1\}, \widetilde{H}^- \leftarrow \widetilde{H} \setminus \widetilde{H}^+$

$\overline{X}_0 \leftarrow e(V^+, V^-) - \sum_{v\in\widetilde{H}^+} \widetilde{e}(v, V^-) - \sum_{v\in\widetilde{H}^-} \widetilde{e}(V^+, v)$

$\overline{X}_1 \leftarrow e(V^+, V^+) + e(V^-, V^-) - \sum_{v\in\widetilde{H}^+}(\widetilde{e}(v, V^+) + \widetilde{e}(V^+, v)) - \sum_{v\in\widetilde{H}^-}(\widetilde{e}(v, V^-) + \widetilde{e}(V^-, v))$

$\overline{X}_2 \leftarrow e(V^-, V^+) - \sum_{v\in\widetilde{H}^+} \widetilde{e}(V^-, v) - \sum_{v\in\widetilde{H}^-} \widetilde{e}(v, V^+)$

$\overline{k}_{\widetilde{L}} \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2}+\varepsilon)^2\overline{X}_0 - (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon)\overline{X}_1 + (\frac{1}{2}-\varepsilon)^2\overline{X}_2]$

for each $v \in \widetilde{H}$ **do**

  $\overline{k}_{S^*,v} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{e}(V^+, v) + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{e}(V^-, v)$

  $\overline{k}_{v,T^*} \leftarrow (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{e}(v, V^+) + (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{e}(v, V^-)$

return $\overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16}$

$$
\begin{aligned}
&\le \frac{\mathrm{Var}[X] \cdot \varepsilon^{16}\delta^6\eta^8 m^2}{\beta^2 \deg^2(v)} \\
&\le \frac{\varepsilon^8\delta^3\eta^4 m}{\beta \deg(v)} && \triangleright\ \mathrm{Var}[X] \le \frac{\beta \deg(v)}{\varepsilon^8\delta^3\eta^4 m} \\
&\le \frac{4096\varepsilon^4\delta^2\eta^2}{\beta}. && \triangleright\ \deg(v) \ge \theta m
\end{aligned}
$$

Recall that $|H| \le \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{8192}{\varepsilon^4\delta\eta^2}$. By union bound, with probability at least $1 - \frac{8192}{\varepsilon^4\delta\eta^2} \cdot \frac{4096\varepsilon^4\delta^2\eta^2}{\beta} = 1 - \frac{\delta}{8}$ (since $\beta = 2^{28}$), $\widetilde{H}$ contains all high-degree vertices H in G. □

Lemma B.9. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0, 1)$. For all vertices $v \in \widetilde{H}$, we have*

(i) $\widetilde{e}(V^+, v) \ge e(V^+, v)$, $\widetilde{e}(v, V^+) \ge e(v, V^+)$, $\widetilde{e}(V^-, v) \ge e(V^-, v)$, $\widetilde{e}(v, V^-) \ge e(v, V^-)$;

(ii) With probability at least $1 - \frac{\delta}{8}$, all of the following inequalities hold: $\widetilde{e}(V^+, v) \le e(V^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^+) \le e(v, V^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(V^-, v) \le e(V^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^-) \le e(v, V^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$.

Proof. (i) By Theorem A.1, for all $v \in V$, we have $\widetilde{e}(V^+, v) \geq e(V^+, v), \widetilde{e}(v, V^+) \geq e(v, V^+), \widetilde{e}(V^-, v) \geq e(V^-, v)$, and $\widetilde{e}(v, V^-) \geq e(v, V^-)$.

(ii) By Theorem A.1, for any $v \in V$, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, each of the following four inequalities holds: $\widetilde{e}(V^+, v) \leq e(V^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^+) \leq e(v, V^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(V^-, v) \leq e(V^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, and $\widetilde{e}(v, V^-) \leq e(v, V^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$. Since $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, by applying the union bound we have that, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta} \cdot \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \cdot 4 \geq 1 - \frac{\delta}{8}$, for all $v \in \widetilde{H}$, all of the above four inequalities hold. □

Proof of Theorem B.7. Given the above lemmas, we are ready to prove Theorem B.7.

Proof. We can assume that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$, as otherwise we can store the whole graph in $O(\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta))$ space.

In Algorithm 3, we maintain four CountMin sketches with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{28}$. By Theorem A.1, this takes $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words of space. Also, we use reservoir sampling to uniformly sample $O(\frac{1}{\varepsilon^8\delta^3\eta^4})$ edges, which takes $O(\frac{1}{\varepsilon^8\delta^3\eta^4})$ words of space. Therefore, the total space complexity of Algorithm 3 is $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words.

Next, we analyze the approximation ratio. Since $V^+ = \widetilde{H}^+ \cup \widetilde{L}^+$ and $V^- = \widetilde{H}^- \cup \widetilde{L}^-$, for any $v \in V$, we have

$$e(V^+, v) = e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v), \qquad e(v, V^+) = e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+),$$
$$e(V^-, v) = e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v), \qquad e(v, V^-) = e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-).$$

By Lemma B.9, with probability at least $1 - \frac{\delta}{8}$, for all vertices $v \in \widetilde{H}$, all of the following inequalities hold:

$$e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v) \leq \widetilde{e}(V^+, v) \leq e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+) \leq \widetilde{e}(v, V^+) \leq e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v) \leq \widetilde{e}(V^-, v) \leq e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-) \leq \widetilde{e}(v, V^-) \leq e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m.$$

Since $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$ and $\widetilde{H}$ is the set of endpoints of edges in F, we have that $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Thus, it holds that $0 \leq e(\widetilde{H}^+, v), e(v, \widetilde{H}^+), e(\widetilde{H}^-, v), e(v, \widetilde{H}^-) \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Therefore,

$$0 \leq \widetilde{e}(V^+, v) - e(\widetilde{L}^+, v) \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{2}$$
$$0 \leq \widetilde{e}(v, V^+) - e(v, \widetilde{L}^+) \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{3}$$
$$0 \leq \widetilde{e}(V^-, v) - e(\widetilde{L}^-, v) \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{4}$$
$$0 \leq \widetilde{e}(v, V^-) - e(v, \widetilde{L}^-) \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m. \tag{5}$$

For each edge with both endpoints in $\widetilde{L}$, let $\widetilde{X}_0, \widetilde{X}_1$, and $\widetilde{X}_2$ denote, respectively, the numbers of edges with (i) $(Y_u, Y_v) = (+1, -1)$, (ii) $Y_u = Y_v$, and (iii) $(Y_u, Y_v) = (-1, +1)$. We have

$$\widetilde{X}_0 = e(\widetilde{L}^+, \widetilde{L}^-) = e(V^+, V^-) - e(\widetilde{H}^+, \widetilde{H}^-) - \sum_{v \in \widetilde{H}^+} e(v, \widetilde{L}^-) - \sum_{v \in \widetilde{H}^-} e(\widetilde{L}^+, v),$$

$$
\begin{aligned}
\widetilde{X}_1 &= e(\widetilde{L}^+, \widetilde{L}^+) + e(\widetilde{L}^-, \widetilde{L}^-) \\
&= e(V^+, V^+) + e(V^-, V^-) - e(\widetilde{H}^+, \widetilde{H}^+) - e(\widetilde{H}^-, \widetilde{H}^-) \\
&\quad - \sum_{v \in \widetilde{H}^+} (e(v, \widetilde{L}^+) + e(\widetilde{L}^+, v)) - \sum_{v \in \widetilde{H}^-} (e(v, \widetilde{L}^-) + e(\widetilde{L}^-, v)), \\
\widetilde{X}_2 &= e(\widetilde{L}^-, \widetilde{L}^+) = e(V^-, V^+) - e(\widetilde{H}^-, \widetilde{H}^+) - \sum_{v \in \widetilde{H}^+} e(\widetilde{L}^-, v) - \sum_{v \in \widetilde{H}^-} e(v, \widetilde{L}^+).
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\overline{X}_0 &= e(V^+, V^-) - \sum_{v \in \widetilde{H}^+} \widetilde{e}(v, V^-) - \sum_{v \in \widetilde{H}^-} \widetilde{e}(V^+, v), \\
\overline{X}_1 &= e(V^+, V^+) + e(V^-, V^-) - \sum_{v \in \widetilde{H}^+} (\widetilde{e}(v, V^+) + \widetilde{e}(V^+, v)) - \sum_{v \in \widetilde{H}^-} (\widetilde{e}(v, V^-) + \widetilde{e}(V^-, v)), \\
\overline{X}_2 &= e(V^-, V^+) - \sum_{v \in \widetilde{H}^+} \widetilde{e}(V^-, v) - \sum_{v \in \widetilde{H}^-} \widetilde{e}(v, V^+).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\overline{X}_0 - \widetilde{X}_0 &= e(\widetilde{H}^+, \widetilde{H}^-) + \sum_{v \in \widetilde{H}^+} (e(v, \widetilde{L}^-) - \widetilde{e}(v, V^-)) + \sum_{v \in \widetilde{H}^-} (e(\widetilde{L}^+, v) - \widetilde{e}(V^+, v)), \\
\overline{X}_1 - \widetilde{X}_1 &= e(\widetilde{H}^+, \widetilde{H}^+) + e(\widetilde{H}^-, \widetilde{H}^-) + \sum_{v \in \widetilde{H}^+} \left[(e(v, \widetilde{L}^+) - \widetilde{e}(v, V^+)) + (e(\widetilde{L}^+, v) - \widetilde{e}(V^+, v)) \right] \\
&\quad + \sum_{v \in \widetilde{H}^-} \left[(e(v, \widetilde{L}^-) - \widetilde{e}(v, V^-)) + (e(\widetilde{L}^-, v) - \widetilde{e}(V^-, v)) \right], \\
\overline{X}_2 - \widetilde{X}_2 &= e(\widetilde{H}^-, \widetilde{H}^+) + \sum_{v \in \widetilde{H}^+} (e(\widetilde{L}^-, v) - \widetilde{e}(V^-, v)) + \sum_{v \in \widetilde{H}^-} (e(v, \widetilde{L}^+) - \widetilde{e}(v, V^+)).
\end{aligned}
$$

Since $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8 \delta^3 \eta^4}$, we have $|\widetilde{H}^+|, |\widetilde{H}^-| \leq \frac{2\beta}{\varepsilon^8 \delta^3 \eta^4}$, and $e(\widetilde{H}^+, \widetilde{H}^-), e(\widetilde{H}^-, \widetilde{H}^+), e(\widetilde{H}^+, \widetilde{H}^+), e(\widetilde{H}^-, \widetilde{H}^-) \leq \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}$. Moreover, combining inequalities (2)–(5), we have

$$
\begin{aligned}
\widetilde{X}_0 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \leq \overline{X}_0 \leq \widetilde{X}_0 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}, \\
\widetilde{X}_1 - \left(\frac{16\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{16} \right) \leq \overline{X}_1 \leq \widetilde{X}_1 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}, \\
\widetilde{X}_2 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \leq \overline{X}_2 \leq \widetilde{X}_2 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}.
\end{aligned}
$$

Let $\widetilde{k}_{\widetilde{L}} := \frac{1}{4\varepsilon^2} [(\frac{1}{2} + \varepsilon)^2 \widetilde{X}_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon) \widetilde{X}_1 + (\frac{1}{2} - \varepsilon)^2 \widetilde{X}_2]$. Now we are ready to show that $\overline{k}_{\widetilde{L}}$ is close to $\widetilde{k}_{\widetilde{L}}$ up to an additive error of $\mathrm{poly}(\eta) \cdot m$ with probability at least $1 - \frac{\delta}{8}$. We have

$$
\begin{aligned}
\overline{k}_{\widetilde{L}} &= \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \overline{X}_0 - \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \overline{X}_1 + \left(\frac{1}{2} - \varepsilon \right)^2 \overline{X}_2 \right] \\
&\geq \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \left(\widetilde{X}_0 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \right) - \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \left(\widetilde{X}_1 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8} \right) \right. \\
&\quad \left. + \left(\frac{1}{2} - \varepsilon \right)^2 \left(\widetilde{X}_2 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \right) \right] \\
&= \widetilde{k}_{\widetilde{L}} - \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) + \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8} \right.
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\frac{1}{2}-\varepsilon\right)^2\left(\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\Bigg]\\
&=\widetilde{k}_{\widetilde{L}}-\left(\frac{5\beta^2}{4\varepsilon^{18}\delta^6\eta^8}+\frac{3\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}+\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\geq\widetilde{k}_{\widetilde{L}}-\frac{\eta m}{32},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}_{\widetilde{L}}&=\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1+\left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right]\\
&\leq\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0+\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1-\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)\right)\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2+\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right]\\
&=\widetilde{k}_{\widetilde{L}}+\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)+\left(\frac{1}{2}-\varepsilon\right)^2\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right]\\
&=\widetilde{k}_{\widetilde{L}}+\left(\frac{3\beta^2}{2\varepsilon^{18}\delta^6\eta^8}-\frac{2\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}-\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\leq\widetilde{k}_{\widetilde{L}}+\frac{\eta m}{64},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$.

For each $v\in\widetilde{H}$, let $\widetilde{k}_{S^*,v}:=(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot e(\widetilde{L}^+,v)+(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot e(\widetilde{L}^-,v)$ and $\widetilde{k}_{v,T^*}:=(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot e(v,\widetilde{L}^+)+(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot e(v,\widetilde{L}^-)$. We have

$$
\begin{aligned}
\overline{k}_{S^*,v}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^-,v)\\
&\geq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot e(\widetilde{L}^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\left(e(\widetilde{L}^-,v)+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}-\left(\frac{1}{4\varepsilon}-\frac{1}{2}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}-\left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4}-\frac{\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^9\delta^5\eta^6 m}{512\beta}-\frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right)\\
&\geq\widetilde{k}_{S^*,v}-\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}_{S^*,v}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^-,v)\\
&\leq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\left(e(\widetilde{L}^+,v)+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot e(\widetilde{L}^-,v)\\
&=\widetilde{k}_{S^*,v}+\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}+\left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4}+\frac{\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^9\delta^5\eta^6 m}{512\beta}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right)\\
&\leq\widetilde{k}_{S^*,v}+\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Therefore, $|\overline{k}_{S^*,v} - \widetilde{k}_{S^*,v}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$. Similarly, $|\overline{k}_{v,T^*} - \widetilde{k}_{v,T^*}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$.

Let $\widetilde{\mathrm{ALG}} := \widetilde{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{32}$ denote the output of the algorithm obtained by running Algorithm 2 with L and H replaced by $\widetilde{L}$ and $\widetilde{H}$, respectively. Analogous to the proof of Theorem B.1, we can show that $(1-\frac{\eta}{2})\cdot\mathrm{OPT} \le \widetilde{\mathrm{ALG}} \le \mathrm{OPT}$ with probability at least $1-\frac{3\delta}{4}$. The argument relies primarily on two properties of the partition $(\widetilde{L},\widetilde{H})$: (i) every vertex in $\widetilde{L}$ has degree at most θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$ (since $\widetilde{L}\subseteq L$), and (ii) $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$.

Let ALG denote the output of Algorithm 3. By applying the union bound on Lemma B.8 and Lemma B.9, with probability at least $1-\frac{\delta}{8}-\frac{\delta}{8}-\frac{3\delta}{4} = 1-\delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\le \widetilde{k}_{\widetilde{L}} + \frac{\eta m}{64} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} + |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\le \widetilde{\mathrm{ALG}} + \frac{\eta m}{32} + \frac{\eta m}{64} + \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\le \mathrm{OPT},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\ge \widetilde{k}_{\widetilde{L}} - \frac{\eta m}{32} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{8} \\
&\ge \left(1-\frac{\eta}{2}\right)\cdot\mathrm{OPT} - \frac{\eta}{2}\cdot\mathrm{OPT} && \triangleright\ m \le 4\cdot\mathrm{OPT} \\
&= (1-\eta)\cdot\mathrm{OPT}.
\end{aligned}$$

□

B.3 Extension to dynamic streams

In this subsection, we present our algorithm for MAX-DICUT in dynamic streams. Similar to [DPV25]'s method, our algorithm in dynamic streams can be easily obtained by modifying our algorithm in insertion-only streams (i.e., Algorithm 3). Specifically, we use ℓ_0-sampling [JST11] to compute a small set $\widetilde{H}$ at the end of the stream such that $\widetilde{H}\supseteq H$. The algorithm is formally given in Algorithm 4.

We have the following theoretical guarantee stating its approximation ratio and space complexity.

Theorem B.10. *Let $\varepsilon\in(0,\frac{1}{2}]$ and $\delta\in(0,1)$. For any $\eta\in(0,1)$, given an ε-accurate oracle $\mathcal{O}$ for the* MAX-DICUT *of G, there exists a single-pass $(1-\eta)$-approximation algorithm for estimating the* MAX-DICUT *value of G in dynamic streams. The algorithm uses $O(\frac{\log^2 n}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ bits of space. The approximation holds with probability at least $1-\delta$.*

B.3.1 Proof of Theorem B.10

The proof follows a similar structure to that of Theorem B.7. First, we show that ℓ_0-sampling in dynamic streams can effectively replace reservoir sampling in insertion-only streams, allowing us to retrieve a set $\widetilde{H}\supseteq H$ with high probability by the end of the stream.

Algorithm 4 Estimating the MAX-DICUT value in dynamic streams

Input: Directed graph G as a dynamic stream of edges, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$

Output: Estimate of the MAX-DICUT value of G

▷ **Pre-processing phase**

initialize $e(V^+, V^+) \leftarrow 0, e(V^-, V^-) \leftarrow 0, e(V^+, V^-) \leftarrow 0, e(V^-, V^+) \leftarrow 0, m \leftarrow 0$

initialize four CountMin sketches $\mathbf{CM}^{\text{in}}[V^+]$, $\mathbf{CM}^{\text{out}}[V^+]$, $\mathbf{CM}^{\text{in}}[V^-]$ and $\mathbf{CM}^{\text{out}}[V^-]$ with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{29} \cdot 25$.

▷ **Streaming phase**

for each item $((u, v), \Delta_{(u,v)})$ in the dynamic stream **do**

  $Y_u = \mathcal{O}(u), Y_v = \mathcal{O}(v)$

  if $\Delta_{(u,v)} = 1$ **then**

    $m \leftarrow m + 1$

    if $Y_u = 1$ and $Y_v = -1$ **then** $e(V^+, V^-) \leftarrow e(V^+, V^-) + 1$

    if $Y_u = Y_v = 1$ **then** $e(V^+, V^+) \leftarrow e(V^+, V^+) + 1$

    if $Y_u = Y_v = -1$ **then** $e(V^-, V^-) \leftarrow e(V^-, V^-) + 1$

    if $Y_u = -1$ and $Y_v = 1$ **then** $e(V^-, V^+) \leftarrow e(V^-, V^+) + 1$

  else

    $m \leftarrow m - 1$

    if $Y_u = 1$ and $Y_v = -1$ **then** $e(V^+, V^-) \leftarrow e(V^+, V^-) - 1$

    if $Y_u = Y_v = 1$ **then** $e(V^+, V^+) \leftarrow e(V^+, V^+) - 1$

    if $Y_u = Y_v = -1$ **then** $e(V^-, V^-) \leftarrow e(V^-, V^-) - 1$

    if $Y_u = -1$ and $Y_v = 1$ **then** $e(V^-, V^+) \leftarrow e(V^-, V^+) - 1$

  UPDATE$(\mathbf{CM}^{\text{in}}[V^{Y_u}][v], \Delta_{(u,v)})$, UPDATE$(\mathbf{CM}^{\text{out}}[V^{Y_v}][u], \Delta_{(u,v)})$

In parallel, maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers (with failure probability $\frac{\varepsilon^8\delta^4\eta^4}{32\beta}$) to obtain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different edges F.

▷ **Post-processing phase**

for each $v \in V$ **do**

  $\widetilde{e}(V^+, v) \leftarrow$ QUERY$(\mathbf{CM}^{\text{in}}[V^+][v])$, $\widetilde{e}(v, V^+) \leftarrow$ QUERY$(\mathbf{CM}^{\text{out}}[V^+][v])$

  $\widetilde{e}(V^-, v) \leftarrow$ QUERY$(\mathbf{CM}^{\text{in}}[V^-][v])$, $\widetilde{e}(v, V^-) \leftarrow$ QUERY$(\mathbf{CM}^{\text{out}}[V^-][v])$

$\widetilde{H} \leftarrow \bigcup_{(u,v)\in F}\{u, v\}, \widetilde{L} \leftarrow V \setminus \widetilde{H}$

$\widetilde{H}^+ \leftarrow \{v \in \widetilde{H} : Y_v = 1\}, \widetilde{H}^- \leftarrow \widetilde{H} \setminus \widetilde{H}^+$

$\overline{X}_0 \leftarrow e(V^+, V^-) - \sum_{v\in\widetilde{H}^+} \widetilde{e}(v, V^-) - \sum_{v\in\widetilde{H}^-} \widetilde{e}(V^+, v)$

$\overline{X}_1 \leftarrow e(V^+, V^+) + e(V^-, V^-) - \sum_{v\in\widetilde{H}^+}(\widetilde{e}(v, V^+) + \widetilde{e}(V^+, v)) - \sum_{v\in\widetilde{H}^-}(\widetilde{e}(v, V^-) + \widetilde{e}(V^-, v))$

$\overline{X}_2 \leftarrow e(V^-, V^+) - \sum_{v\in\widetilde{H}^+} \widetilde{e}(V^-, v) - \sum_{v\in\widetilde{H}^-} \widetilde{e}(v, V^+)$

$\overline{k}_{\widetilde{L}} \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2} + \varepsilon)^2\overline{X}_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon)\overline{X}_1 + (\frac{1}{2} - \varepsilon)^2\overline{X}_2]$

for each $v \in \widetilde{H}$ **do**

  $\overline{k}_{S^*,v} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{e}(V^+, v) + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{e}(V^-, v)$

  $\overline{k}_{v,T^*} \leftarrow (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{e}(v, V^+) + (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{e}(v, V^-)$

return $\overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16}$

Lemma B.11. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0, 1)$. With probability at least $1 - \frac{\delta}{16}$, Line 18 of Algorithm 4 returns a set F which contains $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different edges.*

Proof. The error is introduced by the ℓ_0-samplers and the probability that sampling $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ uniform random edges in parallel does not yield $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different edges.

Since we maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers, by Theorem A.3 and applying the union bound, the failure probability introduced by these ℓ_0-samplers is at most $\frac{\beta}{\varepsilon^8\delta^3\eta^4} \cdot \frac{\varepsilon^8\delta^4\eta^4}{32\beta} = \frac{\delta}{32}$.

Suppose that all the ℓ_0-samplers succeed, i.e., $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$. Now we bound the probability that there exist two identical edges in F. Consider two fixed edges $e_1, e_2 \in F$. Then $\Pr[e_1 = e_2] = \frac{1}{m}$. By union bound, the probability that there exists at least one pair of identical edges is at most $\binom{\beta/\varepsilon^8\delta^3\eta^4}{2} \cdot \frac{1}{m} \le \frac{\beta^2}{\varepsilon^{16}\delta^6\eta^8 m}$.

Therefore, the probability that F contains $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different edges is at least $1 - \frac{\delta}{32} - \frac{\beta^2}{\varepsilon^{16}\delta^6\eta^8 m} \ge 1 - \frac{\delta}{32} - \frac{\delta}{32} = 1 - \frac{\delta}{16}$ (since we can assume $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$). □

Lemma B.12. *Let $\delta \in (0,1)$. Then $\widetilde{H}$ contains all high-degree vertices H in G (i.e., $H \subseteq \widetilde{H}$) with probability at least $1 - \frac{\delta}{16}$.*

Proof. Let $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$. Consider a vertex $v \in H$, so $\deg(v) \ge \theta m$. For each directed edge e_i incident to v, define an indicator random variable

$$X_i = \begin{cases} 1 & \text{if } e_i \in F, \\ 0 & \text{otherwise.} \end{cases}$$

So, $\mathbb{E}[X_i] = \Pr[X_i = 1] = 1 - (1 - \frac{1}{m})^{\beta/\varepsilon^8\delta^3\eta^4} \le \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. On the other hand, $\mathbb{E}[X_i] = 1 - (1 - \frac{1}{m})^{\beta/\varepsilon^8\delta^3\eta^4} \ge 1 - \exp\Big(-\frac{\beta}{\varepsilon^8\delta^3\eta^4 m}\Big) \ge \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} - \frac{\beta^2}{2\varepsilon^{16}\delta^6\eta^8 m^2} \ge \frac{\beta}{5\varepsilon^8\delta^3\eta^4 m}$ since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. We define $X := \sum_{i\in[\deg(v)]} X_i$ to denote the number of edges incident to v that are sampled by the end of the stream. Then, $\mathbb{E}[X] = \sum_{i\in[\deg(v)]} \mathbb{E}[X_i] \in [\frac{\beta\deg(v)}{5\varepsilon^8\delta^3\eta^4 m}, \frac{\beta\deg(v)}{\varepsilon^8\delta^3\eta^4 m}]$.

