

# Efficient Yao Graph Construction

Daniel Funke ✉

Karlsruhe Institute of Technology, Germany

Peter Sanders ✉

Karlsruhe Institute of Technology, Germany

---

## Abstract

Yao graphs are geometric spanners that connect each point of a given point set to its nearest neighbor in each of  $k$  cones drawn around it. Yao graphs were introduced to construct minimum spanning trees in  $d$  dimensional spaces. Moreover, they are used for instance in topology control in wireless networks. An optimal  $\mathcal{O}(n \log n)$ -time algorithm to construct Yao graphs for a given point set has been proposed in the literature but – to the best of our knowledge – never been implemented. Instead, algorithms with a quadratic complexity are used in popular packages to construct these graphs. In this paper we present the first implementation of the optimal Yao graph algorithm. We engineer the data structures required to achieve the  $\mathcal{O}(n \log n)$  time bound and detail algorithmic adaptations necessary to take the original algorithm from theory to practice. We propose a priority queue data structure that separates static and dynamic events and might be of independent interest for other sweep-line algorithms. Additionally, we propose a new Yao graph algorithm based on a uniform grid data structure that performs well for medium-sized inputs. We evaluate our implementations on a wide variety of synthetic and real-world datasets and show that our implementation outperforms current publicly available implementations by at least an order of magnitude.

**2012 ACM Subject Classification** Theory of computation → Sparsification and spanners

**Keywords and phrases** computational geometry, geometric spanners, Yao graphs, sweep-line algorithms, optimal algorithms

**Digital Object Identifier** 10.4230/LIPIcs.SEA.2023.20

**Related Version** *Technical report*: [arXiv:2303.07858](https://arxiv.org/abs/2303.07858) [11]

**Supplementary Material** *Software*: <https://github.com/dfunke/YaoGraph>  
archived at `swh:1:dir:c9682e5265d00eef03757c009b3aaefdeaa5288`

## 1 Introduction

Yao graphs are geometric spanners that connect each point of a given point set to its nearest neighbor in each of  $k$  cones, refer to Figure 1 for an example. A  $t$ -spanner is a weighted graph, where for any pair of vertices there exists a  $t$ -path between them, which is a path with weight at most  $t$  times their spatial distance. The parameter  $t$  is known as the *stretch factor* of the spanner. Upper bounds on the stretch factor of Yao graphs have been the subject of extensive research. While the stretch factor of Yao graphs with  $k \leq 3$  cones is proved to be unbounded, bounds have been established for all graphs with  $k \geq 4$  cones [3]. Whereas for  $k \geq 7$  cones the stretch factor is proved to be bounded by the general formula  $(1 + \sqrt{2 - 2 \cos(2\pi/k)}) / (2 \cos(2\pi/k) - 1)$ , bounds on Yao graphs with 4 to 6 cones require complex individual arguments [3, 8].

Yao introduced this kind of graphs to construct minimum spanning trees in  $d$ -dimensional space [17]. Moreover, they are used for instance in topology control in wireless networks [14, 18]. Chang et al. [6] present an optimal algorithm to construct these graphs in  $\mathcal{O}(n \log n)$  time. Due to the intricate nature of their algorithm and the reliance on expensive geometric constructions, to the best of our knowledge, there is no implementation of their algorithm available. Instead, an algorithm with an inferior  $\mathcal{O}(n^2)$  time bound is used in the cone-based spanners package of the popular CGAL library [15].



© Daniel Funke and Peter Sanders;

licensed under Creative Commons License CC-BY 4.0

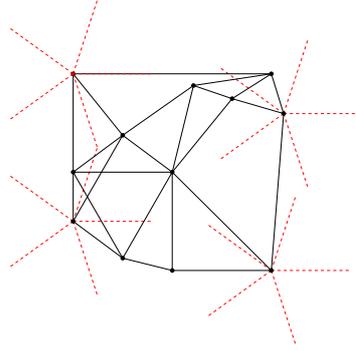
21st International Symposium on Experimental Algorithms (SEA 2023).

Editor: Loukas Georgiadis; Article No. 20; pp. 20:1–20:20

Leibniz International Proceedings in Informatics



LIPIC Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** Yao graph for ten points and  $k = 5$  cones. The five cones are illustrated as red dashed lines around four example points.

**Contribution.** In this paper we present the first publicly available implementation of Chang et al.’s optimal algorithm for Yao graph construction. We take their algorithm from theory to practice by engineering the data structures required to achieve the  $\mathcal{O}(n \log n)$  time bound and provide detailed descriptions of all operations of the algorithm that are missing in the original paper, such as input point ordering, handling of composite boundaries and enclosing region search. In our event queue, we separate static (input point) events and dynamic (intersection point) events. This greatly improves the efficiency of priority queue operations and might be a useful technique for other sweepline algorithms. We test our algorithm on a wide range of synthetic and real-world datasets. We show that, despite the intricate nature of the algorithm and the use of expensive geometric constructions, our implementation achieves a speedup of an order of magnitude over other currently available implementations. Additionally, we develop a new Yao graph algorithm based on a uniform grid data structure that is simple to implement, easy to parallelize, and performs well for medium-sized inputs.

**Outline.** In Section 2 we review related work on the construction of Yao graphs. Section 3 presents the optimal algorithm of Chang et al. and algorithmic adaptations necessary for its implementation. Further implementation details such as data structures and geometric operations are described in Section 4. We evaluate our implementation and compare against its competitors in Section 5. Section 6 summarizes our paper and presents an outlook on future work.

