2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)

# Truly Optimal Euclidean Spanners

Hung Le
*Department of Computer Science*
*University of Victoria*
*Victoria, BC, Canada*
*hungle@uvic.ca*

Shay Solomon
*School of Electrical Engineering*
*Tel Aviv University*
*Tel Aviv, Israel*
*solo.shay@gmail.com*

***Abstract*—Euclidean spanners are important geometric structures, having found numerous applications over the years. Cornerstone results in this area from the late 80s and early 90s state that for any $d$-dimensional $n$-point Euclidean space, there exists a $(1+\epsilon)$-spanner with $O(n\epsilon^{-d+1})$ edges and lightness (normalized weight) $O(\epsilon^{-2d})$.[1] Surprisingly, the fundamental question of whether or not these dependencies on $\epsilon$ and $d$ for small $d$ can be improved has remained elusive, even for $d=2$. This question naturally arises in any application of Euclidean spanners where precision is a necessity (thus $\epsilon$ is tiny). In the most extreme case $\epsilon$ is inverse polynomial in $n$, and then one could potentially improve the size and lightness bounds by factors that are polynomial in $n$.**

**The state-of-the-art bounds $O(n\epsilon^{-d+1})$ and $O(\epsilon^{-2d})$ on the size and lightness of spanners are realized by the *greedy* spanner. In 2016, Filtser and Solomon [25] proved that, in low dimensional spaces, the greedy spanner is "near-optimal"; informally, their result states that the greedy spanner for dimension $d$ is just as sparse and light as any other spanner *but for dimension larger by a constant factor*. Hence the question of whether the greedy spanner is truly optimal remained open to date.**

**The contribution of this paper is two-fold.**

1) **We resolve these longstanding questions by nailing down the exact dependencies on $\epsilon$ and $d$ and showing that the greedy spanner is truly optimal. Specifically, for any $d = O(1), \epsilon = \Omega(n^{-\frac{1}{d-1}})$:**
   - **We show that any $(1+\epsilon)$-spanner must have $\Omega(n\epsilon^{-d+1})$ edges, implying that the greedy (and other) spanners achieve the optimal size.**
   - **We show that any $(1+\epsilon)$-spanner must have lightness $\Omega(\epsilon^{-d})$, and then improve the upper bound on the lightness of the greedy spanner from $O(\epsilon^{-2d})$ to $\tilde{O}_\epsilon(\epsilon^{-d})$.**
2) **We then complement our negative result for the size of spanners with a rather counterintuitive positive result: Steiner points lead to a quadratic improvement in the size of spanners! Our bound for the size of Steiner spanners is tight as well (up to lower-order terms).**

*Keywords*-**Euclidean spanners; light spanners; Steiner spanners; spherical codes;**

## I. Introduction

### *1.1 Background and motivation:*

*Sparse spanners:* Let $P$ be a set of $n$ points in $\mathbb{R}^d, d \geq 2$, and consider the complete weighted graph $G_P = (P, \binom{P}{2})$ induced by $P$, where the weight of any edge $(x,y) \in \binom{P}{2}$ is the Euclidean distance $|xy|$ between its endpoints. Let $H = (P,E)$ be a spanning subgraph of $G_P$, with $E \subseteq \binom{P}{2}$, where, as in $G_P$, the weight function is given by the Euclidean distances. For any $t \geq 1$, $H$ is called a $t$-*spanner* for $P$ if for every $x,y \in P$, the distance $d_H(x,y)$ between $x$ and $y$ in $H$ is at most $t|xy|$; the parameter $t$ is called the *stretch* of the spanner and the most basic goal is to get it down to $1+\epsilon$, for arbitrarily small $\epsilon > 0$, without using too many edges. Euclidean spanners were introduced in the pioneering SoCG'86 paper of Chew [16], who showed that $O(n)$ edges can be achieved with stretch $\sqrt{10}$, and later improved the stretch bound to 2 [17]. The first Euclidean spanners with stretch $1+\epsilon$, for an arbitrarily small $\epsilon > 0$, were presented independently in the seminal works of Clarkson [18] (FOCS'87) and Keil [38] (see also [39]), which introduced the $\Theta$-*graph* in $\mathbb{R}^2$ and $\mathbb{R}^3$, and soon afterwards was generalized for any $\mathbb{R}^d$ in [45], [2]. The $\Theta$-graph is a natural variant of the *Yao graph*, introduced by Yao [51] in 1982, where, roughly speaking, the space $\mathbb{R}^d$ around each point $p \in P$ is partitioned into cones of angle $\Theta$ each, and then edges are added between each point $p \in P$ and its closest points in each of the cones centered around it. The $\Theta$-graph is defined similarly, where, instead of connecting $p$ to its closest point in each cone, we connect it to a point whose orthogonal projection to some fixed ray contained in the cone is closest to $p$. Taking $\Theta$ to be $c\epsilon$, for small enough constant $c$, one obtains a $(1+\epsilon)$-spanner with $O(n\epsilon^{-d+1})$ edges.

Euclidean spanners turned out to be a fundamental geometric construct, which evolved into an important research area [38], [39], [2], [20], [3], [21], [4], [44], [31], [26], [1], [12], [29], [13], [22], [49], [24], with a plethora of applications, such as in geometric approximation algorithms [44], [32], [33], [34], geometric distance oracles [32], [33], [35], [34], network design [36], [41] and machine learning [28]. (See the book by Narasimhan and Smid [42] for an excellent account on Euclidean spanners and some of their applications.)

The tradeoff between stretch $1+\epsilon$ and $O(n\epsilon^{-d+1})$ edges

[1] The lightness of a spanner is the ratio of its weight and the MST weight.

2575-8454/19/$31.00 ©2019 IEEE
DOI 10.1109/FOCS.2019.00069

is the current state-of-the-art, and is also achieved by other spanner constructions, including the *path-greedy* (abbreviated as "greedy") spanner [2], [14], [42] and the gap-greedy spanner [46], [4]. Surprisingly, despite the extensive body of work on Euclidean spanners since the 80s, the following fundamental question remained open, even for $d = 2$.

**Question 1.** *Is the tradeoff between stretch $1 + \epsilon$ and $n \cdot O(\epsilon^{-d+1})$ edges tight?*

We remark that the $\Theta$-graph and its variants provide stretch $1+\epsilon$ only for sufficiently small angle $\Theta$. These graphs have also been studied for fixed values of $\Theta$; see [10], [6], [11], [5], [40], [37], [9], and the references therein. The general goal here is to determine the best possible stretch for small values of $\Theta$. E.g., it was shown in SODA'19 [9] that the $\Theta$ graph with 4 cones, $\Theta_4$, has stretch $\leq 17$. This line of work is somewhat orthogonal to Question 1, which concerns the *asymptotic* behavior of the tradoeff.

*Light spanners:* Another basic property of spanners, important for various applications, is *lightness*, defined as the ratio of the spanner *weight* (i.e., the sum of all edge weights in it) to the weight $w(\mathrm{MST}(P))$ of the minimum spanning tree $\mathrm{MST}(P)$ for $P$. ($\mathrm{MST}(P)$ is also the minimum spanning tree of graph $G_P$.) In SoCG'93, Das et al. [19] showed that the aforementioned greedy spanner of [2] has constant lightness in $\mathbb{R}^3$, which was generalized in SODA'95 [21] to $\mathbb{R}^d$ for any constant $d$; the dependencies on $\epsilon$ and $d$ in the lightness bound were not explicated in [2], [19], [21]. Later, in their seminal STOC'98 paper on approximating TSP in $\mathbb{R}^d$ using light spanners, Rao and Smith [44] showed that the greedy spanner has lightness $\epsilon^{-O(d)}$ in $\mathbb{R}^d$ for any constant $d$. In the open problems section of their paper [44], they raised the question of determining the exact constant hiding in the $O$-notation $O(d)$ in the exponent of their upper bound. [2] All the proofs in [2], [19], [21], [44] had many missing details. The first complete proof was given in the book of [42], where a 60-page chapter was devoted to it, showing that the greedy $(1 + \epsilon)$-spanner has lightness $O(\epsilon^{-2d})$. In SODA'19, Borradaile, Le and Wulff-Nilsen [8] presented a much shorter and arguably simpler alternative proof that, in fact, applies to the wider family of *doubling metrics* (see also [30]), but the lightness bound of $O(\epsilon^{-2d})$ remains the state-of-the-art.[3] Therefore, the following question remained open all these years, even for $d = 2$.

**Question 2.** *Is the tradeoff between stretch $1 + \epsilon$ and lightness $O(\epsilon^{-2d})$ tight?*

*Existential near-optimality:* In PODC'16, Filtser and Solomon [25] studied the optimality of the greedy spanner in *doubling metrics*, which is wider than the family of low-dimensional Euclidean spaces. They showed that the greedy spanner is *existentially near-optimal* with respect to both the size and the lightness. Roughly speaking, the greedy spanner is said to be existentially optimal for a graph family $\mathcal{G}$ if its worst performance (in terms of size and/or lightness) over all graphs in $\mathcal{G}$ is just as good as the worst performance of an optimal spanner over all graphs in $\mathcal{G}$. For doubling metrics, the loss encapsulated by the "near-optimality" guarantee comes into play with the dimension $d$: one compares the greedy $(1 + \epsilon)$-spanner over metrics with doubling dimension $d$ with any other $(1 + \epsilon)$-spanner, but over metrics with doubling dimension $2d$. This loss in the dimension becomes more significant if we restrict the attention to Euclidean spaces, as then the comparison is between Euclidean dimension $d$ and doubling dimension $2d$, but spanners for metrics with doubling dimension $2d$ (or even $d$) tend to admit significantly weaker guarantees (as a function of $\epsilon$ and $d$) than the corresponding ones for $d$-dimensional Euclidean spaces.

Consequently, this result by [25] maybe not resolve Questions 1 and 2 for two reasons. First, it only implies near-optimality of the greedy spanner, which, as mentioned, comes with a constant factor loss in the dimension, and this constant factor slack appears in the exponents of the size and lightness bounds. Second, and more importantly, even if we knew that the greedy spanner is truly optimal – which is what we show in the current work – this still does not unveil the tight dependencies on $\epsilon$ and $d$.

*1.2 Our contribution:* Throughout we assume that $\epsilon \ll 1$. Our starting point is a surprisingly simple observation regarding evenly spaced point sets on the $d$-dimensional sphere, using which we prove:

**Theorem I.1.** *For any constant $d$ and any $n$ and $\epsilon$ such that $\epsilon = \Omega(n^{-\frac{1}{d-1}})$, there is a set $P$ of $n$ points in $\mathbb{R}^d$ such that any $(1 + \epsilon)$-spanner for $P$ must have lightness $\Omega(\epsilon^{-d})$ and $\Omega(n\epsilon^{-d+1})$ edges.*

Theorem I.1 immediately resolves Question 1 in the affirmative, and it also shows that the greedy spanner is truly optimal with respect to the size parameter.

We then improve the lightness bound of the greedy spanner to match our lower bound.

**Theorem I.2.** *The greedy $(1+\epsilon)$-spanner in $\mathbb{R}^d$ has lightness $\tilde{O}_\epsilon\left(\epsilon^{-d}\right)$ where $\tilde{O}_\epsilon(.)$ notation hides the poly-logarithmic factor of $\frac{1}{\epsilon}$.*

Theorem I.2 answers Question 2 in the negative, and it also shows that the greedy spanner is truly optimal with

[2] In the full (unpublished) version of their paper, Rao and Smith remarked that in the Euclidean plane, a lightness bound of $O(\epsilon^{-2})$ is optimal by pointing out that any $(1+\epsilon)$-spanner of a set of $\Theta(\frac{1}{\epsilon})$ points evenly placed on the boundary of a circle has lightness $\Omega(\frac{1}{\epsilon^2})$; this statement was not accompanied with a proof. In general, the full unpublished version of [44] contains several claims on light spanners, but the proofs for most of these claims are either missing, incomplete or incorrect.

[3] The *doubling dimension* of a metric space $(X, \delta)$ is the smallest value $d$ such that every ball $B$ in the metric space can be covered by at most $2^d$ balls of half the radius of $B$. This notion generalizes the Euclidean dimension, since the doubling dimension of the Euclidean space $\mathbb{R}^d$ is $\Theta(d)$. A metric space is called *doubling* if its doubling dimension is constant.

respect to the lightness parameter. The proof of Theorem I.2 is intricate.

Our lightness analysis of the greedy algorithm builds on exciting developments on light spanners from recent years, which started from the works of Gottlieb [30] and Chechik and Wulff-Nilsen [15] on non-greedy spanners. Using the result of [25], the framework of [15] was refined in the works of Borradaile, Le and Wulff-Nilsen [7], [8]. As mentioned, it was shown in [8] that the greedy spanner in metrics of doubling dimension $d$ has lightness $\epsilon^{-O(d)}$. We demonstrate that, by adapting the analysis in [8] to Euclidean spaces and applying a few tweaks, one can obtain a lightness bound of $\epsilon^{-(d+2)}$. To shave the remaining slack of $\epsilon^{-2}$ factor, we introduce several highly nontrivial geometric insights to the analysis.

*Sparse Steiner spanners: Steiner points* are additional points that are not part of the input point set. A standard usage of Steiner points is for reducing the weight of the tree, with the Steiner Minimum Tree (SMT) problem serving as a prime example: In any metric, the Steiner ratio, which is the ratio of the SMT weight to the MST weight, is at least $\frac{1}{2}$ (by the triangle inequality) and at most 1 (by definition). In $\mathbb{R}^2$ the Steiner ratio is known to be between $\approx 0.824$ and $\frac{\sqrt{3}}{2} \approx 0.866$, and the famous (still open) "Gilbert-Pollak Conjecture" is that the upper bound $\frac{\sqrt{3}}{2}$ is tight [27]. As another example, a spanning tree that simultaneously approximates a shortest-path tree and a minimum spanning tree is called a *shallow-light tree* (shortly, SLT). In FOCS'11, Elkin and Solomon [23] showed that in general metric spaces, Steiner points can be used to get an exponential improvement to the lightness of SLTs. The construction of [23] does not apply to Euclidean spaces, but Solomon [48] showed that Steiner points can be used to get a quadratic improvement to the lightness of SLTs in $\mathbb{R}^d$ for $d = O(1)$.

Although these examples demonstrate that Steiner points could be very useful for reducing the weight of tree structures, note that the resulting Steiner trees must contain more edges than the original trees by definition. Broadly speaking, it seems counterintuitive that Steiner points could be used as means for reducing the number of edges of graph structures such as spanners. And indeed, essentially all the prior work in this context only support this intuition; in particular, Althöfer et al. [2] asserts that, in general metrics, Steiner points do not help much in reducing the spanner size (see Theorems 6-8 therein), and this result was strengthened in [43] (see Theorem 1.2 therein). We remark that these hardness results of [2], [43] are based on girth arguments, and are not applicable in low-dimensional Euclidean spaces.

The size lower bound provided by Theorem I.1 implies that cornerstone spanner constructions from the 80s, such as the $\Theta$-graph and the greedy spanner, cannot be improved in size. We contrast this negative message with a positive and counterintuitive one: Steiner points can be used to obtain a quadratic improvement on the size of spanners! We'll focus on the Euclidean plane $\mathbb{R}^2$, but we get this quadratic improvement in any dimension $d \geq 2$: $n \cdot \frac{\epsilon^{-o(1)}}{\epsilon^{(d-1)/2}}$ edges using Steiner points versus $n \cdot \Omega(\epsilon^{-d+1})$ edges without using them. We use $\tilde{O}_\epsilon$ and $\tilde{\Omega}_\epsilon$ to suppress poly-logarithmic factors of $\log \frac{1}{\epsilon}$.

**Theorem I.3.** *For any set of $n$ points $P$ in $\mathbb{R}^2$, there is a Steiner $(1+\epsilon)$-spanner for $P$ with $\frac{n}{\sqrt{\epsilon}} \cdot \epsilon^{-o(1)}$ edges. In general, there is a Steiner spanner with $n \cdot \frac{\epsilon^{-o(1)}}{\epsilon^{(d-1)/2}}$ edges, for any set of $n$ points $P$ in $\mathbb{R}^d$.*

**Remarks.** (1) The exact upper bound is $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log \frac{1}{\epsilon}})}$ in $\mathbb{R}^2$ and $\frac{n}{\epsilon^{(d-1)/2}} 2^{O(\sqrt{\log \frac{1}{\epsilon}})}$ in $\mathbb{R}^d$. ; we did not try to optimize the lower-order term $\epsilon^{-o(1)}$. (2) Our construction can be implemented in near-linear time; this implementation lies outside the scope of the current paper, which aims to study the combinatorial tradeoffs between the stretch and size/lightness of spanners.

The following lower bound shows that our construction of sparse Steiner spanners (Theorem I.3) is optimal to within the lower-order term $\epsilon^{-o(1)}$ for 2-dimensional Euclidean spaces.

**Theorem I.4.** *For any $n$ and $\epsilon$ such that $\epsilon = \tilde{\Omega}(\frac{1}{n^2})$, there exists a set of $n$ points in $\mathbb{R}^2$ such that any Steiner $(1+\epsilon)$-spanner must have at least $\tilde{\Omega}_\epsilon(\frac{n}{\sqrt{\epsilon}})$ edges and lightness at least $\tilde{\Omega}_\epsilon(\frac{1}{\epsilon})$.*

**Remark.** Since the SMT and MST weights are the same up to a small constant, we can define the lightness of Steiner spanners with respect to the MST weight, just as with non-Steiner spanners.

To prove our upper and lower bounds for Steiner spanners (Theorems I.3 and I.4), we come up with novel geometric insights, which may be of independent interest, as discussed in the next section.

*1.3 Proof overview, comparison with prior work, and technical highlights:* The starting point of this work is in making a remarkably simple observation regarding a set of evenly spaced points along the boundary of a circle, which suffices for getting the lower bound for the size and lightness of spanners in $\mathbb{R}^2$ (Theorem I.1). Numerous papers have identified this point set as a natural candidate for lower bounds (see, e.g., [44], [23], [49]), yet we are not aware of any paper that managed to rigourously prove such a result. The $d$-dimensional analogue is a set of evenly spaced points along the sphere, providing a set of $\Theta(\epsilon^{-d+1})$ points corresponding to the codewords of a spherical code in $\mathbb{R}^d$. The distance between any two codewords is $\Omega(\epsilon)$, using which we show that for any two points $x, y$ with $|xy| = \Theta(1)$, any $(1+\epsilon)$-spanner must take $xy$ as an edge. Since $P$ has $\Theta(\epsilon^{-2d+2})$ pairs of points of distance $\Theta(1)$, any spanner for $P$ must have lightness $O(\epsilon^{-d})$ and $\Omega(\epsilon^{-2d+2})$ edges. The lightness bound immediately follows, whereas to obtain the size lower bound we consider multiple copies of the same point set that lie sufficiently far from each other, so that each point set must be handled with a separate vertex-disjoint spanner; see Section III.

We bypass the size lower bound by using Steiner points. For simplicity of presentation, we mostly focus on the Euclidean plane $\mathbb{R}^2$, but the argument can be naturally generalized for $\mathbb{R}^d, d > 2$. We start with constructing Steiner spanners for point sets of bounded spread $\Delta$, with $\frac{n}{\sqrt{\epsilon}} 2^{O(\sqrt{\log \Delta})}$ edges.[4] Our construction is recursive, and the main idea is to partition each bounding square in the current recursion level into *overlapping subsquares*. The observation is that short distances are handled by deeper levels of the recursion (due to the overlapping regions) while long distances can be handled with relatively few edges (using Steiner points) in the current recursion level. Using this observation carefully, we are able to construct a sparse (as a function of $\Delta$) Steiner spanner. We then show a reduction from a general point set (of possibly huge spread) to a point set of spread $O(\frac{1}{\epsilon})$. This reduction builds on the standard net-tree spanner (see, e.g., [26], [13], [29]) in a novel way, using a notion that we shall refer to as a *ring spanner*. A $t$-ring spanner of a point set, for $t \geq 1$, is a spanner that preserves (to within a factor of $t$) distances between every pair of points $p, q$ such that $q$ belongs to a ring (or annulus) around $p$. The net-tree spanner is obtained, in fact, as a union of $\Theta(\log \Delta)$ ring 1-spanners, where the inner and outer radii of the annulus are within a factor of $1/\epsilon$. Using this fact, we are able to reduce the problem of constructing a Steiner spanner for a general point set to the problem of constructing $\Theta(\log \Delta)$ ring $(1+\epsilon)$-spanners for point sets of spread $O(\frac{1}{\epsilon})$ each, and further show how to reduce it to the construction of just one such ring $(1+\epsilon)$-spanner. Our strategy of constructing Steiner spanners by building on the net-tree spanner is somewhat surprising, since all known (non-Steiner) net-tree spanners have $\Omega(\frac{1}{\epsilon}^2)$ edges, which exceeds the optimal bound $O(\frac{1}{\epsilon})$ obtained by other spanners (such as the $\Theta$-graph) by a factor of $1/\epsilon$. However, by looking at the net-tree spanner through the lens of ring spanners and, of course, through the use of Steiner points, we are able to achieve the improved size bound; the details appear in Section IV.