Next, we bound $\mathrm{Var}[X]$. For any $i \in [\deg(v)]$, $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \mathbb{E}[X_i] - (\mathbb{E}[X_i])^2 \le \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. Since $(X_i)_{i\in[\deg(v)]}$ are negatively correlated, $\mathrm{Var}[X] \le \sum_{i\in[\deg(v)]} \mathrm{Var}[X_i] \le \deg(v) \cdot \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} = \frac{\beta\deg(v)}{\varepsilon^8\delta^3\eta^4 m}$. Therefore, by Chebyshev's inequality,

$$\begin{aligned}
\Pr\Big[v \notin \widetilde{H}\Big] = \Pr[X \le 0] &\le \Pr\left[X \le \mathbb{E}[X] - \frac{\beta\deg(v)}{5\varepsilon^8\delta^3\eta^4 m}\right] \\
&\le \frac{\mathrm{Var}[X] \cdot 25\varepsilon^{16}\delta^6\eta^8 m^2}{\beta^2\deg^2(v)} \\
&\le \frac{25\varepsilon^8\delta^3\eta^4 m}{\beta\deg(v)} && \triangleright\ \mathrm{Var}[X] \le \frac{\beta\deg(v)}{\varepsilon^8\delta^3\eta^4 m} \\
&\le \frac{25\cdot 4096\varepsilon^4\delta^2\eta^2}{\beta}. && \triangleright\ \deg(v) \ge \theta m
\end{aligned}$$

Recall that $|H| \le \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{8192}{\varepsilon^4\delta\eta^2}$. By union bound, with probability at least $1 - \frac{8192}{\varepsilon^4\delta\eta^2} \cdot \frac{25\cdot 4096\varepsilon^4\delta^2\eta^2}{\beta} = 1 - \frac{\delta}{16}$ (since $\beta = 2^{29}\cdot 25$), $\widetilde{H}$ contains all high-degree vertices H in G. □

Next, we show that for all vertices $v \in \widetilde{H}$, our estimates of the number of edges from v to V^+, from v to V^-, from V^+ to v, and from V^- to v have additive error at most $\mathrm{poly}(\varepsilon,\delta,\eta)\cdot m$, with high probability.

Lemma B.13. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For all vertices $v \in \widetilde{H}$, we have*

(i) $\widetilde{e}(V^+, v) \ge e(V^+, v)$, $\widetilde{e}(v, V^+) \ge e(v, V^+)$, $\widetilde{e}(V^-, v) \ge e(V^-, v)$, $\widetilde{e}(v, V^-) \ge e(v, V^-)$;

(ii) With probability at least $1 - \frac{\delta}{8}$, all of the following inequalities hold: $\widetilde{e}(V^+, v) \le e(V^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^+) \le e(v, V^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(V^-, v) \le e(V^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^-) \le e(v, V^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$.

Proof. (i) By Theorem A.1, for all $v \in V$, we have $\widetilde{e}(V^+, v) \ge e(V^+, v), \widetilde{e}(v, V^+) \ge e(v, V^+), \widetilde{e}(V^-, v) \ge e(V^-, v)$, and $\widetilde{e}(v, V^-) \ge e(v, V^-)$.

(ii) By Theorem A.1, for any $v \in V$, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, each of the following four inequalities holds: $\widetilde{e}(V^+, v) \le e(V^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(v, V^+) \le e(v, V^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{e}(V^-, v) \le e(V^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, and $\widetilde{e}(v, V^-) \le e(v, V^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$. Since $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, by applying the union bound we have that, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta} \cdot \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \cdot 4 \ge 1 - \frac{\delta}{8}$, for all $v \in \widetilde{H}$, all of the above four inequalities hold. □

Proof of Theorem B.10. Given the above lemmas, we are ready to prove Theorem B.10.

Proof. We can assume that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$, as otherwise we can store the whole graph in $O(\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta))$ space.

In Algorithm 4, we maintain four CountMin sketches with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{29} \cdot 25$. By Theorem A.1, this takes $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words of space. Also, we maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers, each with failure probability $\frac{\varepsilon^8\delta^4\eta^4}{32\beta}$. By Theorem A.3, this takes $O(\frac{\log^2 n}{\varepsilon^8\delta^3\eta^4}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{\log^2 n}{\varepsilon^8\delta^3\eta^4}\log\frac{1}{\varepsilon\delta\eta})$ bits of space. Therefore, the total space complexity of Algorithm 4 is $O(\frac{\log^2 n}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ bits.

Next, we analyze the approximation ratio. Since $V^+ = \widetilde{H}^+ \cup \widetilde{L}^+$ and $V^- = \widetilde{H}^- \cup \widetilde{L}^-$, for any $v \in V$, we have

$$e(V^+, v) = e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v), \qquad e(v, V^+) = e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+),$$
$$e(V^-, v) = e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v), \qquad e(v, V^-) = e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-).$$

By Lemma B.13, with probability at least $1 - \frac{\delta}{8}$, for all vertices $v \in \widetilde{H}$, all of the following inequalities hold:

$$e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v) \le \widetilde{e}(V^+, v) \le e(\widetilde{H}^+, v) + e(\widetilde{L}^+, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+) \le \widetilde{e}(v, V^+) \le e(v, \widetilde{H}^+) + e(v, \widetilde{L}^+) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v) \le \widetilde{e}(V^-, v) \le e(\widetilde{H}^-, v) + e(\widetilde{L}^-, v) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-) \le \widetilde{e}(v, V^-) \le e(v, \widetilde{H}^-) + e(v, \widetilde{L}^-) + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m.$$

Since $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$ and $\widetilde{H}$ is the set of endpoints of edges in F, we have that $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Thus, it holds that $0 \le e(\widetilde{H}^+, v), e(v, \widetilde{H}^+), e(\widetilde{H}^-, v), e(v, \widetilde{H}^-) \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Therefore,

$$0 \le \widetilde{e}(V^+, v) - e(\widetilde{L}^+, v) \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{6}$$
$$0 \le \widetilde{e}(v, V^+) - e(v, \widetilde{L}^+) \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{7}$$
$$0 \le \widetilde{e}(V^-, v) - e(\widetilde{L}^-, v) \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{8}$$
$$0 \le \widetilde{e}(v, V^-) - e(v, \widetilde{L}^-) \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m. \tag{9}$$

For each edge with both endpoints in $\widetilde{L}$, let $\widetilde{X}_0, \widetilde{X}_1$, and $\widetilde{X}_2$ denote, respectively, the numbers of edges with (i) $(Y_u, Y_v) = (+1, -1)$, (ii) $Y_u = Y_v$, and (iii) $(Y_u, Y_v) = (-1, +1)$. We have

$$\widetilde{X}_0 = e(\widetilde{L}^+, \widetilde{L}^-) = e(V^+, V^-) - e(\widetilde{H}^+, \widetilde{H}^-) - \sum_{v \in \widetilde{H}^+} e(v, \widetilde{L}^-) - \sum_{v \in \widetilde{H}^-} e(\widetilde{L}^+, v),$$

$$
\begin{aligned}
\widetilde{X}_1 &= e(\widetilde{L}^+, \widetilde{L}^+) + e(\widetilde{L}^-, \widetilde{L}^-) \\
&= e(V^+, V^+) + e(V^-, V^-) - e(\widetilde{H}^+, \widetilde{H}^+) - e(\widetilde{H}^-, \widetilde{H}^-) \\
&\quad - \sum_{v \in \widetilde{H}^+} (e(v, \widetilde{L}^+) + e(\widetilde{L}^+, v)) - \sum_{v \in \widetilde{H}^-} (e(v, \widetilde{L}^-) + e(\widetilde{L}^-, v)), \\
\widetilde{X}_2 &= e(\widetilde{L}^-, \widetilde{L}^+) = e(V^-, V^+) - e(\widetilde{H}^-, \widetilde{H}^+) - \sum_{v \in \widetilde{H}^+} e(\widetilde{L}^-, v) - \sum_{v \in \widetilde{H}^-} e(v, \widetilde{L}^+).
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\overline{X}_0 &= e(V^+, V^-) - \sum_{v \in \widetilde{H}^+} \widetilde{e}(v, V^-) - \sum_{v \in \widetilde{H}^-} \widetilde{e}(V^+, v), \\
\overline{X}_1 &= e(V^+, V^+) + e(V^-, V^-) - \sum_{v \in \widetilde{H}^+} (\widetilde{e}(v, V^+) + \widetilde{e}(V^+, v)) - \sum_{v \in \widetilde{H}^-} (\widetilde{e}(v, V^-) + \widetilde{e}(V^-, v)), \\
\overline{X}_2 &= e(V^-, V^+) - \sum_{v \in \widetilde{H}^+} \widetilde{e}(V^-, v) - \sum_{v \in \widetilde{H}^-} \widetilde{e}(v, V^+).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\overline{X}_0 - \widetilde{X}_0 &= e(\widetilde{H}^+, \widetilde{H}^-) + \sum_{v \in \widetilde{H}^+} (e(v, \widetilde{L}^-) - \widetilde{e}(v, V^-)) + \sum_{v \in \widetilde{H}^-} (e(\widetilde{L}^+, v) - \widetilde{e}(V^+, v)), \\
\overline{X}_1 - \widetilde{X}_1 &= e(\widetilde{H}^+, \widetilde{H}^+) + e(\widetilde{H}^-, \widetilde{H}^-) + \sum_{v \in \widetilde{H}^+} \left[(e(v, \widetilde{L}^+) - \widetilde{e}(v, V^+)) + (e(\widetilde{L}^+, v) - \widetilde{e}(V^+, v)) \right] \\
&\quad + \sum_{v \in \widetilde{H}^-} \left[(e(v, \widetilde{L}^-) - \widetilde{e}(v, V^-)) + (e(\widetilde{L}^-, v) - \widetilde{e}(V^-, v)) \right], \\
\overline{X}_2 - \widetilde{X}_2 &= e(\widetilde{H}^-, \widetilde{H}^+) + \sum_{v \in \widetilde{H}^+} (e(\widetilde{L}^-, v) - \widetilde{e}(V^-, v)) + \sum_{v \in \widetilde{H}^-} (e(v, \widetilde{L}^+) - \widetilde{e}(v, V^+)).
\end{aligned}
$$

Since $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8 \delta^3 \eta^4}$, we have $|\widetilde{H}^+|, |\widetilde{H}^-| \le \frac{2\beta}{\varepsilon^8 \delta^3 \eta^4}$, and $e(\widetilde{H}^+, \widetilde{H}^-), e(\widetilde{H}^-, \widetilde{H}^+), e(\widetilde{H}^+, \widetilde{H}^+) + e(\widetilde{H}^-, \widetilde{H}^-) \le \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}$. Moreover, combining inequalities (6)–(9), we have

$$
\begin{aligned}
\widetilde{X}_0 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \le \overline{X}_0 \le \widetilde{X}_0 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}, \\
\widetilde{X}_1 - \left(\frac{16\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{16} \right) \le \overline{X}_1 \le \widetilde{X}_1 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}, \\
\widetilde{X}_2 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \le \overline{X}_2 \le \widetilde{X}_2 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8}.
\end{aligned}
$$

Let $\widetilde{k}_{\widetilde{L}} := \frac{1}{4\varepsilon^2} [(\frac{1}{2} + \varepsilon)^2 \widetilde{X}_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon) \widetilde{X}_1 + (\frac{1}{2} - \varepsilon)^2 \widetilde{X}_2]$. Now we are ready to show that $\overline{k}_{\widetilde{L}}$ is close to $\widetilde{k}_{\widetilde{L}}$ up to an additive error of $\mathrm{poly}(\eta) \cdot m$ with probability at least $1 - \frac{\delta}{8}$. We have

$$
\begin{aligned}
\overline{k}_{\widetilde{L}} &= \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \overline{X}_0 - \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \overline{X}_1 + \left(\frac{1}{2} - \varepsilon \right)^2 \overline{X}_2 \right] \\
&\ge \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \left(\widetilde{X}_0 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \right) - \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \left(\widetilde{X}_1 + \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8} \right) \right. \\
&\quad \left. + \left(\frac{1}{2} - \varepsilon \right)^2 \left(\widetilde{X}_2 - \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) \right) \right] \\
&= \widetilde{k}_{\widetilde{L}} - \frac{1}{4\varepsilon^2} \left[\left(\frac{1}{2} + \varepsilon \right)^2 \left(\frac{8\beta^2}{\varepsilon^{16} \delta^6 \eta^8} + \frac{\varepsilon^2 \delta^2 \eta^2 m}{32} \right) + \left(\frac{1}{2} + \varepsilon \right) \left(\frac{1}{2} - \varepsilon \right) \frac{4\beta^2}{\varepsilon^{16} \delta^6 \eta^8} \right.
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\frac{1}{2}-\varepsilon\right)^2\left(\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\Bigg]\\
&=\widetilde{k}_{\widetilde{L}}-\left(\frac{5\beta^2}{4\varepsilon^{18}\delta^6\eta^8}+\frac{3\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}+\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\geq\widetilde{k}_{\widetilde{L}}-\frac{\eta m}{32},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}_{\widetilde{L}}&=\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1+\left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right]\\
&\leq\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0+\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1-\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)\right)\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2+\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right]\\
&=\widetilde{k}_{\widetilde{L}}+\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)+\left(\frac{1}{2}-\varepsilon\right)^2\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right]\\
&=\widetilde{k}_{\widetilde{L}}+\left(\frac{3\beta^2}{2\varepsilon^{18}\delta^6\eta^8}-\frac{2\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}-\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\leq\widetilde{k}_{\widetilde{L}}+\frac{\eta m}{64},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$.

For each $v\in\widetilde{H}$, let $\widetilde{k}_{S^*,v}:=(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot e(\widetilde{L}^+,v)+(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot e(\widetilde{L}^-,v)$ and $\widetilde{k}_{v,T^*}:=(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot e(v,\widetilde{L}^+)+(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot e(v,\widetilde{L}^-)$. We have

$$
\begin{aligned}
\overline{k}_{S^*,v}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^-,v)\\
&\geq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot e(\widetilde{L}^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\left(e(\widetilde{L}^-,v)+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}-\left(\frac{1}{4\varepsilon}-\frac{1}{2}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}-\left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4}-\frac{\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^9\delta^5\eta^6 m}{512\beta}-\frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right)\\
&\geq\widetilde{k}_{S^*,v}-\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}_{S^*,v}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^+,v)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{e}(V^-,v)\\
&\leq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\left(e(\widetilde{L}^+,v)+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot e(\widetilde{L}^-,v)\\
&=\widetilde{k}_{S^*,v}+\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}_{S^*,v}+\left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4}+\frac{\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^9\delta^5\eta^6 m}{512\beta}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right)\\
&\leq\widetilde{k}_{S^*,v}+\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Therefore, $|\overline{k}_{S^*,v} - \widetilde{k}_{S^*,v}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$. Similarly, $|\overline{k}_{v,T^*} - \widetilde{k}_{v,T^*}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$.

Let $\widetilde{\mathrm{ALG}} := \widetilde{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{32}$ denote the output of the algorithm obtained by running Algorithm 2 with L and H replaced by $\widetilde{L}$ and $\widetilde{H}$, respectively. Analogous to the proof of Theorem B.1, we can show that $(1-\frac{\eta}{2})\cdot\mathrm{OPT} \le \widetilde{\mathrm{ALG}} \le \mathrm{OPT}$ with probability at least $1-\frac{3\delta}{4}$. The argument relies primarily on two properties of the partition $(\widetilde{L},\widetilde{H})$: (i) every vertex in $\widetilde{L}$ has degree at most θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$ (since $\widetilde{L}\subseteq L$), and (ii) $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$.

Let ALG denote the output of Algorithm 4. By applying the union bound on Lemma B.11, Lemma B.12 and Lemma B.13, with probability at least $1-\frac{\delta}{16}-\frac{\delta}{16}-\frac{\delta}{8}-\frac{3\delta}{4} = 1-\delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\le \widetilde{k}_{\widetilde{L}} + \frac{\eta m}{64} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} + |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\le \widetilde{\mathrm{ALG}} + \frac{\eta m}{32} + \frac{\eta m}{64} + \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\le \mathrm{OPT},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{v\in\widetilde{H}} \max\{\overline{k}_{S^*,v}, \overline{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\ge \widetilde{k}_{\widetilde{L}} - \frac{\eta m}{32} + \sum_{v\in\widetilde{H}} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{8} \\
&\ge \left(1-\frac{\eta}{2}\right)\cdot\mathrm{OPT} - \frac{\eta}{2}\cdot\mathrm{OPT} && \triangleright\ m \le 4\mathrm{OPT} \\
&= (1-\eta)\cdot\mathrm{OPT}.
\end{aligned}$$

□

C Our algorithms for MAX-kAND

In this section, we extend our approach to MAX-kAND. In Section C.1, we provide the remaining proofs and streaming implementations for MAX-2AND. For MAX-kAND, Section C.2 gives a bounded-degree warm-up, and Section C.3 gives the unbounded-degree algorithm used in the main theorem. Here the notion of variable degree is analogous to vertex degree in graphs: for each $i \in [n]$, we (slightly abusing graph terminology) let $\deg(i)$ denote the number of constraints in which x_i appears, either positively as x_i or negatively as $-x_i$.

C.1 $k = 2$

C.1.1 Deferred proofs from Section 4.2

Proof of Lemma 4.2. Let $\Psi_L := \{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in L\}$ denote the set of constraints whose associated variables are both in L. Let $m_L := |\Psi_L|$. For each constraint $((j_1,j_2),(b_1,b_2)) \in \Psi_L$, define the following indicator random variables:

$$X_{0,j_1j_2} = \begin{cases} 1 & \text{if } b_1Y_{j_1} = b_2Y_{j_2} = 1, \\ 0 & \text{otherwise,} \end{cases} \quad X_{1,j_1j_2} = \begin{cases} 1 & \text{if } b_1Y_{j_1} \ne b_2Y_{j_2}, \\ 0 & \text{otherwise,} \end{cases} \quad X_{2,j_1j_2} = \begin{cases} 1 & \text{if } b_1Y_{j_1} = b_2Y_{j_2} = -1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that for $i \in \{0,1,2\}$, we have $\mathbb{E}[X_{i,j_1j_2}] = \Pr[X_{i,j_1j_2}=1]$ and $X_i = \sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L} X_{i,j_1j_2}$. Therefore,

$$\begin{aligned}
\mathbb{E}[X_i^2] &= \mathbb{E}[(\sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L} X_{i,j_1j_2})^2] \\
&= \mathbb{E}[\sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L}\ \sum_{((j_1',j_2'),(b_1',b_2'))\in\Psi_L} X_{i,j_1j_2}X_{i,j_1'j_2'}] \\
&= \sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L}\ \sum_{((j_1',j_2'),(b_1',b_2'))\in\Psi_L} \mathbb{E}[X_{i,j_1j_2}X_{i,j_1'j_2'}] \\
&= \sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L}\ \sum_{((j_1',j_2'),(b_1',b_2'))\in\Psi_L} \Pr\big[X_{i,j_1j_2}=1 \wedge X_{i,j_1'j_2'}=1\big] \\
&= \sum_{\substack{((j_1,j_2),(b_1,b_2))\in\Psi_L,\\((j_1',j_2'),(b_1',b_2'))\in\Psi_L,\\ \{j_1,j_2\}\cap\{j_1',j_2'\}=\emptyset}} \Pr\big[X_{i,j_1j_2}=1 \wedge X_{i,j_1'j_2'}=1\big] + \sum_{\substack{((j_1,j_2),(b_1,b_2))\in\Psi_L,\\((j_1',j_2'),(b_1',b_2'))\in\Psi_L,\\ \{j_1,j_2\}\cap\{j_1',j_2'\}\neq\emptyset}} \Pr\big[X_{i,j_1j_2}=1 \wedge X_{i,j_1'j_2'}=1\big] \\
&\le \sum_{\substack{((j_1,j_2),(b_1,b_2))\in\Psi_L,\\((j_1',j_2'),(b_1',b_2'))\in\Psi_L,\\ \{j_1,j_2\}\cap\{j_1',j_2'\}=\emptyset}} \Pr[X_{i,j_1j_2}=1]\Pr\big[X_{i,j_1'j_2'}=1\big] + \sum_{\substack{((j_1,j_2),(b_1,b_2))\in\Psi_L,\\((j_1',j_2'),(b_1',b_2'))\in\Psi_L,\\ \{j_1,j_2\}\cap\{j_1',j_2'\}\neq\emptyset}} 1 \\
&\le \sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L}\ \sum_{((j_1',j_2'),(b_1',b_2'))\in\Psi_L} \mathbb{E}[X_{i,j_1j_2}]\mathbb{E}[X_{i,j_1'j_2'}] + m_L\cdot 2\cdot\theta m \\
&\le (\mathbb{E}[\sum_{((j_1,j_2),(b_1,b_2))\in\Psi_L} X_{i,j_1j_2}])^2 + 2\theta m^2 \\
&= (\mathbb{E}[X_i])^2 + 2\theta m^2,
\end{aligned}$$

where the third-to-last step follows from $\deg(i) < \theta m$ for $i \in L$, and the second-to-last step follows from $m_L \le m$.

Then we have $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \le 2\theta m^2$. By Chebyshev's inequality,

$$\begin{aligned}
\Pr\left[|X_i - \mathbb{E}[X_i]| > \frac{\varepsilon^2\eta m}{8}\right] &\le \frac{\mathrm{Var}[X_i]}{(\varepsilon^2\eta m/8)^2} \le \frac{128\theta m^2}{\varepsilon^4\eta^2 m^2} && \triangleright\ \mathrm{Var}[X_i] \le 2\theta m^2 \\
&= \frac{128\theta}{\varepsilon^4\eta^2} < \frac{\delta}{6}. && \triangleright\ \theta < \frac{\varepsilon^4\delta\eta^2}{768}
\end{aligned}$$

□

Proof of Lemma 4.3. Let $k_0 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in L,\ b_1\sigma^*_{j_1} = b_2\sigma^*_{j_2} = 1\}|$ denote the number of constraints in Ψ whose associated variables are both in L and for which both literals are satisfied by $\boldsymbol{\sigma}^*$. Note that $k_0 = k_L$.

Let $k_1 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in L,\ b_1\sigma^*_{j_1} \neq b_2\sigma^*_{j_2}\}|$ denote the number of constraints in Ψ whose associated variables are both in L and for which exactly one literal is satisfied by $\boldsymbol{\sigma}^*$.

Let $k_2 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in L,\ b_1\sigma^*_{j_1} = b_2\sigma^*_{j_2} = -1\}|$ denote the number of constraints in Ψ whose associated variables are both in L and for which neither literal is satisfied by $\boldsymbol{\sigma}^*$.

According to Algorithm 1, we have

$$\mathbb{E}[X_0] = \left(\frac{1}{2}+\varepsilon\right)^2\cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_1 + \left(\frac{1}{2}-\varepsilon\right)^2\cdot k_2,$$

$$\mathbb{E}[X_1] = 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_0 + \left[\left(\frac{1}{2}+\varepsilon\right)^2 + \left(\frac{1}{2}-\varepsilon\right)^2\right]\cdot k_1 + 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_2,$$

$$\mathbb{E}[X_2] = \left(\frac{1}{2}-\varepsilon\right)^2 \cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot k_1 + \left(\frac{1}{2}+\varepsilon\right)^2 \cdot k_2.$$

Therefore, by Gaussian elimination, we have

$$k_L = k_0 = \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2 \cdot \mathbb{E}[X_0] - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) \cdot \mathbb{E}[X_1] + \left(\frac{1}{2}-\varepsilon\right)^2 \cdot \mathbb{E}[X_2]\right] \quad (10)$$

By applying union bound on Lemma 4.2, we have that with probability at least $1 - \frac{\delta}{6} \cdot 3 = 1 - \frac{\delta}{2}$, $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^2 \eta m}{8}$ for all $i \in \{0,1,2\}$. Therefore,

$$\begin{aligned}
\widetilde{k}_L &= \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2 X_0 - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) X_1 + \left(\frac{1}{2}-\varepsilon\right)^2 X_2\right] \\
&\leq \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2 \left(\mathbb{E}[X_0] + \frac{\varepsilon^2\eta m}{8}\right) - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\mathbb{E}[X_1] - \frac{\varepsilon^2\eta m}{8}\right)\right. \\
&\quad \left. + \left(\frac{1}{2}-\varepsilon\right)^2 \left(\mathbb{E}[X_2] + \frac{\varepsilon^2\eta m}{8}\right)\right] \\
&= k_L + \frac{\eta m}{32}\left[\left(\frac{1}{2}+\varepsilon\right)^2 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) + \left(\frac{1}{2}-\varepsilon\right)^2\right] \\
&= k_L + \frac{\eta m}{32}\left(\frac{3}{4}+\varepsilon^2\right) \\
&\leq k_L + \frac{\eta m}{32},
\end{aligned}$$

where the third-to-last step follows from Equation (10). Similarly, $\widetilde{k}_L \geq k_L - \frac{\eta m}{32}$. □

Proof of Lemma 4.4. We prove the first inequality below; the second inequality follows by the same argument.

