A SIZE-SENSITIVE DISCREPANCY BOUND FOR SET SYSTEMS OF BOUNDED PRIMAL SHATTER DIMENSION∗

ESTHER EZRA†

Abstract. Let (X, S) be a set system on an n-point set X. The discrepancy of S is defined as the minimum of the largest deviation from an even split, over all subsets of S ∈ S and two-colorings χ on X. We consider the scenario where, for any subset X′ ⊆ X of size m ≤ n and for any parameter 1 ≤ k ≤ m, the number of restrictions of the sets of S to X′ of size at most k is only O(mk−d−1) for fixed integers d ≥ 0 and 1 ≤ d1 ≤ d (this generalizes the standard notion of bounded primal shatter dimension when d1 = d). In this case we show that there exists a coloring χ with discrepancy bound O*(|S|1/2−d1/(2d)n(d1−1)/(2d)), for each S ∈ S, where O*(·) hides a polylogarithmic factor in n. This bound is tight up to a polylogarithmic factor [J. Matoušek, Discrete Comput. Geom., 13 (1995), pp. 593–601, Geometric Discrepancy, Algorithms Combin. 18, Springer-Verlag, Heidelberg, 1999], and the corresponding coloring χ can be computed in expected polynomial time using the very recent machinery of Lovett and Meka [Proceedings of the 53 Annual IEEE Symposium on Foundations of Computer Science, 2012, pp. 61–67] for constructive discrepancy minimization. Our bound improves and generalizes the bounds obtained from the machinery of Har-Peled and Sharir [Output-Sensitive Tools for Range Searching in Higher Dimensions, unpublished manuscript, 2011; available online from www.cs.tau.ac.il/thesis/thesis/zaban.pdf]) for points and halfspaces in d-space for d ≥ 3. Last but not least, we show that our bound yields improved bounds for the size of relative (ε, δ)-approximations for set systems of the above kind.

Key words. geometric discrepancy, set systems of bounded primal shatter dimension, δ-packing, relative approximations, partial coloring, entropy

AMS subject classifications. AUTHOR MUST PROVIDE

DOI. 10.1137/140977746

1. Introduction. Let (X, S) be a finite set system with n = |X|. A two-coloring of X is a mapping χ : X → {−1, +1}. For a set S ∈ S we define χ(S) := ∑x∈S χ(x). The discrepancy of S is then defined as

\[ \text{disc}(S) := \min_{\chi} \max_{S \in S} |\chi(S)|. \]

In other words, the discrepancy of the set system (X, S) is the minimum over all colorings χ of the largest deviation from an even split, over all sets in S.

Our goal in this paper is to derive discrepancy bounds for (X, S) in the scenario where (X, S) admits a polynomially bounded primal shatter function and has some additional favorable properties. In the bounds that we derive, the discrepancy for each S ∈ S is sensitive to its cardinality |S|. Let us first recall the definition of set systems of this kind (in this definition, the ground set X can be infinite).

Definition 1.1 (primal shatter function [23, 28]). The primal shatter function

∗Received by the editors July 16, 2014; accepted for publication (in revised form) November 11, 2015; published electronically DATE. This work was supported by the NSF under grant CCF-12-16689. A preliminary version of this paper has appeared in Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 2014, pp. 1378–1388.
†School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 (eezra3@math.gatech.edu).
of a set system \((X, S)\) is a function, denoted by \(\pi_S\), whose value at \(m\) is defined by

\[
\pi_S(m) = \max_{Y \subseteq X, |Y| = m} |S|_Y,
\]

where \(S|_Y\) is the collection of all sets in \(S\) projected onto (that is, restricted to) \(Y\). In other words, \(\pi_S(m)\) is the maximum possible number of distinct intersections of the sets of \(S\) with an \(m\)-point subset of \(X\).

From now on we say that a set system \((X, S)\) with \(|X| = n\) (where \(n\) can be assumed to be arbitrarily large) has a primal shatter dimension \(d\) if \(\pi_S(m) \leq Cm^d\), for all \(m \leq n\), where \(d > 1\) and \(C > 0\) are constants.

A typical family of set systems that arise in geometry with bounded primal shatter dimension consists of set systems \((X, S)\) of points in some low-dimensional space \(\mathbb{R}^d\), and \(S\) is a collection of certain simply shaped regions, e.g., halfspaces, balls, or simplices (where \(d > 0\) is assumed to be a constant). In such cases, the primal shatter function is \(m^{O(d)}\); see, e.g., [19] for more details. In fact, set systems of this kind are part of a more general family, referred to as set systems of finite VC-dimension [36]; the reader is referred to [19, 23] for the exact definition. Although the formal definition of finite VC-dimension is different, it suffices to have the same requirement as for set systems of polynomially bounded primal shatter function. It is also known that the VC-dimension is finite if and only if the primal shatter dimension is finite, although they do not necessarily have the same value; see, e.g., [19] for more details. From now on we make only the assumption about having a finite primal shatter dimension; in particular, this is the case in our construction and analysis, and the VC-dimension is mentioned here only for the sake of completeness of the presentation.

We note for the nonexpert reader that a central motivation to study set systems with a polynomially bounded primal shatter function (or just geometric set systems) is the fact that the discrepancy bounds in this case tend to be tremendously better than those achieved in abstract set systems. Specifically, in the latter case these bounds tend to be of order \(\sqrt{n}\), which is sometimes a consequence of having an exponential primal shatter function (see our discussion below for related results in abstract set systems). For example, this is the case for an \(n\)-point set system with \(n\) random sets.\(^2\) We state below a major result of Matoušek [26] (see also [28, 30]) addressing the discrepancy bounds when the primal shatter dimension is bounded.

**Theorem 1.2 (see Matoušek [28]).** Let \((X, S)\) be a set system as above with \(|X| = n\), \(\pi_S(m) \leq Cm^d\), for all \(m \leq n\), where \(d > 1\) and \(C > 0\) are constants. Then

\[
\text{disc}(S) = O(n^{1/2 - 1/(2d)}),
\]

where the constant of proportionality depends on \(d\) and \(C\).

This bound is known to be tight in the worst case (see [28] and our discussion on related work).

**Relative \((\varepsilon, \delta)\)-approximations and \(\varepsilon\)-nets.** The motivation to set up size-sensitive discrepancy bounds of the above kind stems from their application in the construction of relative \((\varepsilon, \delta)\)-approximations, introduced by Har-Peled and Sharir [20]\(^3\) based on the work of Li, Long, and Srinivasan [24] in the context of machine learning theory.

\(^1\)There are set systems for which these bounds are achieved (see, e.g., the construction in [28, Chapter 4]).

\(^2\)Roughly speaking, this is due to the fact that the number of sets restricted to a subset \(Y\) of \(\log n\) points is \(\Omega(n)\), with constant probability. We leave these details to be verified by the reader.

\(^3\)In [20] they were introduced, with a slightly different notation, as relative \((p, \varepsilon)\)-approximations.
We recall the definition from [20]: For a set system \((X, S)\) (with \(X\) finite), the measure of a set \(S \in S\) is the quantity \(\overline{X}(S) = \frac{|S \cap X|}{|X|}\). Given a set system \((X, S)\) and two parameters, \(0 < \varepsilon < 1\) and \(0 < \delta < 1\), we say that a subset \(Z \subseteq X\) is a relative \((\varepsilon, \delta)\)-approximation if it satisfies, for each set \(S \in S\),
\[
\overline{X}(S)(1 - \delta) \leq \underline{Z}(S) \leq \overline{X}(S)(1 + \delta) \quad \text{if} \quad \overline{X}(S) \geq \varepsilon,
\]
\[
\overline{X}(S) - \delta \varepsilon \leq \underline{Z}(S) \leq \overline{X}(S) + \delta \varepsilon \quad \text{otherwise}.
\]

A strongly related notion is the so-called \((\nu, \alpha)\)-sample [19, 21, 24], in which case the subset \(Z \subseteq X\) satisfies, for each set \(S \in S\),
\[
d_{\nu}(\overline{X}(S), \underline{Z}(S)) := \frac{|Z(S) - \overline{X}(S)|}{\overline{X}(S) + \underline{X}(S) + \nu} < \alpha.
\]

As observed by Har-Peled and Sharir [20], relative \((\varepsilon, \delta)\)-approximations and \((\nu, \alpha)\)-samples are equivalent with an appropriate relation between \(\varepsilon\), \(\delta\) and \(\nu\), \(\alpha\) (roughly speaking, they are equivalent up to some constant factor). Due to this observation, they conclude that the analysis of Li, Long, and Srinivasan [24] (that shows a bound on the size of \((\nu, \alpha)\)-samples) implies that for set systems of finite VC-dimension \(d\) there exist relative \((\varepsilon, \delta)\)-approximations of size \(\frac{c \log(1/\varepsilon)}{\delta^2}\), where \(c > 0\) is an absolute constant. In fact, any random sample of these many elements of \(X\) is a relative \((\varepsilon, \delta)\)-approximation with constant probability. More specifically, success with probability at least \(1 - q\) is guaranteed if one samples \(\frac{c(d \log(1/\varepsilon) + \log(1/q))}{\delta^2}\) elements of \(X\).\footnote{We note that although in the original analysis for this bound \(d\) is the VC-dimension, this assumption can be replaced by having just a primal shatter dimension \(d\); see, e.g., [19] for the details of the analysis.}

It was also observed in [20] that \(\varepsilon\)-nets arise as a special case of relative \((\varepsilon, \delta)\)-approximations. Specifically, an \(\varepsilon\)-net is a subset \(N \subseteq X\) with the property that any set \(S \in S\) with \(|S \cap X| \geq \varepsilon |X|\) contains an element of \(N\); in other words, \(N\) is a hitting set for all the “heavy” sets. In this case, if we set \(\delta\) to be some constant fraction, say, \(1/4\), then a relative \((\varepsilon, 1/4)\)-approximation becomes an \(\varepsilon\)-net. Moreover, a random sample of \(X\) of size \(O\left(\frac{\log(d/\varepsilon) + \log(1/q)}{\varepsilon^2}\right)\), with an appropriate choice of the constant of proportionality, is an \(\varepsilon\)-net with probability at least \(1 - q\); see [20] for further details. Our analysis exploits these two structures and the relation between them; see below.