To prove the lower bound on the size of Steiner spanners in $\mathbb{R}^2$, we can use the same point set used for our lower bounds for non-Steiner spanners, of evenly spaced points along the boundary of a circle. The argument here, however, is significantly more intricate. It is technically more convenient to work with a similar point set $P$, where the points are evenly spaced along two opposite sides of a unit square $U$, denoted by $N$ ("north") and $S$ ("south"). The distance between any two consecutive points along $N$ and along $S$ will be $\Theta(\sqrt{\epsilon \log \frac{1}{\epsilon}})$, so that $|P| = \Theta_\epsilon(\frac{1}{\sqrt{\epsilon}})$. Our goal is to show that any Steiner spanner must use roughly $\Theta(|P|^2)$ edges to preserve the distances for all pairs of points from $N$ and $S$ to within a factor of $1+\epsilon$, and then taking multiple copies of the same point set that are sufficiently far from each other would complete the proof. Instead of proving the size lower bound directly, we show that any Steiner spanner for $P$ must incur a weight of $\Omega(|P|^2)$; the size lower bound would follow easily, as the distance between any pair of points in $P$ is $O(1)$. We then demonstrate that the problem of lower bounding the spanner weight for $P$ boils down to the problem of determining the lengths of intersecting shortest paths in the spanner. Next, we say that the *intersecting pattern* of two shortest paths of two pairs of points is "good" if the total length of all intersecting subpaths between them is small; the smaller the intersection is, the "better" the pattern is. Determining the "quality" of intersecting patterns of arbitrary pairs of shortest paths is challenging. To this end, define the *distance* between two pairs of points $\{x_1, x_2\}, \{y_1, y_2\}$, where $x_1, y_1 \in N$ and $x_2, y_2 \in S$, to be $\max(|x_1y_1|, |x_2y_2|)$, and denote it by $d(\{x_1, x_2\}, \{y_1, y_2\})$. Let $Q_x$ and $Q_y$ denote fixed shortest paths between the pairs $x_1, x_2$ and $y_1, y_2$ in the Steiner spanner, respectively; the key ingredient in our proof is establishing an inverse-quadratic relationship between $w(Q_x \cap Q_y)$ and $d(\{x_1, x_2\}, \{y_1, y_2\})$: $w(Q_x \cap Q_y) = O(\frac{\epsilon}{d(\{x_1,x_2\},\{y_1,y_2\})^2})$. A charging argument that employs this relationship is then applied to derive the weight lower bound; see Section V.

As mentioned, our lightness analysis of the greedy algorithm builds on several earlier works. In particular, the framework of [15] was refined in the works of Borradaile, Le and Wulff-Nilsen [7], [8]; in what follows, BLW shall be used as a shortcut for the approach of Borradaile, Le and Wulff-Nilsen [8], though we emphasize that some of the credit that we attribute to BLW (for brevity reasons) should be attributed to the aforementioned previous works. In BLW, the first step is to construct a hierarchical clustering $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_L$. Clusters in $\mathcal{C}_i$ have diameter roughly $\Theta(L_i)$ where $L_i = \frac{L_{i-1}}{\epsilon}$ and $L_0 = \frac{w(\text{MST})}{n-1}$. The edge set of the greedy spanner, denoted by $E$, is also partitioned according to the clustering hierarchy, $E = E_0 \cup \ldots \cup E_L$, where the edges in $E_i$ have length $\Theta(L_i)$. Credit is then allocated to the clusters in $\mathcal{C}_0$ for a total amount of $c(\epsilon)w(\text{MST})$ for some constant $c(\epsilon)$ depending on $\epsilon$ and $d$, which will ultimately be the lightness bound. Clusters in $\mathcal{C}_0$ spend their credits in two different ways: (1) they give the clusters in $\mathcal{C}_1$ a (major) part of their credit and (2) use the remaining credit to pay for the spanner edges in $L_1$. Clusters in $\mathcal{C}_1$, after getting the credit from $\mathcal{C}_0$, also spend their credit in the same way: they give clusters in $\mathcal{C}_2$ a part of their credit and use the remaining credit to pay for the spanner edges in $L_2$. Inductively, clusters in $\mathcal{C}_{i-1}$, after being given credit by the clusters in $\mathcal{C}_{i-2}$, give the clusters in $\mathcal{C}_i$ a part of their credit and use the remaining credit to pay for the edges in $L_i$.

[4] The *spread* of a point set $P$, denoted by $\Delta(P)$, is the ratio of the largest to the smallest pairwise distance.

BLW showed roughly that for all $2 \leq i \leq L$:
(a) Each cluster $C \in \mathcal{C}_{i-1}$ would get roughly $\Theta(c(\epsilon)L_{i-1})$ credits from clusters in $\mathcal{C}_{i-2}$.
(b) Each cluster $C \in \mathcal{C}_{i-1}$, after giving their credit to clusters in $\mathcal{C}_i$, has $\Omega(\epsilon^{O(1)}c(\epsilon)L_{i-1})$ leftover credits.

Using a standard packing argument, BLW showed that each cluster in $C \in \mathcal{C}_{i-1}$ is incident to $O(\epsilon^{-O(d)})$ edges in $E_i$ (of length $\Theta(L_i) = \Theta(L_{i-1}/\epsilon)$). Thus, by choosing $c(\epsilon) = \epsilon^{-c_0 d}$ for some constant $c_0$, $C$ can pay for its incident spanner edges in $E_i$. Inductively, every spanner edge will be paid at the end. Since only $c(\epsilon)w(\mathrm{MST})$ credits are allocated at the beginning (to $\mathcal{C}_0$), the total weight of all spanner edges is $O(c(\epsilon)) = O(\epsilon^{-O(d)})$. We first observe that the packing argument in $\mathbb{R}^d$ gives an upper bound $O(\epsilon^{-d})$ in the number of edges in $E_i$ incident to a cluster $C \in \mathcal{C}_{i-1}$. Furthermore, the bound in (b) can be made as good as $\epsilon c(\epsilon)L_{i-1}$. Thus, if we are careful, choosing $c(\epsilon) = \Theta(\epsilon^{-d+2})$ suffices, which as a result, gives us lightness bound $O(c(\epsilon)) = O(\epsilon^{-(d+2)})$. To shave the extra $\epsilon^{-2}$ factor, we introduce two new ideas. Firstly, by carefully constructing the hierarchical clustering and partitioning the edge set $E$, we can reduce the worst case bound on the number of edges in $E_i$ incident to a cluster $C \in \mathcal{C}_{i-1}$ from $O(\epsilon^{-d})$ to $O(\epsilon^{d-1})$. This shaves the first $\epsilon^{-1}$ factor. Secondly, we show that in most cases, each cluster $C \in \mathcal{C}_{i-1}$, after giving its credit to clusters in $\mathcal{C}_i$, has at least $\Omega(c(\epsilon)L_{i-1})$ leftover credits. Note that the leftover credit bound in BLW argument is $\Omega(c(\epsilon)\epsilon L_{i-1})$. Thus, the second idea helps us in shaving another $\epsilon^{-1}$ factor.

The major technical difficulty that we are faced with is in realizing the second idea. Achieving the weaker credit leftover bound $\Omega(c(\epsilon)\epsilon L_{i-1})$ (as done in BLW) is already a challenge and, in fact, sometimes impossible. This is because the credit argument has several subtleties in the way credit is distributed; a more detailed explanation is provided in Section VI. To achieve the stronger $\Omega(c(\epsilon)L_{i-1})$ leftover credit bound, we study the geometric arrangement of clusters in $\mathcal{C}_{i-1}$. We then prove a structural lemma that determines the relationship between the curvature of the curve following the arrangement of clusters in $\mathcal{C}_{i-1}$ and the number of edges in $E_i$ connecting $\mathcal{C}_{i-1}$, which we believe to be of independent interest. If we observe more edges in $E_i$ connecting the clusters in $\mathcal{C}_{i-1}$ on the curve, the curvature of the curve must be higher, which, in turn, implies that the clusters in $\mathcal{C}_{i-1}$ can save more credits as leftover.

The details of this argument are presented in Section VI. We remark that our argument for obtaining the optimal lightness bound is elaborate and intricate, but this should be acceptable, given that the previous lightness bound of $O(\epsilon^{-2d})$ required an intricate proof, spreading over a 60-paged chapter in [42].

## II. Preliminaries

For a pair $x, y$ of points in $\mathbb{R}^d$, we denote by $xy$ the line segment between $x$ and $y$. The distance between $x$ and $y$ will be denoted by $|xy|$. We use $B_d(x, r)$ to denote the ball of radius $r$ centered at $x$ in $\mathbb{R}^d$.

Let $G$ be a weighted graph with vertex set $V$. We shall denote the distance between $x$ and $y$ in $G$ by $d_G(x, y)$. Whenever $G$ is clear from the context, we may omit the subscript $G$ in the distance notation. We use $V(G)$ and $E(G)$ to denote the vertex set and edge set of $G$. Sometimes, the vertex set of $G$ is a set of points in $\mathbb{R}^d$ and the weights of edges are given by the corresponding Euclidean distances. In this case, we use the term *vertex* and *point* interchangeably.

The *spread* (or *aspect ratio*) of a point set $P$, denoted by $\Delta(P)$, is the ratio of the largest pairwise distance to the smallest pairwise distance, i.e.,

$$\Delta(P) \;=\; \frac{\max\{|xy| : x, y \in P\}}{\min\{|xy| : x \neq y \in P\}} \qquad (1)$$

The *distance* between a pair $X, Y$ of point sets, denoted by $d(X, Y)$, is the minimum distance between a point in $X$ and a point in $Y$.

We call a subset $N \subseteq P$ an $\epsilon$-*cover* of $P$ if for any $x \in P$, there is a point $y \in N$ such that $|xy| \leq \epsilon$. We say $N$ is an $\epsilon$-*net* if it is an $\epsilon$-cover and for any two points $x \neq y \in N$, $|xy| \geq \epsilon$.

We use $[n]$ and $[0, n]$ to denote the sets $\{1, 2, \ldots, n\}$ and $\{0, 1, \ldots, n\}$, respectively. We will use the following inequalities:

$$\begin{array}{rclcl} x/2 & \leq & \sin(x) & \leq & x \qquad \text{when } 0 \leq x \leq \pi/2 \\ 1 - x^2 & \leq & \cos(x) & \leq & 1 - x^2/3 \qquad \text{when } 0 \leq x \leq \pi/2 \end{array} \qquad (2)$$

In this work, we are mainly interested in (Steiner) spanners with stretch $(1 + \epsilon)$ for some constant $\epsilon$ sufficiently smaller than 1. This is without loss of generality because a (Steiner) $(1+\epsilon)$-spanner is also a (Steiner) $(1+2\epsilon)$-spanner. We use $\epsilon \ll 1/c$, for some constant $c \geq 1$, to indicate the fact that we are assuming $\epsilon$ is sufficiently smaller than $\frac{1}{c}$.

Given a point set $P$, we use $S_{\mathrm{grd}}(P)$ to denote the greedy $(1+\epsilon)$-spanner of $P$. $S_{\mathrm{grd}}(P)$ is obtained by considering all pairs of points in $P$ in increasing distance order and adding to the spanner edge $xy$ whenever the distance between $x$ and $y$ in the current spanner is at least $(1 + \epsilon)|xy|$. When $P$ is clear from the context, we simply denote the greedy $(1 + \epsilon)$-spanner of $P$ by $S_{\mathrm{grd}}$.

## III. Lower bounds for spanners

In this section we provide our lower bounds for spanners, which are tight for both size and lightness for any $d = O(1)$. We start with the lower bound for $\mathbb{R}^2$, which is our main focus, and then generalize the argument for higher constant dimension $d = O(1)$. For simplicity of presentation, let us consider stretch $1 + c\epsilon$ for some constant $c \leq 1$ independent of $\epsilon$; the same lower bounds for stretch $1 + \epsilon$ follow by scaling.

*Lower bounds for spanners in $\mathbb{R}^2$:* Let $C$ be an unit circle on the plane $\mathbb{R}^2$ and let $P$ be a set of points of size $k = \frac{1}{\epsilon}$ evenly placed on the boundary of $C$. The $\mathrm{MST}$ of $P$ has weight at most the circumference of $C$ which is at most $2\pi$. We shall use the fact that $|pq| \geq 2\pi\epsilon$, for every $p \neq q \in P$, to argue that:

**Claim III.1.** *Let $x, y \in P$ with $|xy| = \Omega(1)$. For any $z \in P$, we have $|xz| + |zy| \geq (1 + \Omega(\epsilon))|xy|$.*

*Proof:* Let $\alpha = \angle yxz$ and $\beta = \angle xyz$. We have $|xy| = |xz|\cos\alpha + |yz|\cos\beta$. Note that for any $0 \leq \alpha, \beta \leq \pi/2$, by Equation 2, $\cos\alpha \leq 1 - \alpha^2/3$ and $\cos\beta \leq 1 - \beta^2/3$. Clearly, $|xz|, |yz| \leq 2$. We have:

$$\begin{aligned}\frac{|xz|+|yz|}{|xy|} &= \frac{|xz|+|yz|}{|xz|\cos\alpha + yz\cos\beta}\\ &\geq \frac{|xz|+|yz|}{|xz|(1-\alpha^2/3)+yz(1-\beta^2/3)}\\ &> 1 + \frac{|xz|\alpha^2/3+|yz|\beta^2/3}{|xz|+|yz|}\\ &\geq 1+|xz|\alpha^2/12+|yz|\beta^2/12\end{aligned}$$

By the triangle inequality, $\max(|xz|, |yz|) \geq \frac{|xy|}{2} = \Omega(1)$. Since $\alpha \geq |yz|/2$ and $\beta \geq |xz|/2$, we have $\max(\alpha^2/12, \beta^2/12) = \Omega(1)$. Thus, we have $\frac{|xz|+|yz|}{|xy|} \geq 1 + \min(|xz|\Omega(1), |yz|\Omega(1)) \geq 1 + \Omega(\epsilon)$. ▮

■

**Corollary III.2.** *Any $(1 + c\epsilon)$-spanner of $P$ must have at least $\Omega(\frac{1}{\epsilon^2})$ edges and weight at least $\Omega(\frac{1}{\epsilon^2})$, for some constant $c < 1$ independent of $\epsilon$.*

*Proof:* Fix an arbitrary point $x \in P$ and let $F(x)$ be the set of $\frac{1}{2\epsilon}$ furthest points from $x$ in $P$. Let $y \in F(x)$ and note that $|xy| \geq \sqrt{2}$. By Claim III.1, $|xz|+|zy| > (1+c\epsilon)|xy|$, for any point $z \in P \setminus x, y$ and some constant $c < 1$ independent of $\epsilon$. Thus, any $(1 + c\epsilon)$-spanner $S$ of $P$ must include all edges $(xy)$, for all $y \in F(x)$. Summing over all $1/\epsilon$ points $x \in P$, there are overall $(\frac{1}{\epsilon} \cdot \frac{1}{2\epsilon})/2 = \frac{1}{4\epsilon^2}$ such edges $(x, y)$ with $x \in P, y \in F(x)$, each with weight $\Omega(1)$, thus the corollary follows. ▮

■

We now prove Theorem I.1 for $d = 2$. In the following we assume that $\epsilon \geq \frac{1}{n}$.

*Lightness bound.* Let $P_n^*$ be any set of $n$ points obtained from the aforementioned set $P$ by adding $n - \frac{1}{\epsilon}$ points at the same locations of points of $P$. The weight of $\mathrm{MST}(P_n^*)$ remains unchanged, i.e., $O(1)$. By Corollary III.2, any $(1 + c\epsilon)$-spanner for $P_n^*$ must have weight $\Omega(\frac{1}{\epsilon^2})$, yielding the lightness bound.

*Sparsity bound.* Let $n' = n\epsilon$, and take $n'$ vertex-disjoint copies of the aforementioned point set $P$, denoted by $P_1, P_2, \ldots, P_{n'}$, where each $P_i$ is defined with respect to a separate unit circle $C_i$ and the $n'$ circles are sufficiently far from each other; it suffices for the circles to be horizontally aligned so that any consecutive circles are at distance 3 from each other. Let $Q_n^* = P_1 \cup P_2 \cup \ldots \cup P_{n'}$.

Let $S := S(Q_n^*)$ be any $(1 + c\epsilon)$-spanner for $Q_n^*$. For each $i \in [n']$, let $S[P_i]$ be the induced subgraph of $S$ on $P_i$. Since the circles are sufficiently far from each other, for each $i \in [n']$, no $(1+\epsilon)$-spanner path between any pair $x, y \in P_i$ in $S$ may contain a point in $P_j$, for any $j \neq i$, hence $S[P_i]$ is an $(1+c\epsilon)$-spanner of $P_i$. By Corollary III.2, we conclude that $|E(S)| \geq \sum_{i\in[n']} |E(S[P_i])| \geq n' \cdot \Omega(\frac{1}{\epsilon^2}) = \Omega(\frac{n}{\epsilon})$.

*Lower bounds for spanners in higher dimensional spaces:* Let $\mathbb{S}_d$ be a $d$-dimensional unit sphere centered at the origin. A *$(d, \theta)$-spherical code $C$* is the set of unit vectors $c_1, c_2, \ldots, c_k \in \mathbb{S}_d$, called *codewords*, such that the angle between any two vectors is at least $\theta$. Let $A(d, \theta)$ be the size of the largest $(d, \theta)$-spherical code. A classic bound on $A(d, \theta)$ (see [47] or [50]) is:

$$A(d, \theta) \;\geq\; (1 + o(1))\sqrt{2\pi d}\frac{\cos\theta}{\sin^{d-1}\theta} \tag{3}$$

Let $P$ be a $(d, 2\pi\epsilon)$-spherical code of maximum size. Observe that $P = \Theta(\epsilon^{-d+1})$; indeed, Equation (3) yields $P = \Omega(\epsilon^{-d+1})$ and $P = O(\epsilon^{-d+1})$ follows from a standard volume argument.

Consider any $x, y \in P$ with $|xy| = \Omega(1)$. Let $z$ be any point in $P \setminus \{x, y\}$ and let $\tilde{C}$ be the circle that goes through $x, y$ and $z$. Since $|xy| = \Omega(1)$, $\tilde{C}$ has radius $\Theta(1)$. For any $z \in P \setminus \{x, y\}$, we have $\min(|xz|, |yz|) \geq 2\pi\epsilon$, hence we can apply Claim III.1 to obtain $|xz|+|zy| \geq (1+\Omega(\epsilon))|xy|$, where the constant hiding in the $\Omega$-notation might be smaller than that in the claim statement, since the claim is stated w.r.t. a unit circle whereas $\tilde{C}$ has radius $\Theta(1)$. It follows that any edge $(x, y)$ with $|xy| = \Omega(1)$ must be included in any $(1+\tilde{c}\epsilon)$-spanner for $P$, for some constant $\tilde{c} < 1$ independent of $d$ and $\epsilon$. Note also that for each point $x \in P$, there are $\Omega(|P|)$ points in $P$ lying on the hemisphere opposite to $x$.