**Definitions.** Given a set  $\mathbf{P}$  of points in two-dimensional Euclidean space and an integer parameter  $k > 1$ , the Yao graph  $G_k = (\mathbf{P}, \mathbf{E})$  is a directed graph, connecting every point  $p \in \mathbf{P}$  with its nearest neighbor in each of  $k$  cones [17]. Every cone  $\mathcal{C}_i = (\theta_L, \theta_R)$ ,  $0 \leq i < k$ , is defined by its two limiting rays with angles  $\theta_L = \frac{2i\pi}{k}$  and  $\theta_R = \frac{2(i+1)\pi}{k}$ . We denote the cone  $\mathcal{C}_i$  with apex at point  $p$  as  $\mathcal{C}_i^p$ . We furthermore define, that the *left* – or counterclockwise – boundary ray with angle  $\theta_L$  belongs to a cone  $\mathcal{C}$ , whereas the *right* one does not, i. e. for a given point  $p \in \mathbf{P}$  and cone  $\mathcal{C}_i^p$  we define the set of points  $\mathbf{P} \cap \mathcal{C}_i^p := \{q \in \mathbf{P} : \angle(p, q) \in [\theta_L, \theta_R)\}$ , with  $\angle(p, q)$  denoting the angle between  $p$  and  $q$ . Then the edge set  $\mathbf{E}$  of the Yao graph  $G_k = (\mathbf{P}, \mathbf{E})$  can be formally defined as  $\mathbf{E} := \{(p, q) : \forall i \in [0, k), \forall p \in \mathbf{P}, q = \arg \min_{v \in \mathbf{P} \cap \mathcal{C}_i^p} (d(p, v))\}$ , with  $d(\cdot, \cdot)$  denoting the Euclidean distance function.

## 2 Related Work

Yao presents an  $\mathcal{O}(n^{5/3} \log n)$ -time algorithm to compute a solution to the *Eight-Neighbors Problem* – a Yao graph with  $k = 8$ . It is based on a tessellation of the Euclidean space into cells. For a given point and cone, each cell of the tessellation is characterized whether it can contain nearest neighbor candidates to reduce the number of necessary distance computations. The problem is solved optimally by Chang et al., who present a  $\mathcal{O}(n \log n)$ -time algorithm for constructing the Yao graph of a given point set and parameter  $k$  [6]. Their algorithm follows the same structure as Fortune’s algorithm for constructing the Voronoi diagram of a point set [10], using the sweepline technique originally introduced by Bentley and Ottmann for computing line-segment intersections [4]. However, even though there are many implementations of Fortune’s algorithm available, there is no implementation of Chang et al.’s Yao graph algorithm that we are aware of. Instead, for instance the CGAL library’s cone-based spanners package implements a less efficient  $\mathcal{O}(n^2)$ -time algorithm [15]. Their algorithm is an adaption of a sweepline algorithm for constructing  $\Theta$ -graphs [13].  $\Theta$ -graphs are defined similarly to Yao graphs, except that the nearest neighbor in each cone is not defined by Euclidean distance but by the projection distance onto the cone’s internal angle bisector. This allows for a  $\mathcal{O}(n \log n)$ -time sweepline algorithm, that uses a balanced search tree as sweepline data structure to answer one-dimensional range queries [13]. For Yao graphs, such a reduction in dimensionality is not possible, thus, CGAL’s algorithm employs linear search within the sweepline data structure to find the nearest neighbor, leading to the  $\mathcal{O}(n^2)$  time bound. However, CGAL’s algorithm is much simpler to implement than the optimal algorithm proposed by [6] and does not require geometric constructions, just predicates. Table 1 in Section 4 provides an overview of the required geometric operations by both algorithms.

## 3 Algorithm

In their 1990 paper, Chang et al. present an  $\mathcal{O}(n \log n)$ -time sweepline algorithm to compute the *oriented Voronoi diagram* (OVD) of a point set. Through a small modification, their algorithm can compute the *geographic neighborhood graph* – or Yao graph – of a point set within the same, optimal, bound [6, Theorem 3.2, Theorem 4.1].<sup>1</sup> To construct the Yao graph  $G_k = (\mathbf{P}, \mathbf{E})$  with  $k$  cones for point set  $\mathbf{P}$ ,  $k$  sweepline passes are required, each considering a specific cone  $\mathcal{C} = (\theta_L, \theta_R)$ . The sweepline for a cone  $\mathcal{C}$  proceeds in direction  $\tau = \frac{\theta_L + \theta_R}{2} + \pi$ , i. e. opposite to the cone’s internal angle bisector. Input points are swept in the order of their projection onto  $\tau$  – represented as blue dashed line in Figure 2 – given by sorting

$$\rho_\tau(p) := \begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} \cos \tau \\ \sin \tau \end{pmatrix} \quad \forall p = (x \ y)^T \in \mathbf{P}. \quad (1)$$

All input points are inserted into an event priority queue  $Q$  with priority  $\rho_\tau(p)$  and event type *input point*. Each input point  $p$  is the origin of a cone  $\mathcal{C}^p$  with boundary rays  $B_L$  and  $B_R$ . Cone  $\mathcal{C}^p$  defines the *region*  $R_p$  of the plane, where point  $p$  is the nearest neighbor with respect to cone  $\mathcal{C}^p$  for any point being swept after  $p$ . The invariant of the algorithm is that once a point has been swept, its nearest neighbor in cone  $\mathcal{C}$  has been determined. For instance, in Figure 2  $p_2$  is in the region of  $p_1$ , therefore  $p_1$  is the nearest neighbor for  $p_2$  with respect to cone  $\mathcal{C}$ . A boundary ray  $B_\square$  always separates the regions of two input

<sup>1</sup> Our implementation computes the Yao graph but can easily be modified to compute the OVD.