For each $i \in H$, let $X^{\text{pos}}_{i,L(+)}, X^{\text{pos}}_{i,L(-)}, X^{\text{neg}}_{i,L(+)}$, and $X^{\text{neg}}_{i,L(-)}$ denote the random variables given by the final values of the corresponding counters at the end of the algorithm. We have

$$\begin{aligned}
\mathbb{E}[X^{\text{pos}}_{i,L(+)}] &= \left(\frac{1}{2}+\varepsilon\right) \cdot k^{\text{pos}}_{i,L(+)} + \left(\frac{1}{2}-\varepsilon\right) \cdot k^{\text{pos}}_{i,L(-)}, \\
\mathbb{E}[X^{\text{pos}}_{i,L(-)}] &= \left(\frac{1}{2}-\varepsilon\right) \cdot k^{\text{pos}}_{i,L(+)} + \left(\frac{1}{2}+\varepsilon\right) \cdot k^{\text{pos}}_{i,L(-)}, \\
\mathbb{E}[X^{\text{neg}}_{i,L(+)}] &= \left(\frac{1}{2}+\varepsilon\right) \cdot k^{\text{neg}}_{i,L(+)} + \left(\frac{1}{2}-\varepsilon\right) \cdot k^{\text{neg}}_{i,L(-)}, \\
\mathbb{E}[X^{\text{neg}}_{i,L(-)}] &= \left(\frac{1}{2}-\varepsilon\right) \cdot k^{\text{neg}}_{i,L(+)} + \left(\frac{1}{2}+\varepsilon\right) \cdot k^{\text{neg}}_{i,L(-)}.
\end{aligned}$$

By solving the above system of linear equations, we obtain

$$\begin{aligned}
k^{\text{pos}}_{i,L(+)} &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) \cdot \mathbb{E}[X^{\text{pos}}_{i,L(+)}] + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) \cdot \mathbb{E}[X^{\text{pos}}_{i,L(-)}], \\
k^{\text{neg}}_{i,L(+)} &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) \cdot \mathbb{E}[X^{\text{neg}}_{i,L(+)}] + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) \cdot \mathbb{E}[X^{\text{neg}}_{i,L(-)}].
\end{aligned}$$

According to Line 22 of Algorithm 1 and by linearity of expectation, we have

$$\begin{aligned}
\mathbb{E}[\widetilde{k}^{\text{pos}}_{i,L(+)}] &= \mathbb{E}\left[\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) X^{\text{pos}}_{i,L(+)} + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) X^{\text{pos}}_{i,L(-)}\right] \\
&= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) \mathbb{E}[X^{\text{pos}}_{i,L(+)}] + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) \mathbb{E}[X^{\text{pos}}_{i,L(-)}]
\end{aligned}$$

$$= k^{\mathrm{pos}}_{i,L(+)}.$$

Next, we calculate $\mathrm{Var}[\widetilde{k}^{\mathrm{pos}}_{i,L(+)}]$. Note that

$$\begin{aligned}
\widetilde{k}^{\mathrm{pos}}_{i,L(+)} &= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) X^{\mathrm{pos}}_{i,L(+)} + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) X^{\mathrm{pos}}_{i,L(-)} \\
&= \left(\frac{1}{2}+\frac{1}{4\varepsilon}\right) X^{\mathrm{pos}}_{i,L(+)} + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) (c^{\mathrm{pos}}_{i,L} - X^{\mathrm{pos}}_{i,L(+)}) \\
&= \frac{1}{2\varepsilon} X^{\mathrm{pos}}_{i,L(+)} + \left(\frac{1}{2}-\frac{1}{4\varepsilon}\right) c^{\mathrm{pos}}_{i,L},
\end{aligned}$$

where $c^{\mathrm{pos}}_{i,L} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in L, b_i = 1\}|$ denotes the number of constraints involving x_i and some low-degree variable x_j such that the literal of x_i appears positively. Therefore, $\mathrm{Var}[\widetilde{k}^{\mathrm{pos}}_{i,L(+)}] = \frac{1}{4\varepsilon^2}\mathrm{Var}[X^{\mathrm{pos}}_{i,L(+)}]$.

Given $i \in H$, for each constraint $((i,j),(b_i,b_j)) \in \Psi$ with $j \in L$ and $b_i = 1$, define an indicator random variable X_{ij} as follows:

$$X_{ij} = \begin{cases} 1 & \text{if } b_j Y_j = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $X^{\mathrm{pos}}_{i,L(+)} = \sum_{((i,j),(b_i,b_j))\in\Psi : j\in L, b_i=1} X_{ij}$ and $\mathbb{E}[X_{ij}] = \Pr[X_{ij}=1]$. Then $\mathrm{Var}[X_{ij}] = \mathbb{E}[X_{ij}^2] - (\mathbb{E}[X_{ij}])^2 = \mathbb{E}[X_{ij}]\cdot(1-\mathbb{E}[X_{ij}]) = (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon) = \frac{1}{4}-\varepsilon^2 \le \frac{1}{4}$. Since X_{ij}'s are negatively correlated, we have

$$\mathrm{Var}[X^{\mathrm{pos}}_{i,L(+)}] \le \sum_{\substack{((i,j),(b_i,b_j))\in\Psi: \\ j\in L, b_i=1}} \mathrm{Var}[X_{ij}] \le \frac{m}{4}.$$

So, $\mathrm{Var}[\widetilde{k}^{\mathrm{pos}}_{i,L(+)}] = \frac{1}{4\varepsilon^2}\mathrm{Var}[X^{\mathrm{pos}}_{i,L(+)}] \le \frac{m}{16\varepsilon^2}$. By Chebyshev's inequality,

$$\begin{aligned}
\Pr\left[|\widetilde{k}^{\mathrm{pos}}_{i,L(+)} - k^{\mathrm{pos}}_{i,L(+)}| \ge \frac{\varepsilon^4\delta\eta^3 m}{49152}\right] &= \Pr\left[|\widetilde{k}^{\mathrm{pos}}_{i,L(+)} - \mathbb{E}[\widetilde{k}^{\mathrm{pos}}_{i,L(+)}]| \ge \frac{\varepsilon^4\delta\eta^3 m}{49152}\right] \\
&\le \frac{\mathrm{Var}[\widetilde{k}^{\mathrm{pos}}_{i,L(+)}]}{(\varepsilon^4\delta\eta^3 m/49152)^2} \le \frac{49152^2/16}{\varepsilon^{10}\delta^2\eta^6 m} \le \frac{\varepsilon^4\delta^2\eta^2}{6144},
\end{aligned}$$

where the last step follows from our assumption that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. □

C.1.2 Our algorithm in insertion-only streams

In this section we show how to implement the offline estimator (Algorithm 1) for MAX-2AND in the streaming model. Suppose the constraints arrive one by one in an insertion-only stream, in an arbitrary order.

Recall that Algorithm 1 (i) partitions the variables into low-degree L and high-degree H, (ii) maintains constraint-type counts within L induced by the predictions and approximately compute (via a small linear system) an estimate of the optimal contribution from L, and (iii) for each $i \in H$, maintains incidence counts for constraints involving i and variables in L, which are used to assign i that maximizes i's estimated contribution. The main challenge in the streaming setting is that we do *not* know H and L in advance, and we cannot afford to maintain degrees for all variables explicitly.

Our streaming implementation is built around three ingredients that together let us "simulate" the offline estimator:

1. We maintain a uniform sample F of $\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$ constraints using reservoir sampling. At the end of the stream we set $\widetilde{H}$ to be the set of associated variables of constraints in F. With high probability, every truly high-degree variable appears in F (hence $H \subseteq \widetilde{H}$), while $|\widetilde{H}| = \mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta)$ remains small. We then define $\widetilde{L} = V \setminus \widetilde{H}$.

Algorithm 5 Estimating the MAX-2AND value in insertion-only streams

Input: MAX-2AND instance Ψ as an insertion-only stream of constraints, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$

Output: Estimate of the maximum value of Ψ

▷ **Pre-processing phase**

initialize $X_{+,+} \leftarrow 0, X_{+,-} \leftarrow 0, X_{-,-} \leftarrow 0, m \leftarrow 0$

initialize four CountMin sketches $\mathbf{CM}^{\mathrm{pos}}[+], \mathbf{CM}^{\mathrm{neg}}[+], \mathbf{CM}^{\mathrm{pos}}[-]$ and $\mathbf{CM}^{\mathrm{neg}}[-]$ with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{28}$.

▷ **Streaming phase**

for each constraint $((j_1, j_2), (b_1, b_2))$ in the stream **do**

  $m \leftarrow m + 1,\ Y_{j_1} = \mathcal{O}(j_1), Y_{j_2} = \mathcal{O}(j_2)$

  if $b_1 Y_{j_1} = b_2 Y_{j_2} = 1$ **then** $X_{+,+} \leftarrow X_{+,+} + 1$

  if $b_1 Y_{j_1} \neq b_2 Y_{j_2}$ **then** $X_{+,-} \leftarrow X_{+,-} + 1$

  if $b_1 Y_{j_1} = b_2 Y_{j_2} = -1$ **then** $X_{-,-} \leftarrow X_{-,-} + 1$

  if $b_1 = 1$ **then** UPDATE$(\mathbf{CM}^{\mathrm{pos}}[b_2 Y_{j_2}][j_1])$ **else** UPDATE$(\mathbf{CM}^{\mathrm{neg}}[b_2 Y_{j_2}][j_1])$

  if $b_2 = 1$ **then** UPDATE$(\mathbf{CM}^{\mathrm{pos}}[b_1 Y_{j_1}][j_2])$ **else** UPDATE$(\mathbf{CM}^{\mathrm{neg}}[b_1 Y_{j_1}][j_2])$

In parallel, uniformly sample $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ constraints F in the stream via reservoir sampling.

▷ **Post-processing phase**

for each $i \in [n]$ **do**

  $\widetilde{X}^{\mathrm{pos}}_{i,+} \leftarrow$ QUERY$(\mathbf{CM}^{\mathrm{pos}}[+][i]), \quad \widetilde{X}^{\mathrm{neg}}_{i,+} \leftarrow$ QUERY$(\mathbf{CM}^{\mathrm{neg}}[+][i])$

  $\widetilde{X}^{\mathrm{pos}}_{i,-} \leftarrow$ QUERY$(\mathbf{CM}^{\mathrm{pos}}[-][i]), \quad \widetilde{X}^{\mathrm{neg}}_{i,-} \leftarrow$ QUERY$(\mathbf{CM}^{\mathrm{neg}}[-][i])$

$\widetilde{H} \leftarrow \bigcup_{((j_1,j_2),(b_1,b_2))\in F}\{j_1, j_2\}, \widetilde{L} \leftarrow [n] \setminus \widetilde{H}$

$\widetilde{H}^+ \leftarrow \{i \in \widetilde{H} : Y_i = 1\}, \widetilde{H}^- \leftarrow \widetilde{H} \setminus \widetilde{H}^+$

$\overline{X}_0 \leftarrow X_{+,+} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{pos}}_{i,+} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{neg}}_{i,+}$

$\overline{X}_1 \leftarrow X_{+,-} - \sum_{i\in\widetilde{H}^+} (\widetilde{X}^{\mathrm{pos}}_{i,-} + \widetilde{X}^{\mathrm{neg}}_{i,+}) - \sum_{i\in\widetilde{H}^-} (\widetilde{X}^{\mathrm{pos}}_{i,+} + \widetilde{X}^{\mathrm{neg}}_{i,-})$

$\overline{X}_2 \leftarrow X_{-,-} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{neg}}_{i,-} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{pos}}_{i,-}$

$\overline{k}_{\widetilde{L}} \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2}+\varepsilon)^2\overline{X}_0 - (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon)\overline{X}_1 + (\frac{1}{2}-\varepsilon)^2\overline{X}_2]$

for each $i \in \widetilde{H}$ **do**

  $\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} \leftarrow (\frac{1}{2}+\frac{1}{4\varepsilon})\cdot\widetilde{X}^{\mathrm{pos}}_{i,+} + (\frac{1}{2}-\frac{1}{4\varepsilon})\cdot\widetilde{X}^{\mathrm{pos}}_{i,-}, \quad \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)} \leftarrow (\frac{1}{2}+\frac{1}{4\varepsilon})\cdot\widetilde{X}^{\mathrm{neg}}_{i,+} + (\frac{1}{2}-\frac{1}{4\varepsilon})\cdot\widetilde{X}^{\mathrm{neg}}_{i,-}$

return $\overline{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16}$

2. Since we cannot store per-variable incident counts, we maintain four CountMin sketches $\mathbf{CM}^{\mathrm{pos}}[+]$, $\mathbf{CM}^{\mathrm{neg}}[+]$, $\mathbf{CM}^{\mathrm{pos}}[-]$, and $\mathbf{CM}^{\mathrm{neg}}[-]$. Intuitively, they estimate (for each variable i) the numbers of constraints in which i appears positively or negatively and the other literal is satisfied or not satisfied by the predicted assignment. When a constraint $((j_1, j_2), (b_1, b_2))$ arrives, we query both variables and update the appropriate sketches: we update $\mathbf{CM}^{\mathrm{pos}}[b_2\mathcal{O}(j_2)][j_1]$ if $b_1 = 1$ and $\mathbf{CM}^{\mathrm{neg}}[b_2\mathcal{O}(j_2)][j_1]$ if $b_1 = -1$; symmetrically, we update $\mathbf{CM}^{\mathrm{pos}}[b_1\mathcal{O}(j_1)][j_2]$ if $b_2 = 1$ and $\mathbf{CM}^{\mathrm{neg}}[b_1\mathcal{O}(j_1)][j_2]$ if $b_2 = -1$. This suffices to recover (up to a small additive error) the quantities $X^{\mathrm{pos}}_{i,\widetilde{L}(+)}, X^{\mathrm{neg}}_{i,\widetilde{L}(+)}, X^{\mathrm{pos}}_{i,\widetilde{L}(-)}, X^{\mathrm{neg}}_{i,\widetilde{L}(-)}$ for each $i \in \widetilde{H}$ at the end of the stream.

3. In parallel, we maintain exact counters over the entire stream for (i) $X_{+,+}$: the number of such constraints for which both literals are satisfied by the predicted assignment, (ii) $X_{+,-}$: the number for which exactly one literal is satisfied, and (iii) $X_{-,-}$: the number for which neither literal is satisfied. These provide the global constraint-type statistics induced by the predictions.

The algorithm is formally described in Algorithm 5. We provide the following theoretical guarantee on its approximation ratio and space complexity.

Theorem C.1. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta \in (0, 1)$. For any $\eta \in (0, 1)$, given an ε-accurate oracle $\mathcal{O}$ for a* MAX-2AND *instance Ψ, there exists a single-pass $(1-\eta)$-approximation algorithm for estimating the maximum value of Ψ in insertion-only streams. The algorithm uses $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words of space. The approximation holds with probability at least $1-\delta$.*

We prove Theorem C.1 by establishing several lemmas that justify the correctness of our streaming simulation of Algorithm 1 using the high-degree threshold θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$, i.e., $H := \{i \in [n] : \deg(i) \geq \theta m\}$. As we will see, this choice of threshold guarantees that our streaming-constructed partition $(\widetilde{L}, \widetilde{H})$ satisfies the key properties required for the offline analysis, and yields the desired success probability. We begin by showing that reservoir sampling identifies a superset $\widetilde{H}$ of the high-degree variables H with high probability. Recall that $\widetilde{H}$ is the set of associated variables of all constraints in the uniformly sampled constraint set F.

Lemma C.2. *Let $\delta \in (0,1)$. Then $\widetilde{H}$ contains all high-degree variables H in Ψ (i.e., $H \subseteq \widetilde{H}$) with probability at least $1 - \frac{\delta}{8}$.*

Proof. Let $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$. Consider any $i \in H$. Then, by definition of H, we have $\deg(i) \geq \theta m$. For each constraint C_j involving variable x_i, define an indicator random variable

$$X_j = \begin{cases} 1 & \text{if } C_j \in F, \\ 0 & \text{otherwise.} \end{cases}$$

So, $\mathbb{E}[X_j] = \Pr[X_j = 1] = \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. We define $X := \sum_{j\in[\deg(i)]} X_j$ to denote the number of constraints involving variable x_i that are sampled by the end of the stream. Then, $\mathbb{E}[X] = \sum_{j\in[\deg(i)]} \mathbb{E}[X_j] = \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m}$.

Next, we bound $\mathrm{Var}[X]$. For any $j \in [\deg(i)]$, $\mathrm{Var}[X_j] = \mathbb{E}[X_j^2] - (\mathbb{E}[X_j])^2 = \mathbb{E}[X_j] - (\mathbb{E}[X_j])^2 \leq \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. Since $(X_j)_{j\in[\deg(i)]}$ are negatively correlated, $\mathrm{Var}[X] \leq \sum_{j\in[\deg(i)]} \mathrm{Var}[X_j] \leq \deg(i) \cdot \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} = \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m}$. Therefore, by Chebyshev's inequality,

$$\begin{aligned}
\Pr\left[i \notin \widetilde{H}\right] = \Pr[X \leq 0] &= \Pr\left[X \leq \mathbb{E}[X] - \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m}\right] \\
&\leq \frac{\mathrm{Var}[X] \cdot \varepsilon^{16}\delta^6\eta^8 m^2}{\beta^2\deg^2(i)} \\
&\leq \frac{\varepsilon^8\delta^3\eta^4 m}{\beta\deg(i)} && \triangleright\ \mathrm{Var}[X] \leq \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m} \\
&\leq \frac{4096\varepsilon^4\delta^2\eta^2}{\beta}. && \triangleright\ \deg(i) \geq \theta m
\end{aligned}$$

Recall that $|H| \leq \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{8192}{\varepsilon^4\delta\eta^2}$. By union bound, with probability at least $1 - \frac{8192}{\varepsilon^4\delta\eta^2} \cdot \frac{4096\varepsilon^4\delta^2\eta^2}{\beta} = 1 - \frac{\delta}{8}$ (since $\beta = 2^{28}$), $\widetilde{H}$ contains all high-degree variables H in Ψ. □

For each $i \in [n]$, define

1. $X_{i,+}^{\mathrm{pos}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = 1, b_j Y_j = 1\}|$ to be the number of constraints involving x_i such that the literal of x_i appears positively (i.e., $b_i = 1$) and the other literal is satisfied by the prediction $\mathbf{Y}$ (i.e., $b_j Y_j = 1$). The other quantities are defined analogously:
2. $X_{i,-}^{\mathrm{pos}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = 1, b_j Y_j = -1\}|$.
3. $X_{i,+}^{\mathrm{neg}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = -1, b_j Y_j = 1\}|$.
4. $X_{i,-}^{\mathrm{neg}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = -1, b_j Y_j = -1\}|$.

Lemma C.3. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For all $i \in \widetilde{H}$, we have*

(i) $\widetilde{X}_{i,+}^{\mathrm{pos}} \geq X_{i,+}^{\mathrm{pos}}$, $\widetilde{X}_{i,-}^{\mathrm{pos}} \geq X_{i,-}^{\mathrm{pos}}$, $\widetilde{X}_{i,+}^{\mathrm{neg}} \geq X_{i,+}^{\mathrm{neg}}$, $\widetilde{X}_{i,-}^{\mathrm{neg}} \geq X_{i,-}^{\mathrm{neg}}$;

(ii) With probability at least $1 - \frac{\delta}{8}$, all of the following inequalities hold: $\widetilde{X}_{i,+}^{\mathrm{pos}} \leq X_{i,+}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{X}_{i,-}^{\mathrm{pos}} \leq X_{i,-}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{X}_{i,+}^{\mathrm{neg}} \leq X_{i,+}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{X}_{i,-}^{\mathrm{neg}} \leq X_{i,-}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$.

Proof. (i) By Theorem A.1, for all $i \in [n]$, we have $\widetilde{X}^{\mathrm{pos}}_{i,+} \geq X^{\mathrm{pos}}_{i,+}$, $\widetilde{X}^{\mathrm{pos}}_{i,-} \geq X^{\mathrm{pos}}_{i,-}$, $\widetilde{X}^{\mathrm{neg}}_{i,+} \geq X^{\mathrm{neg}}_{i,+}$, and $\widetilde{X}^{\mathrm{neg}}_{i,-} \geq X^{\mathrm{neg}}_{i,-}$.

(ii) By Theorem A.1, for any $i \in [n]$, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, each of the following four inequalities holds: $\widetilde{X}^{\mathrm{pos}}_{i,+} \leq X^{\mathrm{pos}}_{i,+} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{X}^{\mathrm{pos}}_{i,-} \leq X^{\mathrm{pos}}_{i,-} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, $\widetilde{X}^{\mathrm{neg}}_{i,+} \leq X^{\mathrm{neg}}_{i,+} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, and $\widetilde{X}^{\mathrm{neg}}_{i,-} \leq X^{\mathrm{neg}}_{i,-} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$. Since $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, by applying the union bound we have that, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta} \cdot \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \cdot 4 \geq 1 - \frac{\delta}{8}$, for all $i \in \widetilde{H}$, all of the above four inequalities hold. □

Proof of Theorem C.1. Given the above lemmas, we are ready to prove Theorem C.1.

Proof. We can assume that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$, as otherwise we can store the whole instance in $O(\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta))$ space.

In Algorithm 5, we maintain four CountMin sketches with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{28}$. By Theorem A.1, this takes $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words of space. Also, we use reservoir sampling to uniformly sample $O(\frac{1}{\varepsilon^8\delta^3\eta^4})$ constraints, which takes $O(\frac{1}{\varepsilon^8\delta^3\eta^4})$ words of space. Therefore, the total space complexity of Algorithm 5 is $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words.

Next, we analyze the approximation ratio. Since $[n] = \widetilde{H} \cup \widetilde{L}$, we have

1. $X^{\mathrm{pos}}_{i,+} = X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)}$, where $X^{\mathrm{pos}}_{i,\widetilde{H}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = 1, b_jY_j = 1\}|$ and $X^{\mathrm{pos}}_{i,\widetilde{L}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = 1, b_jY_j = 1\}|$.
2. $X^{\mathrm{pos}}_{i,-} = X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)}$, where $X^{\mathrm{pos}}_{i,\widetilde{H}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = 1, b_jY_j = -1\}|$ and $X^{\mathrm{pos}}_{i,\widetilde{L}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = 1, b_jY_j = -1\}|$.
3. $X^{\mathrm{neg}}_{i,+} = X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)}$, where $X^{\mathrm{neg}}_{i,\widetilde{H}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = -1, b_jY_j = 1\}|$ and $X^{\mathrm{neg}}_{i,\widetilde{L}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = -1, b_jY_j = 1\}|$.
4. $X^{\mathrm{neg}}_{i,-} = X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)}$, where $X^{\mathrm{neg}}_{i,\widetilde{H}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = -1, b_jY_j = -1\}|$ and $X^{\mathrm{neg}}_{i,\widetilde{L}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = -1, b_jY_j = -1\}|$.

By Lemma C.3, with probability at least $1 - \frac{\delta}{8}$, for all $i \in \widetilde{H}$, all of the following inequalities hold:

$$X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)} \leq \widetilde{X}^{\mathrm{pos}}_{i,+} \leq X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)} \leq \widetilde{X}^{\mathrm{pos}}_{i,-} \leq X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)} \leq \widetilde{X}^{\mathrm{neg}}_{i,+} \leq X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,$$
$$X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)} \leq \widetilde{X}^{\mathrm{neg}}_{i,-} \leq X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m.$$

Since $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$ and $\widetilde{H}$ is the set of associated variables of constraints in F, we have that $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Thus, it holds that $0 \leq X^{\mathrm{pos}}_{i,\widetilde{H}(+)}, X^{\mathrm{pos}}_{i,\widetilde{H}(-)}, X^{\mathrm{neg}}_{i,\widetilde{H}(+)}, X^{\mathrm{neg}}_{i,\widetilde{H}(-)} \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Therefore,

$$0 \leq \widetilde{X}^{\mathrm{pos}}_{i,+} - X^{\mathrm{pos}}_{i,\widetilde{L}(+)} \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{11}$$

$$0 \leq \widetilde{X}^{\mathrm{pos}}_{i,-} - X^{\mathrm{pos}}_{i,\widetilde{L}(-)} \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{12}$$

$$0 \leq \widetilde{X}^{\mathrm{neg}}_{i,+} - X^{\mathrm{neg}}_{i,\widetilde{L}(+)} \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m, \tag{13}$$

$$0 \leq \widetilde{X}^{\mathrm{neg}}_{i,-} - X^{\mathrm{neg}}_{i,\widetilde{L}(-)} \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m. \tag{14}$$

We define

1. $\widetilde{X}_0 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{L},\ b_1Y_{j_1} = b_2Y_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which both literals are satisfied by the predictions $\mathbf{Y}$.
2. $\widetilde{X}_1 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{L},\ b_1Y_{j_1} \neq b_2Y_{j_2}\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which exactly one literal is satisfied by $\mathbf{Y}$.
3. $\widetilde{X}_2 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{L},\ b_1Y_{j_1} = b_2Y_{j_2} = -1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which neither literal is satisfied by $\mathbf{Y}$.