This enables us to generalize Corollary III.2: Any $(1+\tilde{c}\epsilon)$-spanner $S$ of $P$ must have $\Omega(|P|^2) = \Omega(\epsilon^{-2d+2})$ edges and weight $w(S) = \Omega(|P|^2) = \Omega(\epsilon^{-2d+2})$, for some constant $\tilde{c} < 1$ independent of $d$ and $\epsilon$. Since the distance between any two nearby points in $P$ is $O(\epsilon)$, $\mathrm{MST}(P) = O(\epsilon|P|)$, and the lightness of $S$ is $\Omega\left(\frac{\epsilon^{-2d+2}}{\epsilon|P|}\right) = \Omega\left(\epsilon^{-d}\right)$. For the size bound, we again use the trick of taking $n/|P|$ copies of the same point set $P$ that are sufficiently far from each other, and get that the spanner size is $\Omega(\frac{n}{|P|} \cdot |P|^2) = \Omega(n \cdot \epsilon^{-d+1})$. For the size and lightness bounds to apply to $n$-point sets, we assume that $n \geq |P|$, i.e., $\epsilon = \Omega(n^{-\frac{1}{d-1}})$.

## IV. Sparse Steiner spanners

In this section, we prove Theorem I.3. Our proof strategy consists of two steps. In the first step we prove a relaxed version of Theorem I.3, where the size of the spanner depends on the spread $\Delta$ of the point set, $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log\Delta})}$.

In the second step, we reduce the general problem to the relaxed one proved in the first step. For the reduction, we solve $\log\Delta$ spanner construction problems, for point sets of spread $O(\frac{1}{\epsilon})$ each, and demonstrate that no dependency on $\Delta$, even a logarithmic one, is incurred. This reduction employs the standard net-tree spanner, based on a hierarchical net structure, which consists of $\log\Delta$ edge sets $E_0, E_1, \ldots, E_{\log\Delta-1}$ that we refer to as *ring spanners*. Each ring spanner $E_i$ connects pairs of points at distance in the range $(r_i, O(r_i/\epsilon))$, where $r_i$ grows geometrically with $i$. A central ingredient of the reduction is a careful replacement of the ring spanners $R_i$ by much sparser *Steiner* ring spanners.

### *A. Steiner spanners for point sets of bounded spread*

In this section we handle point sets of bounded spread, which constitutes a central ingredient in the proof of Theorem I.3. Specifically, we prove the following statement.

**Proposition IV.1.** *For any set $P$ of $n$ points in $\mathbb{R}^2$ with spread $\le \Delta$ and any $\epsilon = \tilde{\Omega}(\frac{1}{n^2})$, there is a Steiner spanner of size $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log\Delta})}$.*

*1) An auxiliary lemma:* We shall assume that $\epsilon$ is sufficiently smaller than 1. In what follows $P$ is an arbitrary (fixed) set of $n$ points in $\mathbb{R}^2$. Let $X$ be a point set in $\mathrm{R}^2$; abusing notation, when $X$ is of infinite size (such as a rectangle or any other polygonal shape), we may use $X$ as a shortcut for $X \cap P$, i.e., to denote the set of points in $X$ that belong to $P$; we may henceforth use $|X|$ as a shortcut for $|X \cap P|$. Let $R_1$ and $R_2$ be two rectangles of the same length and width whose sides are parallel to the $x$ and $y$ axis. We say that $R_1, R_2$ are *horizontally* (respectively, *vertically*) *parallel* if there is a vertical (respectively, horizontal) line going through the left (respectively, top) sides of both rectangles The following lemma is crucial in our proof.

**Lemma IV.2.** *Let $R_1, R_2$ be two horizontally (respectively, vertically) parallel rectangles of width (resp., length) $W$ and the same length (resp., width). Let $d(R_1, R_2) = \frac{W}{\ell}$ where $\ell > 1$. There is a Steiner spanner $S$ with $O(\frac{\ell}{\sqrt{\epsilon}}(|R_1|+|R_2|))$ edges such that for any point $p \in R_1, q \in R_2$, $d_S(p,q) \le (1+\epsilon)|pq|$, assuming $\epsilon$ is sufficiently smaller than* 1.

*Proof:* By symmetry, it suffices to prove the lemma for two horizontally parallel rectangles. By scaling, we may assume that $W = 1$. Let $\mathtt{L_{left}}, \mathtt{L_{right}}$ be two vertical lines that contain the left sides and right sides of $R_1$ and $R_2$, respectively. Let $L$ be the horizontal line segment of length 1 with endpoints touching $\mathtt{L_{left}}$ and $\mathtt{L_{right}}$ and that is within distance $1/2\ell$ from both $R_1$ and $R_2$ (see Figure 1). We then place a set $X$ of $\frac{\ell}{\sqrt{\epsilon}}$ evenly spaced Steiner points along $L$, and take to the Steiner spanner $S$ all edges that connect each of the Steiner points with all the points in $R_1 \cup R_2$. The vertex set of $S$ is $X \cup R_1 \cup R_2$ and its edge set is of size $|X|(|R_1| + |R_2|) = \frac{\ell}{\sqrt{\epsilon}}(|R_1| + |R_2|)$.

We next prove the stretch bound for an arbitrary pair $p, q$ of points with $p \in R_1, q \in R_2$, assuming $\epsilon$ is sufficiently smaller than 1. Let $y$ be the intersection of the line segments $pq$ and $L$ and let $x$ be the closest point of $X$ to $y$. Since the distance between consecutive points of $X$ along $L$ is $\frac{\sqrt{\epsilon}}{\ell}$, we have $|xy| \le \frac{\sqrt{\epsilon}}{\ell}$. Note also that $|py| \ge \frac{d(R_1,R_2)}{2} = \frac{1}{2\ell}$. Defining $\alpha = \angle xpy$, we conclude that

$$\sin(\alpha) \;=\; \frac{|xy|\sin\angle pxy}{|py|} \;\le\; \frac{|xy|}{|py|} \le 2\ell|xy| \;\le\; 2\sqrt{\epsilon},$$

hence $\alpha \le 4\sqrt{\epsilon}$ by Equation 2. Let $x'$ be the projection of $x$ onto $pq$. We have:

$$|px| \;=\; \frac{|px'|}{\cos(\alpha)} \;\le\; \frac{|px'|}{1-\alpha^2} \le \frac{|px'|}{1-16\epsilon} \;\le\; (1+O(\epsilon))|px'|.$$

By symmetry, we have $|xq| \le (1+O(\epsilon))|x'q|$. Since $S$ contains both edges $(p,x)$ and $(x,q)$, it follows that

$$\begin{aligned} d_S(p,q) \;&\le\; |px| + |xq| \;\le\; (1+O(\epsilon))(|px'| + |x'q|) \\ &=\; (1+O(\epsilon))|pq| \;\le\; (1+\epsilon')|pq|, \end{aligned}$$

where $c\epsilon = \epsilon'$ and $c$ is the constant hiding in the $O$-notation above. Thus we obtain a spanner with stretch $1+\epsilon'$ and $O(\frac{\ell}{\sqrt{\epsilon'}}(|R_1| + |R_2|))$ edges, and the required result now follows by scaling. ■

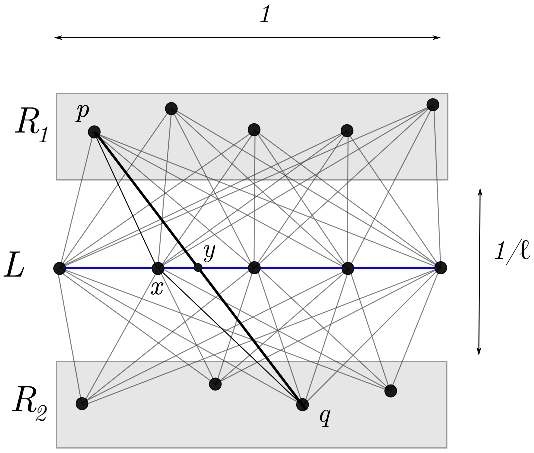


Figure 1. Illustration for the proof of Lemma IV.2. The solid blue line is $L$.

*2) Proof of Proposition IV.1:* This section is devoted to the proof of Proposition IV.1.

Since the spread is $\le \Delta$, we may assume without loss of generality that the point set $P$ is contained in a unit axis-parallel square, $Q_0$, and the minimum pairwise distance is $\ge \frac{1}{c\Delta}$ for some constant $c$.

We employ recursion to construct the Steiner spanner that proves the proposition. The high-level idea is to divide the current square into *overlapping* subsquares, where the overlap between adjacent (i.e., neighboring) squares is a small constant fraction of their side length. This overlap is useful since any pair of points that is not contained in any subsquare must be far apart, and specifically, at distance longer than the overlap between adjacent squares. We can thus apply Lemma IV.2 to get a sparse Steiner spanner for all the long pairwise distances, and then recurse on each subsquare. Note that each short distance is "internal" to at least one subsquare, so it is handled in deeper levels of the recursion. The top level of the recursion is applied to (the point set in) the original square, denoted by $Q_0$, with side length 1.

We now describe the recursive construction in detail.

Let $\eta$ be a parameter to be determined later and consider an arbitrary level $d$ of the recursion. All level $d$ squares will be handled in exactly the same way, hence it suffices to consider an arbitrary such square, denoted by $Q_d$. Let $X_d = \frac{\beta_d}{2^{\alpha_d \cdot \eta}}$ be the side length of $Q_d$. Initially, $d = 0$, $\beta_0 = 1$ and $\alpha_0 = 0$, which gives $X_0 = 1$ as required. We partition $Q_d$ into $2^{2\eta}$ disjoint subsquares $Y_{i,j}, i, j \in [2^\eta]$, each of side length $\frac{X_d}{2^\eta}$. Then, we extend each of the subsquares $Y_{i,j}$ equally in four directions by a $(1 + 2\rho)$ factor, for some constant $\rho < \frac{1}{2}$ to be determined later, so that each subsquare has side length $X_{d+1} = \frac{X_d(1+2\rho)}{2^\eta}$ (see Figures (2a) and (2b)). Let $\mathcal{Q}_{d+1}$ be the set of extended (overlapping) subsquares. We recurse on each extended subsquare $Q_{d+1} \in \mathcal{Q}_{d+1}$ to obtain a Steiner spanner, denoted by $S_{d+1}(Q_{d+1})$, for the point set in $Q_{d+1}$. By construction, it holds that $\beta_{d+1} = (1 + 2\rho)\beta_d$ and $\alpha_{d+1} = \alpha_d + 1$, which inductively resolves to $\beta_d = (1 + 2\rho)^d$ and $\alpha_d = d$. Observe that the overlapping region between any two adjacent subsquares is a rectangle of side lengths $\frac{2\rho X_d}{2^\eta}$ and $\frac{(1+2\rho)X_d}{2^\eta}$.

Next we handle the pairwise distances that are not taken care of by the recursion, i.e., those corresponding to pairs of points $p, q$ such that there is no subsquare of $\mathcal{Q}_{d+1}$ to which both $p$ and $q$ belong. Without loss of generality we assume that the subsquares $Y_{i,j}, i, j \in [2^\eta]$ are indexed from left to right and from top to bottom. For each $i \in [2^\eta]$, we call $H_i := \cup_{j \in [2^\eta]} Y_{i,j}$ a *horizontal band* of $Q_d$. Similarly, for each $j \in [2^\eta]$, we call $V_j = \cup_{i \in [2^\eta]} Y_{i,j}$ a *vertical band* of $Q_d$. To handle pairwise distances corresponding to pairs of points that belong to different horizontal bands, we use *horizontal Steiner spanners*. Consider a pair of horizontal bands $H_i, H_j$, where $i < j$. The horizontal Steiner spanner corresponding to the pair $H_i, H_j$, denoted by $HS_{i,j}$, is constructed as follows. We distinguish between two cases:

*Case 1:* $j \geq i + 2$. Note that $H_i$ and $H_j$ have width $X_d$ and $d(H_i, H_j) \geq (j - i - 1)\frac{X_d}{2^\eta}$. Thus we can apply Lemma IV.2 with $R_1 = H_i, R_2 = H_j, W = X_d$ to obtain a Steiner spanner $HS_{i,j}$ satisfying:

$$|E(HS_{i,j})| = O\left(\frac{2^\eta(|H_i| + |H_j|)}{(j - i - 1)\sqrt{\epsilon}}\right). \tag{4}$$

*Case 2:* $j = i + 1$. Let $H'_i$ and $H'_{i+1}$ be the bands obtained from $H_i$ and $H_{i+1}$, respectively, by removing the overlapping region of all the subsquares in $\mathcal{Q}_{d+1}$ extended from subsquares in $H_i$ and $H_{i+1}$ (see Figure 2(c)); we will refer to any such band $H'_i$ as a *truncated band*. Note that $H'_i$ and $H'_{i+1}$ have width $X_d$ and $d(H'_i, H'_{i+1}) = \frac{2\rho X_d}{2^\eta}$. Thus we can apply Lemma IV.2 with $R_1 = H'_i, R_2 = H'_{i+1}, W = X_d$ to obtain a Steiner spanner $HS_{i,i+1}$ satisfying:

$$\begin{aligned}|E(HS_{i,i+1})| &= O\left(\frac{2^\eta(|H'_i| + |H'_{i+1}|)}{\sqrt{\epsilon}2\rho}\right)\\ &= O\left(\frac{2^\eta(|H_i| + |H_{i+1}|)}{\sqrt{\epsilon}2\rho}\right).\end{aligned} \tag{5}$$

To handle pairwise distances corresponding to pairs of points that belong to different vertical bands, we use *vertical Steiner spanners* $VS_{i,j}$, which are constructed (using Lemma IV.2) in the same way.

Overall, the recursive spanner construction for $Q_d$, denoted by $S_d(Q_d)$, is given by:

$$\begin{aligned}S_d(Q_d) = (&\bigcup_{Q_{d+1} \in \mathcal{Q}_{d+1}} S_{d+1}(Q_{d+1}))\\ &\bigcup(\bigcup_{1 \leq i < j \leq 2^\eta} HS_{i,j})\bigcup(\bigcup_{1 \leq i < j \leq 2^\eta} VS_{i,j}),\end{aligned} \tag{6}$$

and the recursion bottoms once a square contains at most one point. In particular, the recursion may proceed to deeper recursion levels only as long as the side length of a square is $> \frac{1}{2c\Delta}$ (where $\Delta$ is the spread of the original point set $P$), as no two points of $P$ may belong to a square of side length $\frac{1}{2c\Delta}$. Denoting by $D$ the recursion depth, we thus have

$$X_D = \frac{(1 + 2\rho)^D}{2^{D \cdot \eta}} \leq \frac{1}{2c\Delta}. \tag{7}$$

Since $\rho < \frac{1}{2}$, choosing $D = \sqrt{\log \Delta}$ and $\eta = \Theta(\sqrt{\log \Delta})$ satisfies Equation 7. (When there is just one point in $P$, the aspect ratio $\Delta$ is defined as 1, and then $D = \sqrt{\log \Delta} = 0$.)

The ultimate Steiner spanner for the entire point set $P$, denoted by $SS(P)$, is given by $S_0(Q_0)$. Unwinding this recursive construction, $SS(P)$ is obtained as the union of all horizontal and vertical Steiner spanners constructed for all level $d$ extended subsquares, taken over all levels $d = 0, 1, \ldots, D = \sqrt{\log \Delta}$.

*Stretch analysis:* The proof is by induction on the recursion depth $D$, where $D = \sqrt{\log \Delta}$, with a trivial basis $D = 0$. For the induction step, let $p, q$ be two points of $P$ in some level $d$ square $Q_d$. Assume first that $p, q$ are in the same subsquare of $\mathcal{Q}_{d+1}$; in this case, the distance between $p$ and $q$ is preserved by the recursive construction for that subsquare. We henceforth assume that there is no subsquare of $\mathcal{Q}_{d+1}$ that contains both $p$ and $q$. If $p$ and $q$ belong to two non-adjacent horizontal or vertical bands, namely, $H_i$ and $H_j$ or $V_i$ and $V_j$ with $j \geq i+2$, respectively, then, by Lemma IV.2, the distance between $p$ and $q$ in the spanner is in check due to the horizontal or vertical Steiner spanners $HS_{i,j}$ or $VS_{i,j}$ constructed in level $d$ of

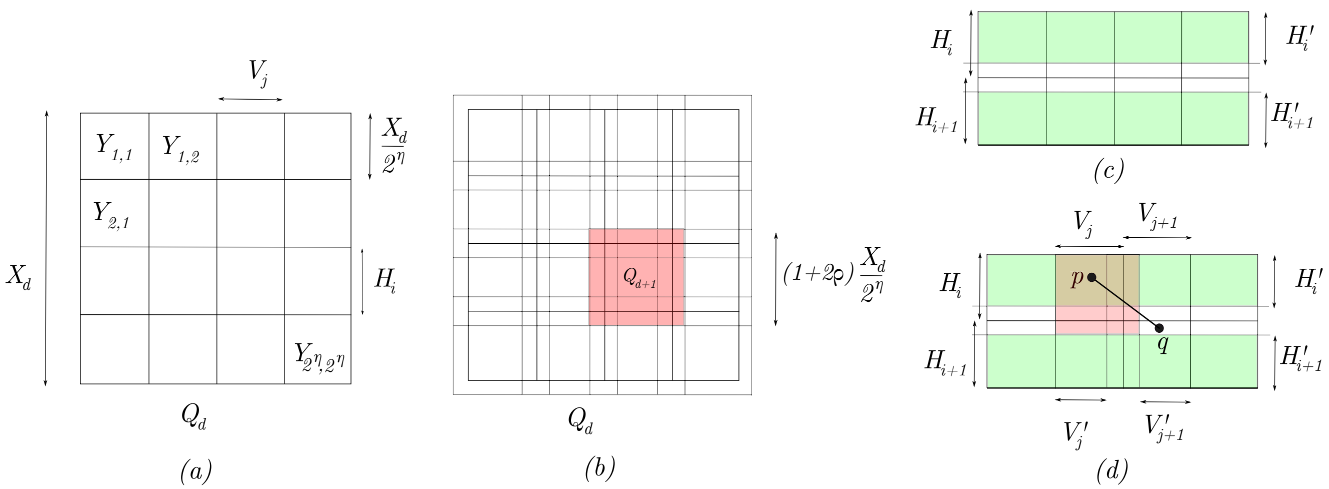


Figure 2. (a) A partition of the square $Q_d$ into $2^{2\eta}$ subsquares $Y_{i,j}$ where $i, j \in [2^\eta]$, each of side length $\frac{X_d}{2^\eta}$. (b) Each subsquare $Y_{ij}$ is extended to a larger square of side length $(1 + 2\rho)\frac{X_d}{2^\eta}$ (one of which is red shaded in the figure). To avoid boundary cases in our analysis, we extend the subsquares on the boundary of $Q_d$ to the region outside $Q_d$. (c) Two adjacent horizontal bands $H_i$ and $H_{i+1}$ and the corresponding truncated bands $H_i'$ and $H_{i+1}'$ (green shaded), obtained by removing the overlapping region of their extended subsquares. (d) A case in the stretch analysis of two points $p, q$. Pairs of points in the red shaded square will be handled recursively.

the recursion. Otherwise $p$ and $q$ belong to two adjacent horizontal and vertical bands, and let $i, j, i', j'$ be the indices for which $p \in H_i \cap V_j$ and $q \in H_{i'} \cap V_{j'}$, such that $|i'-i| \le 1, |j'-j| \le 1$. In this case either $p$ and $q$ belong to the truncated horizontal bands $H_i'$ and $H_{i'}'$ with $|i'-i| = 1$ or they belong to the truncated vertical bands $V_j'$ and $V_{j'}'$ with $|j' - j| = 1$; indeed, otherwise there must be an extended subsquare in $\mathcal{Q}_{d+1}$ to which both $p$ and $q$ belong. In the former (respectively, latter) case the distance between $p$ and $q$ in the spanner is in check, again by Lemma IV.2, due to the horizontal (resp., vertical) Steiner spanner $HS_{i,i+1}$ or $HS_{i-1,i}$ (resp., $VS_{j,j+1}$ or $VS_{j-1,j}$) constructed in level $d$ of the recursion, depending on whether $i' = i+1$ or $i' = i-1$ (resp., $j' = j + 1$ or $j' = j - 1$). (See Figure 2(d) for an illustration of the latter case, where $p$ belongs to $V_j'$ and $q$ belongs to $V_{j+1}'$, with $i' = i + 1, j' = j + 1$.)