## 20:4 Efficient Yao Graph Construction

■ **Algorithm 1** Sweepline algorithm for cone defined by  $(\theta_L, \theta_R)$ .  $L(p, \theta, R_a, R_b)$  and  $L(\overleftarrow{pv}, R_a, R_b)$  denote the ray originating at  $p$  with angle  $\theta$  and the line segment from  $p$  to  $v$ , respectively, both separating regions  $R_a$  and  $R_b$ .

**Input:** points  $\mathbf{P} = \{p_1, \dots, p_n\}$  with  $p_i \in \mathbb{R}^2$ , cone  $(\theta_L, \theta_R)$

**Output:** GNG  $G = (V, E)$

---

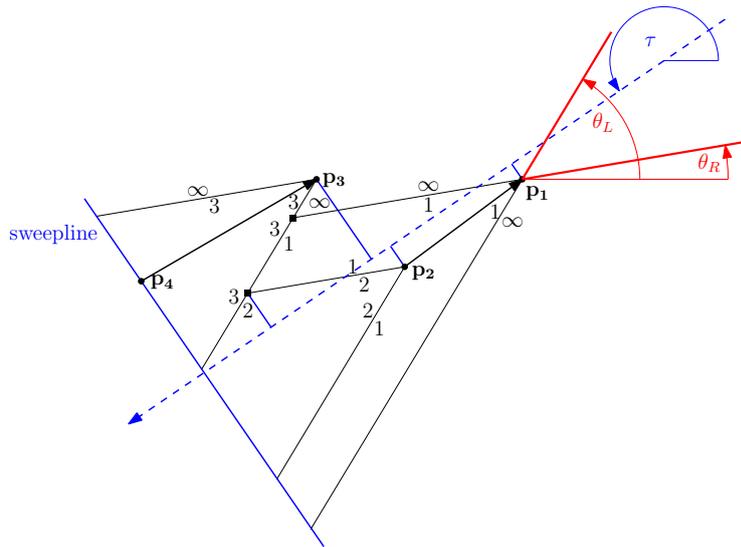
```

1:  $\tau \leftarrow \frac{\theta_L + \theta_R}{2} + \pi$  ▷ opposite of cone's internal angle bisector
2:  $Q \leftarrow \{(\rho_\tau(p), p, I) : p \in \mathbf{P}\}$  ▷ initialize PQ with input points
3:  $SL \leftarrow \emptyset$ 
4:  $G = (V, E) \leftarrow (\mathbf{P}, \emptyset)$ 

5: while  $p \leftarrow \text{popMin}(Q)$  do
6:   if  $p$  is input point then
7:      $B_L, B_R \leftarrow \text{findRegion}(p, SL)$  ▷  $B_L$  and  $B_R$  enclose  $p$ 
8:      $E \cup = (p, B_L^r)$  ▷ assert ( $B_L^r == B_R^l$ )
9:     if  $B_L \cap B_R = v$  then delete  $v$  from  $Q$ 
10:     $B_L^* \leftarrow L(p, \theta_L + \pi, B_L^r, R_p)$ 
11:     $B_R^* \leftarrow L(p, \theta_R + \pi, R_p, B_R^l)$ 
12:    insert  $[B_L^*, B_R^*]$  into  $SL$  between  $B_L$  and  $B_R$ 
13:    if  $B_L \cap B_L^* = v$  then  $Q \cup = (\rho_\tau(v), v)$ 
14:    if  $B_R \cap B_R^* = v$  then  $Q \cup = (\rho_\tau(v), v)$ 
15:  if  $p$  is intersection then
16:     $B_L, B_R \leftarrow$  intersecting rays at  $p$ 
17:     $a \leftarrow B_L^l$     $b \leftarrow B_R^r$ 
18:    if  $B_L \cap \text{prev}(B_L) = v$  then delete  $v$  from  $Q$  ▷ left neighbor boundary on  $SL$ 
19:    if  $B_R \cap \text{succ}(B_R) = v$  then delete  $v$  from  $Q$  ▷ right neighbor boundary on  $SL$ 
20:     $L_{BS} \leftarrow L(\frac{a+b}{2}, \sphericalangle(a, b) + \frac{\pi}{2}, R_a, R_b)$  ▷ bisector of line segment  $\overleftrightarrow{ab}$ 
21:     $L_L \leftarrow L(p, \theta_L + \pi, R_a, R_b)$ 
22:     $L_R \leftarrow L(p, \theta_R + \pi, R_a, R_b)$ 
23:    if  $L_L \cap L_{BS} = \emptyset = L_{BS} \cap L_R$  then ▷ bisector intersects no line from  $p$ 
24:       $B^* = L\left(p, \pi + \begin{cases} \theta_L & \text{if } \rho_\tau(a) < \rho_\tau(b) \\ \theta_R & \text{else} \end{cases}, R_a, R_b\right)$ 
25:    if  $L_L \cap L_{BS} = p = L_{BS} \cap L_R$  then ▷ bisector intersects both lines in  $p$ 
26:       $B^* = L(p, \sphericalangle(a, b) + \frac{\pi}{2}, R_a, R_b)$ 
27:    if  $L_L \cap L_{BS} = v$  or  $L_{BS} \cap L_R = v$  then ▷ bisector intersects one line in  $v$ 
28:       $B^* = L(\overleftarrow{pv}, R_a, R_b) + L(v, \sphericalangle(a, b) + \frac{\pi}{2}, R_a, R_b)$ 
29:       $Q \cup = (\rho_\tau(v), v)$  ▷ deletion event
30:    replace  $[B_L, B_R]$  in  $SL$  with  $B^*$ 
31:    if  $B^* \cap \text{prev}(B^*) = v$  then  $Q \cup = (\rho_\tau(v), v)$ 
32:    if  $B^* \cap \text{succ}(B^*) = v$  then  $Q \cup = (\rho_\tau(v), v)$ 
33:  if  $p$  is deletion point then
34:     $B \leftarrow$  ray belonging to  $p$  ▷  $B = L(\overleftarrow{pv}, R_a, R_b) + L(v, \sphericalangle(a, b) + \frac{\pi}{2}, R_a, R_b)$ 
35:    replace  $B$  in  $SL$  with  $L(v, \sphericalangle(a, b) + \frac{\pi}{2}, R_a, R_b)$ 
36: return  $G$ 