We note that relative approximations appear as a major tool in approximate range counting machinery, where we are given a geometric set system \((X, S)\) (that is, \(X\) is a set of points in some low-dimensional space, and \(S\) is a collection of simply shaped regions, as above), and the goal is to preprocess \(X\) into a data structure that supports efficient queries of the following form: Given \(S \in S\), compute a number \(t\) such that \((1 - \delta)\overline{X}(S) \leq t/|X| \leq (1 + \delta)\overline{X}(S)\), where \(\delta > 0\) is the relative error parameter. This relation has been addressed in [5, 20]; see also [1, 4].

Related work. There is a rich body of literature in discrepancy theory, with numerous bounds and results. It is beyond the scope of this paper to mention all results, and we just list those that are most relevant to our work. We refer the reader to the book of Chazelle [13] for an overview of discrepancy theory and the book of Matoušek [28] for various results in geometric discrepancy. In particular, our work is based on the techniques overviewed in the latter.

We first briefly overview previous results for an abstract set system on an \(n\)-point set \(X\), with \(m = |S|\) sets. The celebrated “six standard deviations” result of
Spencer [34], an extension to the partial coloring method of Beck [8], implies that for such set systems there exists a coloring \( \chi \) such that \( \chi(S) \leq K \sqrt{n(1 + \log(m/n))} \), for each \( S \in \mathcal{S} \), where \( K > 0 \) is a universal constant. In particular, when \( m = n \) we have \( K = 6 \), in which case the discrepancy bound becomes \( 6\sqrt{n} \). In contrast, one can easily show that using simple probabilistic considerations, a random coloring yields a (suboptimal) discrepancy bound of \( \sqrt{n}\log m \) (or \( \sqrt{n}\log n \) if \( m = O(n) \)).

A long-standing open problem was whether the result of Spencer [34] can be made constructive, and this has recently been answered in the positive by Bansal [7] for the case \( m = O(n) \) (for the general case his bound is slightly suboptimal with respect to the bound in [34]). In a follow-up work, Lovett and Meka [25] have shown a new elementary constructive proof of Spencer’s result, resulting in similar asymptotic discrepancy bounds to those in [34], for arbitrary values of \( m \). Very recently, this result has been further generalized by Rothvoss [32].

In geometric set systems, upper bounds were first shown by Beck, where the Lebesgue measure of a class of geometric shapes is approximated by a discrete point set; see, e.g., [9] and the book by Beck and Chen [10] for a bound of \( O(n^{1/4}\sqrt{\log n}) \) for disks as well as rotated rectangles in the plane. For arbitrary points sets, Matoušek, Welzl, and Wernisch [30] have addressed the case of points and halfspaces in \( d \)-space, and showed an almost tight bound of \( O(n^{1/2−1/(2d)}\log^{1/2+1/(2d)} n) \), which has later been improved to \( O(n^{1/2−1/(2d)}) \) [26, 28] (Theorem 1.2). Concerning lower bounds, there is a rich literature where several such bounds are obtained in geometric set systems. We only mention the lower bound \( \Omega(n^{1/2−1/(2d)}) \) of Alexander [2] for set systems of points and halfspaces in \( d \)-space. For further results we refer the reader to [26] and the references therein.

The extension of discrepancy bounds for points and halfspaces in \( d \)-space to be size-sensitive has been addressed in the work of Har-Peled and Sharir [20], who showed a bound of \( O(|S|^{1/4}\log n) \) in the two-dimensional case, using an intricate extension of the technique of Welzl [38] (see also [14]) for constructing spanning trees with low crossing numbers. Nevertheless, their technique cannot be applied in higher dimensions, because already at \( d = 3 \) they showed a counterexample to their construction. The follow-up work of Sharir and Zaban [33] (based on the construction in [20]) addresses these cases, establishing the bound

\[
O\left(n^{(d-2)/2(d-1)}|S|^{1/(2d(d-1))}\log^{(d+1)/(2(d-1))} n\right).
\]

The size-sensitive discrepancy bounds in both studies [20, 33] imply improved bounds for relative \((\varepsilon, \delta)\)-approximations for points and halfspaces in \( d \)-space (roughly speaking, these bounds are better than the bound of Li, Long, and Srinivasan [24] when \( \delta \) is not too large with respect to \( \varepsilon \)). Specifically, Har-Peled and Sharir [20] showed how to derive such bounds due to a “halving technique,” where in two dimensions the resulting bound is \( O\left(\log^{1/2(1/(d\varepsilon))}\right) \). Their result in three and higher dimensions is somewhat restricted, as they obtained a sequence of \( O(\log (1/\varepsilon)) \) samples, where each halfspace is associated with one such sample that constitutes its relative approximation. The follow-up work in [33] overcomes this difficulty and shows how to construct a single sample (for \( d \geq 3 \)) of size \( O\left(\log^{(\log(1/\varepsilon))}\right) \), where \( \gamma = \frac{2(d-1)-1}{(d-1)(d+1)} \), \( \mu = \frac{2d}{d+1} \), and \( \eta = \frac{d}{d+1} \).

Our results. In this paper we refine the bound in Theorem 1.2 so that it becomes sensitive to the size of the sets \( S \in \mathcal{S} \) in several favorable cases. Specifically, we assume that for any (finite) set system projected onto a ground set of size \( m \leq n \) and for
any parameter $1 \leq k \leq m$ the number of sets of size at most $k$ is only $O(m^{d_1}k^{d-d_1})$, where $d$ is the primal shatter dimension and $1 \leq d_1 \leq d$ is an integer parameter.\footnote{We ignore the cases where $d_1 = 0$ or $d_1$ takes fractional values, as they do not seem to appear in natural set systems and, in particular, in the geometric set systems that we consider.}

By assumption, when $k = m$ we obtain $O(m^d)$ sets in total, in accordance with the assumption that the primal shatter dimension is $d$, but the above bound is also sensitive to the size $k$ of the set.

We show that for set systems of this kind there exists a coloring $\chi$ such that

$$\chi(S) = O^* (|S|^{1/2 - d_1/(2d)}n^{(d_1 - 1)/(2d)})$$

where $O^*(\cdot)$ hides a polylogarithmic factor in $n$. This bound almost matches (up to the polylogarithmic factor) the optimal discrepancy bound in Theorem 1.2 when $|S| = n$, but it is more general than this bound, as it yields a range of bounds for $1 \leq d_1 \leq d$. Specifically, when $d_1 = d$ the number of sets is just $O(m^d)$ (that is, this bound is not sensitive to their size), in which case we just have a set system of bounded primal shatter dimension, and then, once again, our discrepancy bound almost matches the optimal bound in Theorem 1.2. In the other extreme scenario, when $d_1 = 1$, the set system is what we call “well-behaved,” and our discrepancy bound then becomes $O^*(|S|^{1/2 - 1/2d})$, which depends only on $|S|$ (up to the polylogarithmic factor). In particular, set systems of points and halfspaces in $d$ dimensions are well-behaved with $d = 2$ and $d = 3$, respectively. In the plane, the resulting bound is slightly suboptimal with respect to the sensitive bound of Har-Peled and Sharir [20] (our hidden polylogarithmic factor is slightly larger than their log $n$ factor). For points and halfspaces in three dimensions and higher, our bound considerably improves the bound in [33] (which extends the construction in [20]). In particular, the bound in the three-dimensional case in [33] is not purely sensitive in $|S|$ but also contains an additional sublinear term in $n$ (whereas the original technique of Har-Peled and Sharir [20] failed to yield such a bound).