*Size analysis:* We next analyze the size of $SS(P) = S_0(Q_0)$. As mentioned, unwinding this recursive construction, $SS(P)$ is obtained as the union of all horizontal and vertical Steiner spanners constructed for all level $d$ extended (sub)squares, taken over all levels $d = 0, 1, \ldots, D = \sqrt{\log \Delta}$. Fix an arbitrary such level $d$ and recall that $\mathcal{Q}_d$ denotes the set of all such squares at level $d$ of the recursion. Since the squares are overlapping, any point in $P$ may belong to many squares in $\mathcal{Q}_d$. To bound the size of the spanner, we will first bound the total number of points (with multiplicities) over all squares in $\mathcal{Q}_d$: $\sum_{Q_d \in \mathcal{Q}_d} |Q_d|$, and then bound the number of edges in all the horizontal and vertical spanners of each $Q_d \in \mathcal{Q}_d$.

First, we bound the total number of points over all squares in $\mathcal{Q}_d$.

**Claim IV.3.** *$\sum_{Q_d \in \mathcal{Q}_d} |Q_d| \le n2^{2d}$, assuming $\rho \le \frac{1}{4}$.*

*Proof:* The *multiplicity* of a point $p$ at level $d$ is the number of squares in $\mathcal{Q}_d$ that contain $p$. It suffices to show that the multiplicity of Each point at level $d$ is $\le 2^{2d}$, which we prove by induction on $d$, where $d = 0, 1, \ldots, D = \sqrt{\log \Delta}$. The basis $d = 0$ is immediate, since $\mathcal{Q}_0 = \{Q_0\}$, hence the multiplicity of each point at level 0 is 1. For the induction step, assume that the multiplicity of each point at level $d$ is $\le 2^{2d}$. Assuming $\rho \le \frac{1}{4}$, any point may belong to at most four overlapping subsquares of $Q_d$ by construction (see Figure 2(d)). Thus the multiplicity of each point at level $d + 1$ is $\le 4 \cdot 2^{2d} = 2^{2(d+1)}$. ■

We next bound the number of edges in all the horizontal Steiner spanners constructed for an arbitrary level $d$ square $Q_d \in \mathcal{Q}_d$. By symmetry, the same bounds apply to the vertical Steiner spanners.

**Claim IV.4.** *$\sum_{i=1}^{2^\eta-2} \sum_{j=i+2}^{2^\eta} |E(HS_{i,j})| = O\left(\frac{2^\eta \eta |Q_d|}{\sqrt{\epsilon}}\right)$.*

*Proof:* Let $m = 2^\eta$. By Equation 4, we have:

$$\sum_{i=1}^{m-2} \sum_{j=i+2}^{m} |E(HS_{i,j})| = O\left(\frac{m}{\sqrt{\epsilon}}\right) \sum_{i=1}^{m-2} \sum_{j=i+2}^{m} \frac{|H_i| + |H_j|}{j - i - 1} \tag{8}$$

Since each point of $Q_d$ may belong to at most two horizontal bands (more specifically, to bands $H_i$ and $H_{i+1}$, for some $i = [m - 1]$), it follows that $\sum_{i=1}^{m} |H_i| \le 2|Q_d|$.

We thus have

$$\begin{aligned}\sum_{i=1}^{m-2}\sum_{j=i+2}^{m}\frac{|H_i|}{j-i-1} &= \sum_{i=1}^{m-2}|H_i|\sum_{j=i+2}^{m}\frac{1}{j-i-1}\\ &\le \sum_{i=1}^{m}|H_i|\ln m \;\le\; 2\ln m|Q_d|,\end{aligned} \tag{9}$$

where the penultimate inequality holds since $\sum_{j=i+2}^{m}\frac{1}{j-i-1} \le \sum_{j=1}^{m}\frac{1}{j} \le \ln m$. We also have

$$\begin{aligned}\sum_{i=1}^{m-2}\sum_{j=i+2}^{m}\frac{|H_j|}{j-i-1} &= \sum_{j=3}^{m}\sum_{i=1}^{j-2}\frac{|H_j|}{j-i-1}\\ &= \sum_{j=3}^{m}|H_j|\sum_{i=1}^{j-2}\frac{1}{j-i-1}\\ &\le 2\ln m|Q_d|.\end{aligned} \tag{10}$$

Plugging Equations (9) and (10) in Equation (8) completes the proof of the claim. ■

**Claim IV.5.** $\sum_{i=1}^{2^\eta-1}|E(HS_{i,i+1})| = O\left(\frac{2^\eta|Q_d|}{\rho\sqrt{\epsilon}}\right)$.

*Proof:* Again let $m = 2^\eta$ and recall that $\sum_{i=1}^{m}|H_i| \le 2|Q_d|$. By Equation 5, we have:

$$\begin{aligned}\sum_{i=1}^{m-1}|E(HS_{i,i+1})| &= \sum_{i=1}^{m-1}O\left(\frac{2^\eta(|H_i|+|H_{i+1}|)}{\sqrt{\epsilon}2\rho}\right)\\ &= O\left(\frac{2\cdot 2^\eta}{\sqrt{\epsilon}2\rho}\right)\sum_{i=1}^{m}|H_i|\\ &= O\left(\frac{2^\eta|Q_d|}{\rho\sqrt{\epsilon}}\right).\end{aligned} \tag{11}$$

■

*Concluding the proof of Proposition IV.1:* Set $\rho = \frac{1}{4}$. By Claims IV.4 and IV.5, $\sum_{i=1}^{2^\eta-1}\sum_{j=i+1}^{2^\eta}|E(HS_{i,j})| = O(\frac{2^\eta\eta|Q_d|}{\sqrt{\epsilon}})$; by symmetry, $\sum_{i=1}^{2^\eta-1}\sum_{j=i+1}^{2^\eta}|E(VS_{i,j})| = O(\frac{2^\eta\eta|Q_d|}{\sqrt{\epsilon}})$. Claim IV.3 yields $\sum_{Q_d\in\mathcal{Q}_d}|Q_d| \le n2^{2d}$, thus the total number of edges added to the spanner (due to all horizontal and vertical Steiner spanners) at level $d$ of the recursion is $\sum_{Q_d\in\mathcal{Q}_d}O(\frac{2^\eta\eta|Q_d|}{\sqrt{\epsilon}}) = O(\frac{2^\eta\eta 2^{2d}n}{\sqrt{\epsilon}})$. Recalling that $\eta = \Theta(\sqrt{\log\Delta})$ and $d = 0,1,\ldots,D = \sqrt{\log\Delta}$, we conclude that the total number of edges of the Steiner spanner $S_0(Q_0) = SS(P)$ over all recursion levels is at most $\sum_{d=0}^{D}O(\frac{2^\eta\eta 2^{2d}n}{\sqrt{\epsilon}}) = \frac{n2^{O(\sqrt{\log\Delta})}}{\sqrt{\epsilon}}$.

### *B. From bounded spread to general point sets*

In this section we provide the reduction from the general case to point sets of bounded spread. We start (Section IV-B1) with an overview of the net-tree spanner construction, on which our reduction builds.

*1) The net-tree spanner: A short overview:* In this section we describe the *net-tree spanner* construction, which has several variants (see, e.g., [26], [13], [29]). For concreteness we consider the constructions of [26], [13], which were discovered independently but are similar to each other, and apply to the wider family of doubling metrics. (The construction of [26] was presented for Euclidean spaces.) We will later (Section IV-B2) demonstrate that, in Euclidean spaces, the constructions of [26], [13] can be strengthened via the usage of Steiner points, to obtain a quadratic improvement to the spanner size. As will be shown, this improvement requires several new insights.

Let $(X,\delta)$ be a metric of doubling dimension $d$. By scaling, we assume that the minimum pairwise distance in $X$ is 1, thus the spread $\Delta = \Delta(X)$ coincides with the *diameter* of $X$, i.e., $\Delta = \max_{u,v\in X}\delta(u,v)$. Fix any $r > 0$ and any set $Y \subseteq X$; $Y$ is called an *r-net* for $X$ if (1) $\delta(y,y') \ge r$, for any $y \ne y' \in Y$, and (2) for each $x \in X$, there is $y \in Y$ with $\delta(x,y) \le r$; such a net can be constructed by a greedy algorithm.

*Hierarchical Nets:* Write $\ell = \lceil\log_2\Delta\rceil + 1$, and let $\{N_i\}_{i\ge 0}^{\ell}$ be a sequence of *hierarchical nets*, where $N_0 = X$, and for each $i \in [\ell]$, $N_i$ is an $2^i$-net for $N_{i-1}$. For each $i \in [0,\ell]$, $N_i$ is called the *i-level net*. Note that $N_0 = X \supseteq N_1 \supseteq \ldots \supseteq N_\ell$, and $N_\ell$ contains exactly one point. The same point of $X$ may have instances in many nets (any point of $N_i$ is necessarily also a point of $N_j$, for each $j \in [0,i]$).

The hierarchical nets induce a hierarchical tree $T = T(X)$, called *net-tree*; this tree is not required for the construction itself, but is rather used in the stretch analysis; we refer to [26], [13] for the details.

*Spanner via Cross Edges:* The spanner $H = H(X)$ of [26], [13] is obtained by adding, for each $i \in [0,\ell-1]$, a set $E_i$ of edges between all points of $N_i$ that are within distance $(4+\frac{32}{\epsilon})2^i$ from each other, called *cross edges*. That is, $E_i := \{(p,q) \mid p,q \in N_i, \delta(p,q) \le \left(4+\frac{32}{\epsilon}\right)2^i\}$ and $H = \bigcup_{i=0}^{\ell-1}E_i$.

It was shown in [26], [13] that $H$ is a $(1+\epsilon)$-spanner for $X$, having $n\cdot\epsilon^{-O(d)}$ edges.

A stronger bound of $O(n\cdot\epsilon^{-d})$ on the size was established in [26] for Euclidean spaces, and this bound carries over to arbitrary doubling metrics. It is weaker, however, than the state-of-the-art in Euclidean spaces by a factor of $1/\epsilon$. Our goal is to obtain a quadratic improvement over the state-of-the-art size bound in Euclidean spaces, namely $O(n\cdot\epsilon^{-d+1})$, using Steiner points.

*2) The reduction:* To improve the size of the net-tree spanner $H$, we will improve the size bound of each edge set $E_i$. We shall focus on 2-dimensional point sets, but our construction naturally generalizes for higher dimensions, as discussed at the end of this section. Packing arguments yield $|E_i| \le |N_i|\cdot\epsilon^{-2}$, and our goal is to replace $E_i$ by an edge set $E_i'$ of size $\le |N_i|\cdot\frac{1}{\sqrt{\epsilon}}\cdot\epsilon^{-o(1)}$ without increasing the pairwise distances by much.

*Ring Steiner spanners:* Fix $c_1, c_2$ such that $0 < c_1 < c_2$ and let $t \geq 1$ be a stretch parameter. We say that a spanner (i.e., an edge set) $R$ is a *$(c_1, c_2)$-ring $t$-spanner* for a point set $P$ if $d_R(p,q) \leq t|pq|$, for any $p, q \in P$ with $c_1 \leq |pq| \leq c_2$. That is, for every $p \in P$, the ring spanner $R$ preserves distances to within a factor of $t$ between $p$ and every point $q$ in the annulus (or ring) $B_2(p, c_2) \setminus B_2(p, c_1)$.

Note that the edge set $E_i$, which handles pairwise distances in the range $[2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i]$, provides a $(2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i)$-ring 1-spanner for $N_i$. Recall that $|E_i| \leq |N_i| \cdot \epsilon^{-2}$. We next show that $E_i$ can be replaced by a significantly sparser set $E'_i$ that uses Steiner points, by building on the result for bounded spread.

**Lemma IV.6.** *For each $i$, there exist subsets $N_i^1, N_i^2, \ldots, N_i^k$ of $N_i$ that form a covering, i.e., $\bigcup_{j=1}^k N_i^j = N_i$, such that (1) for each $j \in [k]$, $N_i^j$ has spread $O(1/\epsilon)$, (2) each point of $N_i$ belongs to at most four subsets from $N_i^1, N_i^2, \ldots, N_i^k$, and (3) for each edge $(p,q) \in E_i$, there exists an index $j$ such that $p, q \in N_i^j$.*

*Proof:* Let $B$ be the bounding box of $N_i$, define $\tau_i = \left(4 + \frac{32}{\epsilon}\right) 2^i$, and assume without loss of generality that the side lengths of $B$ are divisible by $2\tau_i$. We partition $B$ into squares of side length $2\tau_i$ each. We extend each square equally in four directions to obtain a square of side length $3\tau_i$. Let $N_i^1, N_i^2, \ldots, N_i^k$ be the nonempty point sets lying in the extended squares. Clearly, $N_i^1, N_i^2, \ldots, N_i^k$ form a covering of $N_i$. Since $N_i$ is a $2^i$-net for $N_{i-1}$, every two points in $N_i$ are at distance at least $2^i$ from each other; thus for each $j \in [k]$, every two points in $N_i^j$ are at distance at least $2^i$ and at most $\sqrt{2} \cdot 3\tau_i$ from each other, and item (1) holds. Note that the overlapping region of any pair of neighboring extended squares is a rectangle of side lengths $\tau_i$ and $3\tau_i$, which implies not only item (2), but also the fact that for any pair of points within distance $\tau_i$ from each other, there is at least one extended square to which they both belong; since $|pq| \leq \tau_i$ for each $(p,q) \in E_i$, item (3) holds as well. ■

Fix any $i \in [0, \ell - 1]$, and consider the subsets $N_i^1, N_i^2, \ldots, N_i^k$ of $N_i$ guaranteed by Lemma IV.6. For each $j \in [k]$, we construct a spanner $S_i^j$ for $N_i^j$ as follows. Let $c$ be the constant hiding in the $O$-notation of the upper bound provided by Proposition IV.1. If $|N_i^j| \leq \frac{2}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$, we take $S_i^j$ to be the complete graph over $N_i^j$, and get a 1-spanner for $N_i^j$ (without Steiner points) with $\binom{|N_i^j|}{2} \leq \frac{|N_i^j|}{\sqrt{\epsilon}} 2^{c\sqrt{\log \frac{1}{\epsilon}}}$ edges. Otherwise $|N_i^j| > \frac{2}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$, and we take $S_i^j$ to be the Steiner $(1+\epsilon)$-spanner for $N_i^j$ provided by Proposition IV.1. Item (1) of Lemma IV.6 implies that the spread of each $N_i^j$ is $O(1/\epsilon)$, thus by Proposition IV.1 the number of edges in $S_i^j$ in this case is also bounded by $\frac{|N_i^j|}{\sqrt{\epsilon}} 2^{c\sqrt{\log \frac{1}{\epsilon}}}$. Define $E'(S_i^j)$ to be the edge set of $S_i^j$, and let $E'_i$ be the set of edges in the union of all the spanners $S_i^j$, i.e., $E'_i = \bigcup_{j=1}^k E'(S_i^j)$.

**Corollary IV.7.** *The edge set $E'_i$ is defined over a superset $N'_i = N_i \cup S_i$ of $N_i$, where $S_i$ is a set of Steiner points, such that $|E'_i| \leq \frac{4|N_i|}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$ and $E'_i$ is a $(2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i)$-ring $(1+\epsilon)$-spanner for $N_i$.*

*Proof:* By item (2) of Lemma IV.6, we have $\sum_{j=1}^k |N_i^j| \leq 4|N_i|$. It follows that

$$\begin{aligned} |E'_i| &= \sum_{j=1}^k |E'(S_i^j)| \leq \sum_{j=1}^k \frac{|N_i^j|}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}} \\ &\leq \frac{4|N_i|}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}. \end{aligned} \tag{12}$$

To show that $E'_i$ is a $(2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i)$-ring $(1+\epsilon)$-spanner for $N_i$, consider any pair $p, q \in N_i$ such that $2^i \leq |pq| \leq \left(4 + \frac{32}{\epsilon}\right) 2^i$. We have $(p,q) \in E_i$ by construction, thus item (3) of Lemma IV.6 implies that there exists an index $j$ such that $p, q \in N_i^j$. Hence there is a $(1+\epsilon)$-spanner path between $p$ and $q$ in the spanner $S_i^j$ for $N_i^j$, and thus also in the superset $E'_i$ of $E(S_i^j)$, i.e., $d_{E'_i}(p,q) \leq (1+\epsilon)|pq|$. ■

*A sparser Steiner spanner via ring Steiner spanners:* Denote by $H'$ the spanner obtained as the union of all the Steiner ring spanners $E'_i$, i.e., $H' = \bigcup_{i=0}^{\ell-1} E'_i$. To complete the reduction from the general case to the case of bounded spread, thus finishing the proof of Theorem I.3, we argue that $H'$ is a Steiner $(1+O(\epsilon))$-spanner for $P$ with $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log \frac{1}{\epsilon}})}$ edges. (One can reduce the stretch down to $1+\epsilon$ by scaling.)

*Stretch analysis.* For each $0 \leq i \leq \ell - 1$, $E_i$ is a $(2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i)$-ring 1-spanner for $N_i$, i.e., all distances in $[2^i, \left(4 + \frac{32}{\epsilon}\right) 2^i]$ are preserved precisely by $E_i$; on the other hand, the stretch bound of $E'_i$ is $(1+\epsilon)$. Since $H$, obtained as the union of all ring 1-spanners $E_i$, is a $(1+\epsilon)$-spanner for $P$, the stretch of $H'$, the spanner obtained as the union of all ring $(1+\epsilon)$-spanners $E'_i$, will be bounded by $(1+\epsilon)^2 = 1 + O(\epsilon)$.

*Size analysis.* We next prove that the size bound of $H'$ is in check.

**Lemma IV.8.** *$H'$ consists of at most $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log \frac{1}{\epsilon}})}$ edges.*

*Proof:* Denote the edge set of $H'$ by $E'$. We apply a charging argument; it will be instructive to consider another edge set $\tilde{E}$ with $|\tilde{E}| \geq |E'|$, which has two properties that are useful for analysis purposes.

Consider the edge set $E'_i = \bigcup_{j=1}^k E'(S_i^j)$, for any level $0 \leq i \leq \ell - 1$. We focus on an arbitrary index $j \in [k]$, and recall that $S_i^j$ is the $(1+\epsilon)$-spanner for $N_i^j$ defined above, having at most $\frac{|N_i^j|}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$ edges. This size bound on $S_i^j$ implies that the average degree of a point in $N_i^j$ due to edges of $E'(S_i^j)$ is at most $\frac{2}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$. A priori, however, the *maximum* degree of a point of $N_i^j$ due to edges of $E'(S_i^j)$

could be huge and, moreover, there could be many edges in $S_i^j$ that are incident on Steiner points. We will consider another edge set $\tilde{E}(S_i^j)$ of at least the same size as $E'(S_i^j)$, defined over $N_i^j$ (i.e., with no Steiner points), where the maximum degree of a point in $N_i^j$ due to edges of $\tilde{E}(S_i^j)$ does not exceed $\frac{2}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$.

Define $D = \frac{2}{\sqrt{\epsilon}} \cdot 2^{c\sqrt{\log \frac{1}{\epsilon}}}$. We distinguish between two cases. In the case that $|N_i^j| \leq D$, recall that $S_i^j$ is the complete graph over $N_i^j$, of maximum degree $|N_i^j| - 1 < D$, hence we can take $\tilde{E}(S_i^j)$ to be $E'(S_i^j) = \binom{N_i^j}{2}$. Otherwise $|N_i^j| > D$, and we take $\tilde{E}(S_i^j)$ to be any edge set over $N_i^j$ that induces a $D$-regular graph; such an edge set clearly exists, and its size is no smaller than that of the original edge set $E'(S_i^j)$. (It suffices for all vertices to have degree $\Theta(D)$, i.e., strict regularity is not needed.)

Observe that each edge of $\tilde{E}(S_i^j)$ has both endpoints in $N_i^j$, which are at distance at most $\sqrt{2} \cdot 3\tau_i$ from each other by construction, where $\tau_i = \left(4 + \frac{32}{\epsilon}\right) 2^i$ is defined in the proof of Lemma IV.6.

Define $\tilde{E}_i = \bigcup_{j=1}^{k} \tilde{E}(S_i^j)$. Although $|\tilde{E}(S_i^j)| \geq |E'(S_i^j)|$ for each $j \in [k]$, it may a priori be that $|\tilde{E}_i| < |E_i'|$, due to potential intersections between the different edge sets $\tilde{E}(S_i^j)$, $j \in [k]$. To overcome this technicality, we consider $\tilde{E}_i$ as a *mutli-graph*, in which edges may appear multiple times. By Item (2) of Lemma IV.6, each point of $N_i$ belongs to at most four edge sets from $\tilde{E}(S_i^1), \tilde{E}(S_i^2), \ldots, \tilde{E}(S_i^k)$, hence the degree of each point due to all edges (with all their multiplicities) of $\tilde{E}_i$ is at most $4D$.