```

---



■ **Figure 2** Example state of the sweepline algorithm for cone  $C = (\theta_L, \theta_R)$  (marked in red). Input points (circles, bold labels) are numbered in the order they are swept by the sweepline, with their projections on the cone’s internal angle bisector shown in blue. Rays are labeled with the regions they are separating. Intersection events are marked with a square. Already determined edges of the Yao graph are indicated by arrows.

points, thus it is defined by its point of origin  $B_{\square}^p$  and angle  $B_{\square}^{\theta}$  as well as its left and right region,  $B_{\square}^l$  and  $B_{\square}^r$  respectively. The region outside any point’s cone is labeled with infinity. Due to intersecting boundary rays, boundaries between regions can also be the union of a line segment and ray as described in detail in Section 3.2. However, for simplicity of presentation we still refer to these composite boundaries as boundary rays, unless this distinction is of relevance. The algorithm maintains an ordered data structure  $SL$  of rays currently intersecting the sweepline. The rays are ordered left-to-right and the data structure needs to support insert, remove and find operations in  $\mathcal{O}(\log n)$  time, as well as access to the left and right neighbors of a given ray. In Section 4.2 we describe a balanced binary search tree that supports these operations and is tuned for our application. Algorithm 1 presents our algorithm, which is described in detail in the following. An example execution of the algorithm is depicted in Figure 11 in the appendix.

### 3.1 Event Types

There are three *event types*: 1) input points, 2) intersection points, and 3) deletion points. In the following, we describe how each event is handled by the algorithm. Through the execution of the algorithm, priority queue  $Q$  contains all unprocessed input points, the intersection points of the boundaries of adjacent regions as well as deletion points for composite boundaries, ordered according to  $\rho_{\tau}(p)$ . If several events coincide, their processing order can be arbitrary.

**1) Input points.** All points of the given set  $\mathbf{P}$  are inserted into the event priority queue at the beginning of the algorithm. For an input point event with associated point  $p$ , the sweepline data structure  $SL$  is searched for the region  $R_q$  containing  $p$ . This region is defined

by its two bounding rays  $B_L$  and  $B_R$  and their associated regions  $B_L^r = B_R^l = R_q$ .<sup>2</sup> We can then add edge  $(p, q)$  to the edge set  $\mathbf{E}$  of  $G_k$ , as proven in [6, Lemma 3.1]. The point  $p$  is the apex of region  $R_p$ , bounded to the left by  $B_L^* = (p, \theta_L + \pi)$ , separating regions  $R_q$  and  $R_p$ , and bounded to the right by  $B_R^* = (p, \theta_R + \pi)$ , separating regions  $R_p$  and  $R_q$ . These new rays are inserted into  $SL$  between  $B_L$  and  $B_R$ , forming the sequence  $[B_L, B_L^*, B_R^*, B_R]$ . Lastly, intersection points of the considered rays need to be addressed. If  $B_L$  and  $B_R$  intersect in point  $v$ , its associated intersection point event needs to be removed from the priority queue  $Q$ , as  $B_L$  and  $B_R$  are no longer neighboring rays. Instead, possible intersection points between  $B_L$  and  $B_L^*$  as well as  $B_R^*$  and  $B_R$  are added to  $Q$  for future processing.

**2) Intersection points.** An intersection point  $v$  is associated with its two intersecting rays  $B_L$  and  $B_R$ . They separate regions  $R_p := B_L^l$ ,  $B_L^r = R_m = B_R^l$  and  $B_R^r := R_q$ , refer to Figure 3. Region  $R_m$  terminates at intersection point  $v$  and a new boundary  $B^*$  between  $R_p$  and  $R_q$  originates at  $v$ . The shape of  $B^*$  depends on the configuration of points  $p$  and  $q$  and can either be a simple ray or a union of a line segment and a ray. Section 3.2 describes in more detail how  $B^*$  is determined.  $B^*$  then replaces the sequence  $[B_L, B_R]$  in the sweepline data structure. Again, intersection points of the considered rays need to be addressed. If  $B_L$  has an intersection point with its left neighbor or if  $B_R$  has an intersection point with its right neighbor, the associated intersection point events need to be removed from the event queue  $Q$ . Correspondingly, if  $B^*$  intersects its left or right neighbor, the appropriate intersection point events are added to  $Q$ .

**3) Deletion event.** Deletion events are not part of the original algorithm described by Chang et al. [6], as the authors do not specify how to handle composite boundaries in their paper. We use them to implement boundaries consisting of a line segment and a ray. The deletion event marks the end of the line segment and the beginning of the ray. It does not change the actual state of the sweepline data structure.