As a major application, we conclude that our new discrepancy bounds yield improved bounds on the size of relative $(\varepsilon, \delta)$-approximations for the corresponding set systems. Specifically, using a variant of the “halving technique” in [20], we obtain a relative $(\varepsilon, \delta)$-approximation of size $O\left(\frac{\log^2 \frac{1}{\varepsilon \delta}}{\log \frac{1}{\varepsilon \delta}}\right)$ for points and halfspaces in three dimensions, which almost matches the bound $O\left(\frac{\log^{3/2} \frac{1}{\varepsilon \delta}}{\log \frac{1}{\varepsilon \delta}}\right)$ of Har-Peled and Sharir [20] obtained as a sequence of $O(\log \frac{1}{\varepsilon})$ samples, whereas we obtain a single sample. In higher dimensions we obtain a bound of $O\left(\frac{\log \frac{1}{\varepsilon \delta}}{\log \frac{1}{\varepsilon \delta}}\right)$ (which once again corresponds to a single sample) that is a considerable improvement over the bounds in [20, 33]. In particular, for arbitrarily large values of $d$ the exponent in the term $(1/\varepsilon)$ is close to $3/2$, whereas in the previous bound [33] this exponent is close to 2. See Theorem 3.9 for more details concerning these bounds.

On the algorithmic front, we use the recent result of Lovett and Meka [25] for constructive discrepancy minimization in order to conclude that one can construct in expected polynomial time a coloring $\chi$ achieving the above discrepancy bounds. As a result, it follows that a relative $(\varepsilon, \delta)$-approximation with the aforementioned bound can be constructed in expected polynomial time as well. See Corollary 3.7 and Theorem 3.9.

Our construction uses a machinery different from that in [20, 33] and is a variant of the construction of Matoušek [28] (see also Matoušek [26] for the special case of points and halfspaces in $d$-space), based on Beck’s partial coloring and on the
entropy method, and is combined with the property that set systems with bounded primal shatter dimension admit a small packing (see below for details concerning these notions). Our assumption about the set system (that is, the bound on the number of sets of size at most $k \leq m \leq n$ stated above) enables us to refine the analysis in [28] to be size sensitive, which eventually leads to the desired bound.

2. Preliminaries. We now briefly review some of the tools mentioned in section 1, which are applied by our analysis. Then, we present the construction.

Partial coloring and entropy. Let $X$ be a set as above. A partial coloring of $X$ is any mapping $\chi : X \rightarrow \{-1, 0, +1\}$. A point $x \in X$ with $\chi(x) = 1$ or $\chi(x) = -1$ is said to be colored by $\chi$, whereas $\chi(x) = 0$ means that $x$ is uncolored.

A method originated by Beck [8], which was subsequently elaborated by others and is known by now as the entropy method (see, e.g., [3, 29, 34]), shows that under some favorable assumptions there exists a “substantial partial coloring” of $(X, S)$ (that is, at least some constant fraction of $X$ is colored) with good discrepancy bounds. In our analysis we apply the entropy method as a black box, and thus only present the constructive version of Beck’s partial coloring lemma, as has recently been formulated by Lovett and Meka [25]; see section 1 and the references therein for more details. We also refer the reader to [27] for the original (nonconstructive) version of the entropy method. One of the differences in the approach of Lovett and Meka [25] with respect to previous studies is the fact that they construct a (partial) coloring function $\chi : X \rightarrow [-1, +1]$, that is, $\chi(\cdot)$ assigns fractional values in $[-1, +1]$, with the property that at least half of the points in $X$ are assigned $\{-1, +1\}$. Specifically, they show the following.

**Proposition 2.1** (see Lovett and Meka [25]). Let $(X, S)$ be a set system with $|X| = n$, and let $\Delta_S > 0$ be some real number assigned to $S$, for each $S \in S$. Suppose that

$$\sum_{S \in S} \exp\left(-\frac{\Delta_S^2}{16|S|}\right) \leq \frac{n}{16}.$$  

Then there exists a partial coloring $\chi : X \rightarrow [-1, +1]$, with $|\chi(x)| = 1$, for at least $n/2$ points $x \in X$, so that $\chi(S) \leq \Delta_S + 1/n^c$, for each $S \in S$, where $c > 0$ is an arbitrarily large constant. Moreover, $\chi$ can be computed in expected polynomial time in $n$ and $|S|$ (where the degree of the polynomial bound depends on $c$).

$\delta$-packing. Let $(X, S)$ be a set system as above, and let $\delta > 0$ be a given integer parameter. A subcollection $P \subseteq S$ is $\delta$-separated if the symmetric difference distance $|S_1 \Delta S_2|$ between any pair of sets $S_1, S_2 \in P$ is strictly greater than $\delta$. A $\delta$-packing for $(X, S)$ is an inclusion-maximal $\delta$-separated subcollection of $S$.

A key property, shown by Haussler [22] (see also [12, 14, 16, 28, 37]), is that set systems of bounded primal shatter dimension admit small $\delta$-packings. That is, we have the following.

**Theorem 2.2** (packing lemma [12, 22]). Let $d > 1$ be a constant, and let $(X, S)$ be a set system on an $n$-point set with primal shatter dimension $d$. Let $1 \leq \delta \leq n$ be

---

6 This follows from a random-walk process, initialized at the origin, where the partial coloring obtained in each step is a vector in $[-1, +1]^n$ and, at each consecutive step, the walk resumes from the vector obtained at the previous iteration.

7 We note that in the original formulation in [22] the assumption is that the set system has a finite VC-dimension. However, a closer inspection of the analysis in [22] shows that this assumption can be replaced with that of having bounded primal shatter dimension.
an integer parameter, and let \( \mathcal{P} \subseteq \mathcal{S} \) be \( \delta \)-separated. Then \( |\mathcal{P}| = O((n/\delta)^d) \), where the constant of proportionality depends on \( d \).

3. The construction. We are now ready to present our construction. Let \((X, \mathcal{S})\) be a set system of bounded primal shatter dimension \( d \), with the additional property that the number of sets of size at most \( k \) in the projection of \((X, \mathcal{S})\) onto any \( m \)-point subset \( X' \subseteq X \) is \( O(m^{d_1}k^{d-d_1}) \) for any \( m \leq n \). We construct a decomposition of each set \( S \in \mathcal{S} \) as a Boolean combination of “canonical sets” obtained from a sequence of packings that we build for \( S \). This decomposition is a variant of that presented in [28, Chapter 5] (see also [26]) and is also referred to as chaining in the literature (see, e.g., [19]).

Chaining. For the sake of completeness, we repeat some of the details in [28, Chapter 5] and use similar notation for ease of presentation. Without loss of generality, we assume that \( \log n \) is an integer and put \( k := \log n + 1 \). For \( j = 0, \ldots, k \), let \( \mathcal{F}_j \subseteq \mathcal{S} \) be an \((n/2^j)\)-packing for \((X, \mathcal{S})\). By definition, this implies that for any pair of sets in \( \mathcal{F}_j \), their distance is larger than \( n/2^j \) and that this set is maximal with respect to this property. In particular, we have \( \mathcal{F}_0 = \{\emptyset\} \), and \( \mathcal{F}_k = \mathcal{S} \) by construction.

Observe that for each \( F_j \in \mathcal{F}_j \), there is a set \( F_{j-1} \in \mathcal{F}_{j-1} \) with \( |F_j \Delta F_{j-1}| \leq n/2^{j-1} \). This follows from the inclusion-maximality of \( \mathcal{F}_{j-1} \). That is, consider the set \( F_j \) closest to \( F_{j-1} \) in \( \mathcal{F}_{j-1} \). Either \( F_j = F_{j-1} \) and then the claim is trivial, or \( F_j \neq F_{j-1} \) and then the opposite inequality \( |F_j \Delta F_{j-1}| > n/2^{j-1} \) is impossible, for then \( F_j \) could be added to \( \mathcal{F}_{j-1} \), contradicting its maximality. Let us attach to each \( F_j \), its nearest neighbor (closest set) \( F_{j-1} \in \mathcal{F}_{j-1} \) and put \( A(F_j) := F_j \setminus F_{j-1} \), \( B(F_j) := F_{j-1} \setminus F_j \). We now form the set systems \( A_j := \{A(F_j) \mid F_j \in \mathcal{F}_j\} \), \( B_j := \{B(F_j) \mid F_j \in \mathcal{F}_j\} \), \( j = 1, \ldots, k \). It has been observed in [28, Chapter 5] that each set \( S \in \mathcal{S} \) can be decomposed as

\[
S = (\cdots (((A_1 \setminus B_1) \cup A_2) \setminus B_2) \cup \cdots \cup A_k) \setminus B_k,
\]

where \( \cup \) denotes disjoint union, and \( A_j \in A_j \), \( B_j \in B_j \), for each \( j = 1, \ldots, k \). Indeed, consider the nearest-neighbor “chain” \( S = F_k \rightarrow F_{k-1} \rightarrow \cdots \rightarrow F_0 = \emptyset \) (recall that \( F_0 = \emptyset \) by our assumption on \( \mathcal{F}_0 \), from which it follows that \( A_1 = F_1, B_1 = \emptyset \) ). Each set \( F_j \in \mathcal{F}_j \) on this chain can be turned into its nearest neighbor \( F_{j-1} \in \mathcal{F}_{j-1} \) by adding \( B(F_j) \) and subtracting \( A(F_j) \). Moreover, it is easy to verify using induction on \( j \geq 1 \) that \( F_j = (\cdots (((A_1 \setminus B_1) \cup A_2) \setminus B_2) \cup \cdots \cup A_j) \setminus B_j \), and \( S \) is obtained at \( j = k \) (see also [31] for similar considerations).