Define $\tilde{E} = \bigcup_{i=0}^{\ell-1} \tilde{E}_i$; as before, we consider this edge set $\tilde{E}$ as a multi-graph. It is easy to verify that the resulting edge set $\tilde{E}$ satisfies $|\tilde{E}| \geq |E'|$. Next, we upper bound the size of $\tilde{E}$.

Following [13], for each point $p \in P$, we define $i^*(p) := \max\{i \in [0, \ell] \mid p \in N_i\}$. To upper bound the size of $\tilde{E}$, we orient each edge $(p, q) \in \tilde{E}$ from $p$ towards $q$ if $i^*(p) < i^*(q)$; if $i^*(p) = i^*(q)$, the edge $(p, q)$ is oriented arbitrarily. We next bound the out-degree of an arbitrary point $p$ by all edges of $\tilde{E}$. Let $i$ be the minimum index such that $p$ has at least one outgoing edge in $\tilde{E}_i$, leading to some point $q$. We know that $|pq| \leq \sqrt{2} \cdot 3\tau_i$; take $\mu$ such that $\mu \cdot 2^i = \sqrt{2} \cdot 3\tau_i + 1$, and note that $|pq| < \mu \cdot 2^i, \mu = \Theta(1/\epsilon)$. Since $N_{i+\lceil \log \mu \rceil}$ is a $2^{i+\lceil \log \mu \rceil}$-net (of $N_{i+\lceil \log \mu \rceil - 1}$), any two points in $N_{i+\lceil \log \mu \rceil}$ are at distance at least $\mu \cdot 2^i$ from each other. Since $\mu \cdot 2^i > |pq|$ and $i^*(p) < i^*(q)$, it follows that $p$ cannot belong to $N_{i+\lceil \log \mu \rceil}$.

Thus, $p$ may only belong to the $\lceil \log \mu \rceil$ nets $N_i, N_{i+1}, \ldots, N_{i+\lceil \log \mu \rceil - 1}$. Observe that the $\lceil \log \mu \rceil$ edge sets $\tilde{E}_i, \tilde{E}_{i+1}, \ldots, \tilde{E}_{i+\lceil \log \mu \rceil - 1}$ are defined over the nets $N_i, N_{i+1}, \ldots, N_{i+\lceil \log \mu \rceil - 1}$, respectively, while all the other edge sets of $\tilde{E}$ are defined over different nets. It follows that the out-degree of $p$ may increase only due to these $\lceil \log \mu \rceil$ edge sets. In each of these edge sets the degree of $p$, let alone its out-degree, is at most $4D$, hence the out-degree of $p$ due to the entire edge set $\tilde{E}$ is at most $\lceil \log \mu \rceil \cdot 4D = \frac{1}{\sqrt{\epsilon}} \cdot 2^{O\left(\sqrt{\log \frac{1}{\epsilon}}\right)}$. Having shown that the out-degree of any point $p \in P$ due to the edge set $\tilde{E}$ (defined as a multi-graph) is $\frac{1}{\sqrt{\epsilon}} \cdot 2^{O\left(\sqrt{\log \frac{1}{\epsilon}}\right)}$, we conclude that the size of $\tilde{E}$, and thus of $E'$, is bounded by $\frac{n}{\sqrt{\epsilon}} \cdot 2^{O(\sqrt{\log \frac{1}{\epsilon}})}$. ■

*Extension to any constant dimension:* The 2-dimensional construction presented here naturally generalizes to $\mathbb{R}^d$, for any constant $d$. We shall only highlight the key components of this generalization.

Two hyperrectagles $R_1 = [0, H] \times [0, W]^{d-1}, R_2 = [H + \frac{W}{\ell}, 2H + \frac{W}{\ell}] \times [0, W]^{d-1}$, for some numbers $H, W > 0$, are called *parallel hyperrectangles*; $R_1$ and $R_2$ have $d-1$ sides of length $W$ and another side of length $H$, and the distance between $R_1$ and $R_2$ is $\frac{W}{\ell}$. The $d$-dimensional analogue of Lemma IV.2 is to construct a Steiner spanner that handles all pairs of points from $R_1$ and $R_2$ with at most $O(\frac{\ell}{\epsilon^{(d-1)/2}}(|R_1| + |R_2|))$ edges; we employ the same argument: scale the metric so that $W = 1$, and then place a grid of $O(\frac{\ell}{\epsilon^{(d-1)/2}})$ Steiner points in the $d-1$ dimensional hypercube $L_d = [H + \frac{1}{2\ell}, H + \frac{1}{2\ell}] \times [0, 1]^{d-1}$ that is aligned with $R_1$ and $R_2$ in $(d-1)$ dimensions and separates $R_1$ and $R_2$ in the middle of the remaining dimension, and finally connect all Steiner points with all points in $R_1$ and $R_2$; $L_d$ is the $d$-dimensional analogue of the separating segment $L$ in the proof of Lemma IV.2. Next, by using the same divide and conquer approach, we can construct a Steiner spanner for point sets of spread at most $\Delta$, with at most $\frac{n}{\epsilon^{(d-1)/2}} 2^{O(\sqrt{\log \Delta})}$ edges; this is the $d$-dimensional analogue of Proposition IV.1. Finally, the reduction from the general case to the case of bounded spread is carried out in a very similar way, by building on the net-tree spanner and replacing cross edges by Steiner ring spanners. That is, as before, for every level $i$, we replace each edge set $E_i$, where $E_i := \left\{(p, q) \mid p, q \in N_i, \delta(p, q) \leq \left(4 + \frac{32}{\epsilon}\right) 2^i\right\}$, by a $(2^i, (4 + \frac{32}{\epsilon}) 2^i)$-ring $(1 + \epsilon)$-spanner for $N_i$; the 2-dimensional treatment for this part extends easily to any dimension. As a result, we get Steiner spanners for general point sets with at most $\frac{n}{\epsilon^{(d-1)/2}} 2^{O(\log \frac{1}{\epsilon})}$ edges.

## V. Lower bounds for sparse Steiner spanners

In this section we prove Theorem I.4. Let $U$ be a unit square with four sides $N, E, S, W$. Let $P_1$ be any set of evenly spaced points along $N$ such that the distance between two consecutive points of $P_1$ along $N$ is $c\sqrt{\epsilon \log(\frac{1}{\epsilon})}$, for a sufficiently large constant $c$. To simplify the argument, we remove the two furthest points of $P_1$, so that every point is at distance $\geq c\sqrt{\epsilon \log(\frac{1}{\epsilon})}$ from the corners of $U$. We define the set of points $P_2$ on $S$ similarly. Let $P = P_1 \cup P_2$. See Figure 3(a) for an illustration. Our goal is to show that:

**Proposition V.1.** *Any Steiner $(1+\epsilon)$-spanner $ST_P$ of $P$ must have $w(ST_P) = \Omega_\epsilon(\frac{1}{\epsilon})$, for some sufficiently large constant $c$ independent of $\epsilon$.*

Before proving Proposition V.1, we show that it implies Theorem I.4.

**Claim V.2.** *If Proposition V.1 is true, then Theorem I.4 holds.*

*Proof:* Assume that $w(ST_P) = \Omega_\epsilon(\frac{1}{\epsilon})$. First we show that $|E(ST_P)| = \Omega_\epsilon(\frac{1}{\epsilon})$. Let $x_1, x_2$ be any two points in $P$. Since $|x_1x_2| = O(1)$, the shortest path between $x_1$ and $x_2$ in $ST_P$ must have length $O(1)$. Thus, denoting by $e_{\max}$ the edge of maximum weight in $E(ST_P)$, we have

$$|E(ST_P)| \;\geq\; \frac{w(ST_P)}{w(e_{\max})} \;=\; \Omega_\epsilon\left(\frac{1}{\epsilon}\right). \qquad (13)$$

Recall that $|P| = O(\frac{1}{\sqrt{\epsilon\log(\frac{1}{\epsilon})}})$. Thus, $S$ has more edges than the number of points by $\Omega_\epsilon(\frac{1}{\sqrt{\epsilon}})$ factors. To complete Theorem I.4, we next extend the argument to $n$-point sets for any $n$ and $\epsilon$ with $\epsilon = \tilde{\Omega}(\frac{1}{n^2})$.

Let $g(\epsilon) = \frac{1}{\sqrt{\epsilon\log(\frac{1}{\epsilon})}}$ and $\alpha$ be such that $|P| = \alpha g(\epsilon)$. For simplicity of presentation, we assume that $n$ is divisible by $\alpha g(\epsilon)$, otherwise, we can always increase $n$ by at most $\alpha g(\epsilon)$ to guarantee this property. We make $k = \frac{n}{\alpha g(\epsilon)}$ vertex-disjoint copies of the aforementioned point set $P$, denoted by $P_1, P_2, \ldots, P_k$, where each $P_i$ is defined with respect to a separate unit square $U_i$, where the squares $U_1, \ldots, U_k$ are horizontally aligned so that the distance between any two nearby squares is 3 (see Figure 3(b)). Let $Q = P_1 \cup P_2 \cup \ldots \cup P_k$ and note that $|Q| = n$. Let $S_Q$ be any Steiner $(1+\epsilon)$-spanner of $Q$ and let $S_Q[P_i]$ be any inclusion-wise minimal subgraph of $S_Q$ that provides a $(1+\epsilon)$-spanner for $P_i$, for each $i \in [k]$. Since $d(U_i, U_j) \geq 3$ for every $i \neq j$, $S_Q[P_i]$ and $S_Q[P_j]$ must be vertex-disjoint (let alone edge-disjoint) by their minimality. By Proposition V.1, $w(S_Q[P_i]) = \Omega_\epsilon(\frac{1}{\epsilon})$, thus $w(S_Q) = \Omega_\epsilon(\frac{k}{\epsilon})$. By construction, we have $w(\mathrm{MST}(Q)) \leq O(k)$, thus $w(S_Q) = \Omega_\epsilon(\frac{1}{\epsilon}w(\mathrm{MST}(Q)))$, which proves the lightness bound. For the size bound, Equation (13) yields $|E(S_Q[P_i])| \geq \Omega_\epsilon(\frac{1}{\epsilon})$, thus $|E(S_Q)| \geq k\Omega_\epsilon(\frac{1}{\epsilon}) = \frac{n}{\alpha g(\epsilon)}\Omega_\epsilon(\frac{1}{\epsilon}) = \Omega_\epsilon(\frac{n}{\sqrt{\epsilon}})$. ■

In what follows we prove Proposition V.1. Assume that $ST_P$ is a Steiner $(1+\epsilon)$-spanner for $P$ of minimum weight. Observe that one can "planarize" $ST_P$ without increasing its weight: whenever two edges of $ST_P$ intersect at a point on the plane that is not a point of $P$, we add the crossing point to the set of vertices of $ST_P$. We argue that the spanner must stay inside $U$.

**Claim V.3.** $ST_P \subseteq U$.

*Proof:* For any point $p$, define its *projection onto* $U$, denoted by $\texttt{proj}(p)$, as follows. If $p \in U$, then $\texttt{proj}(p) = p$, otherwise $\texttt{proj}(p)$ is the closest point on the boundary of $U$. Observe that for every pair $p, q$ of points, $|\texttt{proj}(p)\texttt{proj}(q)| \leq |pq|$. Thus the spanner obtained from $ST_P$ by replacing every edge $(p,q) \in ST_P$ with the "projected" edge $(\texttt{proj}(p), \texttt{proj}(q))$ has stretch and weight no greater than those of $ST_P$. ■

Let $x_1$ and $x_2$ be two points in $P_1$ and $P_2$, respectively. We define a *bell* of radius $r$ of the line segment $x_1x_2$, denoted by $\texttt{Bell}(x_1x_2, r)$, to be the set of points in $U$ at distance at most $r$ from $x_1x_2$. We call the boundary line segments of $\texttt{Bell}(x_1x_2, r)$ connecting the $N$ and $S$ sides of $U$ the *long boundaries* of the bell, and the other two boundary line segments are called the *short boundaries*. Since we made sure that every point of $P$ is at distance $\geq c\sqrt{\epsilon\log(\frac{1}{\epsilon})}$ from the corners of $U$, we have $\texttt{Bell}(x_1, x_2) \subseteq U$, which in particular means that all the bells (including the two extreme ones) are of precisely the same size.

Let $Q_x$ be an arbitrary shortest path between $x_1$ and $x_2$ in $ST_P$. We claim that:

**Claim V.4.** $Q_x \subseteq \texttt{Bell}(x_1x_2, 2\sqrt{\epsilon})$.

*Proof:* Suppose for contradiction that $Q_x$ contains a point outside $\texttt{Bell}(x_1x_2, \sqrt{2\epsilon})$. By Claim V.3, $Q_x$ must intersect a long boundary, say $L$, of $\texttt{Bell}(x_1x_2, \sqrt{2\epsilon})$ at a point $t$. Let $p$ be the reflection point of $x_1$ over the line defined by $L$ (see Figure 4(a)). By the triangle inequality, we have

$$\begin{aligned} Q_x \geq\; & |tx_1| + |tx_2| \;=\; |tp| + |tx_2| \geq |px_2| \\ =\; & \sqrt{|x_1x_2|^2 + |x_1p|^2} \;=\; \sqrt{|x_1x_2|^2 + 16\epsilon} \\ \geq\; & |x_1x_2|\sqrt{1+8\epsilon} \qquad \text{since } |x_1x_2| \leq \sqrt{2} \\ >\; & (1+\epsilon)|x_1x_2| \qquad \text{since } \epsilon < 1, \end{aligned}$$

which contradicts the fact that $Q_x$ is a $(1+\epsilon)$-spanner path for the pair $x_1, x_2$. ■

Since the acute angle between $x_1x_2$ and $N$ is at least $\pi/4$, we have:

**Observation V.5.** *The segments* $\texttt{Bell}(x_1x_2, 2\sqrt{\epsilon}) \cap N$, $\texttt{Bell}(x_1x_2, 2\sqrt{\epsilon}) \cap S$ *have length at most* $4\sqrt{2\epsilon}$ *each.*

Let $y_1, y_2$ be two points in $P_1$ and $P_2$, respectively, such that $|\{y_1, y_2\} \cap \{x_1, x_2\}| \leq 1$. By Observation V.5 and since the minimum pairwise distance between points in $P$ is at least $c\sqrt{\epsilon\log(\frac{1}{\epsilon})}$ for some sufficiently big constant $c$, we have:

**Observation V.6.** *If* $x_1x_2 \cap y_1y_2 = \emptyset$, *then* $\texttt{Bell}(x_1x_2, 2\sqrt{\epsilon}) \cap \texttt{Bell}(y_1y_2, 2\sqrt{\epsilon}) = \emptyset$

Let $Q_y$ be an arbitrary shortest path between $y_1$ and $y_2$ in $ST_P$. For a given path $Q$, we will use $Q[a, b]$ to denote the subpath between $a$ and $b$ of $Q$. Recall that $ST_P$ is planarized, and so $Q_x$ and $Q_y$ may only intersect at points that are vertices of $ST_P$. We want to upper bound the sum of weights of all subpaths shared by $Q_x$ and $Q_y$ (if any), denoted by

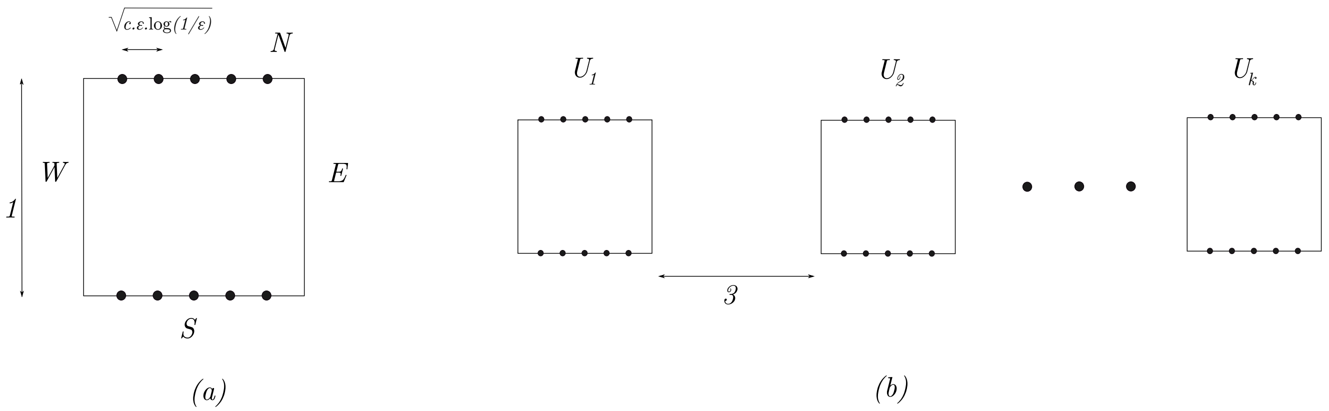


Figure 3. (a) The point set $P$ lying on the $N$ and $S$ sides of the unit square in our lower bound proof. (b) Extending the example in (a) to the case where the point set is arbitrarily large.

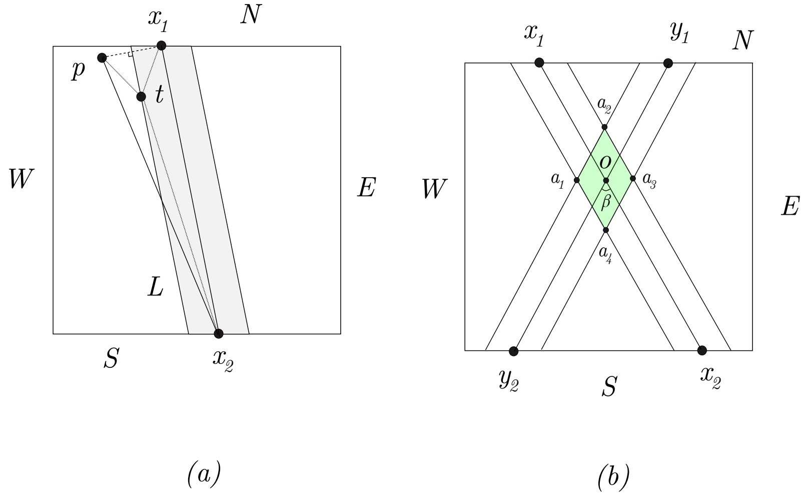


Figure 4. (a) The shaded region is the bell $\texttt{Bell}(x_1x_2, 2\sqrt{\epsilon})$. (b) The parallelogram $A$ in the proof of Lemma V.8 is green shaded.

$w(Q_x \cap Q_y)$. The next observation shows that this sum is maximized when $Q_x$ and $Q_y$ share a single subpath.

**Observation V.7.** *If $(Q_x \cap Q_y) \neq \emptyset$, then $w(Q_x \cap Q_y)$ is maximized when $Q_x \cap Q_y$ is a single path in $ST_P$.*

*Proof:* Let $p$ and $q$ be the first and the last points along $Q_y$ that belong to $Q_x \cap Q_y$, respectively. Since $Q_x$ and $Q_y$ are shortest paths between $x_1$ and $x_2$ and between $y_1$ and $y_2$ in $ST_P$, respectively, we have $Q_x[p,q] = Q_y[p,q]$. Thus, we can replace $Q_y[p,q]$ by $Q_x[p,q]$ to obtain another shortest path $Q'_y$ in $ST_P$ between $y_1$ and $y_2$ such that $Q_x \cap Q'_y$ is a single path and $w(Q_x \cap Q_y) \leq w(Q_x \cap Q'_y)$. ■

We define the distance between two pairs $\{x_1, x_2\}$ and $\{y_1, y_2\}$, denoted by $d(\{x_1, x_2\}, \{y_1, y_2\})$, to be $\max(|x_1y_1|, |x_2y_2|)$. The following Lemma is central to the proof of Proposition V.1.