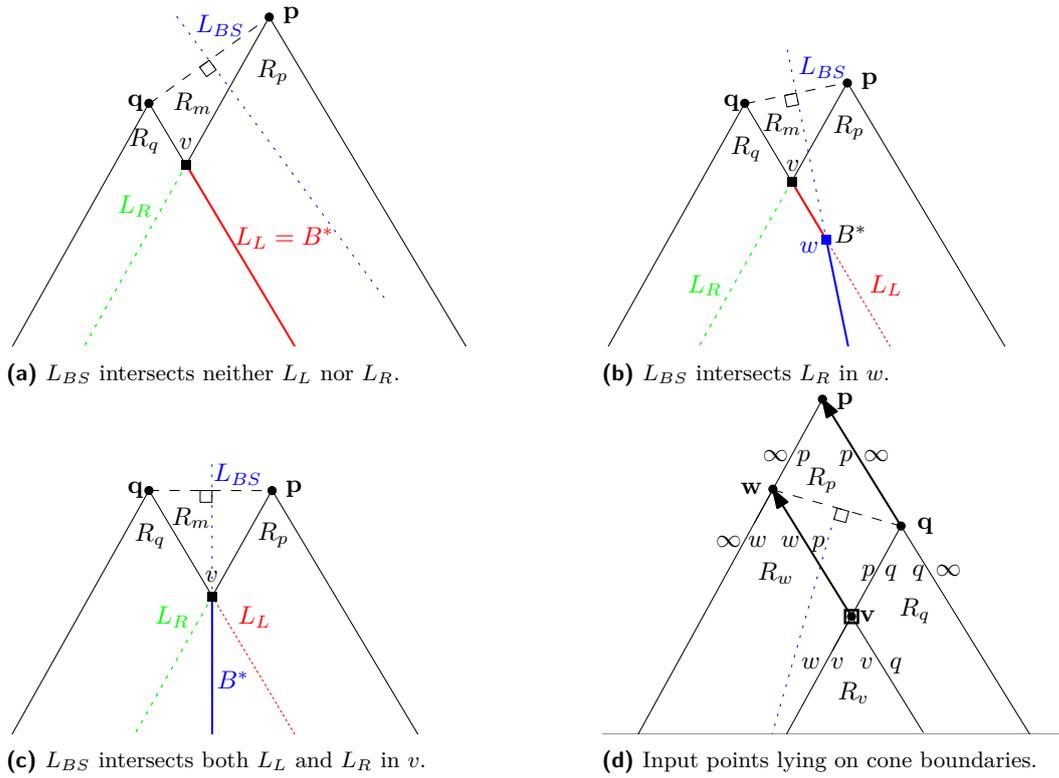
### 3.2 Boundary Determination

As described in the previous section, at an intersection point event  $v$ , the two intersecting rays  $B_L$  and  $B_R$  are merged into a new boundary  $B^*$ , separating regions  $R_p := B_L^r$  and  $R_q := B_R^l$ . The shape of the boundary is determined by the configuration of points  $p$  and  $q$ . The determination is based on the number of intersection points between lines

- $L_L = (v, \theta_L + \pi)$  (green),
- $L_R = (v, \theta_R + \pi)$  (red), and
- the bisector  $L_{BS}$  of line  $\overline{pq}$ ,  $L_{BS} = (\frac{\bar{v} + \bar{q}}{2}, \sphericalangle(p, q) + \pi/2)$  (blue, dashed).

The colors refer to the lines in Figure 3 which illustrates the different cases. If  $L_{BS}$  intersects neither  $L_L$  nor  $L_R$  then  $B^* = L_L = (v, \pi + \theta_R)$  if  $p$  was swept before  $q$ , otherwise  $B^* = L_R = (v, \pi + \theta_L)$  (Figure 3a). Intuitively, the region of the lower point with respect to the sweepline direction continues, whereas the upper region stops at intersection point  $v$ . If  $L_{BS}$  intersects both lines  $L_L$  and  $L_R$ , then the two intersection points must coincide with  $v$  (Figure 3c). In this case,  $B^* = (p, \sphericalangle(p, q) + \pi/2)$ , i. e. the boundary continues with the angle of the bisector from point  $v$ . Otherwise, i. e.  $L_{BS}$  intersects either  $L_L$  or  $L_R$  in a point  $w$ , the resulting boundary  $B^*$  will be the union of the line segment  $v\vec{w}$  and ray  $(w, \sphericalangle(p, q) + \pi/2)$  (Figure 3b). In this case, a deletion point event is added to the priority queue at point  $w$ .

<sup>2</sup> Note that Chang et al. [6] explicitly store rays and regions in their sweepline data structure. However, since a region is identifiable by its two bounding rays, we choose this simpler representation of the sweepline state.



■ **Figure 3** Illustration of the three possible configurations for boundary  $B^*$  following an intersection point event. In all examples input point  $p$  is swept before  $q$ . Lines  $L_L$ ,  $L_R$  and  $L_{BS}$  are dotted, the resulting boundary  $B^*$  is denoted in bold. In Figure d), Yao graph edges are shown as bold arrows.