Refined decomposition. We next refine this decomposition as shown below. The idea of having such a refinement is to obtain, for each set \( S \in \mathcal{S} \), a decomposition that consists of canonical sets, whose size is roughly that of \( S \). That is, the complexity of this decomposition is sensitive to \( |S| \). This is crucial for the analysis, as after a more careful counting of the number of such canonical sets (Theorem 3.3) we will be able to reduce the bound on the entropy function. See also our comment after Corollary 3.2.

We thus proceed as follows. We partition the sets in \( \mathcal{S} \) into \( k \) subfamilies \( \mathcal{S}_1, \ldots, \mathcal{S}_k \), where \( S \in \mathcal{S}_i \) if

\[
\frac{n}{2^i} < |S| \leq \frac{n}{2^{i-1}}
\]

for \( i = 1, \ldots, k \) (by definition, \( \mathcal{S}_k \) contains all singleton sets). Fix an index \( i = 1, \ldots, k \). For each \( S \in \mathcal{S}_i \), we modify (3.1) by iterating from \( k \) down to \( i \); that is, we eliminate \( A_j, B_j \) from the decomposition for \( j = 1, \ldots, i - 1 \) and replace it by \( F_{i-1} \in \mathcal{F}_{i-1} \).
Specifically, we have
\[(3.2) \quad S = \ldots (((F_{i-1} \cup A_j) \setminus B_i) \cup A_{i+1}) \setminus B_{i+1} \cup \ldots \cup A_k) \setminus B_k.\]

Indeed, similarly to decomposition (3.1), it is easy to verify by induction on the index \(j \geq i\) of the sets that \(F_j = \ldots (((F_{i-1} \cup A_i) \setminus B_i) \cup A_{i+1}) \setminus B_{i+1} \cup \ldots \cup A_j) \setminus B_j\), and our claim is obtained when \(j = k\). In particular, when \(i = 1\) we obtain the same decomposition as in (3.1), as \(F_0 = \emptyset\).

Let us now fix an index \(i\) and construct the subsets \(F_j^i\) of the packings \(F_j\), for each \(j = i-1, \ldots, k\), as follows. For each \(S \in S_i\), we follow its nearest-neighbor chain (where at this time we stop at \(F_{i-1}\)) \(S = F_k \rightarrow F_{k-1} \rightarrow \cdots \rightarrow F_j \rightarrow \cdots \rightarrow F_{i-1}\) and put in \(F_j^i\) the corresponding canonical sets \(F_j \in F_j\), \(j = i-1, \ldots, k\). We next show that the size of each of these members, which we now denote as \(F_j^i \in F_j^i\), is at most \(O(|S|) = O(n/2^{i-1})\). First we show the following.

**Claim 3.1.** For each of the sets \(F_j^i \in F_j^i\), \(j = i-1, \ldots, k\), in the nearest-neighbor chain of \(S \in S_i\) we have
\[|S \triangle F_j^i| < \frac{n}{2^{j-1}}.\]

**Proof.** By construction we have \(|F_j^i \triangle F_{j-1}^i| \leq n/2^{j-1}, F_j^i \in F_j^i, \) and \(F_{j-1}^i \in F_j^i\) for each \(j = i, \ldots, k\). We now apply the triangle inequality on the symmetric difference distance\(^8\) and obtain
\[|S \triangle F_j^i| \leq |S \triangle F_{j-1}^i| + |F_{j-1}^i \triangle F_{j-2}^i| + \cdots + |F_{j+1}^i \triangle F_j^i| \leq \frac{n}{2^{k-1}} + \frac{n}{2^{k-2}} + \cdots + \frac{n}{2^j} < \frac{n}{2^{j-1}},\]
as asserted. \(\square\)

Combining this with the fact that \(|S| \leq n/2^{i-1}\) by construction (and the obvious relation \(F_j^i \subseteq S \cup (S \triangle F_j^i)\)), we obtain the following.

**Corollary 3.2.** For each of the sets \(F_j^i \in F_j^i, j = i-1, \ldots, k\), we have
\[|F_j^i| = O\left(\frac{n}{2^{j-1}}\right).\]

**Remark.** We note that by construction \(|F_j^i \setminus F_{j-1}^i|, |F_{j-1}^i \setminus F_j^i|\) are bounded by \(n/2^{i-1}\), and this fact is used later in the entropy method. Nevertheless, the property that the actual sets \(F_j^i\) have size \(O(n/2^{i-1})\) is stronger, and it enables us to prove a tighter bound on the number of the canonical sets \(F_j^i\) (Theorem 3.3) and thus on the number of “pair sets” \(F_j^i \setminus F_{j-1}^i, F_{j-1}^i \setminus F_j^i\). This is crucial for our analysis, as it enables us to reduce the bound on the entropy, from which we will derive the desired discrepancy bound; see below.

**Bounding the size of the packing.** Having a fixed index \(i\) as above, we consider all sets \(S \in S_i\) and the corresponding canonical sets \(F_j^i \in F_j^i\) participating in decomposition (3.2), \(j = i-1, \ldots, k\). For a fixed index \(j = i-1, \ldots, k\), Theorem 2.2 implies that the size of \(F_j^i\) is \(O(2^{jd})\), but our goal is to show that the actual bound can be made sensitive to the size of the sets in \(F_j^i\), that is, to \(O(n/2^{j-1})\).

**Theorem 3.3 (sensitive packing lemma).**
\[|F_j^i| = O\left(\frac{j^{d/2jd}}{2^{(d-d_1)(i-1)}}\right).\]

\(^8\)This follows from the property that for each triple of sets \(X, Y, Z\) we have \(X \triangle Z \subseteq (X \triangle Y) \cup (Y \triangle Z)\).
Proof. We use a variant of the analysis of Dudley [16] and refine it to our scenario. Put \( \delta := n/2^j \). Since the following analysis will restrict sets \( S \in \mathcal{S} \) to subsets of \( X \), we will refer to \( |S| \) more explicitly as \( |S \cap X| \), for the sake of presentation.

We form the set system \( (X, \mathcal{D}) \), where \( \mathcal{D} = \{ S_1 \triangle S_2 \mid S_1, S_2 \in \mathcal{S} \} \). As observed in [16], this set system admits an \( \varepsilon \)-net of size \( O((1/\varepsilon) \log (1/\varepsilon)) \), with a constant of proportionality depending on \( d \) (see our discussion in section 1). In fact, a random sample of that size with a sufficiently large constant of proportionality (that depends on \( d \)) is an \( \varepsilon \)-net with probability at least \( 3/4 \), say.

Set \( \varepsilon := \delta/n = 1/2^j \), and let \( N \subseteq X \) be an \( \varepsilon \)-net for \( (X, \mathcal{D}) \); \( |N| = O(j2^j) \), with a constant of proportionality as above. By the \( \varepsilon \)-net property, whenever the symmetric difference between two sets \( S_1, S_2 \in \mathcal{S} \) contains at least \( \varepsilon n = \delta \) elements, we must have \( (S_1 \triangle S_2) \cap N \neq \emptyset \). We thus must have \( S_1 \cap N \neq S_2 \cap N \) for any pair of such sets. But this implies that each set in \( \mathcal{F}_j \) is mapped to a distinct set in the set system \( (X, \mathcal{S}) \) projected onto \( N \). Let \( (N, \mathcal{S}|_N) \) be this set system. Thus the number of sets in \( \mathcal{F}_j \) is bounded by \( |\mathcal{S}|_N \).

Recall that in view of Corollary 3.2 we are interested only in those sets whose size is \( O(n/2^{j-1}) \). We next show the following.

Claim 3.4. Each set \( S \in \mathcal{S} \) with \( |S \cap X| = O(n/2^{j-1}) \) satisfies \( |S \cap N| = O(j2^j/2^{i-1}) \), with probability at least \( 3/4 \). In particular, due to Corollary 3.2, this is also the bound on the cardinality of the sets \( \mathcal{F}'_j \in \mathcal{F}_j \) restricted to \( N \).