**Lemma V.8.** *If $d(\{x_1, x_2\}, \{y_1, y_2\}) = j\sqrt{\epsilon}$ for some sufficiently large $j$, then $w(Q_x \cap Q_y) = O(\frac{1}{j^2})$.*

*Proof:* We assume without loss of generality that $|x_2y_2| \geq |x_1y_1|$. If $x_1x_2 \cap y_1y_2 = \emptyset$, then $w(Q_x \cap Q_y) = 0$ by Observation V.6, and we are done. We henceforth assume that $x_1x_2 \cap y_1y_2 \neq \emptyset$ and let $o = x_1x_2 \cap y_1y_2$. Since $|x_2y_2| \geq |x_1y_1|$, we have $d(o, S), |ox_2|, |oy_2| \geq \frac{1}{2}$. Let $\beta = \angle x_2oy_2$.

Let $A$ be the parallelogram given by $A = \texttt{Bell}(x_1x_2, 2\sqrt{\epsilon}) \cap \texttt{Bell}(y_1y_2, 2\sqrt{\epsilon})$, and let $a_1, a_2, a_3, a_4$ be the vertices of $A$, where $a_1, a_2, a_3, a_4$ are closest to the left, top ($N$), right and bottom ($S$) sides of the square $U$, respectively (see Figure V.4(b)). We now bound $|oa_4|$. Since $\pi/4 \leq \angle ox_2y_2 \leq \pi/2$, we have:

$$\sin(\beta) = \frac{(\sin \angle ox_2y_2)|x_2y_2|}{|oy_2|} \geq \frac{1}{\sqrt{2}} \frac{|x_2y_2|}{|oy_2|} \geq |x_2y_2|/2 = \frac{j\sqrt{\epsilon}}{2}. \tag{14}$$

and

$$\sin(\beta) \leq \frac{|x_2y_2|}{|oy_2|} \leq 2j\sqrt{\epsilon} \tag{15}$$

Thus, by Equation 2, $\frac{j\sqrt{\epsilon}}{2} \leq \beta \leq 4j\sqrt{\epsilon}$. Let $o_1, o_2$ be the projections of $o$ onto the lines that go through the line segments $a_1a_4$ and $a_4a_3$, respectively. Since $\angle o_1a_4o + \angle oa_4o_2 = \beta$, at least one among $\angle o_1a_4o$ and $\angle oa_4o_2$, without loss of generality $\angle o_1a_4o$, must have degree at least $\beta/2$. Thus $\beta/2 \leq \angle o_1a_4o \leq \pi/2$, and so $\sin\angle o_1a_4o \geq \sin(\beta/2) \geq (\sin\beta)/2$, which implies that

$$\begin{aligned} |oa_4| &= \frac{|oo_1|}{\sin\angle o_1a_4o} \leq \frac{2\sqrt{\epsilon}}{\sin(\beta/2)} \\ &\leq \frac{4\sqrt{\epsilon}}{\sin\beta} \leq \frac{8\sqrt{\epsilon}}{j\sqrt{\epsilon}} = \frac{8}{j}. \end{aligned}$$

By the triangle inequality, we conclude that $d(a_4, S) \geq d(o, S) - |oa_4| \geq \frac{1}{2} - \frac{8}{j} \geq \frac{1}{4}$, for any $j \geq 32$.

By Observation V.7, $Q_x \cap Q_y$ is a single path. Let $p$ and $q$ be its endpoints. By Claim V.4, $p, q \in A$. Thus, $d(p, S) \geq d(a_4, S) \geq \frac{1}{4}$ and $d(q, S) \geq d(a_4, S) \geq 1/4$. If $p = q$, then $w(Q_x \cap Q_y) = 0$ and Lemma V.8 holds. Moreover, if $Q_x[p, q] \leq \sqrt{2}\epsilon$, then again the lemma must hold, since the fact that $|x_1y_1|, |x_2y2| \leq 1$ yields $j \leq \frac{1}{\sqrt{\epsilon}}$. We henceforth assume that $p \neq q$ and $Q_x[p, q] > \sqrt{2}\epsilon$.

Let $L_x$ (respectively, $L_y$) be the line going through $p$ and parallel to $x_1x_2$ (resp., $y_1y_2$). Let $L'_x$ (respectively, $L'_y$) be the line going through $q$ and parallel to $x_1x_2$ (resp., $y_1y_2$). Note that $\angle L_xpL_y = \angle L'_xqL'_y = \beta$. By construction, it is readily verified that all lines $L_x, L_y, L'_x, L'_y$ intersect $S$; we henceforth define $x = L_x \cap S, y = L_y \cap S, x' = L'_x \cap S, y' = L'_y \cap S$.

**Claim V.9.** *All angles $\angle ypy_2, \angle xpx_2, \angle y'qy_2, \angle x'qx_2$ are at most $32\sqrt{2\epsilon}$.*

*Proof:* By symmetry, it suffices to bound $\angle ypy_2$. Since $y \in \texttt{Bell}(y_1y_2, 2\sqrt{\epsilon}) \cap S$, Observation V.5 yields $|y_2y| \leq 4\sqrt{2\epsilon}$. Thus, $\sin(\angle y_2py) \leq \frac{|yy_2|}{|py_2|} \leq \frac{|yy_2|}{d(p,S)} \leq 4|yy_2| = 16\sqrt{2\epsilon}$. By Equation 2, $\angle y_2py \leq 32\sqrt{2\epsilon}$. ■

**Claim V.10.** $\beta/2 \leq \angle x_2py_2, \angle x_2qy_2 \leq 2\beta$.

*Proof:* By symmetry, it suffices to bound $\angle x_2py_2$. Recall that $j\sqrt{\epsilon}/2 \leq \beta$. By Claim V.9, when $j$ is sufficiently large, it holds that: $\angle x_2py_2 \leq \angle x_2px + \angle xpy + \angle y_2py \leq \beta + 64\sqrt{2\epsilon} \leq 2\beta$,
$\angle x_2py_2 \geq \angle xpy - \angle x_2px - \angle y_2py \geq \beta - 64\sqrt{2\epsilon} \geq \beta/2$. ■

Let $L_S$ be the line containing $S$ side of square $U$. We define the *admissible triangle* of $p$ w.r.t. $x_2$ (resp. $y_2$) to be the triangle $w_1pw_2$ such that (a) $w_1, w_2 \in L_S$, (b) $px_2$ (resp. $py_2$) is the bisector of the triangle $w_1pw_2$ and (c) $\angle w_1pw_2 = \beta/4$ (see Figure 5(a)). The admissible triangles of $q$ w.r.t. $x_2$ and $y_2$ are defined in the same way.

Suppose for contradiction that Lemma V.8 does not hold, and specifically, that $Q_x[p, q] > \frac{768\sqrt{2}}{j^2}$.

**Claim V.11.** *If $p \in Q_x[x_1, q]$ (respectively $p \in Q_y[y_1, q]$), then $q$ is in the admissible triangle of $p$ w.r.t. $x_2$ (resp. $y_2$). Symmetrically, if $q \in Q_x[x_1, p]$ (resp. $q \in Q_y[y_1, p]$), then $p$ is in the admissible triangle of $q$ w.r.t. $x_2$ (resp., $y_2$).*

*Proof:* By symmetry, it suffices to assume that $p \in Q_x[x_1, q]$ and prove that $q$ is in the admissible triangle of $p$ w.r.t. $x_2$. Let $q_0$ be the projection of $q$ on $px_2$. We observe that $q_0$ must belong to the line segment $x_2p$ since otherwise, recalling the fact that $Q_x[p, q] > \sqrt{2}\epsilon$ and noting that $|x_1x_2| \leq \sqrt{2}$, we have:

$$\begin{aligned} w(Q_x) &\geq |x_1p| + Q_x[p, q] + |qx_2| \\ &\geq |x_1p| + Q_x[p, q] + |q_0x_2| \\ &\geq |x_1p| + Q_x[p, q] + |px_2| \\ &> |x_1x_2| + \sqrt{2}\epsilon \geq (1 + \epsilon)|x_1x_2| \end{aligned}$$

contradicting the fact that $Q_x$ is a $(1 + \epsilon)$-spanner path for the pair $x_1, x_2$.

Suppose that $q$ is not in the admissible triangle of $p$ w.r.t. $x_2$ (see Figure 5(a)). Then $\gamma = \angle qpx_2 \geq \beta/8 \geq \frac{j\sqrt{\epsilon}}{16}$. We now show that:

$$Q_x[p, q] - |pq_0| \leq \sqrt{2}\epsilon \qquad (16)$$

If Equation 16 is not true, by the triangle inequality and the fact that $|y_1y_2| \leq \sqrt{2}$,

$$\begin{aligned} w(Q_x) &\geq |x_1p| + Q_x[p, q] + |qx_2| \\ &\geq (|x_1p| + |qx_2| + |pq_0|) + (Q_x[p, q] - |pq_0|) \\ &> (|x_1p| + |q_0x_2| + |pq_0|) + \sqrt{2}\epsilon \\ &\geq |x_1x_2| + \sqrt{2}\epsilon \geq (1 + \epsilon)|x_1x_2| \end{aligned}$$

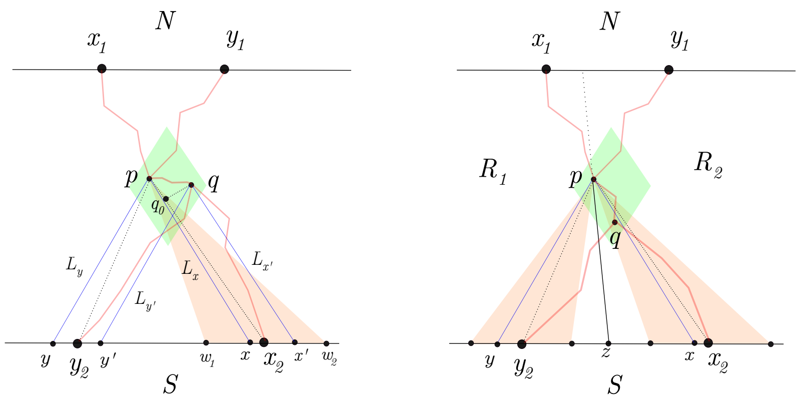
contradicting the fact that $Q_y$ is a $(1 + \epsilon)$-spanner path for the pair $y_1, y_2$. Thus Equation 16 must hold.

Observe that $|pq_0| = |pq|\cos\gamma \leq Q_x[p, q]\cos\gamma$. By Equation 16, we thus have $Q_x[p, q] \leq \sqrt{2}\epsilon + Q_x[p, q]\cos\gamma$, which yields $Q_x[p, q] \leq \frac{\sqrt{2}\epsilon}{1-\cos\gamma} \leq \frac{3\sqrt{2}\epsilon}{\gamma^2} \leq \frac{768\sqrt{2}}{j^2}$, where the last equation holds since $\gamma \geq \frac{j\sqrt{\epsilon}}{16}$, which contradicts our assumption that $Q_x[p, q] > \frac{768\sqrt{2}}{j^2}$. ■

Without loss of generality, we assume that $p \in Q_x[x_1, q]$. By Claim V.11, $q$ is in the admissible triangle of $p$ w.r.t. $x_2$. We consider two cases:

*Case 1:* $p \in Q_y[y_1, q]$. By Claim V.11, $q$ is in the admissible triangle of $p$ w.r.t. $y_2$. Let $pz$ be the bisector of the angle $x_2py_2$ where $z \in S$ (see Figure 5(b)). Let $R_1$ and $R_2$ be the left and the right regions, respectively, of the square separated by the line containing $pz$. By Claim V.10, $\angle zpx_2 = \angle x_2py_2/2 \geq \beta/4$. Since $q$ is in the admissible triangle of $p$ w.r.t. $x_2$, $\angle x_2pq \leq \beta/8$. Thus, $q \in R_2$. By the same argument, since $q$ is in the admissible triangle of $p$ w.r.t. $y_2$, $q$ must be in $R_1$, which is a contradiction.

*Case 2:* $q \in Q_y[y_1, p]$. By Claim V.9 the acute angle between $px_2$ and $S$ is at least $\pi/4 - \angle xpx_2 \geq \pi/4 - 32\sqrt{2\epsilon} > \pi/6$ when $\epsilon$ is sufficiently smaller than 1. Since $\beta \leq \pi/2$ and $q$ is in the admissible triangle of $p$ w.r.t. $x_2$ , $\angle qpx_2 \leq \beta/8 < \pi/6$. It follows that $d(p, N) < d(q, N)$.



(a) (b)

Figure 5. The red paths are $Q_x$ and $Q_y$. (a) The light orange shaded region is the admissible triangle of $p$ w.r.t. $x_2$. (b) A case in the proof of Lemma V.8.

By Claim V.11, $p$ is in the admissible triangle of $q$ w.r.t. $y_2$, thus by a symmetric argument to the above, since $p$ is in the admissible triangle of $q$ w.r.t. $y_2$, we get $d(q, N) < d(p, N)$, which is a contradiction.

This completes the proof of Lemma V.8. ■

*Proof of Proposition V.1:* Consider a pair $x_1 \in P_1, x_2 \in P_2$ of points. Let $Q_x$ be a shortest path between $x_1$ and $x_2$ in $ST_P$. We say that the pair $x_1, x_2$ contributes a *positive cost* of $w(Q_x)$ to $w(ST_P)$, and note that $w(Q_x) \geq 1$. Denote by $\bar{P}(x_1, x_2)$ the set of pairs $y_1 \in P_1, y_2 \in P_2$, such that $|\{y_1, y_2\} \cap \{x_1, x_2\}| \leq 1$. For every pair $y_1, y_2 \in \bar{P}(x_1, x_2)$, we charge a *negative cost* of $w(Q_x \cap Q_y)$ to $w(ST_P)$, and associate it with the pair $x_1, x_2$. The sum of negative costs associated with pair $x_1, x_2$, denoted by $\texttt{negCost}(x_1, x_2)$, is given by $\texttt{negCost}(x_1, x_2) = \sum_{y_1, y_2 \in \bar{P}(x_1, x_2)} w(Q_x \cap Q_y)$.

**Observation V.12.** $w(ST_P) \geq \sum_{x_1 \in P_1, x_2 \in P_2} (w(Q_x) - \texttt{negCost}(x_1, x_2))$.

*Proof:* Observe that $w(ST_P) \geq w(\cup_{x_1 \in P_1, x_2 \in P_2} Q_x)$. By the inclusion-exclusion principle,

$$\begin{aligned} w(ST_P) &\geq \sum_{x_1 \in P_1, x_2 \in P_2} (w(Q_x) \\ &- \sum_{y_1, y_2 \in \bar{P}(x_1, x_2)} w(Q_x \cap Q_y)) \\ &= \sum_{x_1 \in P_1, x_2 \in P_2} (w(Q_x) - \texttt{negCost}(x_1, x_2)) \end{aligned}$$

■

We next upper bound $\texttt{negCost}(x_1, x_2)$. By definition of $P$ and $\bar{P}(x_1, x_2)$, for any pair $y_1, y_2 \in \bar{P}(x_1, x_2)$, we can write $d(\{x_1, x_2\}, \{y_1, y_2\}) = cj\sqrt{\epsilon \log(\frac{1}{\epsilon})}$ for some index $j$ satisfying $1 \leq j \leq \frac{1}{c\sqrt{\epsilon \log(\frac{1}{\epsilon})}}$. Fix an arbitrary index $j$ such that $1 \leq j \leq \frac{1}{c\sqrt{\epsilon \log(\frac{1}{\epsilon})}}$, and note that there are at most $4j$ pairs $y_1, y_2 \in \bar{P}(x_1, x_2)$ such that $d(\{x_1, x_2\}, \{y_1, y_2\}) = cj\sqrt{\epsilon \log(\frac{1}{\epsilon})}$. Take $k$ so that $k\sqrt{\epsilon} = cj\sqrt{\epsilon \log(\frac{1}{\epsilon})}$; by Lemma V.8, the total contribution to $\texttt{negCost}(x_1, x_2)$ by all such pairs is at most

$$\begin{aligned} O\left(\frac{1}{k^2}\right)4j &= O\left(\frac{1}{\log(\frac{1}{\epsilon})} \cdot \frac{1}{c^2 j^2}\right)4j \\ &= O\left(\frac{1}{\log(\frac{1}{\epsilon})} \cdot \frac{1}{c^2 j}\right). \end{aligned}$$

Summing over all possible values of $j$ and using the fact that $c$ is sufficiently large, we get that $\texttt{negCost}(x_1, x_2)$ satisfies

$$\begin{aligned} \texttt{negCost}(x_1, x_2) &\leq O\left(\frac{1}{c^2 \log(\frac{1}{\epsilon})} \sum_{j=1}^{\frac{1}{c\sqrt{\epsilon \log(\frac{1}{\epsilon})}}} \frac{1}{j}\right) \\ &= O\left(\frac{1}{c^2 \log(\frac{1}{\epsilon})} \log\left(\frac{1}{c\sqrt{\epsilon \log(\frac{1}{\epsilon})}}\right)\right) \\ &\leq 1/2, \end{aligned}$$

Hence $w(Q_x) - \texttt{negCost}(x_1, x_2) \geq \frac{1}{2}$, and by Observation V.12

$$\begin{aligned} w(ST_P) &\geq \sum_{x_1 \in P_1, x_2 \in P_2} (w(Q_x) - \texttt{negCost}(x_1, x_2)) \\ &\geq \frac{|P_1||P_2|}{2} = \Omega\left(\frac{1}{\epsilon \log(1/\epsilon)}\right) = \Omega_\epsilon\left(\frac{1}{\epsilon}\right). \end{aligned}$$

## VI. Upper bounds for greedy light spanners the in Euclidean space

In this section, we provide high level ideas of our analysis of the greedy algorithm; the details are deferred to the full version. We follow the lightness analysis framework of the greedy spanners by Borradaile, Le and Wulff-Nilsen [8], [7], that we abbreviate as BLW approach. We would focus more on the presentation of the framework for the metrics of bounded doubling dimension [8]. Doubling metrics can be seen as a discrete version of $\mathbb{R}^d$ in a sense that the (discrete) metric space of any point set in $\mathrm{R}^d$ has doubling dimension

$O(d)$. Thus, one can directly transfer BLW analysis to $\mathbb{R}^d$. However, the lightness bound obtained in [8] is $\epsilon^{-O(d)}$ with an unspecified constant behind the big-$O$. Meanwhile, in our work, the goal is to establish the exact constant in the exponent of $\epsilon$, specifically $O(\epsilon^{-d})$. This bound is optimal by the lower bound in Theorem I.1. To achieve the goal, we reformulate BLW analysis in the doubling metric setting to the geometric setting and bring in several new insights. Our first insight is that by carefully tailoring BLW analysis directly to $\mathbb{R}^d$, we can obtain the lightness bound $O(\epsilon^{-(d+2)})$. This bound is better than the current best upper bound $O(\epsilon^{-2d})$ by Narasimhan and Smid [42] when $d \geq 3$. However, in $\mathrm{R}^2$, which is arguably the most important case, the lightness bound is $\epsilon^{-4}$, which is far from the optimal bound $\epsilon^{-2}$. To shave the factor $\epsilon^{-2}$, we show several geometric properties of the greedy algorithm. Before sketching the high level ideas of our analysis, we briefly review BLW approach, that we tailor to $\mathbb{R}^d$. Recall that $S_{\mathrm{grd}}$ is the greedy spanner of the point set $P$.

### *A. A brief review of BLW approach*

Let $\bar{w} = \frac{w(\mathrm{MST})}{n-1}$ be the average weight of MST edges. They construct a clustering hierarchy $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_L$ where for each $i \in [p]$, each cluster in level-$i$ clustering $\mathcal{C}_i$ is a connected subgraph of $S_{\mathrm{grd}}$ of diameter $O(L_i)$ where $L_i = \frac{\bar{w}}{\epsilon^i})$. (To be precise, they construct $O(\log \frac{1}{\epsilon})$ clustering hierarchies, but for the purpose of simplifying the notation, we assume that there is only one clustering hierarchy.) We can think of $\mathcal{C}_I$ as a $O(L_i)$-cover of the point set. Additionally, clusters in $\mathcal{C}_i$ are constructed by grouping clusters in $\mathcal{C}_{i-1}$.

The set of edges of the spanner is partitioned according to the clustering hierarchy: level-$i$ spanner edges have length $\Theta(L_i)$. This remind us of the net-tree spanner presented in Section IV, with a slight different details, which is, the edge sets are scaled by $\frac{1}{\epsilon}$ instead of 2.

Let $C$ be a level-$(i-1)$ cluster. By the packing argument, they showed that $C$ is incident to at most $\epsilon^{-O(d)}$ level-$i$ edges. We observe that the same packing argument in $\mathbb{R}^d$ gives an upper bound $\epsilon^{-d}$ on the number of level-$i$ edges incident to $C$.