Similarly, we define

1. $\widetilde{X}_{0,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1Y_{j_1} = b_2Y_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which both literals are satisfied by $\mathbf{Y}$.
2. $\widetilde{X}_{1,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1Y_{j_1} \neq b_2Y_{j_2}\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which exactly one literal is satisfied by $\mathbf{Y}$.
3. $\widetilde{X}_{2,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1Y_{j_1} = b_2Y_{j_2} = -1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which neither literal is satisfied by $\mathbf{Y}$.

We have

$$\begin{aligned}
\widetilde{X}_0 &= X_{+,+} - \widetilde{X}_{0,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \sum_{i\in\widetilde{H}^-} X^{\mathrm{neg}}_{i,\widetilde{L}(+)},\\
\widetilde{X}_1 &= X_{+,-} - \widetilde{X}_{1,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} (X^{\mathrm{pos}}_{i,\widetilde{L}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)}) - \sum_{i\in\widetilde{H}^-} (X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)}),\\
\widetilde{X}_2 &= X_{-,-} - \widetilde{X}_{2,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \sum_{i\in\widetilde{H}^-} X^{\mathrm{pos}}_{i,\widetilde{L}(-)}.
\end{aligned}$$

Recall that

$$\begin{aligned}
\overline{X}_0 &= X_{+,+} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{pos}}_{i,+} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{neg}}_{i,+},\\
\overline{X}_1 &= X_{+,-} - \sum_{i\in\widetilde{H}^+} (\widetilde{X}^{\mathrm{pos}}_{i,-} + \widetilde{X}^{\mathrm{neg}}_{i,+}) - \sum_{i\in\widetilde{H}^-} (\widetilde{X}^{\mathrm{pos}}_{i,+} + \widetilde{X}^{\mathrm{neg}}_{i,-}),\\
\overline{X}_2 &= X_{-,-} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{neg}}_{i,-} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{pos}}_{i,-}.
\end{aligned}$$

Therefore,

$$\overline{X}_0 - \widetilde{X}_0 = \widetilde{X}_{0,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} (X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{pos}}_{i,+}) + \sum_{i\in\widetilde{H}^-} (X^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{neg}}_{i,+}),$$
$$\overline{X}_1 - \widetilde{X}_1 = \widetilde{X}_{1,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} \left[(X^{\mathrm{pos}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{pos}}_{i,-}) + (X^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{neg}}_{i,+})\right] + \sum_{i\in\widetilde{H}^-} \left[(X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{pos}}_{i,+}) + (X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{neg}}_{i,-})\right],$$
$$\overline{X}_2 - \widetilde{X}_2 = \widetilde{X}_{2,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} (X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{neg}}_{i,-}) + \sum_{i\in\widetilde{H}^-} (X^{\mathrm{pos}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{pos}}_{i,-}).$$

Since $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, we have $|\widetilde{H}^+|, |\widetilde{H}^-| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, and $\widetilde{X}_{0,\widetilde{H}}, \widetilde{X}_{1,\widetilde{H}}, \widetilde{X}_{2,\widetilde{H}} \le \binom{|\widetilde{H}|}{2}\cdot 2^2 \le \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}$. Moreover, combining inequalities (11)–(14), we have

$$\widetilde{X}_0 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right) \le \overline{X}_0 \le \widetilde{X}_0 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8},$$
$$\widetilde{X}_1 - \left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{16}\right) \le \overline{X}_1 \le \widetilde{X}_1 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8},$$
$$\widetilde{X}_2 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right) \le \overline{X}_2 \le \widetilde{X}_2 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}.$$

Let $\widetilde{k}_{\widetilde{L}} := \frac{1}{4\varepsilon^2}[(\frac{1}{2}+\varepsilon)^2\widetilde{X}_0 - (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon)\widetilde{X}_1 + (\frac{1}{2}-\varepsilon)^2\widetilde{X}_2]$. Now we are ready to show that $\overline{k}_{\widetilde{L}}$ is close to $\widetilde{k}_{\widetilde{L}}$ up to an additive error of $\mathrm{poly}(\eta)\cdot m$ with probability at least $1-\frac{\delta}{8}$. We have

$$\begin{aligned}
\overline{k}_{\widetilde{L}} &= \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0 - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1 + \left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right] \\
&\ge \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right) - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right. \\
&\quad \left. + \left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right)\right] \\
&= \widetilde{k}_{\widetilde{L}} - \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right) + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right. \\
&\quad \left. + \left(\frac{1}{2}-\varepsilon\right)^2\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right] \\
&= \widetilde{k}_{\widetilde{L}} - \left(\frac{5\beta^2}{2\varepsilon^{18}\delta^6\eta^8} + \frac{6\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\delta^2\eta^2 m}{256} + \frac{\varepsilon^2\delta^2\eta^2 m}{64}\right) \\
&\ge \widetilde{k}_{\widetilde{L}} - \frac{\eta m}{32},
\end{aligned}$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$\begin{aligned}
\overline{k}_{\widetilde{L}} &= \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0 - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1 + \left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right] \\
&\le \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right) - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1 - \left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)\right)\right. \\
&\quad \left. + \left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right] \\
&= \widetilde{k}_{\widetilde{L}} + \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{16}\right) + \left(\frac{1}{2}-\varepsilon\right)^2\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right]
\end{aligned}$$

$$= \widetilde{k}_{\widetilde{L}} + \left(\frac{3\beta^2}{\varepsilon^{18}\delta^6\eta^8} - \frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\delta^2\eta^2 m}{256} - \frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)$$
$$\leq \widetilde{k}_{\widetilde{L}} + \frac{\eta m}{64},$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$.

For each $i \in \widetilde{H}$, let $\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} := (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot X^{\mathrm{pos}}_{i,\widetilde{L}(-)}$ and $\widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)} := (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot X^{\mathrm{neg}}_{i,\widetilde{L}(+)} + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot X^{\mathrm{neg}}_{i,\widetilde{L}(-)}$. We have

$$\begin{aligned}
\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} &= \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \cdot \widetilde{X}^{\mathrm{pos}}_{i,+} + \left(\frac{1}{2} - \frac{1}{4\varepsilon}\right) \cdot \widetilde{X}^{\mathrm{pos}}_{i,-} \\
&\geq \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \cdot X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \left(\frac{1}{2} - \frac{1}{4\varepsilon}\right) \left(X^{\mathrm{pos}}_{i,\widetilde{L}(-)} + \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right) \\
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \left(\frac{1}{4\varepsilon} - \frac{1}{2}\right) \left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right) \\
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4} - \frac{\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^9\delta^5\eta^6 m}{512\beta} - \frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right) \\
&\geq \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$\begin{aligned}
\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} &= \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \cdot \widetilde{X}^{\mathrm{pos}}_{i,+} + \left(\frac{1}{2} - \frac{1}{4\varepsilon}\right) \cdot \widetilde{X}^{\mathrm{pos}}_{i,-} \\
&\leq \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \left(X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right) + \left(\frac{1}{2} - \frac{1}{4\varepsilon}\right) \cdot X^{\mathrm{pos}}_{i,\widetilde{L}(-)} \\
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right) \\
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4} + \frac{\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^9\delta^5\eta^6 m}{512\beta} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right) \\
&\leq \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Therefore, $|\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}| \leq \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$. Similarly, $|\overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}| \leq \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$.

Let $\widetilde{\mathrm{ALG}} := \widetilde{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{32}$ denote the output of the algorithm obtained by running Algorithm 1 with L and H replaced by $\widetilde{L}$ and $\widetilde{H}$, respectively. Analogous to the proof of Theorem 4.1, we can show that $(1 - \frac{\eta}{2}) \cdot \mathrm{OPT} \leq \widetilde{\mathrm{ALG}} \leq \mathrm{OPT}$ with probability at least $1 - \frac{3\delta}{4}$. The argument relies primarily on two properties of the partition $(\widetilde{L}, \widetilde{H})$: (i) every variable in $\widetilde{L}$ has degree at most θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$ (since $\widetilde{L} \subseteq L$), and (ii) $|\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$.

Let ALG denote the output of Algorithm 5. By applying the union bound on Lemma C.2 and Lemma C.3, with probability at least $1 - \frac{\delta}{8} - \frac{\delta}{8} - \frac{3\delta}{4} = 1 - \delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16} \\
&\leq \widetilde{k}_{\widetilde{L}} + \frac{\eta m}{64} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} + |\widetilde{H}| \cdot \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\leq \widetilde{\mathrm{ALG}} + \frac{\eta m}{32} + \frac{\eta m}{64} + \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4}
\end{aligned}$$

$$\leq \mathrm{OPT},$$

and

$$\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16} \\
&\geq \widetilde{k}_{\widetilde{L}} - \frac{\eta m}{32} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - |\widetilde{H}| \cdot \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\geq \widetilde{\mathrm{ALG}} - \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \leq \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\geq \widetilde{\mathrm{ALG}} - \frac{\eta m}{8} \\
&\geq \left(1-\frac{\eta}{2}\right)\cdot \mathrm{OPT} - \frac{\eta}{2}\cdot \mathrm{OPT} && \triangleright\ m \leq 4\mathrm{OPT} \\
&= (1-\eta)\cdot \mathrm{OPT}.
\end{aligned}$$

□

C.1.3 Extension to dynamic streams

Similar to Algorithm 4, our algorithm for MAX-2AND in dynamic streams can be easily obtained by modifying our algorithm in insertion-only streams (i.e., Algorithm 5). Specifically, we use ℓ_0-sampling [JST11] to compute a small set $\widetilde{H}$ at the end of the stream such that $\widetilde{H} \supseteq H$.

The algorithm is formally given in Algorithm 6. We have the following theoretical guarantee stating its approximation ratio and space complexity.

Theorem C.4. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta \in (0,1)$. For any $\eta \in (0,1)$, given an ε-accurate oracle $\mathcal{O}$ for a* MAX-2AND *instance Ψ, there exists a single-pass $(1-\eta)$-approximation algorithm for estimating the maximum value of Ψ in dynamic streams. The algorithm uses $O(\frac{\log^2 n}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ bits of space. The approximation holds with probability at least $1-\delta$.*

Lemma C.5. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. With probability at least $1-\frac{\delta}{16}$, Line 23 of Algorithm 6 returns a set F which contains $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different constraints.*

Proof. The error is introduced by the ℓ_0-samplers and the probability that sampling $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ uniform random constraints in parallel does not yield $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different constraints.

Since we maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers, by Theorem A.3 and applying the union bound, the failure probability introduced by these ℓ_0-samplers is at most $\frac{\beta}{\varepsilon^8\delta^3\eta^4}\cdot\frac{\varepsilon^8\delta^4\eta^4}{32\beta} = \frac{\delta}{32}$.

Suppose that all the ℓ_0-samplers succeed, i.e., $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$. Now we bound the probability that there exist two identical constraints in F. Consider two fixed constraints $C_1, C_2 \in F$. Then $\Pr[C_1 = C_2] = \frac{1}{m}$. By union bound, the probability that there exists at least one pair of identical constraints is at most $\binom{\beta/\varepsilon^8\delta^3\eta^4}{2}\cdot\frac{1}{m} \leq \frac{\beta^2}{\varepsilon^{16}\delta^6\eta^8 m}$.

Therefore, the probability that F contains $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different constraints is at least $1-\frac{\delta}{32}-\frac{\beta^2}{\varepsilon^{16}\delta^6\eta^8 m} \geq 1-\frac{\delta}{32}-\frac{\delta}{32} = 1-\frac{\delta}{16}$ (since we can assume $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$). □

Lemma C.6. *Let $\delta \in (0,1)$. Then $\widetilde{H}$ contains all high-degree variables H in Ψ (i.e., $H \subseteq \widetilde{H}$) with probability at least $1-\frac{\delta}{16}$.*

Proof. Let $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$. Consider any $i \in H$. Then, by definition of H, we have $\deg(i) \geq \theta m$. For each constraint C_j involving variable x_i, define an indicator random variable

$$X_j = \begin{cases} 1 & \text{if } C_j \in F, \\ 0 & \text{otherwise.} \end{cases}$$

Algorithm 6 Estimating the MAX-2AND value in dynamic streams

Input: MAX-2AND instance Ψ as a dynamic stream of constraints, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, $\delta, \eta \in (0, 1)$

Output: Estimate of the maximum value of Ψ

▷ **Pre-processing phase**

initialize $X_{+,+} \leftarrow 0, X_{+,-} \leftarrow 0, X_{-,-} \leftarrow 0, m \leftarrow 0$

initialize four CountMin sketches $\mathbf{CM}^{\text{pos}}[+], \mathbf{CM}^{\text{neg}}[+], \mathbf{CM}^{\text{pos}}[-]$ and $\mathbf{CM}^{\text{neg}}[-]$ with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{29} \cdot 25$.

▷ **Streaming phase**

for each item $((\mathbf{j} = (j_1, j_2), \mathbf{b} = (b_1, b_2)), \Delta_{(\mathbf{j},\mathbf{b})})$ in the dynamic stream **do**

  $Y_{j_1} = \mathcal{O}(j_1), Y_{j_2} = \mathcal{O}(j_2)$

  if $\Delta_{(\mathbf{j},\mathbf{b})} = 1$ **then**

    $m \leftarrow m + 1$

    if $b_1 Y_{j_1} = b_2 Y_{j_2} = 1$ **then** $X_{+,+} \leftarrow X_{+,+} + 1$

    if $b_1 Y_{j_1} \neq b_2 Y_{j_2}$ **then** $X_{+,-} \leftarrow X_{+,-} + 1$

    if $b_1 Y_{j_1} = b_2 Y_{j_2} = -1$ **then** $X_{-,-} \leftarrow X_{-,-} + 1$

  else

    $m \leftarrow m - 1$

    if $b_1 Y_{j_1} = b_2 Y_{j_2} = 1$ **then** $X_{+,+} \leftarrow X_{+,+} - 1$

    if $b_1 Y_{j_1} \neq b_2 Y_{j_2}$ **then** $X_{+,-} \leftarrow X_{+,-} - 1$

    if $b_1 Y_{j_1} = b_2 Y_{j_2} = -1$ **then** $X_{-,-} \leftarrow X_{-,-} - 1$

  if $b_1 = 1$ **then**

    UPDATE$(\mathbf{CM}^{\text{pos}}[b_2 Y_{j_2}][j_1], \Delta_{(\mathbf{j},\mathbf{b})})$

  else

    UPDATE$(\mathbf{CM}^{\text{neg}}[b_2 Y_{j_2}][j_1], \Delta_{(\mathbf{j},\mathbf{b})})$

  if $b_2 = 1$ **then**

    UPDATE$(\mathbf{CM}^{\text{pos}}[b_1 Y_{j_1}][j_2], \Delta_{(\mathbf{j},\mathbf{b})})$

  else

    UPDATE$(\mathbf{CM}^{\text{neg}}[b_1 Y_{j_1}][j_2], \Delta_{(\mathbf{j},\mathbf{b})})$

In parallel, maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers (with failure probability $\frac{\varepsilon^8\delta^4\eta^4}{32\beta}$) to obtain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ different constraints F.

▷ **Post-processing phase**

for each $i \in [n]$ **do**

  $\widetilde{X}^{\text{pos}}_{i,+} \leftarrow \text{QUERY}(\mathbf{CM}^{\text{pos}}[+][i]), \quad \widetilde{X}^{\text{neg}}_{i,+} \leftarrow \text{QUERY}(\mathbf{CM}^{\text{neg}}[+][i])$

  $\widetilde{X}^{\text{pos}}_{i,-} \leftarrow \text{QUERY}(\mathbf{CM}^{\text{pos}}[-][i]), \quad \widetilde{X}^{\text{neg}}_{i,-} \leftarrow \text{QUERY}(\mathbf{CM}^{\text{neg}}[-][i])$

$\widetilde{H} \leftarrow \bigcup_{((j_1,j_2),(b_1,b_2)) \in F} \{j_1, j_2\}, \widetilde{L} \leftarrow [n] \setminus \widetilde{H}$

$\widetilde{H}^+ \leftarrow \{i \in \widetilde{H} : Y_i = 1\}, \widetilde{H}^- \leftarrow \widetilde{H} \setminus \widetilde{H}^+$

$\overline{X}_0 \leftarrow X_{+,+} - \sum_{i \in \widetilde{H}^+} \widetilde{X}^{\text{pos}}_{i,+} - \sum_{i \in \widetilde{H}^-} \widetilde{X}^{\text{neg}}_{i,+}$

$\overline{X}_1 \leftarrow X_{+,-} - \sum_{i \in \widetilde{H}^+} (\widetilde{X}^{\text{pos}}_{i,-} + \widetilde{X}^{\text{neg}}_{i,+}) - \sum_{i \in \widetilde{H}^-} (\widetilde{X}^{\text{pos}}_{i,+} + \widetilde{X}^{\text{neg}}_{i,-})$

$\overline{X}_2 \leftarrow X_{-,-} - \sum_{i \in \widetilde{H}^+} \widetilde{X}^{\text{neg}}_{i,-} - \sum_{i \in \widetilde{H}^-} \widetilde{X}^{\text{pos}}_{i,-}$

$\overline{k}_{\widetilde{L}} \leftarrow \frac{1}{4\varepsilon^2}[(\frac{1}{2} + \varepsilon)^2 \overline{X}_0 - (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon)\overline{X}_1 + (\frac{1}{2} - \varepsilon)^2 \overline{X}_2]$

for each $i \in \widetilde{H}$ **do**

  $\overline{k}^{\text{pos}}_{i,\widetilde{L}(+)} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{X}^{\text{pos}}_{i,+} + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{X}^{\text{pos}}_{i,-}, \quad \overline{k}^{\text{neg}}_{i,\widetilde{L}(+)} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon}) \cdot \widetilde{X}^{\text{neg}}_{i,+} + (\frac{1}{2} - \frac{1}{4\varepsilon}) \cdot \widetilde{X}^{\text{neg}}_{i,-}$

return $\overline{k}_{\widetilde{L}} + \sum_{i \in \widetilde{H}} \max\{\overline{k}^{\text{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\text{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16}$

So, $\mathbb{E}[X_j] = \Pr[X_j = 1] = 1 - (1 - \frac{1}{m})^{\beta/\varepsilon^8\delta^3\eta^4} \leq \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. On the other hand, $\mathbb{E}[X_j] = 1 - (1 - \frac{1}{m})^{\beta/\varepsilon^8\delta^3\eta^4} \geq 1 - \exp\left(-\frac{\beta}{\varepsilon^8\delta^3\eta^4 m}\right) \geq \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} - \frac{\beta^2}{2\varepsilon^{16}\delta^6\eta^8 m^2} \geq \frac{\beta}{5\varepsilon^8\delta^3\eta^4 m}$ since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. We define $X := \sum_{j \in [\deg(i)]} X_j$ to denote the number of constraints involving variable x_i that are sampled by the end of the stream. Then, $\mathbb{E}[X] = \sum_{j \in [\deg(i)]} \mathbb{E}[X_j] \in [\frac{\beta \deg(i)}{5\varepsilon^8\delta^3\eta^4 m}, \frac{\beta \deg(i)}{\varepsilon^8\delta^3\eta^4 m}]$.

Next, we bound $\mathrm{Var}[X]$. For any $j \in [\deg(i)]$, $\mathrm{Var}[X_j] = \mathbb{E}[X_j^2] - (\mathbb{E}[X_j])^2 = \mathbb{E}[X_j] - (\mathbb{E}[X_j])^2 \le \frac{\beta}{\varepsilon^8\delta^3\eta^4 m}$. Since $(X_j)_{j\in[\deg(i)]}$ are negatively correlated, $\mathrm{Var}[X] \le \sum_{j\in[\deg(i)]} \mathrm{Var}[X_j] \le \deg(i) \cdot \frac{\beta}{\varepsilon^8\delta^3\eta^4 m} = \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m}$. Therefore, by Chebyshev's inequality,

$$\begin{aligned}
\Pr\left[i \notin \widetilde{H}\right] = \Pr[X \le 0] &\le \Pr\left[X \le \mathbb{E}[X] - \frac{\beta\deg(i)}{5\varepsilon^8\delta^3\eta^4 m}\right] \\
&\le \frac{\mathrm{Var}[X] \cdot 25\varepsilon^{16}\delta^6\eta^8 m^2}{\beta^2\deg^2(i)} \\
&\le \frac{25\varepsilon^8\delta^3\eta^4 m}{\beta\deg(i)} && \triangleright\ \mathrm{Var}[X] \le \frac{\beta\deg(i)}{\varepsilon^8\delta^3\eta^4 m} \\
&\le \frac{25\cdot 4096\varepsilon^4\delta^2\eta^2}{\beta}. && \triangleright\ \deg(i) \ge \theta m
\end{aligned}$$

Recall that $|H| \le \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{8192}{\varepsilon^4\delta\eta^2}$. By union bound, with probability at least $1 - \frac{8192}{\varepsilon^4\delta\eta^2} \cdot \frac{25\cdot 4096\varepsilon^4\delta^2\eta^2}{\beta} = 1 - \frac{\delta}{16}$ (since $\beta = 2^{29}\cdot 25$), $\widetilde{H}$ contains all high-degree variables H in Ψ. □

For each $i \in [n]$, define

1. $X_{i,+}^{\mathrm{pos}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = 1, b_j Y_j = 1\}|$ to be the number of constraints involving x_i such that the literal of x_i appears positively (i.e., $b_i = 1$) and the other literal is satisfied by the prediction $\mathbf{Y}$ (i.e., $b_j Y_j = 1$). The other quantities are defined analogously:
2. $X_{i,-}^{\mathrm{pos}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = 1, b_j Y_j = -1\}|$.
3. $X_{i,+}^{\mathrm{neg}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = -1, b_j Y_j = 1\}|$.
4. $X_{i,-}^{\mathrm{neg}} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in [n], b_i = -1, b_j Y_j = -1\}|$.

Lemma C.7. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For all $i \in \widetilde{H}$, we have*

(i) $\widetilde{X}_{i,+}^{\mathrm{pos}} \ge X_{i,+}^{\mathrm{pos}},\ \widetilde{X}_{i,-}^{\mathrm{pos}} \ge X_{i,-}^{\mathrm{pos}},\ \widetilde{X}_{i,+}^{\mathrm{neg}} \ge X_{i,+}^{\mathrm{neg}},\ \widetilde{X}_{i,-}^{\mathrm{neg}} \ge X_{i,-}^{\mathrm{neg}}$;

(ii) With probability at least $1 - \frac{\delta}{8}$, all of the following inequalities hold: $\widetilde{X}_{i,+}^{\mathrm{pos}} \le X_{i,+}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,\ \widetilde{X}_{i,-}^{\mathrm{pos}} \le X_{i,-}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,\ \widetilde{X}_{i,+}^{\mathrm{neg}} \le X_{i,+}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,\ \widetilde{X}_{i,-}^{\mathrm{neg}} \le X_{i,-}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$.

Proof. (i) By Theorem A.1, for all $i \in [n]$, we have $\widetilde{X}_{i,+}^{\mathrm{pos}} \ge X_{i,+}^{\mathrm{pos}},\ \widetilde{X}_{i,-}^{\mathrm{pos}} \ge X_{i,-}^{\mathrm{pos}},\ \widetilde{X}_{i,+}^{\mathrm{neg}} \ge X_{i,+}^{\mathrm{neg}}$, and $\widetilde{X}_{i,-}^{\mathrm{neg}} \ge X_{i,-}^{\mathrm{neg}}$.