Proof. Indeed, since \( N \) is a random sample of size \( O(j2^j) \) with a sufficiently large constant of proportionality (that depends on \( d \)), it is also a relative \((1/2^j, 1/4)\)-approximation for \( (X, \mathcal{S}) \) with probability at least \( 3/4 \) (see our discussion in section 1 and [20]). In particular, this means that \( N \) is both a \((1/2^j, 1/4)\)-net for \( (X, \mathcal{D}) \) and a relative \((1/2^j, 1/4)\)-approximation for \( (X, \mathcal{S}) \) with probability at least \( 1/2 \) (and thus we obtain a single sample with a “double role”). The latter property implies that

\[
\left| \frac{|S \cap N|}{|N|} - \frac{|S \cap X|}{|X|} \right| \leq \frac{1}{4} \cdot \frac{|S \cap X|}{|X|}
\]

if \( |S \cap X| \geq 1/2^j \) and

\[
\left| \frac{|S \cap N|}{|N|} - \frac{|S \cap X|}{|X|} \right| \leq \frac{1}{4} \cdot \frac{1}{2^j}
\]

otherwise. Combining the fact that \( |S \cap X| = O(n/2^{j-1}) \), \( j \geq i - 1 \), and the bound on \( |N| \), we obtain that \( |S \cap N| = O(j2^j/2^{i-1}) \), as asserted. \( \square \)

Recall that \( \mathcal{S}|_N \) is the family \( \mathcal{S} \) projected onto \( N \), and, due to the double role of \( N \) (as described in the proof of Claim 3.4), each \( \mathcal{F}'_j \in \mathcal{F}_j \) is mapped to a distinct set of \( \mathcal{S}|_N \) of size \( O(j2^j/2^{i-1}) \). By assumption, the number of sets of \( (N, \mathcal{S}|_N) \) of size \( 1 \leq k \leq |N| \) is \( O(|N|d^i k^{d-d_1}) \), and hence the number of its sets of size \( O(j2^j/2^{i-1}) \) is at most

\[
O \left( |N|^{d_1} \left( j2^j/2^{i-1} \right)^{d-d_1} \right) = O \left( \frac{j^2 d^{2j} d^{i} 2^{j(d-d_1)(d-1)}}{2^{j(d-d_1)(1-1)}} \right),
\]

from which we obtain the bound on \( |\mathcal{F}'_j| \). \( \square \)

Applying the entropy method. Let us fix an index \( i \) for the family \( \mathcal{S}_i \) under consideration. Returning to decomposition (3.2), we denote, with a slight abuse of notation, the canonical sets \( A_j, B_j \) by \( A^j_i, B^j_i \), respectively. By construction, \( A^j_i = F^1_j \setminus F^j_{i-1} \),
$B_j^i = F_j^i \setminus F_j^{i-1}$. Let $\mathcal{M}_j^i$ be the collection of these sets $A_j^i$, $B_j^i$ (or $F_j^i$ if $j = i - 1$). We now set a parameter $\Delta_j^i$ for the discrepancy bound (with respect to partial coloring) for each of the canonical sets in $\mathcal{M}_j^i$, where all sets in this subcollection are assigned the same discrepancy bound $\Delta_j^i$. We then use the entropy method on this new set system in order to obtain a partial coloring $\chi$ achieving the preassigned discrepancy bounds. Having these bounds at hand, we can then conclude that the discrepancy of any $S \in \mathcal{S}_i$ with respect to $\chi$ is at most $2 \sum_{j=i-1}^{k} \Delta_j^i$ (using standard arguments; see, e.g., [28]), and this will yield the desired size-sensitive bound for $\chi(8)$.

In order to apply the entropy method as presented in Proposition 2.1 we need to have, for each $j = i - 1, \ldots, k$, (i) a bound on $|\mathcal{M}_j^i|$, (ii) a bound on the size of the canonical sets $\mathcal{M}_j^i$, and (iii) an appropriate choice for the parameter $\Delta_j^i$, uniformly assigned to all these sets. We set each of the bounds in (i)–(iii) for a fixed index $i$ and then sum them up over all $i = 1, \ldots, k$.

For the bound in (i), we observe that each canonical set $A_j^i$ (resp., $B_j^i$) corresponds to a pair $(F_j^i, F_j^{i-1})$, but each of these pairs can uniquely be charged to $F_j^i$, as $F_j^{i-1}$ is the corresponding nearest neighbor in the packing $\mathcal{F}_{j-1}^i$. Thus the number of these canonical sets is $|\mathcal{F}_j^i|$ for $j = i, \ldots, k$. For $j = i - 1$, the bound is trivially $|\mathcal{F}_j^i|$. Thus by applying the sensitive packing lemma (Theorem 3.3) we conclude that the overall number of canonical sets is at most $C \cdot \frac{j^{d+2d}}{2^{(d-1)(d-1/2)}}$, for $j = i - 1, \ldots, k$, where $C > 0$ is an appropriate constant whose choice depends on $d$. By construction, the size of the sets $A_j^i, B_j^i$ is at most $s_i = n/2^{i-1}$ (recall that $|F_j^i \setminus F_j^{i-1}| \leq n/2^{i-1}$) and $|F_{i-1}^i| = O(n/2^{i-1})$ by Corollary 3.2, whence we get the bound for (ii). For the choice in (iii), we set

$$\Delta_j^i := A \cdot \frac{1}{(1 + |j - j_0|)^2} \left( \frac{n^{1/2-1/(2d)}}{2^{(d-1)/2-1/(2d)}} \right) \log^{1/2+1/2d} n,$$

where

$$j_0 := (1/d) \log n + (1 - d_1/d)(i - 1) - (1 + 1/d) \log \log n - B,$$

for an appropriate constant $B > 5 + \log C$, and for a sufficiently large constant of proportionality $A > 0$, whose choice depends on $B$ and will be determined shortly (note that all the three constants $A, B,$ and $C$ depend on $d$).

We first provide a brief justification for our choice of $j_0$ and the appearance of the coefficient $\frac{1}{(1 + j - j_0)^2}$. For the first, our goal is to bound the entropy function, bounded by the sum (3.4) appearing in Proposition 3.5, and at $j = j_0$ the corresponding summands achieve their maximum value (which is some linear function of $n/\log n$ with an appropriate constant of proportionality). Then when $j \geq j_0$ the exponents “take over” this summation, in which case it decreases superexponentially, and when $j < j_0$ the terms $C \cdot \frac{j^{d+2d}}{2^{(d-1)(d-1/2)}}$ representing the packings take over this summation and decrease geometrically. This eventually leads to the bound stated in Proposition 3.5. The coefficient $\frac{1}{(1 + j - j_0)^2}$ in (3.3) guarantees that the sum $\sum_{j=i-1}^{k} \Delta_j^i$ (corresponding to the asymptotic bound on the discrepancy of any $S \in \mathcal{S}_i$) converges to $O\left(\frac{n^{1/2-1/(2d)}}{2^{(d-1)/2-1/(2d)}} \log^{1/2+1/2d} n\right)$. In particular, it does not contain an extra logarithmic factor over the initial bound of $\Delta_j^i$; see below. Similar ideas have been used in [28]. We defer the remaining technical details to Appendix A, where we present the proof of Proposition 3.5.

**Proposition 3.5.** The choice in (3.3), for $A > 0$ sufficiently large (whose choice
depends on \( C \) and thus on \( d \), satisfies

\[
(3.4) \quad \sum_{i=1}^{k} \sum_{j=1}^{k} C \cdot \frac{j^d 2^d}{2^{d-d_i} j(s-1)} \exp \left( -\frac{(\Delta_j^i)^2}{16 s_j} \right) \leq \frac{n}{16}.
\]

**Remark.** Currently, our analysis is slightly suboptimal, as our bound in (3.3) contains an extra fractional logarithmic power, which results from the following reasons: (i) We may overcount in our bound on the entropy function the same set \( A_i \) (or \( B_i \)) when we iterate over \( i = 1, \ldots, k \); recall that for each \( S \in \mathcal{S}_i \) we follow its nearest-neighbor chain and put in each \( \mathcal{F}_j \) the corresponding element from \( \mathcal{F}_j \), and thus an element from \( \mathcal{F}_j \) may appear in multiple layers \( i \). (ii) In the sensitive packing lemma (Theorem 3.3) the size of the random sample \( N \) is \( O((n/\delta) \log (n/\delta)) \), whereas in the analysis of the original packing lemma (Theorem 2.2) it is sufficient to have a sample of \( (n/\delta) \) elements, resulting in the bound \( O((n/\delta)^d) \). Nevertheless, due to the fact that our sample also needs to be a relative approximation (to exploit the property that our packing consists of sets of size at most \( O(n/2^{i-1}) \)), we had to use a slightly larger sample. It is an interesting open problem whether the extra logarithmic factor can be removed from the bound on the size of the sensitive packing. If so, this will imply that the actual bound in Theorem 3.3 is \( O(\frac{n^{2^d}}{n^{d-1}}) \).