They then introduce an amortized approach to bound the weight of all edges via credit. Assume that somehow each level-$(i-1)$ cluster $C$ gets an amount of credit proportional to $L_{i-1}$, say $\Omega(\epsilon^{-d+1}L_{i-1})$ for some constant $c(\epsilon)$. Recall the total weight of (at most $\epsilon^{-d}$) level-$i$ spanner edges incident to $C$ is $O(\epsilon^{-d}L_i) = O(\epsilon^{-d+1}L_{i-1})$. Then $C$'s credit is enough to pay for all of its incident level-$i$ edges.

Their idea to connect the credit with the weight of the MST can be roughly described as follows. They first allocate each cluster in $\mathcal{C}_0$ a credits roughly $c(\epsilon)\bar{w}$, so that the total allocated credit is $c(\epsilon)w(\mathrm{MST})$, where $c(\epsilon)$ is a parameter chosen later. Note here that credits are allocated once during the entire course of the analysis. Therefore, $c(\epsilon)$ is also the lightness upper bound. Recall that $L_0 = \Theta(\bar{w})$. That implies every level-0 cluster $C$ would have at least $\Omega(c(\epsilon)L_0)$ credits. As mentioned above, level-0 clusters pay for level-1 spanner edges, and they can afford the payment for an appropriate value of $c(\epsilon)$. (Level-0 spanner edges have total weight only $O(\epsilon^{-d})w(\mathrm{MST})$, so we can ignore them from the credit argument.). However, to pay for level-2 spanner edges, level-1 clusters need to have credit. Since all the credit is allocated to level-0 clusters, level-1 clusters need to take out partial credit of level-0 clusters. Therefore, level-0 clusters are not allowed to use all of their credit to pay for level-1 spanner edges; they can only use the remaining credit (after level-1 clusters took out some) to do so. The main challenge is to balance the amount of credit level-1 clusters take from level-0 clusters so that they (level-1 clusters) have enough credit to pay for level-2 spanner edges, with the amount of leftover credit of level-0 clusters so that they can pay for its incident level-1 edges. The same challenge arises when they pay for higher level spanner edges.

The key technical contribution of their analysis is to achieve the balance by inductively guaranteeing the following two invariants at all levels: nolistsep,noitemsep

(a) Each cluster $X \in \mathcal{C}_i$ has at least $\Omega(c(\epsilon)L_i)$ credits, which were taken from the clusters in $\mathcal{C}_{i-1}$.

(b) Each cluster $C \in \mathcal{C}_{i-1}$, after its credit was partly taken by the clusters in $\mathcal{C}_i$, has at least $\Omega(c(\epsilon)\epsilon^c L_{i-1})$ leftover credits for some constant $c$. That is, $C$'s remaining credit is at least $\epsilon^c$ fraction of its total credit.

We observe that the bound in invariant (b) in BLW approach can be made as big as $\Omega(c(\epsilon)\epsilon L_{i-1})$, but not bigger. Given this observation and assuming that the two invariants are guaranteed at all level, one can choose $c(\epsilon) = \Theta(\epsilon^{-d+2})$ so that the remaining credit of $C$ is enough to pay for its incident level-$i$ spanner edges. This implies $O(\epsilon^{-d+2})$ lightness upper bound. However, one technical problem is that BLW cannot always guarantee invariant (b) for all the clusters in $\mathcal{C}_{i-1}$; we will elaborate more on this problem in the next section. Our main goal in this paper is to remove $+2$ in the exponent of $\epsilon$ in the lightness bound.

### *B. The high level ideas of our analysis*

Our strategy can be divided into two steps. noitemsep

- (Step 1) We show that each cluster $C \in \mathcal{C}_{i-1}$ is incident to at most $O(\epsilon^{-d+1})$ level-$i$ spanner edges. This shaves an $\frac{1}{\epsilon}$ factor from the packing bound $O(\epsilon^{-d})$.
- (Step 2) We show that each cluster $C \in \mathcal{C}_{i-1}$, *in most cases*, has $\Omega(c(\epsilon)L_{i-1})$ leftover credits. This shaves another $\frac{1}{\epsilon}$ factor from the credit lower bound achieved by BLW approach.

We emphasize in most cases in Step 2 because it is not always possible to achieve this much leftover credit. Recall that in invariant (b) mentioned in Subsection VI-A, even achieving $\Omega(c(\epsilon)\epsilon L_{i-1})$ is highly non-trivial and not always possible. Note that before $C$'s credit is taken out by its parent in $\mathcal{C}_i$, it only has $\Omega(c(\epsilon)L_{i-1})$ credits. Thus, Step 2 is essentially equivalent to showing that $C$ can keep a constant factor of its credit to pay for its level-$i$ spanner edges.

Since each step we shave a $\frac{1}{\epsilon}$ factor, the final lightness upper bound we obtain is $O(\epsilon^{-d})$. We now sketch our ideas to implement both steps.

Let $C$ be a level-$(i-1)$ cluster and $k$ be the number of level-$i$ spanner edges incident to $C$. Let $C_1, \ldots, C_k$ be $C$'s neighbors. (A level-$(i-1)$ cluster $C_j$ is a neighbor of $C$ if there is a level-$i$ spanner edge connecting two points in $C_j$ and $C$, respectively.) For notational convenience, let $C_0 = C$. Since the edges connecting $C_0$ and $C_j$ for all $j \in [k]$ has length $\Theta(L_i)$, we can show that distances between any two clusters $C_p, C_q$, $p \neq q \in \{0, \ldots, k\}$, is $\Omega(\epsilon L_i)$. Thus, the standard packing argument implies that the number of clusters $C_j$, $0 \leq j \leq k$ is at most $O((\frac{L_i}{\epsilon L_i})^d) = O(\epsilon^{-d})$. To shave an $O(\epsilon^{-1})$ factor, we partition the plane into $O(\epsilon^{-d+1})$ cones around a point in $C$. Our insight is that, if the clustering hierarchy is constructed carefully, we can show that in each cone, there is at most one cluster $C_j$ in the cone that could have an edge to $C$; other edges, if any, would not be in the greedy spanner $S_{\mathrm{grd}}$. Thus, the number of neighbors of $C$ is equal to the number of cones, which is $O(\epsilon^{-d+1})$. This complies Step 1.

Before going into details of the implementation Step 2, let us sketch the idea in BLW approach to obtain the leftover credit bound $\Omega(c(\epsilon)\epsilon L_{i-1})$. Recall that a cluster $X$ in $\mathcal{C}_i$ have $|\texttt{child}(X)| = \Theta(\frac{1}{\epsilon})$. The idea is to show that *at least one* child of $X$ whose credit was not taken to maintain credit lower bound for $X$. Such a child has at least $\Omega(c(\epsilon)L_{-1})$ credits by invariant (a). By distributing it's credits to all other children of $X$, each would get $\Omega(c(\epsilon)L_{i-1}/\epsilon) = \Omega(c(\epsilon)\epsilon L_{i-1})$ credits. To realize Step 2, all we need to do is showing that the credit of at least $\Omega(\frac{1}{\epsilon})$ children of $X$ was not taken by $X$ and thus we can distribute their credit to all other children to obtain the desired leftover credit bound.

However, there are two technical subtleties of the credit argument that make the task challenging. To reveal the subtitles, let us examine two simple ideas and explain why they fail. The first simple idea is to allow $X$ to take the credit of a half of its children, say $\frac{1}{2\epsilon}$ children, assuming $|\texttt{child}(X)| = \frac{1}{\epsilon}$. Then $X$ would have at least $\frac{1}{2\epsilon}\Omega(c(\epsilon)L_{i-1}) = \Omega(c(\epsilon)L_i)$ credits. The first subtlety of the credit argument lies in the constant behind $\Omega$. Specifically, each cluster $X \in \mathcal{C}_i$ must have at least $gc(\epsilon)L_i$ credits for some universal constant $g$ for all $i$. Thus, the total credit of $\frac{1}{2\epsilon}$ clusters in $\texttt{child}(X)$ is only worth $\frac{c(\epsilon)L_{i-1}g}{2\epsilon} = \frac{c(\epsilon)gL_i}{2}$, which is less than the credit lower bound $gc(\epsilon)L_i$ for $X$. The second simple idea is to guarantee that $X$ has $\frac{2}{\epsilon}$ children instead of $\frac{1}{\epsilon}$ children. Then, the total credit of $\frac{1}{\epsilon}$ $X$'s children would beat the credit lower bound $gc(\epsilon)L_i$ and we have the credit of at least $\Omega(\frac{1}{\epsilon})$ children left in $X$ as desired. However, the second subtlety of the credit argument is that each cluster $X \in \mathcal{C}_i$ must have at least $c(\epsilon)\textsc{Dm}(X)$ credits, where $\textsc{Dm}(X)$ is the diameter of $X$. If $X$ has more children, its diameter increases and thus, it must take more credits from its children to maintain the credit lower bound $c(\epsilon)\textsc{Dm}(X)$. The precise credit invariant that we need to guarantee is each cluster $X \in \mathcal{C}_i$ has at least $c(\epsilon)\max(\textsc{Dm}(X), L_i/2)$ credits. Here we allow $X$ to have diameter less than $L_i/2$.

Due to the credit lower bound $c(\epsilon)\textsc{Dm}(X)$, even showing the existence of one child in $X$ whose credit is not taken by $X$ is a non-trivial task. A worst case example is when all children of $X$ are disjoint spheres that are horizontally aligned. Then $\textsc{Dm}(X) \geq \sum_{C\in\texttt{child}(X)} \textsc{Dm}(C)$. It means that to maintain the credit lower bound $c(\epsilon)\textsc{Dm}(X)$, it must take the credit of all children since each child $C$ of $X$ is only guaranteed to have $c(\epsilon)\textsc{Dm}(C)$ credits. The main insight of BLW is that in this worst case example, there would be no spanner edge of length $\Theta(L_i)$ connecting two different children of $X$. (There may be edges from $X$'s children to other level-$(i-1)$ clusters not in $X$, but this is not a problem because they can be paid for by the clusters not in $X$.) If there is at least one spanner edge connecting two children of $X$, then $X$'s children would not be horizontally aligned. This insight allow BLW to show that the credit of at least one child in $X$ is not taken by $x$. Recall that our goal is to show that the credit of at least $\Omega(\frac{1}{\epsilon})$ children of $X$ is not taken by $X$. This of course is not always possible as we have pointed out, even when the children of $X$ are not horizontally aligned. To resolve this issue, we bring in two new insights: noitemsep

- Insight (1) We can relax the leftover credit lower bound of each child $C \in X$ to $\Omega(c(\epsilon)\epsilon L_{i-1}\deg_{i-1}(C))$ instead of $\Omega(c(\epsilon)L_{i-1})$ as stated in Step 2.
- Insight (2) There is a relationship between the curvature of the path, say $P$, following the arrangement of $X$'s children and the number of level-$i$ spanner edges connecting them. Intuitively, if there are more edges connecting $X$'s children, $P$ is more deviated from the straight line. That means $\sum_{C\in\texttt{child}(X)} \textsc{Dm}(C) - \textsc{Dm}(X)$ is bigger. Quantitatively, if there are $t$ level-$i$ spanner edges connecting $X$'s children, we are able to show that $\sum_{C\in\texttt{child}(X)} \textsc{Dm}(C) - \textsc{Dm}(X) = \Omega(tL_{i-1})$, which allows us to show the leftover credit bound in insight(1).

Insight (1) leads to two interesting variants of our argument. The first variant is when $d \geq 3$, the worst case bound on the degree of $C \in \mathcal{C}_{i-1}$ is $\Omega(\epsilon^{-d+1}) = \Omega(\epsilon^{-2})$. We distinguish between two cases: if $\deg_{i-1}(C) = c\epsilon^{-1}$ for some sufficiently big constant $c$, we have a simple way to group $C$ and its neighbors (Stage 0 of our construction) so that $C$ has at least $\Omega(c(\epsilon)L_{i-1})$ leftover credits as desired. If $\deg_{i-1}(C) < c\epsilon^{-1}$, we allow $C$ to only has $\Omega(c(\epsilon)\epsilon L_{i-1})$ leftover credits. This is because $C$ only has $O(\epsilon^{-1})$ incident level-$i$ spanner edges of length $L_i$ (for any $d \geq 3$), so choosing $c(\epsilon) = \Omega(\epsilon^{-3})$ suffices. This results in the lightness bound $O(\max(\epsilon^{-3}, \epsilon^{-d})) = O(\epsilon^{-d})$ when $d \geq 3$. We want to emphasize that insight (2) is even not needed when $d \geq 3$. However, the same reasoning when is applied to $\mathbb{R}^2$, the lightness bound we can obtain is $O(\max(\epsilon^{-3}, \epsilon^{-d})) = O(\epsilon^{-3})$, which is not optimal. Arguably, $\mathbb{R}^2$ is also the most important case.

We now restrict our discussion to the case $d = 2$. The goal is somehow to make use of insight (2) to further reduce the lightness bound to $O(\epsilon^{-2})$. The road to get to the point where the spanner has some structure that we can apply insight (2) is quite long. It is because we need to deal with many different types of structural connections between the MST and clusters. However, most of the structures we deal with along the way give us leftover credit lower bound $\Omega(\epsilon L_{i-1})$. There are three stages in our argument and in each stage, we gain more structural understanding of the greedy spanner. (Stage-0 is only used to handle the case $d \geq 3$.)

Before going into deeper details, to avoid confusion, we refer to level-$(i-1)$ clusters as *$\epsilon$-clusters* as BLW did in [8], and we drop $i$ in level-$i$ clusters. Let $\mathcal{T}$ be the tree whose nodes are $\epsilon$-clusters and edges are MST edges connecting $\epsilon$-clusters. By a simple trick, we can guarantee that edges of $\mathcal{T}$ have roughly equal length and are much shorter that the diameter of $\epsilon$-clusters. In this way, we can think of two $\epsilon$-clusters connected by an edge as two circles that are almost touch each other, though in our detailed argument, we do not try to associate any geometric shape with a cluster. Also, $\mathcal{T}$ have weight on both nodes and edges, where the weight on each node is the diameter of the corresponding $\epsilon$-cluster. Thus, the length of each path in $\mathcal{T}$ is the total weight all the nodes and edges on the path. We say a node of $\mathcal{T}$ *branching* if it has degree at least three in $\mathcal{T}$.

*Stage 1: constructing zoom-outs:* In this stage, we try to break $\mathcal{T}$ into subtrees that we call *zoom-outs* (see Figure 6(a)). Zoom-outs have three properties: noitemsep

(i) Every branching node belongs to some zoom-out.
(ii) Each zoom-out has a center node, which is branching, where there are at least three branches from that node of roughly equal length. A branch is simply a path of $\mathcal{T}$.
(iii) Each zoom-out has diameter $O(L_i)$.

Zoom-outs are the basis of our cluster construction in a sense that, once a node of a zoom-out becomes a child of a cluster $X$, then all nodes of the zoom-out belong to $\texttt{child}(X)$. However, one should keep in mind that not every node of $\mathcal{T}$ belongs to a zoom-out. Property (i) only constrains branching nodes. For example, nodes on a long path of $\mathcal{T}$ may not belong to any zoom-out.

The most important property of zoom-out in our credit argument is (ii). Suppose later that a zoom-out $\widehat{\chi}$ is contained in a cluster $X$ (in $\mathcal{C}_i$). For any diameter path $D$ of $X$, the subpath of $D$ going through $\widehat{\chi}$ would not be longer than the total length of the two longest branches of $\widehat{\chi}$. Thus, the total credit of the nodes (or $\epsilon$-clusters) on the two branches is enough to account for this subpath. By that, we mean the total credit is at least $c(\epsilon)$ times the length of the subpath of $D$ going through $\widehat{\chi}$. Thus, the credit of the nodes on the third branch is leftover, and since all three branches have length in a constant time of each other (hence the number of nodes intuitively are roughly equal), each gets distributed by at least $\Omega(c(\epsilon)L_{i-1})$ credits (see Figure 6(b)). (Nodes that are not in the three branches can keep their credit as leftover.) Therefore, the nodes in zoom-outs can pay for their incident evel-$i$ spanner edges and we can safely discard the paid spanner edges from the construction in later stages.

Let $\widehat{\mathcal{T}}$ be the tree obtained from $\mathcal{T}$ by contracting each zoom-out into a single node. Another interesting structural property of zoom-out that follows from our construction is: noitemsep

(iv) Every braching node of $\widehat{\mathcal{T}}$ corresponds to a zoom-out of diameter $\Theta(L_i)$.

This is a very crucial property for the next stage construction, that we will discuss in more details later.

Constructing zoom-out is essentially a clustering problem on tree with nodes and edges weighted. The difficulty of the construction is in guaranteeing both property (i) and (ii), and hence, a simple greedy algorithm such as iteratively removing a branching node out of a tree does not work. We believe that this clustering problem maybe of independent interest.

*Stage 2: constructing Type-I,II,III and tiny clusters:* First, for every zoom-out constructed in Stage 1 of diameter $\Omega(L_i)$, we turn it into a level-$i$ cluster, that we call a *Type-I cluster*. Property (iii) of zoom-outs implies that the cluster has diameter $\Theta(L_i)$. By property (ii) of zoom-outs, and as discussed above, a Type-I cluster $X$ can maintain the credit lower bound $c(\epsilon)\textsc{Dm}(X)$ while each of $X$'s children has at least $\Omega(c(\epsilon)L_{i-1})$ credits. So Type-I clusters are good.

We then remove all the nodes corresponding to Type-I clusters from $\widehat{\mathcal{T}}$. By property (iv), any branching node of $\widehat{\mathcal{T}}$ corresponds to a zoom-out of diameter $\Theta(L_i)$ and thus is removed. What remaining is a set of paths, that we call a *linear forest* $\widehat{\mathcal{F}}$. We then study structures of the paths in $\widehat{\mathcal{F}}$. If there are two subpaths of $\widehat{\mathcal{F}}$ that almost lie along two opposite (long) sides of a fat triangle of side length $\Theta(L_i)$ (see Figure 6(c)), then we form a new cluster $X$ from the subpaths, that we call *Type-II clusters*. The intuition is that the diameter of the fat triangle is a smaller than of the total length of its two long sides by at least a constant factor. Thus, the credit of at least $\Omega(\frac{1}{\epsilon})$ $\epsilon$-clusters in $X$ is leftover after $X$ takes its children's credit to maintain credit low

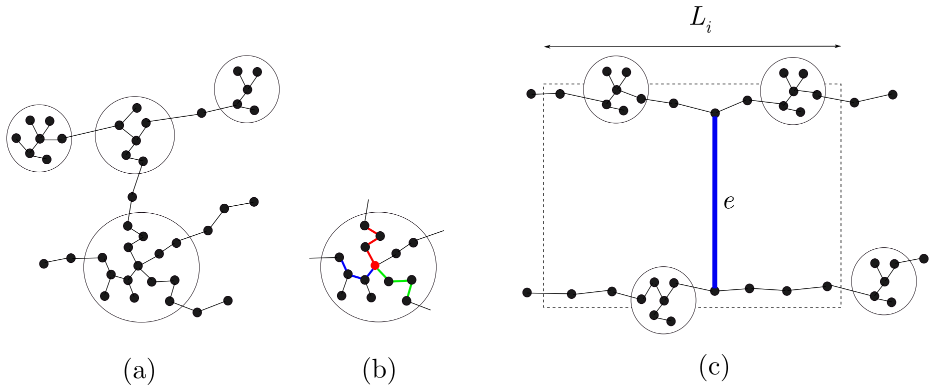


Figure 6. Thin black edges are MST edges while thick blue edges are level-$i$ spanner edges. (a) Zoom-outs, in circles, are constructed from the tree $\mathcal{T}$. Nodes are not in zoom-outs have degree at most 2. (b) Each zoom-out has at least three branches, highlighted by $\{\texttt{red}, \texttt{green}, \texttt{blue}\}$ colors from the center node, highlighted red. The three branches have roughly the same length, up to a constant factor. (c) A typical Type-II cluster formed in Stage 2. Circles are zoom-outs.

bound $c(\epsilon)\textsc{Dm}(X)$. That implies each child of $X$ has at least $\Omega(c(\epsilon)L_{i-1})$ leftover credits.