**Input Points Colinear on Cone Boundaries.** Chang et al. make the assumption that no line between two input points is with angle  $\theta_L$  or  $\theta_R$ . In the following we shall lift this requirement. Input points sharing a common line with angle  $\theta_\square$  become *visible* from each other. This impacts the regions the passing boundary rays are separating. Refer to Figure 3d for a graphical representation of the following discussion. Recall, that the *left* – or counterclockwise – boundary ray with angle  $\theta_L$  belongs to a cone  $\mathcal{C}$ , whereas the *right* one does not. If a boundary ray  $B_\square = (p, \theta_\square, B_\square^l, B_\square^r)$  intersects an input point  $q$  then  $B_\square$  terminates at  $q$  and a new boundary ray  $B_\square^l$  is formed. If  $B_\square$  is a left boundary, i. e.  $\theta_\square = \theta_L$ , then edge  $(p, q)$  is added to  $G_K$  and  $B_\square^l$  separates regions  $B_\square^l$  and  $R_q$ . If it is a right boundary then no edge is added to  $G_k$  and  $B_\square^l$  separates  $R_q$  and  $B_\square^r$ . Refer to example points  $q$  and  $w$  in Figure 3d.

### 3.3 Analysis

The total number of events processed by the sweepline algorithm is the sum of input point events  $N_{\text{input}}$ , intersection point events  $N_{\text{IE}}$  and deletion events  $N_{\text{DE}}$ . In order to bound the number of events processed by the sweepline algorithm, we consider the number of rays that can be present in the sweepline data structure during the execution of the algorithm.

Every input point event adds two rays to the sweepline data structure  $SL$ , resulting in a total of  $2n$  rays. It possibly removes one intersection event from the event queue  $Q$  and may add up to two new such events. Every intersection point event removes two rays from  $SL$  and adds one new ray, thus reducing the sweepline size by one. Therefore, at most  $2n$

■ **Table 1** Comparison of geometric predicates used in algorithms for Yao graph construction.

	Chang et al.	CGAL	Naive	Grid
<b>Complexity</b>	$\mathcal{O}(n \log n)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$
<b>Predicates</b>				
Eucl. distance comp.	X	X	X	X
dist. to line comp.	X	X		
oriented side of line	X		X	X
<b>Constructions</b>				
cone boundaries	X	X	X	X
box construction				X
line projection	X			
ray intersection	X			

intersection events can be processed before all rays are removed from the sweepline. An intersection event possibly removes two additional intersection events aside from itself from  $Q$  and may add one new intersection event.<sup>3</sup> Additionally, one deletion event may be added to  $Q$ . Therefore, at most  $2n$  deletion events may be processed, each of which leaves the number of rays and intersections unchanged. In total,

$$\begin{aligned} N_{\text{events}} &\leq N_{\text{input}} + N_{\text{IE}} + N_{\text{DE}} \\ &\leq n + 2n + 2n = 5n \end{aligned} \tag{2}$$

With a balanced binary search tree as sweepline data structures each event can be processed in  $\mathcal{O}(\log n)$  time, yielding the bound of  $\mathcal{O}(n \log n)$ .

## 4 Implementation Details

In this section we highlight some of the design decisions of our implementation of Chang et al.'s Yao graph algorithm.

### 4.1 Geometric Kernels

The algorithm by Chang et al. [6] requires many different geometric predicates and constructions. We implement our own version of the required predicates and constructions in an *inexact* manner. Additionally, the user can employ kernels provided by the CGAL library. The EPIC kernel provides `exact` `predicates` and `inexact` `constructions`, whereas the EPEC kernel features `exact` `predicates` and `exact` `constructions` [5].

Table 1 lists the geometric predicates used by the different algorithms for Yao graph construction presented in this paper, refer to Section 5 for details on the naive and grid algorithm. Only the sweepline algorithm requires the computation of particularly costly intersections. The naive as well as grid-based Yao graph algorithm require an oriented side of line predicate only if cone boundaries are constructed exactly, in order to determine the cone  $\mathcal{C}_I^p$  a point  $q$  lies in with respect to point  $p$ . Additionally, the grid algorithm could construct

<sup>3</sup> Technically,  $B^*$  can intersect both its neighbors, leading to two intersection events. However, when the first – as defined by  $\rho_r(\cdot)$  – intersection event is processed, it will delete the second event from  $Q$ , as  $B^*$  is removed from  $SL$  and a new boundary ray is inserted by the first event.

the grid data structure using exact computations. However, in our implementation we only use inexact computations to place the input points into grid cells, as we did not encounter the need for exact computation in any of our experiments. Note that the determination whether a grid cell could hold a closer neighbor than the currently found one is done using the (exact) Euclidean distance comparison predicate.

In order to reduce the number of expensive ray-ray intersection calculations, we store all found intersection points in a linear probing hash table, with the two intersecting rays as key, see e.g. Algorithm 1:13. If we need to check whether two rays are intersecting – e.g. Algorithm 1:9 – we merely require a hash table lookup.

## 4.2 Sweepline Data Structure

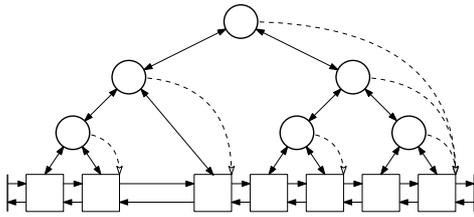
Chang et al. prove an  $\mathcal{O}(n \log n)$ -time complexity for the algorithm [6, Theorem 3.2]. In order to achieve this bound, the data structure maintaining the rays currently intersected by the sweepline must provide the following operations: insert, remove and predecessor search in  $\mathcal{O}(\log n)$  time. Our data structure furthermore provides neighbor access in  $\mathcal{O}(1)$  time.