**Wrapping up.** Incorporating Propositions 2.1 and 3.5, we obtain that there exists a partial coloring \( \chi \), computable in expected polynomial time, which colors at least \( n/2^d \) points of \( X \), such that \( \chi(M_j^i) \leq \Delta_j^i + 1/n^c \), for each \( M_j^i \in \mathcal{M}_j^i \), where \( c > 0 \) is an arbitrarily large constant. But then our choice in (3.3) and our earlier discussion imply that, for each \( S \in \mathcal{S}_i \),

\[
\chi(S) \leq 2 \sum_{j=1}^{k} \Delta_j^i = O \left( \frac{n^{1/2-1/(2d)}}{2^{d/d_i-1/(2d)}} \log^{1/2+1/2d} n \right),
\]

since the sum \( \sum_{j=1}^{k} \frac{1}{1+(1+1/2d)} \) converges. Due to the fact that \( n/2^i \leq |S| < n/2^{i-1} \) (by definition), the latter term is

\[
O \left( |S|^{1/2-d_i/(2d)} n^{(d_i-1)/(2d)} \log^{1/2+1/(2d)} n \right).
\]

Applying the partial coloring procedure (Proposition 2.1) for at most \( \log n \) iterations (which yields a full coloring for \( X \)), we obtain that for each \( S \in \mathcal{S}_i \), \( \chi(S) = O(|S|^{1/2-d_i/(2d)} n^{(d_i-1)/(2d)} \log^{1/2+1/2d} n) \) if \( d_i > 1 \) and

\[
\chi(S) = O \left( |S|^{1/2-1/(2d)} \log^{3/2+1/2d} n \right)
\]

if \( d_i = 1 \) (recall that we assume \( 1 \leq d_i \leq d \) is an integer parameter). This is because, for \( d_i > 1 \), the term \( n^{(d_i-1)/(2d)} \) creates a decreasing geometric sequence over the at most \( \log n \) iterations. We have just shown the following.

**Theorem 3.6.** Let \( (X, \mathcal{S}) \) be a (finite) set system of primal shatter dimension \( d \) with the additional property that in any set system restricted to \( X' \subseteq X \) the number of sets of size \( k \leq |X'| \) is \( O(|X'|^{d_1-k^{d-d_1}}) \), where \( d_1 \) is an integer parameter between 1 and \( d \). Then

\[
\text{disc}(\mathcal{S}) = \begin{cases} 
O \left( |S|^{1/2-d_1/(2d)} n^{(d_1-1)/(2d)} \log^{1/2+1/2d} n \right) & \text{if } d_1 > 1, \\
O \left( |S|^{1/2-1/(2d)} \log^{3/2+1/2d} n \right) & \text{if } d_1 = 1,
\end{cases}
\]
where the constant of proportionality depends on $d$.

Remark. We note that although the number of uncolored points in $X$ decreases by at least a half after applying Proposition 2.1, it does not necessarily guarantee that the size of a set $S \subseteq S_i$ decreases by the same factor, and so we can bound it from above only by $n/2^j$ at each round. Moreover, it may happen that a set $S \subseteq S_i$ from the previous round now lies in a different class at the current partition. Thus at each round we need to resume the process from scratch, due to which we obtain an extra logarithmic factor for $d_1 = 1$, as shown in the bound above.

Algorithmic aspects. In order to apply the randomized algorithm of Lovett and Meka [25], we first need to construct, for each $i = 1, \ldots, k$, the canonical sets in $T^*_j$.

In order to do so for a fixed $i$, we need to construct a $\delta$-packing, for $\delta = n/2^j$, $j = i - 1, \ldots, k$, such that the size of each set in the packing does not exceed $Kn/2^{i-1}$ for an appropriate constant $K > 0$. We thus iterate over $j = i - 1, \ldots, k$, and form a $\delta$-packing $T^*_j$ as above in a brute force manner by initially picking an arbitrary set $F \subseteq S_i$, whose size is at most $Kn/2^{i-1}$, to be included into $T^*_j$, and then keep collecting sets $F' \subseteq S_i$ into $T^*_j$ if (i) $|F'| \leq Kn/2^{i-1}$ and (ii) the symmetric difference distance between $F'$ and each of the elements currently in $T^*_j$ is greater than $\delta$. We stop as soon as there are no leftover sets $F''$ of the above kind. The set just created is inclusion-maximal, and thus according to the sensitive packing lemma (Theorem 3.3) its size is only $O(\frac{\delta^d n^d}{2^d + 2^d n^d})$. It is easy to verify that the construction of each $\delta$-packing $T^*_j$ can be performed in polynomial time due to the fact that the number of sets in $S$ is only $O(n^d)$. Omitting any further details we conclude the following.

Corollary 3.7. A coloring $\chi$ achieving the discrepancy bound in Theorem 3.6 can be computed in expected polynomial time.

The case of points and halfspaces in $d$ dimensions. When $(X, S)$ is a set system of points and halfspaces in $d$-space, it is known that the number of halfspaces containing at most $k$ points of $S$ is $O(n^{d/2}k^{(d/2)})$ (see, e.g., [15]). Thus from Theorem 3.6 and Corollary 3.7 we conclude the following.

Theorem 3.8. Let $(X, S)$ be a set system of points and halfspaces in $d$-space. Then
\[
\text{disc}(S) = O\left(|S|^{1/4}n^{1/4 - 1/(2d)}\log^{1/2 + 1/2d}n\right),
\]
for $d \geq 4$ even, and
\[
\text{disc}(S) = O\left(|S|^{1/4 + 1/(4d)}n^{3/4 - 3/(4d)}\log^{1/2 + 1/2d}n\right),
\]
for $d \geq 5$ odd, where the constant of proportionality in these bounds depends on $d$. In particular, when $d = 3$, the bound is $O(|S|^{1/3}\log^{5/3}n)$. In each of the above cases the corresponding coloring $\chi$ can be computed in expected polynomial time.

Remark. When $d = 2$ the resulting bound is $O(|S|^{1/4}\log^{7/4}n)$. However, this is slightly suboptimal with respect to the bound $O(|S|^{1/4}\log n)$ shown in [20].

3.1. An application to relative ($\epsilon, \delta$)-approximations. We next show how to obtain improved bounds on the size of relative ($\epsilon, \delta$)-approximations, using a variant of the halving technique presented in [20] (which was originally formulated in [30] and is overviewed in detail in [13]). We defer the analysis and the technical details to Appendix B and state our main result below.

Theorem 3.9. Let $(X, S)$ be a (finite) set system of primal shatter dimension $d$ with the additional property that in any set system restricted to $X' \subseteq X$ the number
of sets of size \( k \leq |X'| \) is \( O(|X'|^{d_1} k^{d-d_1}) \), where \( d_1 \) is an integer between 1 and \( d \). Then \((X, S)\) admits a relative \((\varepsilon, \delta)\)-approximation of size

\[
O \left( \frac{\log \frac{1}{\delta}}{\varepsilon^{d+1} \delta^{d+1}} \right),
\]

for \( d_1 > 1 \), and

\[
O \left( \frac{\log \frac{3d+1}{\delta \varepsilon^{d+1}}}{\frac{1}{\varepsilon \delta^{d+1}}} \right).
\]

for \( d_1 = 1 \), where the constant of proportionality depends on \( d \). Moreover, such a sample can be constructed in expected polynomial time.

When \((X, S)\) is a set system of points and halfspaces in \( d \)-space we thus conclude the following.

**Corollary 3.10.** Let \((X, S)\) be a set system of points and halfspaces in \( d \)-space. Then \((X, S)\) admits a relative \((\varepsilon, \delta)\)-approximation of size

\[
O \left( \frac{\log \frac{1}{\delta}}{\varepsilon^{d+1} \delta^{d+1}} \right),
\]

for \( d \geq 4 \), where the constant of proportionality depends on \( d \). When \( d = 3 \) this bound is

\[
O \left( \frac{\log \frac{2}{\delta}}{\varepsilon \delta^{d/2}} \right).
\]

**Concluding remarks and further research.** We note that whereas our construction is a variant of that of Matoušek [26], a key ingredient in our analysis is the sensitive packing lemma (Theorem 3.3), where we restrict each set under consideration to have a bounded size. As our analysis shows, when these sets are relatively small the bound on the size of the packing is considerably smaller than the bound \( O((n/\delta)^d) \) derived in the original packing lemma (Theorem 2.2). This bound is eventually integrated into the entropy method applied in Proposition 3.5 (in addition to our decomposition (3.2)), from which we eventually obtain the discrepancy bound in (3.3).

This study raises several open problems. A main problem, concerning the remarks following Proposition 3.5 and Theorem 3.6, is whether the logarithmic factor in our discrepancy bound can be reduced or even removed completely, in which case it becomes optimal. The appearance of this bound is due to two main ingredients: (i) the application of partial coloring, which results in an extra \( \log n \) factor for \( d_1 = 1 \), and (ii) the extra polylogarithmic factor in the bound obtained by the sensitive packing lemma (Theorem 3.3), which follows from the approach of Dudley [16] that results in a suboptimal packing bound. We believe that resolving (i) is an extremely challenging problem. Concerning (ii), in a follow-up work of the author with Dutta and Ghosh (see [17]), the extra logarithmic factor in the bound on the packing is removed completely. This is achieved by adapting the analysis of Haussler [22], who showed an optimal packing bound (stated in Theorem 2.2).