(ii) By Theorem A.1, for any $i \in [n]$, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, each of the following four inequalities holds: $\widetilde{X}_{i,+}^{\mathrm{pos}} \le X_{i,+}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,\ \widetilde{X}_{i,-}^{\mathrm{pos}} \le X_{i,-}^{\mathrm{pos}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m,\ \widetilde{X}_{i,+}^{\mathrm{neg}} \le X_{i,+}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$, and $\widetilde{X}_{i,-}^{\mathrm{neg}} \le X_{i,-}^{\mathrm{neg}} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}m$. Since $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, by applying the union bound we have that, with probability at least $1 - \frac{\varepsilon^8\delta^4\eta^4}{64\beta} \cdot \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \cdot 4 \ge 1 - \frac{\delta}{8}$, for all $i \in \widetilde{H}$, all of the above four inequalities hold. □

Proof of Theorem C.4. Given the above lemmas, we are ready to prove Theorem C.4.

Proof. We can assume that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$, as otherwise we can store the whole instance in $O(\mathrm{poly}(1/\varepsilon, 1/\delta, 1/\eta))$ space.

In Algorithm 6, we maintain four CountMin sketches with accuracy parameter $\frac{\varepsilon^{10}\delta^5\eta^6}{128\beta}$ and failure probability $\frac{\varepsilon^8\delta^4\eta^4}{64\beta}$, where $\beta = 2^{29}\cdot 25$. By Theorem A.1, this takes $O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{1}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ words of space. Also, we maintain $\frac{\beta}{\varepsilon^8\delta^3\eta^4}$ ℓ_0-samplers, each with failure probability $\frac{\varepsilon^8\delta^4\eta^4}{32\beta}$. By Theorem A.3, this takes $O(\frac{\log^2 n}{\varepsilon^8\delta^3\eta^4}\log\frac{1}{\varepsilon^8\delta^4\eta^4}) = O(\frac{\log^2 n}{\varepsilon^8\delta^3\eta^4}\log\frac{1}{\varepsilon\delta\eta})$ bits of space. Therefore, the total space complexity of Algorithm 6 is $O(\frac{\log^2 n}{\varepsilon^{10}\delta^5\eta^6}\log\frac{1}{\varepsilon\delta\eta})$ bits.

Next, we analyze the approximation ratio. Since $[n] = \widetilde{H} \cup \widetilde{L}$, we have

1. $X^{\mathrm{pos}}_{i,+} = X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)}$, where $X^{\mathrm{pos}}_{i,\widetilde{H}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = 1, b_j Y_j = 1\}|$ and $X^{\mathrm{pos}}_{i,\widetilde{L}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = 1, b_j Y_j = 1\}|$.
2. $X^{\mathrm{pos}}_{i,-} = X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)}$, where $X^{\mathrm{pos}}_{i,\widetilde{H}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = 1, b_j Y_j = -1\}|$ and $X^{\mathrm{pos}}_{i,\widetilde{L}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = 1, b_j Y_j = -1\}|$.
3. $X^{\mathrm{neg}}_{i,+} = X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)}$, where $X^{\mathrm{neg}}_{i,\widetilde{H}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = -1, b_j Y_j = 1\}|$ and $X^{\mathrm{neg}}_{i,\widetilde{L}(+)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = -1, b_j Y_j = 1\}|$.
4. $X^{\mathrm{neg}}_{i,-} = X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)}$, where $X^{\mathrm{neg}}_{i,\widetilde{H}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{H}, b_i = -1, b_j Y_j = -1\}|$ and $X^{\mathrm{neg}}_{i,\widetilde{L}(-)} := |\{((i,j),(b_i,b_j)) \in \Psi : j \in \widetilde{L}, b_i = -1, b_j Y_j = -1\}|$.

By Lemma C.7, with probability at least $1 - \frac{\delta}{8}$, for all $i \in \widetilde{H}$, all of the following inequalities hold:

$$
\begin{aligned}
X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)} \le \widetilde{X}^{\mathrm{pos}}_{i,+} \le X^{\mathrm{pos}}_{i,\widetilde{H}(+)} + X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m,\\
X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)} \le \widetilde{X}^{\mathrm{pos}}_{i,-} \le X^{\mathrm{pos}}_{i,\widetilde{H}(-)} + X^{\mathrm{pos}}_{i,\widetilde{L}(-)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m,\\
X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)} \le \widetilde{X}^{\mathrm{neg}}_{i,+} \le X^{\mathrm{neg}}_{i,\widetilde{H}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m,\\
X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)} \le \widetilde{X}^{\mathrm{neg}}_{i,-} \le X^{\mathrm{neg}}_{i,\widetilde{H}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m.
\end{aligned}
$$

Since $|F| = \frac{\beta}{\varepsilon^8\delta^3\eta^4}$ and $\widetilde{H}$ is the set of associated variables of constraints in F, we have that $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Thus, it holds that $0 \le X^{\mathrm{pos}}_{i,\widetilde{H}(+)}, X^{\mathrm{pos}}_{i,\widetilde{H}(-)}, X^{\mathrm{neg}}_{i,\widetilde{H}(+)}, X^{\mathrm{neg}}_{i,\widetilde{H}(-)} \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$. Therefore,

$$
0 \le \widetilde{X}^{\mathrm{pos}}_{i,+} - X^{\mathrm{pos}}_{i,\widetilde{L}(+)} \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m, \tag{15}
$$

$$
0 \le \widetilde{X}^{\mathrm{pos}}_{i,-} - X^{\mathrm{pos}}_{i,\widetilde{L}(-)} \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m, \tag{16}
$$

$$
0 \le \widetilde{X}^{\mathrm{neg}}_{i,+} - X^{\mathrm{neg}}_{i,\widetilde{L}(+)} \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m, \tag{17}
$$

$$
0 \le \widetilde{X}^{\mathrm{neg}}_{i,-} - X^{\mathrm{neg}}_{i,\widetilde{L}(-)} \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6}{128\beta} m. \tag{18}
$$

We define

1. $\widetilde{X}_0 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in \widetilde{L},\ b_1 Y_{j_1} = b_2 Y_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which both literals are satisfied by the predictions $\mathbf{Y}$.
2. $\widetilde{X}_1 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in \widetilde{L},\ b_1 Y_{j_1} \ne b_2 Y_{j_2}\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which exactly one literal is satisfied by $\mathbf{Y}$.
3. $\widetilde{X}_2 := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1, j_2 \in \widetilde{L},\ b_1 Y_{j_1} = b_2 Y_{j_2} = -1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{L}$ and for which neither literal is satisfied by $\mathbf{Y}$.

Similarly, we define

1. $\widetilde{X}_{0,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1 Y_{j_1} = b_2 Y_{j_2} = 1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which both literals are satisfied by $\mathbf{Y}$.
2. $\widetilde{X}_{1,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1 Y_{j_1} \neq b_2 Y_{j_2}\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which exactly one literal is satisfied by $\mathbf{Y}$.
3. $\widetilde{X}_{2,\widetilde{H}} := |\{((j_1,j_2),(b_1,b_2)) \in \Psi : j_1,j_2 \in \widetilde{H},\ b_1 Y_{j_1} = b_2 Y_{j_2} = -1\}|$ to be the number of constraints in Ψ whose associated variables are both in $\widetilde{H}$ and for which neither literal is satisfied by $\mathbf{Y}$.

We have

$$\begin{aligned}
\widetilde{X}_0 &= X_{+,+} - \widetilde{X}_{0,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \sum_{i\in\widetilde{H}^-} X^{\mathrm{neg}}_{i,\widetilde{L}(+)},\\
\widetilde{X}_1 &= X_{+,-} - \widetilde{X}_{1,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} (X^{\mathrm{pos}}_{i,\widetilde{L}(-)} + X^{\mathrm{neg}}_{i,\widetilde{L}(+)}) - \sum_{i\in\widetilde{H}^-} (X^{\mathrm{pos}}_{i,\widetilde{L}(+)} + X^{\mathrm{neg}}_{i,\widetilde{L}(-)}),\\
\widetilde{X}_2 &= X_{-,-} - \widetilde{X}_{2,\widetilde{H}} - \sum_{i\in\widetilde{H}^+} X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \sum_{i\in\widetilde{H}^-} X^{\mathrm{pos}}_{i,\widetilde{L}(-)}.
\end{aligned}$$

Recall that

$$\begin{aligned}
\overline{X}_0 &= X_{+,+} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{pos}}_{i,+} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{neg}}_{i,+},\\
\overline{X}_1 &= X_{+,-} - \sum_{i\in\widetilde{H}^+} (\widetilde{X}^{\mathrm{pos}}_{i,-} + \widetilde{X}^{\mathrm{neg}}_{i,+}) - \sum_{i\in\widetilde{H}^-} (\widetilde{X}^{\mathrm{pos}}_{i,+} + \widetilde{X}^{\mathrm{neg}}_{i,-}),\\
\overline{X}_2 &= X_{-,-} - \sum_{i\in\widetilde{H}^+} \widetilde{X}^{\mathrm{neg}}_{i,-} - \sum_{i\in\widetilde{H}^-} \widetilde{X}^{\mathrm{pos}}_{i,-}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\overline{X}_0 - \widetilde{X}_0 &= \widetilde{X}_{0,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} (X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{pos}}_{i,+}) + \sum_{i\in\widetilde{H}^-} (X^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{neg}}_{i,+}),\\
\overline{X}_1 - \widetilde{X}_1 &= \widetilde{X}_{1,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} \left[(X^{\mathrm{pos}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{pos}}_{i,-}) + (X^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{neg}}_{i,+})\right] + \sum_{i\in\widetilde{H}^-} \left[(X^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{X}^{\mathrm{pos}}_{i,+}) + (X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{neg}}_{i,-})\right],\\
\overline{X}_2 - \widetilde{X}_2 &= \widetilde{X}_{2,\widetilde{H}} + \sum_{i\in\widetilde{H}^+} (X^{\mathrm{neg}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{neg}}_{i,-}) + \sum_{i\in\widetilde{H}^-} (X^{\mathrm{pos}}_{i,\widetilde{L}(-)} - \widetilde{X}^{\mathrm{pos}}_{i,-}).
\end{aligned}$$

Since $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, we have $|\widetilde{H}^+|, |\widetilde{H}^-| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$, and $\widetilde{X}_{0,\widetilde{H}}, \widetilde{X}_{1,\widetilde{H}}, \widetilde{X}_{2,\widetilde{H}} \le \binom{|\widetilde{H}|}{2}\cdot 2^2 \le \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}$. Moreover, combining inequalities (15)–(18), we have

$$\begin{aligned}
\widetilde{X}_0 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right) \le \overline{X}_0 \le \widetilde{X}_0 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8},\\
\widetilde{X}_1 - \left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{16}\right) \le \overline{X}_1 \le \widetilde{X}_1 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8},\\
\widetilde{X}_2 - \left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8} + \frac{\varepsilon^2\delta^2\eta^2 m}{32}\right) \le \overline{X}_2 \le \widetilde{X}_2 + \frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}.
\end{aligned}$$

Let $\widetilde{k}_{\widetilde{L}} := \frac{1}{4\varepsilon^2}[(\frac{1}{2}+\varepsilon)^2\widetilde{X}_0 - (\frac{1}{2}+\varepsilon)(\frac{1}{2}-\varepsilon)\widetilde{X}_1 + (\frac{1}{2}-\varepsilon)^2\widetilde{X}_2]$. Now we are ready to show that $\overline{k}_{\widetilde{L}}$ is close to $\widetilde{k}_{\widetilde{L}}$ up to an additive error of $\mathrm{poly}(\eta)\cdot m$ with probability at least $1-\frac{\delta}{4}$. We have

$$\overline{k}_{\widetilde{L}} = \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0 - \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1 + \left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right]$$

$$
\begin{aligned}
&\geq \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0-\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right)-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1+\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2-\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right)\right]\\
&=\widetilde{k}_{\widetilde{L}}-\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)+\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\frac{16\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{32}\right)\right]\\
&=\widetilde{k}_{\widetilde{L}}-\left(\frac{5\beta^2}{2\varepsilon^{18}\delta^6\eta^8}+\frac{6\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}+\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\geq \widetilde{k}_{\widetilde{L}}-\frac{\eta m}{32},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}_{\widetilde{L}}&=\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\overline{X}_0-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\overline{X}_1+\left(\frac{1}{2}-\varepsilon\right)^2\overline{X}_2\right]\\
&\leq \frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\left(\widetilde{X}_0+\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)-\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\widetilde{X}_1-\left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)\right)\right.\\
&\quad\left.+\left(\frac{1}{2}-\varepsilon\right)^2\left(\widetilde{X}_2+\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right)\right]\\
&=\widetilde{k}_{\widetilde{L}}+\frac{1}{4\varepsilon^2}\left[\left(\frac{1}{2}+\varepsilon\right)^2\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\left(\frac{32\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\varepsilon^2\delta^2\eta^2 m}{16}\right)+\left(\frac{1}{2}-\varepsilon\right)^2\frac{8\beta^2}{\varepsilon^{16}\delta^6\eta^8}\right]\\
&=\widetilde{k}_{\widetilde{L}}+\left(\frac{3\beta^2}{\varepsilon^{18}\delta^6\eta^8}-\frac{4\beta^2}{\varepsilon^{16}\delta^6\eta^8}+\frac{\delta^2\eta^2 m}{256}-\frac{\varepsilon^2\delta^2\eta^2 m}{64}\right)\\
&\leq \widetilde{k}_{\widetilde{L}}+\frac{\eta m}{64},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$.

For each $i\in\widetilde{H}$, let $\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}:=(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot X^{\mathrm{pos}}_{i,\widetilde{L}(+)}+(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot X^{\mathrm{pos}}_{i,\widetilde{L}(-)}$ and $\widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}:=(\frac{1}{2}+\frac{1}{4\varepsilon})\cdot X^{\mathrm{neg}}_{i,\widetilde{L}(+)}+(\frac{1}{2}-\frac{1}{4\varepsilon})\cdot X^{\mathrm{neg}}_{i,\widetilde{L}(-)}$. We have

$$
\begin{aligned}
\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{X}^{\mathrm{pos}}_{i,+}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{X}^{\mathrm{pos}}_{i,-}\\
&\geq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot X^{\mathrm{pos}}_{i,\widetilde{L}(+)}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\left(X^{\mathrm{pos}}_{i,\widetilde{L}(-)}+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}-\left(\frac{1}{4\varepsilon}-\frac{1}{2}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)\\
&=\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}-\left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4}-\frac{\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^9\delta^5\eta^6 m}{512\beta}-\frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right)\\
&\geq\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}-\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m=\Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. On the other hand,

$$
\begin{aligned}
\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}&=\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\cdot\widetilde{X}^{\mathrm{pos}}_{i,+}+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot\widetilde{X}^{\mathrm{pos}}_{i,-}\\
&\leq\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)\left(X^{\mathrm{pos}}_{i,\widetilde{L}(+)}+\frac{2\beta}{\varepsilon^8\delta^3\eta^4}+\frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right)+\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)\cdot X^{\mathrm{pos}}_{i,\widetilde{L}(-)}
\end{aligned}
$$

$$
\begin{aligned}
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right)\left(\frac{2\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{128\beta}\right) \\
&= \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \left(\frac{\beta}{2\varepsilon^9\delta^3\eta^4} + \frac{\beta}{\varepsilon^8\delta^3\eta^4} + \frac{\varepsilon^9\delta^5\eta^6 m}{512\beta} + \frac{\varepsilon^{10}\delta^5\eta^6 m}{256\beta}\right) \\
&\le \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} + \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta},
\end{aligned}
$$

where the last step follows since $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Therefore, $|\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)} - \widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$. Similarly, $|\overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)} - \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}| \le \frac{\varepsilon^8\delta^3\eta^5 m}{128\beta}$.

Let $\widetilde{\mathrm{ALG}} := \widetilde{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{32}$ denote the output of the algorithm obtained by running Algorithm 1 with L and H replaced by $\widetilde{L}$ and $\widetilde{H}$, respectively. Analogous to the proof of Theorem 4.1, we can show that $(1-\frac{\eta}{2})\cdot\mathrm{OPT} \le \widetilde{\mathrm{ALG}} \le \mathrm{OPT}$ with probability at least $1 - \frac{3\delta}{4}$. The argument relies primarily on two properties of the partition $(\widetilde{L}, \widetilde{H})$: (i) every variable in $\widetilde{L}$ has degree at most θm for $\theta := \frac{\varepsilon^4\delta\eta^2}{4096}$ (since $\widetilde{L} \subseteq L$), and (ii) $|\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4}$.

Let ALG denote the output of Algorithm 6. By applying the union bound on Lemma C.5, Lemma C.6, and Lemma C.7, with probability at least $1 - \frac{\delta}{16} - \frac{\delta}{16} - \frac{\delta}{8} - \frac{3\delta}{4} = 1 - \delta$, we have

$$
\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16} \\
&\le \widetilde{k}_{\widetilde{L}} + \frac{\eta m}{64} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} + |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\le \widetilde{\mathrm{ALG}} + \frac{\eta m}{32} + \frac{\eta m}{64} + \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\le \mathrm{OPT},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{ALG} &= \overline{k}_{\widetilde{L}} + \sum_{i\in\widetilde{H}} \max\{\overline{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \overline{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - \frac{\eta m}{16} \\
&\ge \widetilde{k}_{\widetilde{L}} - \frac{\eta m}{32} + \sum_{i\in\widetilde{H}} \max\{\widetilde{k}^{\mathrm{pos}}_{i,\widetilde{L}(+)}, \widetilde{k}^{\mathrm{neg}}_{i,\widetilde{L}(+)}\} - |\widetilde{H}|\cdot\frac{\varepsilon^8\delta^3\eta^5 m}{128\beta} - \frac{\eta m}{16} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{64} - \frac{\eta m}{16} && \triangleright\ |\widetilde{H}| \le \frac{2\beta}{\varepsilon^8\delta^3\eta^4} \\
&\ge \widetilde{\mathrm{ALG}} - \frac{\eta m}{8} \\
&\ge \left(1 - \frac{\eta}{2}\right)\cdot\mathrm{OPT} - \frac{\eta}{2}\cdot\mathrm{OPT} && \triangleright\ m \le 4\mathrm{OPT} \\
&= (1-\eta)\cdot\mathrm{OPT}.
\end{aligned}
$$

□

C.2 $k \ge 3$: Bounded-degree instances

In this section, we present the streaming algorithm for MAX-kAND with $k \ge 3$ when all variables have bounded degree. Specifically, we maintain $k+1$ constraint-type counters $X_0, X_1, \ldots, X_k$ during the stream. For each $i \in \{0, \ldots, k\}$, X_i records the number of constraints for which exactly i literals are *not* satisfied by the predicted assignment. At the end of the stream, we use these counters to decode an estimate of the optimal MAX-kAND value by solving a linear system. Refer to Algorithm 7 for a formal description.

Note that although the algorithm is presented for insertion-only streams, it extends directly to dynamic streams: when a constraint deletion arrives, we simply decrement the corresponding counters. In

Algorithm 7 A learning-augmented streaming algorithm for estimating the MAX-kAND value for $k \geq 3$ in bounded-degree instances

Input: MAX-kAND instance Ψ with maximum degree $\Delta < \frac{\varepsilon^{2k}\delta\eta^2 m}{4k(k+1)}$ as a stream of constraints, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$
Output: Estimate of the maximum value of Ψ
▷ **Pre-processing phase**
initialize $X_0, X_1, \ldots, X_k \leftarrow 0, m \leftarrow 0$.
▷ **Streaming phase**
for each constraint $((j_1, \ldots, j_k), (b_1, \ldots, b_k))$ in the stream **do**
  $m \leftarrow m + 1$
  $Y_{j_1} = \mathcal{O}(j_1), \ldots, Y_{j_k} = \mathcal{O}(j_k)$
  $i \leftarrow |\{t \in [k] : b_t Y_{j_t} \neq 1\}|$
  $X_i \leftarrow X_i + 1$
▷ **Post-processing phase**
return $\frac{1}{(2\varepsilon)^k} \sum_{i=0}^{k} (\frac{1}{2} + \varepsilon)^{k-i} (-\frac{1}{2} + \varepsilon)^i X_i - \frac{\eta m}{2^{k+1}}$

addition, we only require the bounded-degree assumption to hold at the end of the stream (i.e., for the final instance).

We have the following performance guarantee regarding our streaming algorithms for bounded-degree MAX-kAND instances with $k \geq 3$.

Theorem C.8. *Let $\varepsilon \in (0, \frac{1}{2}], \delta \in (0, 1)$, and $k \geq 3$. For any $\eta \in (0, 1)$, given an ε-accurate oracle $\mathcal{O}$ for a* MAX-kAND *instance Ψ, if the maximum variable degree Δ satisfies $\Delta < \frac{\varepsilon^{2k}\delta\eta^2 m}{4k(k+1)}$, then there exists a single-pass algorithm that, with probability at least $1 - \delta$, outputs a $(1 - \eta)$-approximation to the maximum value of Ψ in both the insertion-only and dynamic streaming models. The algorithm uses $O(k)$ words of space.*

For each $i \in \{0, \ldots, k\}$, let X_i denote the random variable given by the final value of the corresponding counter at the end of the algorithm. We first prove concentration bounds for X_i's around their expectations.

Lemma C.9. *Let $\varepsilon \in (0, \frac{1}{2}]$ and $\delta \in (0, 1)$. For each $i \in \{0, \ldots, k\}$, we have $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^k \eta m}{2}$ with probability at least $1 - \frac{\delta}{k+1}$.*

Proof. For each constraint $(\mathbf{j} = (j_1, \ldots, j_k), \mathbf{b} = (b_1, \ldots, b_k)) \in \Psi$, define the following indicator random variables:

$$X_{i,\mathbf{j}} = \begin{cases} 1 & \text{if } |\{t \in [k] : b_t Y_{j_t} \neq 1\}| = i, \\ 0 & \text{otherwise,} \end{cases}$$

for $i \in \{0, \ldots, k\}$. Note that we have $\mathbb{E}[X_{i,\mathbf{j}}] = \Pr[X_{i,\mathbf{j}} = 1]$ and $X_i = \sum_{(\mathbf{j},\mathbf{b}) \in \Psi} X_{i,\mathbf{j}}$. Therefore,

$$\begin{aligned}
\mathbb{E}[X_i^2] &= \mathbb{E}\left[\left(\sum_{(\mathbf{j},\mathbf{b}) \in \Psi} X_{i,\mathbf{j}}\right)^2\right] = \mathbb{E}\left[\sum_{(\mathbf{j},\mathbf{b}) \in \Psi} \sum_{(\mathbf{j}',\mathbf{b}') \in \Psi} X_{i,\mathbf{j}} X_{i,\mathbf{j}'}\right] \\
&= \sum_{(\mathbf{j},\mathbf{b}) \in \Psi} \sum_{(\mathbf{j}',\mathbf{b}') \in \Psi} \mathbb{E}[X_{i,\mathbf{j}} X_{i,\mathbf{j}'}] \\
&= \sum_{(\mathbf{j},\mathbf{b}) \in \Psi} \sum_{(\mathbf{j}',\mathbf{b}') \in \Psi} \Pr[X_{i,\mathbf{j}} = 1 \wedge X_{i,\mathbf{j}'} = 1] \\
&= \sum_{\substack{(\mathbf{j},\mathbf{b}),(\mathbf{j}',\mathbf{b}') \in \Psi: \\ \{j_1,\ldots,j_k\} \cap \{j'_1,\ldots,j'_k\} = \emptyset}} \Pr[X_{i,\mathbf{j}} = 1 \wedge X_{i,\mathbf{j}'} = 1] + \sum_{\substack{(\mathbf{j},\mathbf{b}),(\mathbf{j}',\mathbf{b}') \in \Psi: \\ \{j_1,\ldots,j_k\} \cap \{j'_1,\ldots,j'_k\} \neq \emptyset}} \Pr[X_{i,\mathbf{j}} = 1 \wedge X_{i,\mathbf{j}'} = 1] \\
&\leq \sum_{\substack{(\mathbf{j},\mathbf{b}),(\mathbf{j}',\mathbf{b}') \in \Psi: \\ \{j_1,\ldots,j_k\} \cap \{j'_1,\ldots,j'_k\} = \emptyset}} \Pr[X_{i,\mathbf{j}} = 1] \Pr[X_{i,\mathbf{j}'} = 1] + \sum_{\substack{(\mathbf{j},\mathbf{b}),(\mathbf{j}',\mathbf{b}') \in \Psi: \\ \{j_1,\ldots,j_k\} \cap \{j'_1,\ldots,j'_k\} \neq \emptyset}} 1
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{(\mathbf{j},\mathbf{b})\in\Psi}\sum_{(\mathbf{j}',\mathbf{b}')\in\Psi}\mathbb{E}[X_{i,\mathbf{j}}]\mathbb{E}[X_{i,\mathbf{j}'}] + m\cdot k\cdot\Delta \\
&\leq \left(\mathbb{E}\left[\sum_{(\mathbf{j},\mathbf{b})\in\Psi_L} X_{i,\mathbf{j}}\right]\right)^2 + k\Delta m \\
&= (\mathbb{E}[X_i])^2 + k\Delta m.
\end{aligned}
$$

Then we have $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \leq k\Delta m$. By Chebyshev's inequality,

$$
\begin{aligned}
\Pr\left[|X_i - \mathbb{E}[X_i]| > \frac{\varepsilon^k\eta m}{2}\right] &\leq \frac{\mathrm{Var}[X_i]}{(\varepsilon^k\eta m/2)^2} \leq \frac{4k\Delta m}{\varepsilon^{2k}\eta^2 m^2} && \triangleright\ \mathrm{Var}[X_i] \leq k\Delta m \\
&= \frac{4k\Delta}{\varepsilon^{2k}\eta^2 m} < \frac{\delta}{k+1}. && \triangleright\ \Delta < \frac{\varepsilon^{2k}\delta\eta^2 m}{4k(k+1)}
\end{aligned}
$$

□

Proof of Theorem C.8. Now we are ready to prove Theorem C.8.