It should be noted that in our formal argument, we do not explicitly quantify how close a path of $\widehat{\mathcal{F}}$ is to a boundary of a fat triangle. We instead look at how paths of $\widehat{\mathcal{F}}$ connected by spanner edges of length $\Theta(L_i)$. This is similar to what BLW did in doubling metrics, but we bring more geometric intuition into our context. Our key insight is that if there is a spanner edge $e$ that connect two zoom-outs or $\epsilon$-clusters *that are far from each other and are not close to the endpoints of the path(s) of $\widehat{\mathcal{F}}$ containing them*, then the two subpaths of $\widehat{\mathcal{F}}$ in the neighborhoods around $e$'s endpoints intuitively look very much like they are along two opposite sides of a fat triangle (of side length $\Theta(L_i)$). But we do not have this nice distribution of zoom-outs around $e$'s endpoints if either of the following cases happens. noitemsep

- Issue (i) At least one of $e$'s endpoints is close to the endpoints of the path containing it.
- Issue (ii) Both endpoints of $e$ are close to each other.

We handle issue (i) in this stage and issue (ii) in the third stage. Figure 7(a) illustrates issue (i) where one endpoint of $e$ is close to the endpoint of a path in $\widehat{\mathcal{F}}$. We will elaborate more on issue (ii) when we sketch the next stage.

To deal with issue (i), for each path $\widehat{\mathcal{P}}$ in $\widehat{\mathcal{F}}$ (after constructing all possible Type-II clusters), we consider its two affices of length roughly $\Theta(L_i)$, say $\widehat{\mathcal{A}}_1, \widehat{\mathcal{A}}_2$ and its position in the tree $\widehat{\mathcal{T}}$. Recall that the branching nodes of $\widehat{\mathcal{T}}$ are removed before the construction of $\widehat{\mathcal{A}}_1$ and $\widehat{\mathcal{A}}_2$. Thus, $\widehat{\mathcal{A}}_1$ and $\widehat{\mathcal{A}}_2$ do not contain any branching node of $\widehat{\mathcal{T}}$. If none of them contain a leaf node of $\widehat{\mathcal{T}}$, then we merge them with nearby Type-I or Type-II clusters.

This is possible because Type-I and Type-II clusters already have enough leftover credits and by adding a subpath to them, we can take all the credit of the added subpath to account for the blow up in diameter, while we still can redistribute leftover credit to the $\epsilon$-clusters of the added subpaths so that each gets at least $\Omega(c(\epsilon)L_{i-1})$ leftover credits. If $\widehat{\mathcal{A}}_1$, say, contains a leaf node of $\mathcal{T}$, then there are no nearby Type-I or Type-II clusters to merge. We form a *Type-III* cluster from $\widehat{\mathcal{A}}_1$. For Type-III clusters, we even cannot guarantee that the credit of a single $\epsilon$-cluster is leftover; a Type-III cluster $X$ may need to take all the credit of its children to maintain credit lower bound $c(\epsilon)\textsc{Dm}(X)$. Thus, children of $X$ do not have leftover credit to pay for its incident spanner edges.

Our key insight to resolve this difficulty is we do not need to pay for level-$i$ spanner edges incident to children of $X$ immediately. Instead, we deposit the weight of these edges to a *debt account* of $X$ and we will pay back the debt in higher level clusterings. If $X$ become a child of a level-$(i+1)$ cluster, say $Y$, then $Y$ must take care of the debt of $X$. If it turns out that $Y$ is also a Type-III cluster and can't pay for its incident spanner edges, then we deposit the debt of $X$ and the total weight of the spanner edges to the debt account of $Y$. In this way, we can ensure that all debts are finally paid for at some point. Our main observation is that the total amount of debt of an $\epsilon$-cluster, accumulated through all levels up to $i-1$, is small. This is due to the fact that $\epsilon$-clusters that have debts are highly structural: only the ones corresponding to leaves of $\mathcal{T}$ have debts. The reason is quite intuitive: a Type-III cluster $X$ contains a leaf of $\mathcal{T}$ and if we look at the cluster tree at higher level, $X$ is a leaf of the tree.

We inductively show that the debt amount of an $\epsilon$-cluster is $O(\epsilon^{-1}L_i)$. Therefore, the total amount of debt of an $\epsilon$-cluster is just a constant factor of the worst case bound on the weight of the level-$i$ spanner edges incident to the $\epsilon$-cluster. All in all, each $\epsilon$-cluster must pay for: (i) its incident level-$i$ spanner edges and (ii) its debts (if any). Our construction of Type-III clusters guarantee that only leaves of $\mathcal{T}$ have non-zero debt.

Finally, what remaining after constructing all possible Type-I, II and III clusters is a linear forest (of zoom-outs and $\epsilon$-clusters) where none of them contains leaves of $\widehat{\mathcal{T}}$.

We break each path in the forest into subpaths of length $\Theta(\zeta L_i)$, for some constant $\zeta \ll 1$, that we call *tiny clusters*. Tiny clusters are the basis of our construction in the third stage.

*Stage 3: constructing Type IV and V clusters:* In this stage, we handle issue (ii) mentioned in Stage 2. This issue arises when there is an edge connecting two zoom-outs that are closed together. In this stage, we have enough structure to apply insight (2) mentioned before.

Note that by the end of Stage 2, all zoom-outs and $\epsilon$-clusters are grouped into tiny clusters. Let $\widehat{C}$ be a tiny cluster. Recall that $\widehat{C}$ has diameter $\Theta(\zeta L_i)$, so it is qualified to be a level-$i$ cluster. If $\widehat{C}$ can take all the credit of its children, then it can maintain credit lower bound $c(\epsilon)\textsc{Dm}(\widehat{C})$. However, it may need to pay for the level-$i$ spanner edges incident to its children. A simple observation is that if every level-$i$ spanner edge incident to a child of $\widehat{C}$ is also incident to a child of a cluster constructed in Stage 2, the edge is already paid for by that Stage 2 cluster. (We still use the term "paid for by a (child of a) Stage 2 cluster" even though it is Type-III; Type-III clusters do not really pay for their incident edges.) Thus, we can safely form a *Type-IV cluster* from $\widehat{C}$ without the need for leftover credit.

However, we may have an edge between two tiny clusters, but we know that endpoints of the edge must be close together (issue (ii) described in Stage 2). The last step of Stage 3 is to group tiny clusters together so that we have leftover credit to pay for their incident edges. In Stage 2, we can guarantee that every cluster has at least $\Omega(\frac{1}{\epsilon})$ $\epsilon$-clusters whose credit is not needed to maintain the credit lower bound, except for Type-III clusters which are allowed to have debts. This is no longer true in this case. Consider a set of zoom-outs and $\epsilon$-clusters that distribute along the boundary of a (very big) circle so that locally any arc, say $A$, of length $\Theta(L_i)$ has small curvature (Figure 7(b)). The small curvature is enough to force the greedy algorithm to add edges between zoom-outs and $\epsilon$-clusters on $A$ to the spanner. So no matter how we group tiny clusters, the new cluster would not have the leftover credit of $\Omega(\frac{1}{\epsilon})$ $\epsilon$-clusters. However, there are not many such edges, depending on the curvature $A$. Intuitively, there is a monotone relationship between the curvature of $A$ and the number of edges between tiny clusters along $A$.

To make this intuition more formal, let $C$ be a $\epsilon$-cluster in a tiny cluster. Let $\widehat{\mathcal{P}}$ be the path containing $C$. Suppose that $C$ has $t$ incident level-$i$ spanner edges connect $C$ to other $\epsilon$-clusters in tiny clusters. Since other endpoints of the edges incident to $C$ are close to $C$, a subpath of $\widehat{\mathcal{P}}$, say $\widehat{\mathcal{Q}}$, of length $\Theta(L_i)$ will contain all the endpoints. We then make $\widehat{\mathcal{Q}}$ a *Type-V cluster*. We are able to show roughly that the different between the length of $\widehat{\mathcal{Q}}$ and the (Euclidean) distance between any two furthest points in $\widehat{\mathcal{Q}}$ is $\Omega(tL_{i-1})$. Note that each $\epsilon$-cluster has diameter $O(L_{i-1})$. That implies the credit of at least $\Omega(t)$ $\epsilon$-clusters are not needed to maintain the credit lower bound of their parent. We distribute the leftover credit to all $\epsilon$-clusters in $\widehat{\mathcal{Q}}$, each gets at least $\Omega(\frac{c(\epsilon)L_{i-1}t}{\epsilon}) = \Omega(c(\epsilon)\epsilon t L_{i-1})$ credits since $\widehat{\mathcal{Q}}$ has roughly $\Theta(\frac{1}{\epsilon})$ $\epsilon$-clusters.

Recall our insight (1) said that it is enough guarantee that each $\epsilon$-cluster $C'$ have $\Omega(\epsilon c(\epsilon)\deg_{i-1}(C')L_{i-1})$ credits. Also, it suffice to assume that every edge of $C'$ is incident to tiny clusters only since otherwise, it would be paid for by the cluster containing the other end point. But in the discussion in the previous paragraph, $C'$ is distributed $\Omega(c(\epsilon)\epsilon t L_{i-1})$ credits, where $t$ is the degree of $C$. To resolve this problem, all we need to do is constructing $\widehat{\mathcal{Q}}$ in a way that $C$ has largest degree among all $\epsilon$-clusters in $\widehat{\mathcal{Q}}$.

## Acknowledgment

Hung Le is supported by a PIMS postdoctoral fellowship and an NSERC grant. Shay Solomon is partially supported by the Israel Science Foundation grant No.1991/19 and by Len Blavatnik and the Blavatnik Family foundation. Shay Solomon is grateful to Michael Elkin, Ofer Neiman and Michiel Smid for fruitful discussions.

## References

[1] P. K. Agarwal, Y. Wang, and P. Yin. Lower bound for sparse Euclidean spanners. In *Proc. of 16th SODA*, pages 670–671, 2005.

[2] I. Althöfer, G. Das, D. Dobkin, D. Joseph, and J. Soares. On sparse spanners of weighted graphs. *Discrete Computational Geometry*, 9(1):81–100, 1993.

[3] S. Arya, G. Das, D. M. Mount, J. S. Salowe, and M. Smid. Euclidean spanners: Short, thin, and lanky. In *Proceedings of the Twenty-seventh Annual ACM Symposium on Theory of Computing*, STOC '95, pages 489–498, 1995.

[4] S. Arya and M. H. M. Smid. Efficient construction of a bounded degree spanner with low weight. *Algorithmica*, 17(1):33–54, 1997.

[5] L. Barba, P. Bose, M. Damian, R. Fagerberg, W. L. Keng, J. O'Rourke, A. van Renssen, P. Taslakian, S. Verdonschot, and G. Xia. New and improved spanning ratios for yao graphs. In *30th Annual Symposium on Computational Geometry, SOCG'14, Kyoto, Japan, June 08 - 11, 2014*, page 30, 2014.

[6] M. Bauer and M. Damian. An infinite class of sparse-yao spanners. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 184–196, 2013.

[7] G. Borradaile, H. Le, and C. Wulff-Nilsen. Minor-free graphs have light spanners. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science*, FOCS '17, pages 767–778, 2017.

[8] G. Borradaile, H. Le, and C. Wulff-Nilsen. greedy spanners are optimal in doubling metrics. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, pages 2371–2379, 2019.

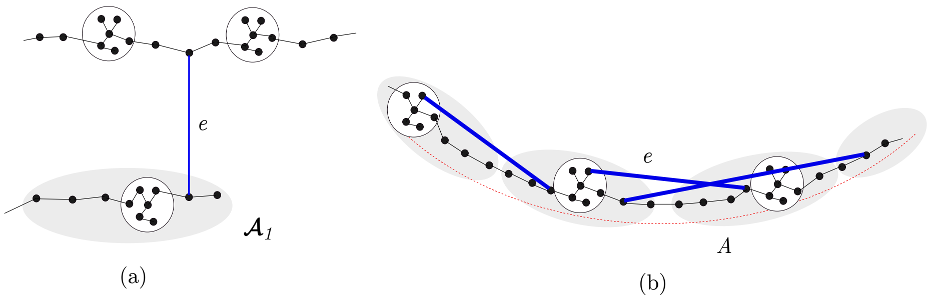


Figure 7. Thin black edges are MST edges while thick blue edges are level-$i$ spanner edges. (a) An edge $e$ has one endpoint close to an endpoint of a path in $\widehat{\mathcal{F}}$, which is a leaf node in $\mathcal{T}$. The affix, say $\mathcal{A}_1$, containing the endpoint of $e$ of length $\Theta(L_i)$ of the path is a Type-III cluster, which is shaded gray. (b) An arc $A$, highlighted by a dashed red curve, of a big circle where zoom-outs and $\epsilon$-clusters lie on. The curvature of $A$ is big enough to force the greedy algorithm to take some edges, highlighted by thick blue lines, to the spanner. The gray-shaded regions are tiny clusters.

[9] P. Bose, J. De Carufel, D. Hill, and M. H. M. Smid. On the spanning and routing ratio of theta-four. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2361–2370, 2019.

[10] P. Bose, M. Damian, K. Douïeb, J. O'Rourke, B. Seamone, M. H. M. Smid, and S. Wuhrer. $\pi/2$-angle yao graphs are spanners. *Int. J. Comput. Geometry Appl.*, 22(1):61–82, 2012.

[11] P. Bose and A. van Renssen. Upper bounds on the spanning ratio of constrained theta-graphs. In *LATIN 2014: Theoretical Informatics - 11th Latin American Symposium, Montevideo, Uruguay, March 31 - April 4, 2014. Proceedings*, pages 108–119, 2014.

[12] H. T.-H. Chan and A. Gupta. Small hop-diameter sparse spanners for doubling metrics. In *Proc. of 17th SODA*, pages 70–78, 2006.

[13] T.-H. Hubert Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. On hierarchical routing in doubling metrics. *ACM Trans. Algorithms*, 12(4):55:1–55:22, 2016. Preliminary version appeared in SODA 2005.

[14] B. Chandra, G. Das, G. Narasimhan, and J. Soares. New sparseness results on graph spanners. In *Proceedings of the Eighth Annual Symposium on Computational Geometry*, 1992.

[15] S. Chechik and C. Wulff-Nilsen. Near-optimal light spanners. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'16, pages 883–892, 2016.

[16] L. P. Chew. There is a planar graph almost as good as the complete graph. In *Proceedings of the Second Annual Symposium on Computational Geometry*, SCG '86, pages 169–177, 1986.

[17] L. P. Chew. There are planar graphs almost as good as the complete graph. *Journal of Computer and System Sciences*, 39(2):205 – 219, 1989.

[18] K. Clarkson. Approximation algorithms for shortest path motion planning. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 56–65, 1987.

[19] G. Das, P. Heffernan, and G. Narasimhan. Optimally sparse spanners in 3-dimensional euclidean space. In *Proceedings of the 9th Annual Symposium on Computational Geometry*, SCG '93, pages 53–62, 1993.

[20] G. Das and G. Narasimhan. A fast algorithm for constructing sparse Euclidean spanners. In *Proc. of 10th SOCG*, pages 132–139, 1994.

[21] G. Das, G. Narasimhan, and J. Salowe. A new way to weigh malnourished euclidean graphs. In *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '95, pages 215–222, 1995.

[22] Y. Dinitz, M. Elkin, and S. Solomon. Shallow-low-light trees, and tight lower bounds for Euclidean spanners. In *Proc. of 49th FOCS*, pages 519–528, 2008.

[23] M. Elkin and S. Solomon. Steiner shallow-light trees are exponentially lighter than spanning ones. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science*, number FOCS '11, pages 373–382, 2011.

[24] M. Elkin and S. Solomon. Optimal euclidean spanners: Really short, thin, and lanky. *J. ACM*, 62(5):35:1–35:45, 2015.

[25] A. Filtser and S. Solomon. The greedy spanner is existentially optimal. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*, PODC '16, pages 9–17, 2016.

[26] J. Gao, L. J. Guibas, and A. Nguyen. Deformable spanners and applications. In *Proc. of 20th SoCG*, pages 190–199, 2004.

[27] E. Gilbert and H. Pollak. Steiner minimal trees. *SIAM Journal on Applied Mathematics*, 16(1):1–29, 1968.

[28] L. Gottlieb, A. Kontorovich, and R. Krauthgamer. Efficient regression in metric spaces via approximate lipschitz extension. *IEEE Trans. Information Theory*, 63(8):4838–4849, 2017.

[29] L. Gottlieb and L. Roditty. An optimal dynamic spanner for doubling metric spaces. In *Proc. of 16th ESA*, pages 478–489, 2008. Another version of this paper is available via http://cs.nyu.edu/~adi/spanner2.pdf.

[30] L. A. Gottlieb. A light metric spanner. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 759–772, 2015.

[31] J. Gudmundsson, C. Levcopoulos, and G. Narasimhan. Fast greedy algorithms for constructing sparse geometric spanners. *SIAM J. Comput.*, 31(5):1479–1500, 2002.

[32] J. Gudmundsson, C. Levcopoulos, G. Narasimhan, and M. H. M. Smid. Approximate distance oracles for geometric graphs. In *Proc. of 13th SODA*, pages 828–837, 2002.

[33] J. Gudmundsson, C. Levcopoulos, G. Narasimhan, and M. H. M. Smid. Approximate distance oracles revisited. In *Proc. of 13th ISAAC*, pages 357–368, 2002.

[34] J. Gudmundsson, C. Levcopoulos, G. Narasimhan, and M. H. M. Smid. Approximate distance oracles for geometric spanners. *ACM Transactions on Algorithms*, 4(1), 2008.

[35] J. Gudmundsson, G. Narasimhan, and M. H. M. Smid. Fast pruning of geometric spanners. In *Proc. of 22nd STACS*, pages 508–520, 2005.

[36] Y. Hassin and D. Peleg. Sparse communication networks and efficient routing in the plane. In *Proc. of 19th PODC*, pages 41–50, 2000.

[37] Y. Jin, J. Li, and W. Zhan. Odd yao-yao graphs are not spanners. In *34th International Symposium on Computational Geometry, SoCG 2018, June 11-14, 2018, Budapest, Hungary*, pages 49:1–49:15, 2018.

[38] J. M. Keil. Approximating the complete euclidean graph. In *Proceedings of the first Scandinavian Workshop on Algorithm Theory*, SWAT '88, pages 208–213, 1988.

[39] J. M. Keil and C. A. Gutwin. Classes of graphs which approximate the complete Euclidean graph. *Discrete and Computational Geometry*, 7(1):13–28, 1992.

[40] J. Li and W. Zhan. Almost all even yao-yao graphs are spanners. In *24th Annual European Symposium on Algorithms, ESA 2016, August 22-24, 2016, Aarhus, Denmark*, pages 62:1–62:13, 2016.

[41] Y. Mansour and D. Peleg. An approximation algorithm for min-cost network design. *DIMACS Series in Discr. Math and TCS*, 53:97–106, 2000.

[42] G. Narasimhan and M. Smid. *Geometric Spanner Networks*. Cambridge University Press, 2007.

[43] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.

[44] S. B. Rao and W. D. Smith. Approximating geometrical graphs via "spanners" and "banyans". In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, STOC '98, pages 540–550, 1998. Full version at http://graphics.stanford.edu/courses/cs468-06-winter/Papers/rs-tsp.pdf.

[45] J. Ruppert and R. Seidel. Approximating the $d$-dimensional complete Euclidean graph. In *Proceedings of the 3rd Canadian Conference on Computational Geometry*, CCCG '91, page 207–210, 1991.

[46] J. S. Salowe. On euclidean spanner graphs with small degree. In *Proceedings of the Eighth Annual Symposium on Computational Geometry, Berlin, Germany, June 10-12, 1992*, pages 186–191, 1992.

[47] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *The Bell System Technical Journal*, 38(3):611–656, 1959.

[48] S. Solomon. Euclidean steiner shallow-light trees. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SoCG '14, pages 454:454–454:463, 2014.

[49] S. Solomon. From hierarchical partitions to hierarchical covers: optimal fault-tolerant spanners for doubling metrics. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 363–372, 2014.

[50] A. D. Wyner. Capabilities of bounded discrepancy decoding. *The Bell System Technical Journal*, 44(6):1061–1122, 1965.

[51] A. C. Yao. On constructing minimum spanning trees in $k$-dimensional spaces and related problems. *SIAM Journal on Computing*, 11(4):721–736, 1982.