Another related problem concerns the discrepancy bounds in geometric set systems of “low degree.” Specifically, in abstract set systems \((X, S)\), this is the so-called Beck–Fiala problem, where each point of \( X \) appears in at most \( t \leq n \) sets of \( S \) (where
n = |X|). In this case, the discrepancy bound is conjectured to be \( O(\sqrt{t}) \) [11]. In fact, using the entropy method, one can obtain a constructive bound of \( O(\sqrt{t} \log n) \) [7, 25] (see also [35]), where the best currently known bound is by Banaszczyk [6], who showed a nonconstructive bound of \( O(\sqrt{t} \log n) \). The question at hand is what the corresponding bounds in geometric set systems should be. Specifically, assume we have a set system of \( n \) corresponding bounds in geometric set systems should be. Specifically, assume we have a set system of \( n \) points and halfspaces in \( d \)-space; is it possible to obtain a bound, which is roughly \( o(\sqrt{t}) \)? In an ongoing work, the author has shown that such bounds exist in two and three dimensions. We note that such bounds do not follow directly from the work on this paper, as in such settings \( S \) may contain a few large sets, whereas the bound on the discrepancy should depend only on the degree \( t \) (say, up to a logarithmic factor in \( n \)).

Last but not least, for further research, we suggest integrating our bound on relative \((\varepsilon, \delta)\)-approximations for points and halfspaces in \( d \)-dimensions (Corollary 3.10) with the existing approximate range counting machinery in [5] for \( d \geq 4 \) and [20] for \( d = 3 \). We hope this will improve the current performance bounds.

Appendix A. Proof of Proposition 3.5. We first note that at \( j_0 \) the above exponent becomes a constant, whereas the size of the packing becomes roughly \( n/\log n \) (for a fixed index \( i \)). Indeed, applying our choice in (3.3), we have

\[
\exp \left( -\frac{(\Delta_j)^2}{16s_j} \right) = \exp \left( -\frac{A^2 \cdot 2^{-1} \log^{1+1/d} n}{16(1 + |j - j_0|)4n^{1/d}(1-d_1/d)(i-1)} \right),
\]

which is \( \exp \left( -\frac{A^2}{16 \cdot 2^{d+1}} \right) \) at \( j = j_0 = (1/d) \log n + (1-d_1/d)(i-1)-(1+1/d) \log \log n-B \). Concerning the bound on the packing size, \( C \cdot \frac{n^{2(d-d_1)(i-1)}}{2^{d-2d_1(i-1)} \log^{d+1} n} \), since \( j \) can always be bounded by \( k = \log n \), at \( j = j_0 \) we obtain

\[
C \cdot \frac{n^{2(d-d_1)(i-1)}}{2^{d-2d_1(i-1)} \log^{d+1} n} \leq C \cdot \frac{n \cdot 2^{d-2d_1(i-1)}}{2^{d-2d_1(i-1)} \log^{d+1} n}.
\]

We now fix an index \( i \), split the summation into the two parts \( j \geq j_0 \) and \( i-1 \leq j < j_0 \), and then bound each part in turn. In the first part, the exponent will “take over” the summation in the sense that it decreases superexponentially, making the other factors (with \( j > j_0 \)) insignificant, and in the second part, the packing size will decrease geometrically. Thus the “peak” of this summation is obtained at \( j = j_0 \) and is decreasing as we go beyond or below \( j \).

For the first part, put \( j := j_0 + l \), for an integer \( l \geq 0 \), and then

\[
\sum_{i=1}^{k} \sum_{j=j_0}^{\infty} C \cdot \frac{j^{d-2d_1}}{2^{(d-d_1)(i-1)}} \exp \left( -\frac{(\Delta_j)^2}{16s_j} \right)
\leq \sum_{i=1}^{k} \sum_{l=0}^{k-j_0} C \cdot \frac{n^{2d}}{2^{dB \log n}} \exp \left(-\frac{A^2 \cdot 2^{l-(B+1)}}{(1+l)^4} \right)
\leq C \cdot n^{2-dB} \sum_{l=0}^{k-j_0} 2^{ld} \exp \left(-\frac{A^2 \cdot 2^{l-(B+1)}}{(1+l)^4} \right),
\]

where the logarithmic factor in the packing size is now eliminated due to the summation over \( i \). The exponents in the above sum decrease superexponentially. Choosing
A sufficiently large (say, \( A > 2^{6+|B+1|+\log d} \)) and having \( B > 5 + \log C \) as above, we can guarantee that the latter sum is strictly smaller than \( n/32 \).

When \( j < j_0 \) put \( j := j_0 - l, \ l > 0 \) as above. We now obtain, by just bounding the exponent from above by 1, and using similar considerations as above,

\[
\sum_{i=1}^{k} \sum_{j=1-i}^{j_0-i} C \cdot \frac{\Delta_i^2}{2^{d-d_i}(n-1)} \exp \left( -\frac{\Delta_i^2}{16n} \right) \leq \sum_{i=1}^{j_0-(l-1)} C \cdot \frac{n}{2^{d+(\epsilon+B)}}.
\]

Once again, our choice for \( B \) guarantees that the above (geometrically decreasing) sum is strictly smaller than \( n/32 \). Thus the entire summation is bounded by \( n/16 \), as asserted.

**Appendix B. Proof of Theorem 3.9.** Following the arguments in [20], it is sufficient to construct a \((\nu, \alpha)\)-sample for \((X, \mathcal{S})\), as such a sample is equivalent to a relative \((\varepsilon, \delta)\)-approximation (where \( \nu, \varepsilon \) and \( \alpha, \delta \) are within some constant factor from each other; see section 1).

For simplicity of presentation, we now assume, without loss of generality, that \( S := X \) is a set in our set system and the coloring induces an even split on \( X \) (recall that we assumed that \( \log n \) is an integer). The latter assumption can be obtained by setting the discrepancy parameter \( \Delta_S \) to zero in this case. This may require us to skip the very last \( O(1) \) partial coloring steps and thus add an \( O(1) \) factor to the discrepancy bound (that is, when \( n \) becomes a constant, the contribution of \( S = X \) to the entropy function may not be negligible). However, this modification does not change the asymptotic bound on the discrepancy obtained in Theorem 3.6, and thus from now on we assume that the coloring halves \( X \).

Our construction proceeds over iterations, where we repeatedly “halve” \( X \) until we obtain a subset of an appropriate size, which we argue comprises the resulting \((\nu, \alpha)\)-sample. Put \( X_0 := X \). Then, at each iteration \( i \geq 1 \), we let \( X_i, X'_i \) be the two corresponding portions of \( X_{i-1} \), where the points in \( X_i \) are, say, colored \( +1 \) and the points in \( X'_i \) are colored \(-1 \). We now keep \( X_i \), remove \( X'_i \), and continue in this “halving” process, and thus the desired \((\nu, \alpha)\)-sample corresponds to some set \( X_i \) obtained in an appropriate iteration. Put \( n_i := |X_i| \). Since the split is even, we have \( n_i = n/2^l \) (recall once again that \( n \) is an integer power of 2).

Applying Theorem 3.6 at iteration \( i \) we obtain, for each set \( S \in \mathcal{S} \),

\[
||S \cap X_i| - |S \cap X'_i|| \leq K \cdot |S \cap X_{i-1}|^{1/2-d_i/(2d)} |X_{i-1}|^{(d_i-1)/(2d)} \log^{1/2+1/(2d)} |X_{i-1}|
\]

for an appropriate constant \( K > 0 \) which depends on \( d \).

We use a variant of the considerations in [20] and proceed as follows. Our goal is to bound, for each \( S \in \mathcal{S} \), the difference \( |S \cap X_{i-1}|/|X_{i-1}| - |S \cap X'_i|/|X'_i| \), which we also denote by \( |X_{i-1}(S) - X_i(S)| \) (recall our notation for a measure of a set from section 1). Since \( |X_i| = |X'_i| = |X_{i-1}|/2 \), this difference is

\[
\left| \frac{|S \cap X_i| + |S \cap X'_i|}{|X_{i-1}|} - 2|S \cap X_i|/|X_{i-1}| \right| = \left| \frac{|S \cap X'_i|}{|X_{i-1}|} - \frac{|S \cap X_i|}{|X_{i-1}|} \right|.
\]

We have thus shown that

\[
|X_{i-1}(S) - X_i(S)| = \left| \frac{|S \cap X'_i| - |S \cap X_i|}{|X_{i-1}|} \right|.
\]

\footnote{For the time being, we assume \( d_i > 1 \) and use the corresponding discrepancy bound in Theorem 3.6. The case \( d_i = 1 \) is handled later on, and its analysis follows almost verbatim; see below.}
We now rewrite the latter term as
\[ |X_{i-1}(S) - X_i(S)| \]
\[ \leq \frac{K \cdot |S \cap X_{i-1}^{1/2-d_i/(2d)}| |X_{i-1}^{(d_i-1)/(2d)} \log^{1/2+1/2d} |X_{i-1}|}. \]