In order to support these operations, we use a doubly linked list of rays, with an AVL search tree on top [1]. Figure 4 shows a graphical representation of our data structure. As the order of the rays along the sweepline is known at the time of insertion – see Algorithm 1:12 – the  $\mathcal{O}(\log n)$ -time search phase of a traditional AVL data structure can be omitted and new rays can be inserted in a bottom-up manner. However, this optimization requires the need for parent pointers in the tree. As always two neighboring rays are inserted into the sweepline data structure at the same time, we implement a special insert operation for this case that only requires one rebalancing operation for both rays. For removal operations, the position of the ray within the sweepline is known as well – refer to Algorithm 1:30. Thus, similar to insert operations, no search phase is required for removals and the operation can be performed in a bottom-up manner. The algorithm always removes the two neighboring rays  $B_L$  and  $B_R$  and replaces them with  $B^*$ .  $B^*$  has the same left neighbor as  $B_L$  and the same right neighbor as  $B_R$ . Therefore, we can simply replace  $B_L$  with  $B^*$  in the data structure and just need to remove  $B_R$ , leading to merely one rebalancing operation.

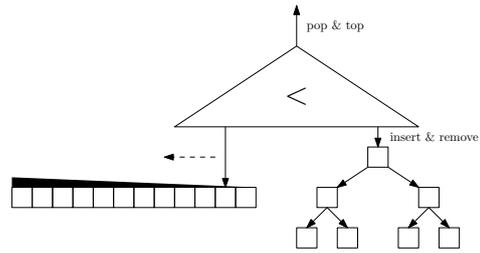
The search for the enclosing region of a point  $p$  – Algorithm 1:7 – is performed by finding the first ray  $B_R$ , currently intersecting the sweepline, that has  $p$  to its right. This requires the evaluation of a oriented side of line predicate at each level of the tree. The left neighbor  $B_L$  of  $B_R$ , must have  $p$  to its left or on it. Therefore  $B_L$  and  $B_R$  enclose  $p$  and  $B_L^r = B_R^l = R_q$  gives the the region  $p$  is contained in. To facilitate searching, each internal node of the tree needs to refer to the rightmost ray in its subtree. As rays are complex objects, we use pointers to the corresponding leaf to save memory. Given the expensive search operations, AVL trees – as strictly height-balancing trees – are preferable to data structures with weaker balancing guarantees, such as red-black trees [16].

## 4.3 Priority Queue

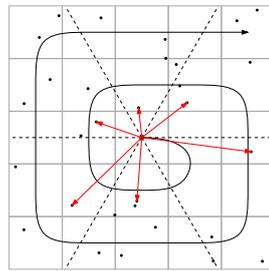
The priority queue (PQ)  $Q$  is initialized with all input points at the beginning of the algorithm. During event processing, intersection and deletion events may be added and removed from  $Q$ , therefore requiring an addressable priority queue. The objects are ordered according to Equation (1), thus keys are (exact) numerical values. Our experiments show that, typically, for  $n$  input points, only about  $\mathcal{O}(\sqrt{n})$  intersection and deletion events are in  $Q$  at any given step. Using the same PQ for all events would result in expensive dynamic PQ operations. As input point events are static in  $Q$ , we can use a two part data structure as shown in



■ **Figure 4** The sweepline data structure is a doubly linked list of rays with an AVL search tree on top. Additionally, each node has a pointer to the rightmost ray in its subtree (dashed).



■ **Figure 5** The priority queue consists of a *static*, sorted array of input points and a *dynamic*, addressable PQ for intersection and deletion events.



■ **Figure 6** Grid-based Yao graph construction algorithm. The cone boundaries are represented by dashed lines. The algorithm visits grid cells in order of the thick curve. Found edges of the Yao graph are labeled in red.

Figure 5. Input point events are stored in an array – sorted by priority in  $Q$  – with a pointer to the smallest unprocessed element. Intersection and deletion events are stored in an actual addressable priority queue. We use an addressable binary heap for this part of the data structure. The TOP operation needs to compare the minimum element of the PQ with the element pointed to in the array and return the minimum of both. POP either performs a regular pop on the PQ or moves the pointer of the array to the next larger element. INSERT and REMOVE operations can access the PQ directly, as only this part of the data structure is dynamic. Thus, the actual dynamic PQ is much smaller resulting in more efficient SIFTDOWN and SIFTUP operations in the binary heap. The smaller heap size not only reduces the tree height but also makes the heap more cache friendly.

This optimization might be of interest for other algorithms that initialize their priority queue with all input points and only have a small number of dynamically added events in their priority queue at any given time. Note that the *total* number of processed intersection and deletion events surpasses the number of input points by far, however only a small number of these events are in the PQ at the same time.

## 5 Evaluation

In this section we evaluate our implementation on a variety of datasets against other algorithms for Yao graph generation.

**Competing Algorithms.** As mentioned before, we are not aware of any previous implementations of Chang et al.’s Yao graph algorithm. Therefore, we evaluate our implementation against other algorithms to construct the Yao graph of a given point set. Our main competitor

is the Yao graph algorithm from the CGAL library’s cone-based spanners package [15]. As we are not aware of any other tuned implementations to construct Yao graphs, we implement two other algorithms ourselves as competition.

First, we implement a *naive*  $\mathcal{O}(n^2)$ -time algorithm that serves as a trivial baseline. The algorithm compares the distance of all point pairs and determines the closest neighbor in each of the  $k$  cones for a point. It requires only two geometric predicates: distance comparison and oriented side of line test. For a point pair  $p$  and  $q$ , the cone  $\mathcal{C}_i^p$  that  $q$  lies in with respect to  $p$  can first be approximated by  $i = \lceil \angle(p,q)/k \rceil$ . Then, two oriented side of line tests suffice to exactly determine the cone  $q$  lies in, independent of the number of cones  $k$ .