Proof. We first analyze the space complexity. In Algorithm 7, we mainly maintain $k+1$ counters: $X_0, \ldots, X_k$, corresponding to the constraints categorized by the predicted signs of their associated variables. Each counter requires $O(\log n)$ bits of space, since $0 \leq X_i \leq m < 2^k n^k$ for $i \in \{0, \ldots, k\}$. Therefore, the total space complexity of the algorithm is $O(k)$ words.

Next, we show that Algorithm 7 with high probability outputs a $(1-\eta)$-approximation of the MAX-kAND value when all variables have bounded degrees.

Let $\boldsymbol{\sigma}^* \in \{\pm 1\}^n$ denote an optimal assignment for the MAX-kAND instance Ψ. For $i \in \{0, \ldots, k\}$, let $k_i := |\{(\mathbf{j} = (j_1, \ldots, j_k), \mathbf{b} = (b_1, \ldots, b_k)) \in \Psi : |\{t \in [k] : b_t\sigma^*_{j_t} \neq 1\}| = i\}|$ denote the number of constraints in Ψ for which exactly i of the k literals are not satisfied by $\boldsymbol{\sigma}^*$. Note that $k_0 = \mathrm{OPT}$. According to Algorithm 7, for each $i \in \{0, \ldots, k\}$, we have

$$
\mathbb{E}[X_i] = \sum_{j=0}^{k} k_j \sum_{s=\max\{0,i-(k-j)\}}^{\min\{i,j\}} \binom{j}{s}\binom{k-j}{i-s}\left(\frac{1}{2}+\varepsilon\right)^{k-(i+j-2s)}\left(\frac{1}{2}-\varepsilon\right)^{i+j-2s}.
$$

Therefore, by Gaussian elimination, we have

$$
\mathrm{OPT} = k_0 = \frac{1}{(2\varepsilon)^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\mathbb{E}[X_i]. \tag{19}
$$

Let ALG denote the output of Algorithm 7. By applying union bound on Lemma C.9, we have that with probability at least $1 - \frac{\delta}{k+1}\cdot(k+1) = 1-\delta$, $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon^k\eta m}{2}$ for all $i \in \{0, \ldots, k\}$. Therefore,

$$
\begin{aligned}
\mathrm{ALG} &= \frac{1}{(2\varepsilon)^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i} X_i - \frac{\eta m}{2^{k+1}} \\
&= \frac{1}{(2\varepsilon)^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\left(\mathbb{E}[X_i] \pm \frac{\varepsilon^k\eta m}{2}\right) - \frac{\eta m}{2^{k+1}} \\
&\leq \mathrm{OPT} + \frac{\varepsilon^k\eta m}{2\cdot(2\varepsilon)^k}\sum_{i=0}^{k}\left|\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\right| - \frac{\eta m}{2^{k+1}} && \triangleright\ \text{Equation (19)} \\
&\leq \mathrm{OPT} + \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} && \triangleright\ \sum_{i=0}^{k}\left|\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\right| \leq 1 \\
&= \mathrm{OPT}.
\end{aligned}
$$

On the other hand,

$$\begin{aligned}
\mathrm{ALG} &= \frac{1}{(2\varepsilon)^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i} X_i - \frac{\eta m}{2^{k+1}} \\
&= \frac{1}{(2\varepsilon)^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\left(\mathbb{E}[X_i] \pm \frac{\varepsilon^k \eta m}{2}\right) - \frac{\eta m}{2^{k+1}} \\
&\geq \mathrm{OPT} - \frac{\varepsilon^k \eta m}{2\cdot(2\varepsilon)^k}\sum_{i=0}^{k}\left|\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\right| - \frac{\eta m}{2^{k+1}} && \triangleright \text{ Equation (19)} \\
&\geq \mathrm{OPT} - \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} && \triangleright \sum_{i=0}^{k}\left|\left(\frac{1}{2}+\varepsilon\right)^{k-i}\left(-\frac{1}{2}+\varepsilon\right)^{i}\right| \leq 1 \\
&= \mathrm{OPT} - \frac{\eta m}{2^k} \\
&\geq (1-\eta)\cdot \mathrm{OPT}. && \triangleright \ \mathrm{OPT} \geq \frac{m}{2^k}
\end{aligned}$$

So, with probability at least $1-\delta$, Algorithm 7 is a $(1-\eta)$-approximation. □

The bounded-degree algorithm above is a warm-up. The following subsection gives the arbitrary-degree MAX-kAND algorithm, still under the simple, unweighted convention from Problem 2.1.

C.3 Unbounded-degree instances

We now remove the bounded-degree assumption for MAX-kAND. The resulting space dependence on k, $1/\varepsilon$, $1/\eta$, and $\log(1/\delta)$ is exponential. However, the space is independent of n in insertion-only streams and polylogarithmic in n in dynamic streams.

We use the same high-degree and low-degree notation as the preceding sections. Whenever $H \subseteq [n]$ is fixed, let $L = [n] \setminus H$. For an assignment $\boldsymbol{\sigma} \in \{\pm 1\}^n$, consider the decoded value

$$\sum_{(\mathbf{j},\mathbf{b})\in\Psi}\ \prod_{\substack{t\in[k]\\ j_t\in H}} \frac{1+b_t\sigma_{j_t}}{2} \prod_{\substack{t\in[k]\\ j_t\in L}} \left(\frac{1}{2}+\frac{b_t Y_{j_t}}{4\varepsilon}\right). \tag{20}$$

The first product evaluates the literals indexed by H exactly under $\boldsymbol{\sigma}$. Each factor in the second product is an unbiased estimate of the corresponding literal under the fixed optimal assignment, because

$$\mathbb{E}\left[\frac{1}{2}+\frac{b_t Y_{j_t}}{4\varepsilon}\right] = \frac{1}{2}+\frac{b_t\sigma^*_{j_t}}{2} = \frac{1+b_t\sigma^*_{j_t}}{2}. \tag{21}$$

The displayed value depends on $\boldsymbol{\sigma}$ only through its coordinates indexed by H.

C.3.1 Offline algorithm

The high-degree set is grown adaptively. When it currently contains $|H|$ variables, another variable is added only if its degree exceeds $\theta m/(|H|+1)$. The harmonic-series bound keeps H small, while the factor $(|H|+1)^{-1}$ makes the decoded values concentrate simultaneously over all assignments to the variables indexed by H.

Theorem C.10. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta, \eta \in (0,1)$, and $k \geq 3$. Consider an instance Ψ of* MAX-kAND *and an ε-accurate oracle $\mathcal{O}$ for Ψ. Then Algorithm 8 outputs a $(1-\eta)$-approximation to the maximum value of Ψ with probability at least $1-\delta$. Its running time is $km \cdot \exp\big(\exp\big(O(k^2\varepsilon^{-2k}\eta^{-2}\log(4/\delta))\big)\big)$.*

Lemma C.11. *Assume $m > 0$, and let H be the set computed by Algorithm 8. Then*

$$|H| \leq \exp\big(4k^2\varepsilon^{-2k}\eta^{-2}\log(4/\delta)\big) - 1.$$

Algorithm 8 Offline learning-augmented estimator for unbounded-degree MAX-kAND

Input: MAX-kAND instance $\Psi = (C_1, \dots, C_m)$, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, and $\delta, \eta \in (0, 1)$

Output: Estimate of the maximum value of Ψ

if $m = 0$ **then**
  return 0
$\theta \leftarrow \frac{\varepsilon^{2k}\eta^2}{4k\log(4/\delta)}$
Compute $\deg(i)$ for every variable appearing in Ψ.
$H \leftarrow \emptyset$.
Repeatedly add any $i \notin H$ satisfying $\deg(i) > \frac{\theta m}{|H|+1}$, until no such variable remains.
$L \leftarrow [n] \setminus H$.
Query $Y_i = \mathcal{O}(i)$ for every $i \in L$ with $\deg(i) > 0$.
return

$$\max_{\boldsymbol{\sigma}\in\{\pm1\}^n} \sum_{(\mathbf{j},\mathbf{b})\in\Psi} \prod_{\substack{t\in[k]\\ j_t\in H}} \frac{1+b_t\sigma_{j_t}}{2} \prod_{\substack{t\in[k]\\ j_t\in L}} \left(\frac{1}{2} + \frac{b_t Y_{j_t}}{4\varepsilon}\right) - \frac{\eta m}{2^{k+1}}.$$

The maximized expression depends only on the coordinates indexed by H, so the maximum is evaluated over $2^{|H|}$ assignments.

Proof. The claim is immediate if $H = \emptyset$. Otherwise, let i_s be the s-th variable added to H, for $s \in [|H|]$. Immediately before i_s is added, the set contains $s-1$ variables, and therefore $\deg(i_s) > \frac{\theta m}{s}$. Every constraint contains exactly k distinct variables, so

$$km = \sum_{i\in[n]} \deg(i) \geq \sum_{s=1}^{|H|} \deg(i_s) > \theta m \sum_{s=1}^{|H|} \frac{1}{s}.$$

Since $m > 0$, the standard harmonic bound gives $\log(|H|+1) \leq \sum_{s=1}^{|H|} \frac{1}{s} < \frac{k}{\theta}$. Substituting the value of θ and exponentiating proves the claim. □

Lemma C.12. *Fix $H \subseteq [n]$, let $L = [n] \setminus H$, and fix an assignment $\boldsymbol{\sigma} \in \{\pm1\}^n$. Then the expectation of (20) is*

$$\sum_{(\mathbf{j},\mathbf{b})\in\Psi} \prod_{\substack{t\in[k]\\ j_t\in H}} \frac{1+b_t\sigma_{j_t}}{2} \prod_{\substack{t\in[k]\\ j_t\in L}} \frac{1+b_t\sigma^*_{j_t}}{2},$$

which is at most OPT. *Equality holds when $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$. Consequently, the maximum over $\boldsymbol{\sigma}$ of the expected decoded value is* OPT.

Proof. The coordinates of every constraint are distinct, and the oracle labels are independent. Hence Equation (21) allows the expectation to pass through the product over the variables indexed by L, which gives the displayed identity. The resulting sum is the value of the assignment that uses $\boldsymbol{\sigma}$ on the variables indexed by H and $\boldsymbol{\sigma}^*$ on the variables indexed by L, so it is at most OPT. Taking $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$ gives equality. □

Lemma C.13. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\eta > 0$, and $\theta > 0$. Let $H \subseteq [n]$ and $L = [n] \setminus H$. Suppose that $\deg(i) \leq \theta m/(|H|+1)$ for every $i \in L$. Then the probability that, for some assignment $\boldsymbol{\sigma} \in \{\pm1\}^n$, the decoded value in (20) differs from its expectation by at least $\eta m/2^{k+1}$ is at most*

$$2^{|H|+1} \exp\left(-\eta^2\varepsilon^{2k}(|H|+1)/(2k\theta)\right).$$

Proof. Fix $\boldsymbol{\sigma}$. The decoded value is a function of the independent oracle labels $(Y_i)_{i\in L}$. Changing Y_i affects only constraints containing x_i. In each such constraint, the factor containing Y_i changes by $1/(2\varepsilon)$, while every other decoded factor has absolute value at most $(2\varepsilon)^{-1}$ and every exact factor has absolute value at most 1. Thus, in the notation of Theorem A.5, one may take $c_i = (2\varepsilon)^{-k}\deg(i)$. The residual degree bound and $\sum_{i\in L}\deg(i) \leq km$ imply

$$\sum_{i\in L} c_i^2 \leq (2\varepsilon)^{-2k} \frac{\theta m}{|H|+1} \sum_{i\in L} \deg(i) \leq \frac{k\theta m^2}{(2\varepsilon)^{2k}(|H|+1)}.$$

Applying Theorem A.5 with deviation $\eta m/2^{k+1}$ gives, for the fixed assignment, failure probability at most $2\exp\left(-\frac{\eta^2\varepsilon^{2k}(|H|+1)}{2k\theta}\right)$. The decoded value depends only on the coordinates of $\boldsymbol{\sigma}$ indexed by H. A union bound over the $2^{|H|}$ possible choices of those coordinates proves the lemma. □

Proof of Theorem C.10. Now we are ready to prove Theorem C.10.

Proof. If $m = 0$, the algorithm returns the exact value. Assume $m > 0$. Lemma C.11 gives the stated number of assignments to enumerate. Computing all degrees takes $O(km)$ time, and each assignment is evaluated by one scan of the m constraints, giving the stated running time.

At termination, every $i \in L$ satisfies $\deg(i) \le \theta m/(|H|+1)$. By Lemma C.13 and the chosen value of θ, the probability that any decoded value differs from its expectation by more than $\eta m/2^{k+1}$ is at most

$$2^{|H|+1}\exp\left(-2(|H|+1)\log(4/\delta)\right) = \left(\frac{\delta^2}{8}\right)^{|H|+1} \le \delta.$$

Condition on the complementary event, and let ALG denote the value returned by the algorithm. By Lemma C.12,

$$\mathrm{ALG} \le \mathrm{OPT} + \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} = \mathrm{OPT}.$$

Taking $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$ in the maximum gives

$$\mathrm{ALG} \ge \mathrm{OPT} - \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} = \mathrm{OPT} - \frac{\eta m}{2^k}.$$

A uniformly random assignment satisfies each MAX-kAND constraint with probability 2^{-k}, so $\mathrm{OPT} \ge m/2^k$. Therefore $\mathrm{ALG} \ge (1-\eta)\mathrm{OPT}$. □

C.3.2 Our algorithm in insertion-only streams

We next implement the offline estimator in one pass. A reservoir sample F provides candidate high-degree variables. A CountMin sketch of the degree vector, written using the frequency notation $f_i = \deg(i)$ from the CountMin preliminaries, supplies estimates $\widetilde{f}_i$ for variables appearing in F. The set $\widetilde{H}$ is then the longest degree-ordered prefix satisfying the adaptive threshold, and $\widetilde{L} = [n] \setminus \widetilde{H}$.

A second CountMin sketch reconstructs the decoded value after $\widetilde{H}$ is known. For every subset of positions in an arriving constraint, its key records the number s of selected positions, the number i of unselected literals not satisfied by the prediction, and the selected signed variables in increasing variable-index order. Thus all reconstruction counts are coordinates of a single sketch.

For a fixed assignment $\boldsymbol{\sigma}$, the value reconstructed from this sketch is

$$\begin{aligned}\sum_{s=0}^{k} \sum_{\substack{j_1,\dots,j_s \in \widetilde{H} \\ j_1<\cdots<j_s}} \sum_{(b_1,\dots,b_s)\in\{\pm1\}^s} &\left[\prod_{\ell=1}^{s} b_\ell\left(\frac{\sigma_{j_\ell}}{2} - \frac{Y_{j_\ell}}{4\varepsilon}\right)\right] \\ &\cdot \sum_{i=0}^{k-s}\left(\frac{1}{2}+\frac{1}{4\varepsilon}\right)^{k-s-i}\left(\frac{1}{2}-\frac{1}{4\varepsilon}\right)^{i} \\ &\cdot \textsc{Query}\left(\mathbf{CM}\left[(s,i,((j_1,b_1),\dots,(j_s,b_s)))\right]\right).\end{aligned} \tag{22}$$

For $s = 0$, the signed-variable tuple and the product are empty, and the product equals 1.

The reservoir capacity is at least $\lceil (8/\theta)\exp(8k/\theta)\log((64/\delta)(k/\theta)\exp(8k/\theta))\rceil$. The degree sketch has accuracy parameter $1/((8k/\theta)\exp(8k/\theta))$ and failure probability equal to $\delta/(16k)$ divided by the reservoir capacity. A union bound over all degree queries for variables appearing in F therefore contributes at most $\delta/16$. The reconstruction sketch $\mathbf{CM}$ has accuracy parameter $1/((8/\eta)(2/\varepsilon)^k(k+1)(1+2\exp(8k/\theta))^k)$ and failure probability $1/((16/\delta)(k+1)(1+2\exp(8k/\theta))^k)$. The reservoir, the two sketches, and the oracle labels use independent randomness.

Algorithm 9 Insertion-only streaming estimator for unbounded-degree MAX-kAND

Input: MAX-kAND instance Ψ as an insertion-only stream of constraints, an ε-accurate oracle $\mathcal{O} \to \{-1, 1\}$, and $\delta, \eta \in (0, 1)$

Output: Estimate of the maximum value of Ψ

▷ **Pre-processing phase**

$m \leftarrow 0$ and $\theta \leftarrow \frac{\varepsilon^{2k}\eta^2}{64k\log(16/\delta)}$.

Maintain a reservoir sample F with the capacity specified above; if fewer constraints arrive, store all of them.

initialize the degree sketch and the reconstruction sketch **CM** with the parameters specified above.

▷ **Streaming phase**

for each constraint $((j_1, \ldots, j_k), (b_1, \ldots, b_k))$ in the stream **do**

    $m \leftarrow m + 1$ and query $Y_{j_t} = \mathcal{O}(j_t)$ for every $t \in [k]$.

    Update the degree sketch once at every coordinate j_t.

    for each $s \in \{0, \ldots, k\}$ and every $1 \le t_1 < \cdots < t_s \le k$ **do**

        $i \leftarrow |\{t \in [k] \setminus \{t_1, \ldots, t_s\} : b_t Y_{j_t} = -1\}|$.

        Sort $(j_{t_1}, b_{t_1}), \ldots, (j_{t_s}, b_{t_s})$ by variable index and update **CM** at the key consisting of s, i, and this ordered tuple.

    Update the reservoir sample F.

▷ **Post-processing phase**

if $m = 0$ **then**

    return 0

For every variable i appearing in a constraint in F, let $\widetilde{f}_i$ be the estimate returned by the degree sketch.

Sort these variables in non-increasing order of $\widetilde{f}_i$.

Let $\widetilde{H}$ be the longest prefix such that the s-th variable in the ordering has estimate greater than $\theta m/(2s)$ for every $s \in [|\widetilde{H}|]$; set $\widetilde{L} \leftarrow [n] \setminus \widetilde{H}$.

Query $Y_i = \mathcal{O}(i)$ for every $i \in \widetilde{H}$.

Evaluate (22) for all $2^{|\widetilde{H}|}$ assignments to the variables indexed by $\widetilde{H}$.

return the maximum reconstructed value minus $\frac{\eta m}{2^{k+1}}$.

Theorem C.14. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta \in (0, 1)$, and $k \ge 3$. For any $\eta \in (0, 1)$, given an ε-accurate oracle $\mathcal{O}$ for a* MAX-kAND *instance Ψ, Algorithm 9 outputs a $(1 - \eta)$-approximation to the maximum value of Ψ in insertion-only streams with probability at least $1 - \delta$. The algorithm uses $\exp\big(O(k^3\varepsilon^{-2k}\eta^{-2}\log(16/\delta))\big)$ words of space.*

Lemma C.15. *Assume the final stream contains $m > 0$ constraints, and let θ be the value used in Algorithm 9. With probability at least $1 - \delta/4$, the set $\widetilde{H}$ computed by the algorithm satisfies $|\widetilde{H}| < \exp(8k/\theta)$ and $\deg(i) \le \theta m/(|\widetilde{H}| + 1)$ for every $i \notin \widetilde{H}$.*

Proof. Consider a variable i with $\deg(i) > m/((8/\theta)\exp(8k/\theta))$. If the stream is shorter than the reservoir capacity, every constraint is stored. Otherwise, the probability that no sampled constraint contains i is at most

$$\left(1 - \frac{\deg(i)}{m}\right)^{|F|} \le \exp\left(-|F|\frac{\deg(i)}{m}\right) \le \frac{\theta\delta}{64k\exp(8k/\theta)}.$$

Because the total variable degree is km, at most $(8k/\theta)\exp(8k/\theta)$ variables exceed the displayed degree threshold. A union bound therefore shows that, with probability at least $1 - \delta/8$, every such variable appears in a sampled constraint.

The reservoir contains at most its stated capacity, so the algorithm queries degree estimates for at most k times that many variables. By Theorem A.1 and the chosen failure probability, with probability at least $1 - \delta/16$ all queried variables simultaneously satisfy

$$\deg(i) \le \widetilde{f}_i \le \deg(i) + m/((8/\theta)\exp(8k/\theta)).$$

Condition on both events. Their joint probability is at least $1 - 3\delta/16 \ge 1 - \delta/4$.

Let i_s be the s-th variable in the non-increasing ordering used to form $\widetilde{H}$. Suppose for contradiction that $|\widetilde{H}| \geq \exp(8k/\theta)$. For every positive integer s satisfying $s < \exp(8k/\theta)$, the prefix rule and the sketch error give

$$\deg(i_s) > \frac{\theta m}{2s} - \frac{\theta m}{8\exp(8k/\theta)} \geq \frac{3\theta m}{8s}.$$

Consequently,

$$km \geq \sum_{1 \leq s < \exp(8k/\theta)} \deg(i_s) > \frac{3\theta m}{8} \sum_{1 \leq s < \exp(8k/\theta)} \frac{1}{s} \geq \frac{3\theta m}{8} \log\left(\exp(8k/\theta)\right) \geq 3km,$$

which is impossible. Hence $|\widetilde{H}| < \exp(8k/\theta)$.

It remains to prove the residual degree bound. If $i \notin \widetilde{H}$ appears in a sampled constraint, maximality of the prefix and the ordering by $\widetilde{f}_i$ imply

$$\deg(i) \leq \widetilde{f}_i \leq \frac{\theta m}{2(|\widetilde{H}|+1)} \leq \frac{\theta m}{|\widetilde{H}|+1}.$$

If i does not appear in a sampled constraint, the first conditioned event gives $\deg(i) \leq \theta m/(8\exp(8k/\theta))$. Since $|\widetilde{H}| < \exp(8k/\theta)$, we have $|\widetilde{H}| + 1 < \exp(8k/\theta) + 1 \leq 2\exp(8k/\theta)$, and the desired bound follows. □

Lemma C.16. *Fix a set $\widetilde{H}$ satisfying $|\widetilde{H}| < \exp(8k/\theta)$, and fix all oracle labels. With probability at least $1 - \delta/4$ over the hash functions of* **CM***, simultaneously for every assignment to the variables indexed by $\widetilde{H}$, the reconstructed value in (22) differs from (20), with $H = \widetilde{H}$ and $L = \widetilde{L}$, by at most $\frac{\eta m}{2^{k+2}}$.*

Proof. Fix an assignment $\boldsymbol{\sigma}$. For every literal whose variable is indexed by $\widetilde{H}$, the exact factor can be written as

$$\frac{1 + b_t \sigma_{j_t}}{2} = \left(\frac{1}{2} + \frac{b_t Y_{j_t}}{4\varepsilon}\right) + b_t \left(\frac{\sigma_{j_t}}{2} - \frac{Y_{j_t}}{4\varepsilon}\right). \tag{23}$$

Expand the product by selecting any number s of the high-variable positions from which the second term in Equation (23) is taken. The remaining $k - s$ factors are determined by their predicted signs: if exactly i of them have $b_t Y_{j_t} = -1$, their product is $(1/2 + 1/(4\varepsilon))^{k-s-i} (1/2 - 1/(4\varepsilon))^i$. During the stream, **CM** receives exactly one update at the corresponding key for this selected tuple of signed variables. Therefore, if every CountMin query in (22) were replaced by its exact frequency, the displayed expansion would be summed once for every constraint and would equal (20).

Each constraint updates exactly 2^k reconstruction keys, so the frequency vector represented by **CM** has total mass $2^k m$. By Theorem A.1 and the chosen accuracy parameter, every fixed query has additive error at most

$$\frac{\eta \varepsilon^k m}{8(k+1)(1 + 2\exp(8k/\theta))^k}$$

with its prescribed failure probability.