We now rewrite the latter term as
\[ \frac{K X_{i-1}(S)^{1/2-d_{i}/(2d)}}{|X_{i-1}|^{1/2+d_{i}/(2d)}} \cdot |X_{i-1}^{(d_i-1)/(2d)} \log^{1/2+1/2d} |X_{i-1}| \]
\[ = \frac{K X_{i-1}(S)^{1/2-d_{i}/(2d)}}{|X_{i-1}|^{1/2+1/(2d)}} \log^{1/2+1/2d} |X_{i-1}| \]
and use the observation that \( x^p \leq (x + y)^{1-p} \) for \( x \geq 0, y > 0 \), and \( 0 \leq p \leq 1 \) (stated in [20]) in order to bound the latter term by (we now replace \( |X_{i-1}| \) with \( n_{i-1} \) and set \( p := 1/2 - d_i/(2d) \)):
\[ \leq \frac{K \cdot \log^{1/2+1/2d} n_{i-1}}{n_{i-1}^{1/2+1/(2d)}} \cdot \frac{X_{i-1}(S) + \nu}{\nu^{1/2+d_i/(2d)}} \cdot \frac{X_i(S) + X_{i-1}(S) + \nu}{\nu^{1/2+d_i/(2d)}}. \]

This implies that
\[ d_{\nu}(X_{i-1}(S), X_i(S)) = \frac{|X_{i-1}(S) - X_i(S)|}{X_{i-1}(S) + X_i(S) + \nu} \leq \frac{K \cdot \log^{1/2+1/2d} n_{i-1}}{\nu^{1/2+d_i/(2d)} n_{i-1}^{1/2+1/(2d)}}. \]

Since \( d_{\nu}(\cdot, \cdot) \) satisfies the triangle inequality (see [24]), we obtain
\[ d_{\nu}(X_0(S), X_i(S)) \leq \sum_{j=1}^{i} d_{\nu}(X_{j-1}(S), X_j(S)) \]
\[ \leq \frac{K}{\nu^{1/2+d_i/(2d)}} \sum_{j=1}^{i} \frac{\log^{1/2+1/(2d)} n_{j-1}}{n_{j-1}^{1/2+1/(2d)}}. \]

Since \( n_j = O(n/2^j) \), \( j = 1, \ldots, i \), we obtain that the latter expression is bounded by
\[ O \left( \frac{\log^{1/2+1/(2d)} n_{i-1}}{\nu^{1/2+d_i/(2d)} n_{i-1}^{1/2+1/(2d)}} \right). \]

We next bound \( d_{\nu}(X_0(S), X_i(S)) \) by the relative error \( \alpha \), in order to conclude that this is valid as long as
\[ n_{i-1} = \Omega \left( \frac{\log 1/\alpha}{\nu^{d+d_i/(d+1)} \alpha^{2d/(d+1)}} \right). \]
We thus stop at that iteration $i$ for which the set $X_{i-1}$ is the smallest that still satisfies this lower bound, from which we obtain the asserted bound on the size of the $(\nu, \alpha)$-sample.

In case $d_1 = 1$, the analysis proceeds almost verbatim, with the slight difference that the bound in (B.1) is now $K \cdot |S \cap X_{i-1}|^{1/2 - 1/(2d)} \log^{3/2 + 1/(2d)} |X_{i-1}|$. Following similar considerations, this leads to the lower bound

$$\Omega \left( \frac{\log^{3d+1} \frac{1}{\nu \alpha}}{\nu \alpha 2d/(d+1)} \right)$$

on $n_{i-1}$.

Acknowledgments. The author wishes to thank Boris Aronov and Micha Sharir for helpful discussions and for their help in writing this paper. The author also wishes to thank Jiri Matoušek for several useful comments about this paper, Shachar Lovett for several discussions concerning discrepancy theory, and the two anonymous referees for their helpful comments.

REFERENCES

[1] P. Afshani and T. M. Chan, On approximate range counting and depth, Discrete Comput. Geom., 42 (2009), pp. 3–21.
[2] R. Alexander, Geometric methods in the theory of uniform distribution, Combinatorica, 10 (1990), pp. 115–136.
[3] N. Alon and J. H. Spencer, The Probabilistic Method, 2nd ed., Wiley-Interscience, New York, 2000.
[4] B. Aronov and S. Har-Peled, On approximating the depth and related problems, SIAM J. Comput., 38 (2008), pp. 899–921.
[5] B. Aronov and M. Sharir, Approximate halfspace range counting, SIAM J. Comput., 39 (2010), pp. 2704–2725.
[6] W. Banaszczyk, Balancing vectors and Gaussian measures of $n$-dimensional convex bodies, Random Structures Algorithms, 12 (1998), pp. 351–360.
[7] N. Bansal, Constructive algorithms for discrepancy minimization, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 3–10.
[8] J. Beck, Roth’s estimate on the discrepancy of integer sequences is nearly sharp, Combinatorica, 1 (1981), pp. 319–325.
[9] J. Beck, Irregularities of distribution I., Acta Math., 159 (1987), pp. 1–49.
[10] J. Beck and W. Chen, Irregularities of Distribution, Cambridge University Press, Cambridge, UK, 1987.
[11] J. Beck and T. Fiala, “Integer making” theorems, Discrete Appl. Math., 3 (1981), pp. 1–8.
[12] B. Chazelle, A Note on Haussler’s Packing Lemma, unpublished manuscript, Princeton University, Princeton, NJ, 1992.
[13] B. Chazelle, The Discrepancy Method: Randomness and Complexity, Cambridge University Press, Cambridge, UK, 2000; paperback, 2001.
[14] B. Chazelle and E. Welzl, Quasi-optimal range searching in spaces of finite VC-dimension, Discrete Comput. Geom., 4 (1989), pp. 467–489.
[15] K. L. Clarkson and P. W. Shor, Applications of random sampling in computational geometry, II., Discrete Comput. Geom., 4 (1989), pp. 387–421.
[16] R. M. Dudley, Central limit theorems for empirical measures, Ann. Probab., 6 (1978), pp. 899–929.
[17] K. Dutta, E. Ezra, and A. Ghosh, Two proofs for shallow packings, in Proceedings of the Symposium on Computational Geometry, 2015, pp. 96–110.
[18] E. Ezra, A size-sensitive discrepancy bound for set systems of bounded primal shatter dimension, in Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 2014, pp. 1378–1388.
[19] S. Har-Peled, Geometric Approximation Algorithms, Math. Surveys Monogr. 173, American Mathematical Society, Providence, RI, 2011.
[20] S. Har-Peled and M. Sharir, Relative $(\mu, \epsilon)$-approximations in geometry, Discrete Comput. Geom., 45 (2011), pp. 462–496.
[21] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, Inform. and Comput., 100 (1992), pp. 78–150.
[22] D. Haussler, Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension, J. Combin. Theory Ser. A, 69 (1995), pp. 217–232.
[23] D. Haussler and E. Welzl, ε-nets and simplex range queries, Discrete Comput. Geom., 2 (1987), pp. 127–151.
[24] Y. Li, P. M. Long, and A. Srinivasan, Improved bounds on the sample complexity of learning, J. Comput. System Sci., 62 (2001), pp. 516–527.
[25] S. Lovett and R. Meek, Constructive discrepancy minimization by walking on the edges, in Proceedings of the 53th Annual IEEE Symposium on Foundations of Computer Science, 2012, pp. 61–67.
[26] J. Matoušek, Tight upper bounds for the discrepancy of halfspaces, Discrete Comput. Geom., 13 (1995), pp. 593–601.
[27] J. Matoušek, An $L_p$ version of the Beck-Fiala conjecture, European J. Combin., 19 (1998), pp. 175–182.
[28] J. Matoušek, Geometric Discrepancy, Algorithms Combin. 18, Springer Verlag, Heidelberg, 1999.
[29] J. Matoušek and J. Spencer, Discrepancy in arithmetic progressions, J. Amer. Math. Soc., 9 (1996), pp. 195–204.
[30] J. Matoušek, E. Welzl, and L. Wernisch, Discrepancy and ε-approximations for bounded VC-dimension, Combinatorica, 13 (1993), pp. 455–466.
[31] S. Muthukrishnan and A. Nikolov, Optimal private halfspace counting via discrepancy, in Proceedings of the 44th Annual ACM Symposium on Theory of Computing, 2012, pp. 1285–1292.
[32] T. Rothvoss, Constructive discrepancy minimization for convex sets, in Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science, 2014, pp. 140–145.
[33] M. Sharir and S. Zaban, Output-Sensitive Tools for Range Searching in Higher Dimensions, unpublished manuscript, 2011; available online from www.cs.tau.ac.il/thesis/thesis/zaban.pdf.
[34] J. Spencer, Six standard deviations suffice, Trans. Amer. Math. Soc., 289 (1985), pp. 679–706.
[35] A. Srinivasan, Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems, in Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, 1997, pp. 692–701.
[36] V. N. Vapnik and A. Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Appl., 16 (1971), pp. 264–280.
[37] E. Welzl, Partition trees for triangle counting and other range searching problems, in Proceedings of the 4th Annual ACM Symposium on Computational Geometry, 1988, pp. 23–33.
[38] E. Welzl, On spanning trees with low crossing numbers, in Data Structures and Efficient Algorithms, Lecture Notes in Comput. Sci. 594, Springer-Verlag, Heidelberg, 1992, pp. 233–249.