The queried keys are independent of the assignment. Their total number is at most

$$\sum_{s=0}^{k} (k - s + 1) \binom{|\widetilde{H}|}{s} 2^s \leq (k+1) \sum_{s=0}^{k} (2|\widetilde{H}|)^s \leq (k+1)(1 + 2\exp(8k/\theta))^k.$$

A union bound and the chosen failure probability show that every query has the preceding error bound simultaneously with probability at least $1 - \delta/16$, which is at least $1 - \delta/4$.

Every factor in the coefficient of a query has absolute value at most $(2\varepsilon)^{-1}$. There are exactly k factors after combining the s selected terms with the $k - s$ unselected terms. On the simultaneous good event, the reconstruction error for every assignment is therefore at most

$$(k+1)(1 + 2\exp(8k/\theta))^k \cdot \frac{\eta \varepsilon^k m}{8(k+1)(1 + 2\exp(8k/\theta))^k} \cdot (2\varepsilon)^{-k} = \frac{\eta m}{2^{k+3}} \leq \frac{\eta m}{2^{k+2}}.$$

□

Proof of Theorem C.14. Now we are ready to prove Theorem C.14.

Proof. If $m = 0$, the algorithm returns the exact value. Assume $m > 0$. By Lemma C.15, with probability at least $1 - \delta/4$ the computed set $\widetilde{H}$ satisfies the stated cardinality bound and $\deg(i) \le \theta m/(|\widetilde{H}| + 1)$ for every $i \in \widetilde{L}$. Condition on the sampling and degree-sketch randomness that determines such a set. This randomness is independent of the oracle, so the labels indexed by $\widetilde{L}$ remain mutually independent and ε-accurate.

Apply Lemma C.13 with $H = \widetilde{H}$, $L = \widetilde{L}$, and $\eta/2$ in place of η. Since $\theta = \varepsilon^{2k}\eta^2/(64k\log(16/\delta))$, the conditional failure probability is at most

$$2^{|\widetilde{H}|+1} \exp\Big(-8(|\widetilde{H}| + 1)\log(16/\delta)\Big) \le \frac{\delta}{2}.$$

Thus, except with conditional probability at most $\delta/2$, every decoded value differs from its expectation by at most $\eta m/2^{k+2}$.

After conditioning also on the oracle labels, Lemma C.16 contributes failure probability at most $\delta/4$. Combining the events, with probability at least $1 - \delta$, the sketch-computed value for every assignment differs from the corresponding expected decoded value by at most $\eta m/2^{k+1}$. By Lemma C.12, the maximum expected decoded value is OPT. Therefore the returned value ALG satisfies

$$\mathrm{ALG} \le \mathrm{OPT} + \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} = \mathrm{OPT}.$$

Taking $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$ gives

$$\mathrm{ALG} \ge \mathrm{OPT} - \frac{\eta m}{2^{k+1}} - \frac{\eta m}{2^{k+1}} = \mathrm{OPT} - \frac{\eta m}{2^k} \ge (1 - \eta)\mathrm{OPT}.$$

It remains to bound the space. The reservoir capacity and the inverse accuracy of the degree sketch are bounded by $\exp\big(O\big(k^2\varepsilon^{-2k}\eta^{-2}\log(16/\delta)\big)\big)$. The inverse accuracy of $\mathbf{CM}$ is $(8/\eta)\cdot(2/\varepsilon)^k(k+1)(1+2\exp(8k/\theta))^k$. Substituting the value of θ, and absorbing all polynomial and logarithmic factors into the displayed exponential, gives the stated bound. □

C.3.3 Our algorithm in dynamic streams

The dynamic algorithm uses the same value of θ, the same sample size, and the same degree and reconstruction sketches as Algorithm 9. The reservoir is replaced by that many independent ℓ_0-samplers on the active constraint vector. In post-processing, every variable appearing among their outputs is treated as a sampled variable; repeated outputs remain independent draws. The active constraint count m, the degree sketch, and $\mathbf{CM}$ are updated linearly using the stream-update sign Δ_i. Post-processing is otherwise identical to the insertion-only algorithm.

Theorem C.17. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta \in (0,1)$, and $k \ge 3$. For any $\eta \in (0,1)$, given an ε-accurate oracle $\mathcal{O}$ for a* MAX-kAND *instance Ψ, there is a single-pass dynamic streaming algorithm that outputs a $(1-\eta)$-approximation to the maximum value of Ψ with probability at least $1-\delta$. The algorithm uses $\exp\big(O(k^3\varepsilon^{-2k}\eta^{-2}\log(16/\delta))\big)\log n$ words of space.*

Proof. Maintain the number of independent ℓ_0-samplers specified above, with individual failure probabilities chosen so that their total failure probability is at most $\delta/16$. Maintain the degree sketch and $\mathbf{CM}$ with the parameters in Algorithm 9, applying every update with the sign Δ_i of the corresponding dynamic-stream update.

If the final active instance is empty, returning 0 is exact. Assume $m > 0$. With probability at least $1 - \delta/16$, every ℓ_0-sampler succeeds. Conditioned on this event, the samplers give independent uniform samples, with replacement, from the final active constraint set. The sampling argument in Lemma C.15 then contributes failure probability at most $\delta/8$, and the degree-sketch event contributes at most $\delta/16$. Hence, with probability at least $1-\delta/4$, the computed set $\widetilde{H}$ satisfies $|\widetilde{H}| < \exp(8k/\theta)$ and $\deg(i) \le \theta m/(|\widetilde{H}| + 1)$ for every $i \in \widetilde{L}$.

Condition on such a fixed $\widetilde{H}$. The samplers and the degree sketch are independent of the oracle labels, so the labels indexed by $\widetilde{L}$ remain mutually independent and ε-accurate. Applying Lemma C.13 with

$\eta/2$ in place of η gives failure probability at most $\delta/2$ for simultaneous additive error $\eta m/2^{k+2}$ over all assignments.

After conditioning on the oracle labels, the final frequency vector represented by $\mathbf{CM}$ is nonnegative. The sketch is linear in the stream updates, so its final table is exactly the CountMin table of this final vector. Therefore Lemma C.16 applies and contributes failure probability at most $\delta/4$. With total probability at least $1-\delta$, every sketch-computed value differs from the corresponding expected decoded value by at most $\eta m/2^{k+1}$. The same comparison as in the insertion-only proof gives

$$(1-\eta)\mathrm{OPT} \le \mathrm{ALG} \le \mathrm{OPT}.$$

The active-constraint universe has size at most $2^k n^k$. By Theorem A.3, each sampler uses $O(k^2 \log^2 n)$ bits times the logarithm of its inverse failure probability. The number of samplers and all dependence on k, $1/\varepsilon$, $1/\eta$, and $1/\delta$ are absorbed by the displayed exponential factor. Dividing the resulting bit bound by the $\Theta(\log n)$ word size gives the stated word bound. The CountMin sketches require no larger asymptotic space. □

D Extension to MAX-kCSPs: Proof of Theorem 1.1

Proof. The proof follows from a standard black-box reduction to the MAX-kAND problem. In particular, we reduce any instance Ψ of MAX-kCSP(f) to an instance $\widetilde{\Psi}$ of MAX-kAND such that the maximum values of Ψ and $\widetilde{\Psi}$ are equal, while increasing the number of constraints by at most a factor 2^k. Given such a reduction, Theorem 1.1 follows from Theorem C.1, Theorem C.4, Theorem C.14, and Theorem C.17; for $k \ge 3$, the 2^k blow-up is absorbed into the displayed k-dependence, and the dynamic-space factor is written as $\mathrm{poly}(\log n)$.

We now explain how the reduction works. The instance $\widetilde{\Psi}$ is over the same set of n variables $x_1, \ldots, x_n$. Consider any constraint $(f, \mathbf{j}, \mathbf{b})$ in Ψ. For every $\mathbf{a} \in \{-1,1\}^k$ such that $f(\mathbf{a}) = 1$, we add the constraint $(\mathrm{AND}, \mathbf{j}, \mathbf{b} \circ \mathbf{a})$ to $\widetilde{\Psi}$, where $\mathbf{b} \circ \mathbf{a} = (b_1 a_1, \ldots, b_k a_k)$.

We now claim that for every $\boldsymbol{\sigma} \in \{-1,1\}^n$,

$$\mathrm{val}_{\Psi}(\boldsymbol{\sigma}) = \mathrm{val}_{\widetilde{\Psi}}(\boldsymbol{\sigma}).$$

Suppose that $(f, \mathbf{j}, \mathbf{b})$ is satisfied by $\boldsymbol{\sigma}$. Then, $f(b_1\sigma_{j_1}, \ldots, b_k\sigma_{j_k}) = 1$ which implies that there exists an $\mathbf{a} \in \{-1,1\}^k$ such that $f(\mathbf{a}) = 1$ and $b_i\sigma_{j_i} = a_i, \forall i \in [k]$. Equivalently $\sigma_{j_i} = b_i a_i, \forall i \in [k]$. Thus, $\mathrm{AND}(\mathbf{j}, \mathbf{b} \circ \mathbf{a}) = 1$ and by definition, no other constraint of the form $(\mathrm{AND}, \mathbf{j}, \mathbf{b} \circ \mathbf{a}')$ where $\mathbf{a}' \neq \mathbf{a}$ can be satisfied by $\boldsymbol{\sigma}$.

On the other hand, if $(f, \mathbf{j}, \mathbf{b})$ is not satisfied by $\boldsymbol{\sigma}$, then there exists an $\mathbf{a} \in \{-1,1\}^k$ such that $f(\mathbf{a}) = 0$ and $b_i\sigma_{j_i} = a_i, \forall i \in [k]$. Equivalently $\sigma_{j_i} = b_i a_i, \forall i \in [k]$. Thus, $\mathrm{AND}(\mathbf{j}, \mathbf{b} \circ \mathbf{a}) = 1$ and by definition, no constraint in $\{(\mathrm{AND}, \mathbf{j}, \mathbf{b} \circ \mathbf{a}') : f(\mathbf{a}') = 1\}$ can be satisfied by $\boldsymbol{\sigma}$.

Thus, if $(f, \mathbf{j}, \mathbf{b})$ is satisfied by $\boldsymbol{\sigma}$, then exactly one of the constraints in $\{(\mathrm{AND}, \mathbf{j}, \mathbf{b} \circ \mathbf{a}) : f(\mathbf{a}) = 1\}$ is satisfied by $\boldsymbol{\sigma}$. On the other hand, if $(f, \mathbf{j}, \mathbf{b})$ is not satisfied by $\boldsymbol{\sigma}$, then $\boldsymbol{\sigma}$ also does not satisfy any constraint in $\{(\mathrm{AND}, \mathbf{j}, \mathbf{b} \circ \mathbf{a}) : f(\mathbf{a}) = 1\}$. It follows that the maximum values of Ψ and $\widetilde{\Psi}$ are equal. Since $|\{\mathbf{a} : f(\mathbf{a}) = 1\}| \le 2^k$, the bounds on the number of constraints and the maximum variable degree in $\widetilde{\Psi}$ follow immediately. □

E The algorithm with an estimated prediction accuracy parameter

In this section, we show that it suffices to know a highly accurate estimate of the prediction accuracy parameter ε for our algorithms to work. Specifically, it is enough to have an estimate $\varepsilon' \in (0, 1/2]$ such that $\varepsilon' = (1 \pm O(\eta)) \cdot \varepsilon$ for MAX-2CSPs; for MAX-kCSPs, the hidden constant may depend on k. The detailed proof below is for the offline MAX-DICUT algorithm; the corresponding MAX-kAND statement is recorded afterwards.

E.1 Offline algorithm for MAX-DICUT estimation with estimated ε

The algorithm is very similar to Algorithm 2. The only changes are as follows. First, we define the high-degree threshold in Line 1 using ε' instead of ε, namely, θm where $\theta := \frac{\varepsilon'^4 \delta \eta^2}{3072}$. Compared with

Algorithm 10 A learning-augmented *offline* algorithm for MAX-DICUT estimation with estimated ε

Input: Directed graph $G = (V, E)$, an oracle $\mathcal{O} \to \{-1, 1\}$ that is ε-accurate for some *unknown* $\varepsilon \in (0, \frac{1}{2}]$, an estimate $\varepsilon' \in (0, \frac{1}{2}]$ satisfying $|\varepsilon' - \varepsilon| \leq \frac{\eta}{256}\varepsilon$, and $\delta, \eta \in (0, 1)$

Output: Estimate of the MAX-DICUT value of G

1: $m \leftarrow |E|, \theta \leftarrow \frac{\varepsilon'^4 \delta \eta^2}{3072}, L \leftarrow \{v \in V : \deg(v) < \theta m\}, H \leftarrow V \setminus L$
2: **initialize** $X_0 \leftarrow 0, X_1 \leftarrow 0, X_2 \leftarrow 0$.
3: **initialize** $X_{v,L^+} \leftarrow 0, X_{v,L^-} \leftarrow 0, X_{L^+,v} \leftarrow 0, X_{L^-,v} \leftarrow 0$ for each $v \in H$.
4: **for each** directed edge $(u, v) \in E$ **do**
5: $\quad Y_u = \mathcal{O}(u), Y_v = \mathcal{O}(v)$
6: $\quad$ **if** $u \in L$ and $v \in L$ **then**
7: $\quad\quad$ **if** $Y_u = 1$ and $Y_v = -1$ **then** $X_0 \leftarrow X_0 + 1$
8: $\quad\quad$ **if** $Y_u = Y_v$ **then** $X_1 \leftarrow X_1 + 1$
9: $\quad\quad$ **if** $Y_u = -1$ and $Y_v = 1$ **then** $X_2 \leftarrow X_2 + 1$
10: $\quad$ **if** $u \in H$ and $v \in L$ **then**
11: $\quad\quad$ **if** $Y_v = 1$ **then** $X_{u,L^+} \leftarrow X_{u,L^+} + 1$ **else** $X_{u,L^-} \leftarrow X_{u,L^-} + 1$
12: $\quad$ **if** $u \in L$ and $v \in H$ **then**
13: $\quad\quad$ **if** $Y_u = 1$ **then** $X_{L^+,v} \leftarrow X_{L^+,v} + 1$ **else** $X_{L^-,v} \leftarrow X_{L^-,v} + 1$
14: $\widetilde{k}_L \leftarrow \frac{1}{4\varepsilon'^2}[(\frac{1}{2} + \varepsilon')^2 X_0 - (\frac{1}{2} + \varepsilon')(\frac{1}{2} - \varepsilon')X_1 + (\frac{1}{2} - \varepsilon')^2 X_2]$
15: **for each** $v \in H$ **do**
16: $\quad \widetilde{k}_{S^*,v} \leftarrow (\frac{1}{2} + \frac{1}{4\varepsilon'})X_{L^+,v} + (\frac{1}{2} - \frac{1}{4\varepsilon'})X_{L^-,v}, \widetilde{k}_{v,T^*} \leftarrow (\frac{1}{2} - \frac{1}{4\varepsilon'})X_{v,L^+} + (\frac{1}{2} + \frac{1}{4\varepsilon'})X_{v,L^-}$
17: **return** $\widetilde{k}_L + \sum_{v \in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16}$

Algorithm 2, the denominator constant is increased from 768 to 3072 to accommodate the additional error caused by replacing ε by ε'. Second, we replace every occurrence of ε in the decoding estimators by ε' (see Line 14 and Line 16). The resulting algorithm is formally given in Algorithm 10.

Next, we analyze the approximation guarantee of Algorithm 10.

Theorem E.1. *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta, \eta \in (0, 1)$, and G be a directed graph with $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$. Given an oracle $\mathcal{O}$ that is ε-accurate for the* MAX-DICUT *of G, and an estimate $\varepsilon' \in (0, \frac{1}{2}]$ satisfying $|\varepsilon' - \varepsilon| \leq \frac{\eta}{256}\varepsilon$, Algorithm 10 provides a $(1 - \eta)$-approximation for the* MAX-DICUT *value of G with probability at least $1 - \delta$.*

E.1.1 Analysis of Algorithm 10

We first reprove concentration bounds for X_0, X_1, X_2 with the new high-degree threshold.

Lemma E.2. *Let $\varepsilon' \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0, 1)$. For each $i \in \{0, 1, 2\}$, we have $|X_i - \mathbb{E}[X_i]| \leq \frac{\varepsilon'^2 \eta m}{16}$ with probability at least $1 - \frac{\delta}{6}$.*

Proof. Let $E_L := \{(u, v) \in E : u, v \in L\}$ denote the set of directed edges with both endpoints in L. Let $m_L := |E_L|$. For each edge $(u, v) \in E_L$, define the following indicator random variables:

$$X_{0,uv} = \begin{cases} 1 & \text{if } Y_u = 1 \wedge Y_v = -1, \\ 0 & \text{otherwise,} \end{cases} \quad X_{1,uv} = \begin{cases} 1 & \text{if } Y_u = Y_v, \\ 0 & \text{otherwise,} \end{cases} \quad X_{2,uv} = \begin{cases} 1 & \text{if } Y_u = -1 \wedge Y_v = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that for $i \in \{0, 1, 2\}$, we have $\mathbb{E}[X_{i,uv}] = \Pr[X_{i,uv} = 1]$ and $X_i = \sum_{(u,v) \in E_L} X_{i,uv}$. Therefore,

$$\begin{aligned} \mathbb{E}\big[X_i^2\big] &= \mathbb{E}\left[\left(\sum_{(u,v) \in E_L} X_{i,uv}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{(u,v) \in E_L} \sum_{(w,z) \in E_L} X_{i,uv} X_{i,wz}\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \mathbb{E}[X_{i,uv}X_{i,wz}] \\
&= \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] \\
&= \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}=\emptyset}} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] + \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}\neq\emptyset}} \Pr[X_{i,uv}=1 \wedge X_{i,wz}=1] \\
&\le \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}=\emptyset}} \Pr[X_{i,uv}=1]\Pr[X_{i,wz}=1] + \sum_{\substack{(u,v),(w,z)\in E_L:\\ \{u,v\}\cap\{w,z\}\neq\emptyset}} 1 \\
&\le \sum_{(u,v)\in E_L} \sum_{(w,z)\in E_L} \mathbb{E}[X_{i,uv}]\mathbb{E}[X_{i,wz}] + m_L\cdot 2\cdot\theta m \\
&\le (\mathbb{E}[\sum_{(u,v)\in E_L} X_{i,uv}])^2 + 2\theta m^2 \\
&= (\mathbb{E}[X_i])^2 + 2\theta m^2,
\end{aligned}
$$

where the third-to-last step follows from $\deg(v) < \theta m$ for $v \in L$, and the second-to-last step follows from $m_L \le m$.

Then we have $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \le 2\theta m^2$. By Chebyshev's inequality,

$$
\begin{aligned}
\Pr\left[|X_i - \mathbb{E}[X_i]| > \frac{\varepsilon'^2\eta m}{16}\right] &\le \frac{\mathrm{Var}[X_i]}{(\varepsilon'^2\eta m/16)^2} \le \frac{512\theta m^2}{\varepsilon'^4\eta^2 m^2} && \triangleright\ \mathrm{Var}[X_i] \le 2\theta m^2 \\
&= \frac{512\theta}{\varepsilon'^4\eta^2} < \frac{\delta}{6}. && \triangleright\ \theta < \frac{\varepsilon'^4\delta\eta^2}{3072}
\end{aligned}
$$

□

Define $k_L := |\{(u,v) \in E : u \in S^* \cap L, v \in T^* \cap L\}|$ to be the number of directed edges with both endpoints in L (i.e., both low-degree) that contribute to the MAX-DICUT value. The next lemma shows that the modified estimator for k_L remains accurate when ε is replaced by ε'.

Lemma E.3. *Let $\delta, \eta \in (0,1)$. We have $|\widetilde{k}_L - k_L| \le \frac{\eta m}{32}$ with probability at least $1 - \frac{\delta}{2}$.*

Proof. Let k_0 denote the number of edges from $S^* \cap L$ to $T^* \cap L$. Let k_1 denote the number of edges whose endpoints both lie in $S^* \cap L$ or both lie in $T^* \cap L$. Let k_2 denote the number of edges from $T^* \cap L$ to $S^* \cap L$. Note that $k_0 = k_L$. We have

$$
\mathbb{E}[X_0] = \left(\frac{1}{2}+\varepsilon\right)^2 \cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_1 + \left(\frac{1}{2}-\varepsilon\right)^2 \cdot k_2, \tag{24}
$$

$$
\mathbb{E}[X_1] = 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_0 + \left[\left(\frac{1}{2}+\varepsilon\right)^2 + \left(\frac{1}{2}-\varepsilon\right)^2\right]\cdot k_1 + 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_2, \tag{25}
$$

$$
\mathbb{E}[X_2] = \left(\frac{1}{2}-\varepsilon\right)^2 \cdot k_0 + \left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)\cdot k_1 + \left(\frac{1}{2}+\varepsilon\right)^2 \cdot k_2. \tag{26}
$$

To bound $|\widetilde{k}_L - k_L|$, we apply the triangle inequality: $|\widetilde{k}_L - k_L| \le |\widetilde{k}_L - \mathbb{E}[\widetilde{k}_L]| + |\mathbb{E}[\widetilde{k}_L] - k_L|$.

We first bound the second term. By Line 14 of Algorithm 10, we have $\widetilde{k}_L := \frac{1}{4\varepsilon'^2}[(\frac{1}{2}+\varepsilon')^2 X_0 - (\frac{1}{2}+\varepsilon')(\frac{1}{2}-\varepsilon')X_1 + (\frac{1}{2}-\varepsilon')^2 X_2]$. Therefore,

$$
\begin{aligned}
\mathbb{E}[\widetilde{k}_L] &= \frac{1}{4\varepsilon'^2}\left[\left(\frac{1}{2}+\varepsilon'\right)^2\mathbb{E}[X_0] - \left(\frac{1}{2}+\varepsilon'\right)\left(\frac{1}{2}-\varepsilon'\right)\mathbb{E}[X_1] + \left(\frac{1}{2}-\varepsilon'\right)^2\mathbb{E}[X_2]\right] \\
&= \frac{(\varepsilon+\varepsilon')^2}{4\varepsilon'^2}\cdot k_0 + \frac{\varepsilon'^2-\varepsilon^2}{4\varepsilon'^2}\cdot k_1 + \frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2}\cdot k_2,
\end{aligned}
$$

where the second equality follows by substituting Equations (24)–(26). Then we have

$$\begin{aligned}|\mathbb{E}[\widetilde{k}_L] - k_L| &= \left|\left(\frac{(\varepsilon+\varepsilon')^2}{4\varepsilon'^2} - 1\right)\cdot k_0 + \frac{\varepsilon'^2-\varepsilon^2}{4\varepsilon'^2}\cdot k_1 + \frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2}\cdot k_2\right| \\ &\le \left|\frac{(\varepsilon+\varepsilon')^2}{4\varepsilon'^2} - 1\right|\cdot k_0 + \left|\frac{\varepsilon'^2-\varepsilon^2}{4\varepsilon'^2}\right|\cdot k_1 + \left|\frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2}\right|\cdot k_2.\end{aligned}$$

Let $q := \frac{|\varepsilon-\varepsilon'|}{\varepsilon'}$. We have $q \le \frac{\frac{\eta}{256}\varepsilon}{(1-\frac{\eta}{256})\varepsilon} = \frac{\eta}{256-\eta} < \frac{\eta}{255}$. Then

$$\begin{aligned}\left|\frac{(\varepsilon+\varepsilon')^2}{4\varepsilon'^2} - 1\right| &= \left|\frac{\varepsilon-\varepsilon'}{\varepsilon'} + \frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2}\right| \le \frac{|\varepsilon-\varepsilon'|}{\varepsilon'} + \frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2} = q + \frac{q^2}{4} < 2q, \\ \left|\frac{\varepsilon'^2-\varepsilon^2}{4\varepsilon'^2}\right| &= \frac{|\varepsilon-\varepsilon'|\cdot|\varepsilon+\varepsilon'|}{4\varepsilon'^2} \le \frac{q\varepsilon'\cdot(|\varepsilon-\varepsilon'|+2\varepsilon')}{4\varepsilon'^2} = \frac{q(2+q)}{4} < 2q, \\ \left|\frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2}\right| &= \frac{(\varepsilon-\varepsilon')^2}{4\varepsilon'^2} = \frac{q^2}{4} < 2q.\end{aligned}$$

Therefore, $|\mathbb{E}[\widetilde{k}_L] - k_L| < 2q(k_0+k_1+k_2) \le 2qm < \frac{2\eta m}{255} < \frac{\eta m}{64}$.

Next, we bound $|\widetilde{k}_L - \mathbb{E}[\widetilde{k}_L]|$. By applying union bound on Lemma E.2, we have that with probability at least $1 - \frac{\delta}{6}\cdot 3 = 1 - \frac{\delta}{2}$, $|X_i - \mathbb{E}[X_i]| \le \frac{\varepsilon'^2\eta m}{16}$ for all $i \in \{0,1,2\}$. Therefore,

$$\begin{aligned}&|\widetilde{k}_L - \mathbb{E}[\widetilde{k}_L]| \\ &\le \frac{1}{4\varepsilon'^2}\left[\left(\frac{1}{2}+\varepsilon'\right)^2|X_0 - \mathbb{E}[X_0]| + \left(\frac{1}{2}+\varepsilon'\right)\left(\frac{1}{2}-\varepsilon'\right)|X_1 - \mathbb{E}[X_1]| + \left(\frac{1}{2}-\varepsilon'\right)^2|X_2 - \mathbb{E}[X_2]|\right] \\ &\le \frac{1}{4\varepsilon'^2}\left[\left(\frac{1}{2}+\varepsilon'\right)^2 + \left(\frac{1}{2}+\varepsilon'\right)\left(\frac{1}{2}-\varepsilon'\right) + \left(\frac{1}{2}-\varepsilon'\right)^2\right]\cdot\frac{\varepsilon'^2\eta m}{16} \\ &= \frac{1}{4}\cdot\left(\frac{3}{4}+\varepsilon'^2\right)\cdot\frac{\eta m}{16} \le \frac{\eta m}{64}.\end{aligned}$$

Therefore, we have $|\widetilde{k}_L - k_L| \le |\widetilde{k}_L - \mathbb{E}[\widetilde{k}_L]| + |\mathbb{E}[\widetilde{k}_L] - k_L| \le \frac{\eta m}{64} + \frac{\eta m}{64} = \frac{\eta m}{32}$ with probability at least $1 - \frac{\delta}{2}$. □

For each high-degree vertex $v \in H$, define

1. $k_{S^*,v} := |\{(u,v) \in E : u \in S^* \cap L\}|$ to be the number of directed edges from $S^* \cap L$ to v.
2. $k_{T^*,v} := |\{(u,v) \in E : u \in T^* \cap L\}|$ to be the number of directed edges from $T^* \cap L$ to v.
3. $k_{v,S^*} := |\{(v,u) \in E : u \in S^* \cap L\}|$ to be the number of directed edges from v to $S^* \cap L$.
4. $k_{v,T^*} := |\{(v,u) \in E : u \in T^* \cap L\}|$ to be the number of directed edges from v to $T^* \cap L$.

Lemma E.4. *Let $\varepsilon' \in (0, \frac{1}{2}]$ and $\delta, \eta \in (0,1)$. For each vertex $v \in H$, each of the following inequalities holds with probability at least $1 - \frac{\varepsilon'^4\delta^2\eta^2}{24576}$:*

(1) $|\widetilde{k}_{S^*,v} - k_{S^*,v}| \le \frac{\varepsilon'^4\delta\eta^3 m}{393216} + \frac{q}{2}\cdot e(L,v)$,

(2) $|\widetilde{k}_{v,T^*} - k_{v,T^*}| \le \frac{\varepsilon'^4\delta\eta^3 m}{393216} + \frac{q}{2}\cdot e(v,L)$,

where $q := \frac{|\varepsilon-\varepsilon'|}{\varepsilon'}$.

Proof. We prove the first inequality below; the second inequality follows by the same argument.

We have

$$\mathbb{E}[X_{L^+,v}] = \left(\frac{1}{2}+\varepsilon\right)\cdot k_{S^*,v} + \left(\frac{1}{2}-\varepsilon\right)\cdot k_{T^*,v}, \tag{27}$$

$$\mathbb{E}[X_{L^-,v}] = \left(\frac{1}{2} - \varepsilon\right) \cdot k_{S^*,v} + \left(\frac{1}{2} + \varepsilon\right) \cdot k_{T^*,v}. \tag{28}$$

To bound $|\widetilde{k}_{S^*,v} - k_{S^*,v}|$, we apply the triangle inequality: $|\widetilde{k}_{S^*,v} - k_{S^*,v}| \le |\widetilde{k}_{S^*,v} - \mathbb{E}[\widetilde{k}_{S^*,v}]| + |\mathbb{E}[\widetilde{k}_{S^*,v}] - k_{S^*,v}|$.

We first bound the second term. By Line 16 of Algorithm 10, we have $\widetilde{k}_{S^*,v} := (\frac{1}{2} + \frac{1}{4\varepsilon'})X_{L^+,v} + (\frac{1}{2} - \frac{1}{4\varepsilon'})X_{L^-,v}$. Hence,

$$\begin{aligned}
\mathbb{E}[\widetilde{k}_{S^*,v}] &= \left(\frac{1}{2} + \frac{1}{4\varepsilon'}\right)\mathbb{E}[X_{L^+,v}] + \left(\frac{1}{2} - \frac{1}{4\varepsilon'}\right)\mathbb{E}[X_{L^-,v}] \\
&= \frac{\varepsilon + \varepsilon'}{2\varepsilon'} \cdot k_{S^*,v} + \frac{\varepsilon' - \varepsilon}{2\varepsilon'} \cdot k_{T^*,v} \\
&= k_{S^*,v} + \frac{\varepsilon - \varepsilon'}{2\varepsilon'} \cdot (k_{S^*,v} - k_{T^*,v}).
\end{aligned}$$

Therefore, we have

$$|\mathbb{E}[\widetilde{k}_{S^*,v}] - k_{S^*,v}| = \left|\frac{\varepsilon - \varepsilon'}{2\varepsilon'} \cdot (k_{S^*,v} - k_{T^*,v})\right| = \left|\frac{\varepsilon - \varepsilon'}{2\varepsilon'}\right| \cdot |k_{S^*,v} - k_{T^*,v}| \le \frac{q}{2} \cdot e(L, v).$$

Next, we bound $|\widetilde{k}_{S^*,v} - \mathbb{E}[\widetilde{k}_{S^*,v}]|$. Note that

$$\begin{aligned}
\widetilde{k}_{S^*,v} &= \left(\frac{1}{2} + \frac{1}{4\varepsilon'}\right) X_{L^+,v} + \left(\frac{1}{2} - \frac{1}{4\varepsilon'}\right) X_{L^-,v} \\
&= \left(\frac{1}{2} + \frac{1}{4\varepsilon'}\right) X_{L^+,v} + \left(\frac{1}{2} - \frac{1}{4\varepsilon'}\right)(e(L, v) - X_{L^+,v}) \\
&= \frac{1}{2\varepsilon'} X_{L^+,v} + \left(\frac{1}{2} - \frac{1}{4\varepsilon'}\right) e(L, v).
\end{aligned}$$

Therefore, $\mathrm{Var}[\widetilde{k}_{S^*,v}] = \frac{1}{4\varepsilon'^2}\mathrm{Var}[X_{L^+,v}]$.

Given a high-degree vertex $v \in H$, for each incoming edge $(u, v) \in E$ with $u \in L$, define an indicator random variable X_{uv} as follows:

$$X_{uv} = \begin{cases} 1 & \text{if } Y_u = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $X_{L^+,v} = \sum_{(u,v)\in E: u\in L} X_{uv}$ and $\mathbb{E}[X_{uv}] = \Pr[X_{uv} = 1]$. Then $\mathrm{Var}[X_{uv}] = \mathbb{E}[X_{uv}^2] - (\mathbb{E}[X_{uv}])^2 = \mathbb{E}[X_{uv}] \cdot (1 - \mathbb{E}[X_{uv}]) = (\frac{1}{2} + \varepsilon)(\frac{1}{2} - \varepsilon) = \frac{1}{4} - \varepsilon^2 \le \frac{1}{4}$. Since X_{uv}'s are independent, we have

$$\mathrm{Var}[X_{L^+,v}] = \sum_{(u,v)\in E: u\in L} \mathrm{Var}[X_{uv}] \le \frac{m}{4}.$$

So, $\mathrm{Var}[\widetilde{k}_{S^*,v}] = \frac{1}{4\varepsilon'^2}\mathrm{Var}[X_{L^+,v}] \le \frac{m}{16\varepsilon'^2}$. By Chebyshev's inequality,

$$\Pr\left[|\widetilde{k}_{S^*,v} - \mathbb{E}[\widetilde{k}_{S^*,v}]| \ge \frac{\varepsilon'^4\delta\eta^3 m}{393216}\right] \le \frac{\mathrm{Var}[\widetilde{k}_{S^*,v}]}{(\varepsilon'^4\delta\eta^3 m/393216)^2} \le \frac{393216^2/16}{\varepsilon'^{10}\delta^2\eta^6 m} \le \frac{\varepsilon'^4\delta^2\eta^2}{24576},$$

where the last step follows from our assumption that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$ and $\varepsilon' = \Theta(\varepsilon)$.

Therefore, we have $|\widetilde{k}_{S^*,v} - k_{S^*,v}| \le |\widetilde{k}_{S^*,v} - \mathbb{E}[\widetilde{k}_{S^*,v}]| + |\mathbb{E}[\widetilde{k}_{S^*,v}] - k_{S^*,v}| \le \frac{\varepsilon'^4\delta\eta^3 m}{393216} + \frac{q}{2} \cdot e(L, v)$ with probability at least $1 - \frac{\varepsilon'^4\delta^2\eta^2}{24576}$. □

Define $k_H := |\{(u, v) \in E : u \in S^* \cap H, v \in T^* \cap H\}|$ to be the number of directed edges with both endpoints in H (i.e., both high-degree) that contribute to the MAX-DICUT value. Note that $|H| \le \frac{2m}{\theta m} = \frac{2}{\theta} = \frac{6144}{\varepsilon'^4\delta\eta^2}$. Thus, we have the following observation.

Fact E.5. *It holds that* $k_H < |H|^2 \le \frac{6144^2}{\varepsilon'^8\delta^2\eta^4}$.

E.1.2 Proof of Theorem E.1

Proof. Recall that k_L (resp. k_H) is the number of directed edges with both endpoints in L (resp. H) that contribute to the MAX-DICUT value.

Furthermore, define

$$k_{L,H} := |\{(u,v) \in E : u \in S^* \cap L, v \in T^* \cap H\}|$$

to be the number of directed edges from L to H that contribute to the MAX-DICUT value, and

$$k_{H,L} := |\{(u,v) \in E : u \in S^* \cap H, v \in T^* \cap L\}|$$

to be the number of directed edges from H to L that contribute to the MAX-DICUT value.

Therefore, we have $\mathrm{OPT} = k_L + k_H + k_{L,H} + k_{H,L}$.

For each $v \in H$, define $c_v := \frac{\varepsilon'^4 \delta \eta^3 m}{393216} + \frac{q}{2} \cdot (e(L,v) + e(v,L))$. By Lemma E.4, we have that for each vertex $v \in H$, each of the following inequalities holds with probability at least $1 - \frac{\varepsilon'^4 \delta^2 \eta^2}{24576}$:

$$|\widetilde{k}_{S^*,v} - k_{S^*,v}| \le \frac{\varepsilon'^4 \delta \eta^3 m}{393216} + \frac{q}{2} \cdot e(L,v) \le c_v,$$
$$|\widetilde{k}_{v,T^*} - k_{v,T^*}| \le \frac{\varepsilon'^4 \delta \eta^3 m}{393216} + \frac{q}{2} \cdot e(v,L) \le c_v.$$

Hence, $|\max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \max\{k_{S^*,v}, k_{v,T^*}\}| \le c_v$. Therefore, with probability at least $1 - \frac{\varepsilon'^4 \delta^2 \eta^2}{24576} \cdot 2 \cdot |H| \ge 1 - \frac{\delta}{2}$, we have

$$\begin{aligned}
&\left| \sum_{v \in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \sum_{v \in H} \max\{k_{S^*,v}, k_{v,T^*}\} \right| \\
&\le \sum_{v \in H} |\max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \max\{k_{S^*,v}, k_{v,T^*}\}| \\
&\le \sum_{v \in H} c_v \\
&= |H| \cdot \frac{\varepsilon'^4 \delta \eta^3 m}{393216} + \frac{q}{2} \sum_{v \in H} (e(L,v) + e(v,L)) \\
&\le \frac{\eta m}{64} + \frac{qm}{2} && \triangleright \sum_{v \in H} (e(L,v) + e(v,L)) \le m \\
&< \frac{\eta m}{64} + \frac{\eta m}{510} && \triangleright\, q := \frac{|\varepsilon - \varepsilon'|}{\varepsilon'} = \frac{\eta}{256 - \eta} < \frac{\eta}{255} \\
&< \frac{\eta m}{32}.
\end{aligned} \tag{29}$$

Let ALG denote the output of Algorithm 10. By union bound, with probability at least $1 - \frac{\delta}{2} - \frac{\delta}{2} = 1 - \delta$, we have

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{v \in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\le k_L + \frac{\eta m}{32} + \sum_{v \in H} \max\{k_{S^*,v}, k_{v,T^*}\} + \frac{\eta m}{32} - \frac{\eta m}{16} && \triangleright \text{Lemma E.3 and Inequality (29)} \\
&= k_L + \sum_{v \in H} \max\{k_{S^*,v}, k_{v,T^*}\} \\
&\le \mathrm{OPT},
\end{aligned}$$

where the last step follows since $k_L + \sum_{v \in H} \max\{k_{S^*,v}, k_{v,T^*}\}$ is a lower bound on the value of a feasible dicut of G. Specifically, we fix the restriction on L to be $(S^* \cap L, T^* \cap L)$, and then assign each vertex $v \in H$ to the side that maximizes the resulting dicut value.

Furthermore,

$$\begin{aligned}
\mathrm{ALG} &= \widetilde{k}_L + \sum_{v\in H} \max\{\widetilde{k}_{S^*,v}, \widetilde{k}_{v,T^*}\} - \frac{\eta m}{16} \\
&\geq k_L - \frac{\eta m}{32} + \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} - \frac{\eta m}{32} - \frac{\eta m}{16} && \triangleright \text{ Lemma E.3 and Inequality (29)} \\
&\geq k_L + k_{L,H} + k_{H,L} - \frac{\eta m}{8} && \triangleright \sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} \geq k_{L,H} + k_{H,L} \\
&= \mathrm{OPT} - k_H - \frac{\eta m}{8} && \triangleright \mathrm{OPT} = k_L + k_H + k_{L,H} + k_{H,L} \\
&\geq \mathrm{OPT} - \frac{6144^2}{\varepsilon'^8\delta^2\eta^4} - \frac{\eta}{2}\cdot \mathrm{OPT} && \triangleright \text{ Fact E.5 and } m \leq 4\mathrm{OPT} \\
&\geq (1-\eta)\cdot \mathrm{OPT}, && \triangleright \frac{6144^2}{\varepsilon'^8\delta^2\eta^4} \leq \frac{\eta m}{8} \leq \frac{\eta}{2}\cdot \mathrm{OPT}
\end{aligned}$$

where the third step follows from

$$\begin{aligned}
\sum_{v\in H} \max\{k_{S^*,v}, k_{v,T^*}\} &= \sum_{v\in H\cap S^*} \max\{k_{S^*,v}, k_{v,T^*}\} + \sum_{v\in H\cap T^*} \max\{k_{S^*,v}, k_{v,T^*}\} \\
&\geq \sum_{v\in H\cap S^*} k_{v,T^*} + \sum_{v\in H\cap T^*} k_{S^*,v} \\
&= k_{H,L} + k_{L,H},
\end{aligned}$$

and the last step follows from our assumption that $m = \Omega(\varepsilon^{-26}\delta^{-10}\eta^{-14})$ and $\varepsilon' = \Theta(\varepsilon)$.

□

E.2 Estimated ε for MAX-kAND

We also use the same estimated-ε modification for the MAX-kAND algorithms. Namely, replace ε by ε' in every decoder coefficient and in every threshold or sketch parameter, run the corresponding exact-ε algorithm with accuracy parameter $\eta/8$, and use final subtractive slack $\eta m/2^{k+2}$. This applies to:

1. the bounded-degree algorithm Algorithm 7, with the bounded-degree assumption stated using ε' and $\eta/8$;
2. the offline unbounded-degree algorithm Algorithm 8, with its adaptive threshold stated using ε' and $\eta/8$;
3. the insertion-only and dynamic unbounded-degree algorithms in Theorems C.14 and C.17, with all sketch parameters stated using ε' and $\eta/8$.

The required accuracy of ε' is as follows.

Corollary E.6 (Estimated ε for MAX-kAND)**.** *Let $\varepsilon \in (0, \frac{1}{2}]$, $\delta, \eta \in (0,1)$, and $k \geq 3$. If the oracle is ε-accurate and*

$$\left|\frac{\varepsilon' - \varepsilon}{2\varepsilon'}\right| \leq \frac{\eta}{2^{k+5}k},$$

then each modified algorithm above remains a $(1-\eta)$-approximation algorithm under the same assumptions as its exact-ε version. The running-time and space bounds are unchanged up to constants, with ε' in place of ε. Equivalently, it suffices to have $\varepsilon' = (1 \pm O(\eta/(k2^k)))\varepsilon$.

Proof. Fix one of the modified algorithms and fix the high set used by it; for Algorithm 7 we take $H = \emptyset$ and $L = [n]$. Fix an assignment $\boldsymbol{\tau}_H \in \{\pm 1\}^H$, and let $\boldsymbol{\tau}$ be the assignment that agrees with $\boldsymbol{\tau}_H$ on H and with $\boldsymbol{\sigma}^*$ on L. We compare the modified decoded value with $\mathrm{val}_\Psi(\boldsymbol{\tau})$.

We first record the one-literal calculation. Since $\mathbb{E}[Y_i] = 2\varepsilon\sigma_i^*$,

$$\mathbb{E}\left[\frac{1}{2} + \frac{bY_i}{4\varepsilon'}\right] = \frac{1}{2} + \frac{b\mathbb{E}[Y_i]}{4\varepsilon'} = \frac{1}{2} + \frac{b\sigma_i^*\varepsilon}{2\varepsilon'} = \begin{cases} 1 - \frac{\varepsilon'-\varepsilon}{2\varepsilon'}, & \text{if } b\sigma_i^* = 1, \\ \frac{\varepsilon'-\varepsilon}{2\varepsilon'}, & \text{if } b\sigma_i^* = -1. \end{cases}$$

The bounded-degree counter decoder is exactly the product of these one-literal decoders. Indeed, if a constraint has exactly i literals not satisfied by the predicted assignment, then

$$\prod_{t=1}^{k}\left(\frac{1}{2}+\frac{b_t Y_{j_t}}{4\varepsilon'}\right)=\left(\frac{1}{2}+\frac{1}{4\varepsilon'}\right)^{k-i}\left(\frac{1}{2}-\frac{1}{4\varepsilon'}\right)^{i}=\frac{1}{(2\varepsilon')^k}\left(\frac{1}{2}+\varepsilon'\right)^{k-i}\left(-\frac{1}{2}+\varepsilon'\right)^{i},$$

which is the coefficient used by Algorithm 7 after replacing ε by ε'.

Now consider a constraint $C=((j_1,\dots,j_k),(b_1,\dots,b_k))$. If some literal with $j_t\in H$ is false under $\boldsymbol{\tau}_H$, then the contribution of C to both the decoded expression and $\mathrm{val}_\Psi(\boldsymbol{\tau})$ is zero. Otherwise, suppose that there are ℓ low variables in C, and exactly s of their literals are false under $\boldsymbol{\sigma}^*$. Since the coordinates of C are distinct, the corresponding oracle labels are independent, and the expected decoded contribution of C is

$$\left(1-\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right)^{\ell-s}\left(\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right)^{s}.$$

The true contribution of C to $\mathrm{val}_\Psi(\boldsymbol{\tau})$ is 1 if $s=0$ and 0 otherwise. Under the assumption $\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\le 1/(2k)$, if $s=0$ then

$$\left|\left(1-\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right)^{\ell}-1\right|\le \ell\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\left(1+\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\right)^{\ell-1}\le 2k\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|,$$

and if $s\ge 1$ then

$$\left|\left(1-\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right)^{\ell-s}\left(\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right)^{s}\right|\le\left(1+\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\right)^{k-1}\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\le 2\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\le 2k\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|.$$

Thus, for every fixed $\boldsymbol{\tau}_H$,

$$\left|\mathbb{E}[\widetilde{\mathrm{val}}_{\varepsilon'}(\boldsymbol{\tau}_H)]-\mathrm{val}_\Psi(\boldsymbol{\tau})\right|\le 2km\left|\frac{\varepsilon'-\varepsilon}{2\varepsilon'}\right|\le\frac{\eta m}{2^{k+4}}. \tag{30}$$

Here $\widetilde{\mathrm{val}}_{\varepsilon'}$ denotes the decoded value after replacing ε by ε'.

Next we bound the centered error around the new expectation. For the bounded-degree algorithm, applying Lemma C.9 with ε' and internal accuracy parameter $\eta/8$ gives, after a union bound,

$$|X_i-\mathbb{E}[X_i]|\le\frac{(\varepsilon')^k\eta m}{16}\quad\text{for all } i\in\{0,\dots,k\}.$$

Therefore the centered error of the decoded counter value is at most

$$\frac{(\varepsilon')^k\eta m}{16(2\varepsilon')^k}\sum_{i=0}^{k}\left(\frac{1}{2}+\varepsilon'\right)^{k-i}\left(\frac{1}{2}-\varepsilon'\right)^{i}\le\frac{\eta m}{2^{k+4}}, \tag{31}$$

where we used $\sum_{i=0}^{k}(\frac{1}{2}+\varepsilon')^{k-i}(\frac{1}{2}-\varepsilon')^i\le 1$.

For the unbounded-degree offline algorithm, the same bounded-differences calculation as in Lemma C.13 applies with ε' in place of ε. Let $r=|H|$. By the stopping condition, every $i\in L$ satisfies $\deg(i)\le\theta m/(r+1)$. For a fixed high assignment, changing the low-variable oracle label Y_i changes one factor by $1/(2\varepsilon')$, and all other decoded factors have absolute value at most $(2\varepsilon')^{-1}$. Hence changing Y_i changes the decoded value by at most

$$c_i=(2\varepsilon')^{-k}\deg(i),$$

and

$$\sum_{i\in L}c_i^2=(2\varepsilon')^{-2k}\sum_{i\in L}\deg(i)^2\le(2\varepsilon')^{-2k}\frac{\theta m}{r+1}\sum_{i\in L}\deg(i)\le\frac{k\theta m^2}{(2\varepsilon')^{2k}(r+1)}.$$

With internal accuracy $\eta/8$, the threshold is chosen so that the bounded-differences inequality and a union bound over the 2^r high assignments give

$$\Pr\left[\exists\boldsymbol{\tau}_H:\left|\widetilde{\mathrm{val}}_{\varepsilon'}(\boldsymbol{\tau}_H)-\mathbb{E}[\widetilde{\mathrm{val}}_{\varepsilon'}(\boldsymbol{\tau}_H)]\right|>\frac{\eta m}{2^{k+4}}\right]\le\delta. \tag{32}$$

In the streaming algorithms, the reconstruction proofs are algebraic and use only the coefficient bound $(2\varepsilon')^{-k}$; after the same replacement of parameters and with internal accuracy $\eta/8$, the additional sketching error is also at most $\eta m/2^{k+4}$ uniformly over all candidates.

Combining Equations (30) to (32) and the streaming reconstruction bound, every decoded candidate value differs from the value of its corresponding feasible assignment by at most $\eta m/2^{k+2}$. The maximum of these feasible-assignment values is OPT: choosing $\boldsymbol{\tau}_H = \boldsymbol{\sigma}^*|_H$ gives OPT, while every other choice is still a feasible assignment. Hence, if M denotes the maximum decoded candidate value, then

$$|M - \mathrm{OPT}| \le \frac{\eta m}{2^{k+2}}.$$

The modified algorithms output $\mathrm{ALG} = M - \eta m/2^{k+2}$. Therefore

$$\mathrm{ALG} \le \mathrm{OPT},$$

and, by taking $\boldsymbol{\tau}_H = \boldsymbol{\sigma}^*|_H$ in the maximum,

$$\mathrm{ALG} \ge \mathrm{OPT} - \frac{\eta m}{2^{k+2}} - \frac{\eta m}{2^{k+2}} \ge \mathrm{OPT} - \frac{\eta m}{2^k} \ge (1-\eta)\mathrm{OPT},$$

where the last step uses $\mathrm{OPT} \ge m/2^k$ for MAX-kAND. The proof uses independence only among the oracle labels of distinct variables in this single oracle vector, exactly as in the exact-ε proofs; it does not use independence between repeated executions of the algorithm. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. These claims accurately reflect our theoretical results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We explicitly state the assumptions underlying each of our theoretical results.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides the full set of assumptions and a complete (and correct) proof for each theoretical result, either in the main paper or the appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: The paper does not include experiments requiring code.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conforms, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper is purely theoretical, with contributions consisting of analyses and proofs. It does not propose a deployable system or involve data or human subjects. We do not foresee direct societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: See the AI Disclosure section for details on the use of LLMs.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.