Second, we implement a *grid*-based algorithm. The algorithm places all points in a uniform grid data structure [2] that splits the bounding box of all input points in  $\mathcal{O}(n)$  equal-sized cells. For each input point  $p$ , the algorithm first visits  $p$ ’s own grid cell and computes for each point  $q$  in the cell its distance to  $p$  and the cone  $q$  lies in with respect to  $p$ . The algorithm then visits the grid cells surrounding  $p$ ’s cell in a spiraling manner, refer to Figure 6. For each visited cell, the algorithm computes the distances and cones for the points contained in it with respect to  $p$  until all cones of point  $p$  are settled. A cone is settled, if a neighbor  $v$  has been found within that cone and no point in adjacent grid cells can be closer to  $p$  than  $v$ . Note that some cones may remain unsettled until all grid cells have been visited if no other input points lie within that cone for point  $p$ . While the algorithm still has a  $\mathcal{O}(n^2)$  worst case time complexity, it performs much better in practice.

**Experimental Setup.** We test all algorithms on a variety of synthetic and real-world datasets. We use input point sets distributed uniformly and normally in the unit square, as well as points lying on the circumference of a circle and at the intersections of a grid [12] – the former being a worst case input for the grid algorithm, the latter being a bad case for numerical stability. We furthermore use two real-world datasets: intersections in road networks and star catalogs. As road networks we use graphs from the 9<sup>th</sup> DIMACS implementation challenge [9]. To generate a road network of a desired size  $n$  from the FULL USA graph, we use a random location and grow a rectangular area around it until at least  $n$  nodes are within the area. US cities feature many points on a grid and therefore present a challenge for numerical stability. We furthermore use the Gaia DR2 star catalog [7], which contains celestial positions for approximately 1.3 billion stars. We use a similar technique as for road networks to generate subgraphs of a desired size. Here, we grow a cube around a random starting location until the desired number of stars fall within it. We then project all stars onto the  $xy$ -plane as 2D input for our experiments. Figure 10 in the appendix shows examples of our input datasets and resulting Yao graphs.

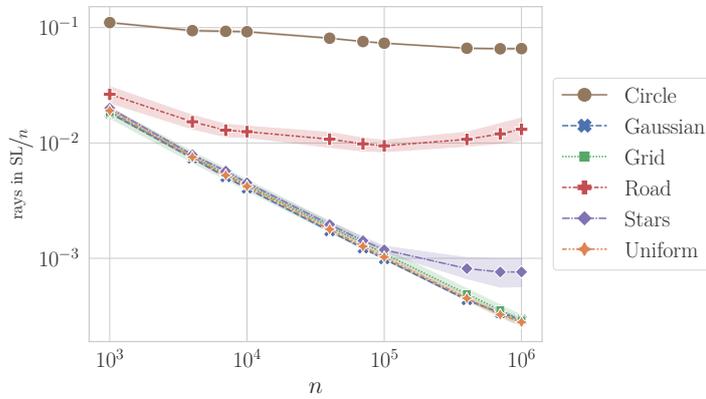
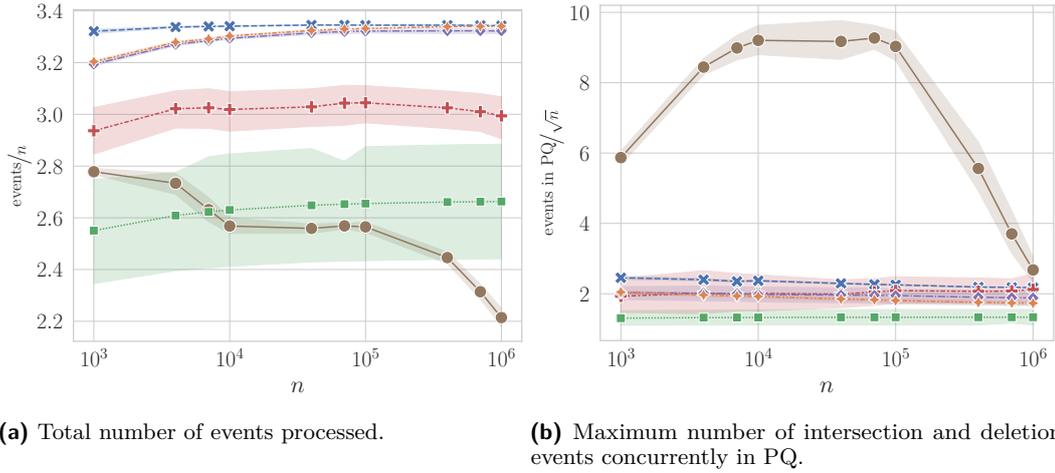
We implemented all algorithms in our C++ framework YAOGRAPH, available on Github.<sup>4</sup> Our code was compiled using GCC 12.1.0 with CGAL version 5.0.2. All experiments were run on a server with an Intel Xeon Gold 6314U CPU with 32 cores and 512 GiB of RAM. For experiments we used three different random seeds and  $k = 6$  unless otherwise specified.

## 5.1 Algorithmic Metrics

Firstly, we discuss relevant properties of the sweepline algorithm. Figure 7a shows the number of events processed per input point by the algorithm. Each input point has one input point event and generates 2.3 intersection and/or deletion events on average, with very little

<sup>4</sup> <https://github.com/dfunke/YaoGraph>

## 20:12 Efficient Yao Graph Construction

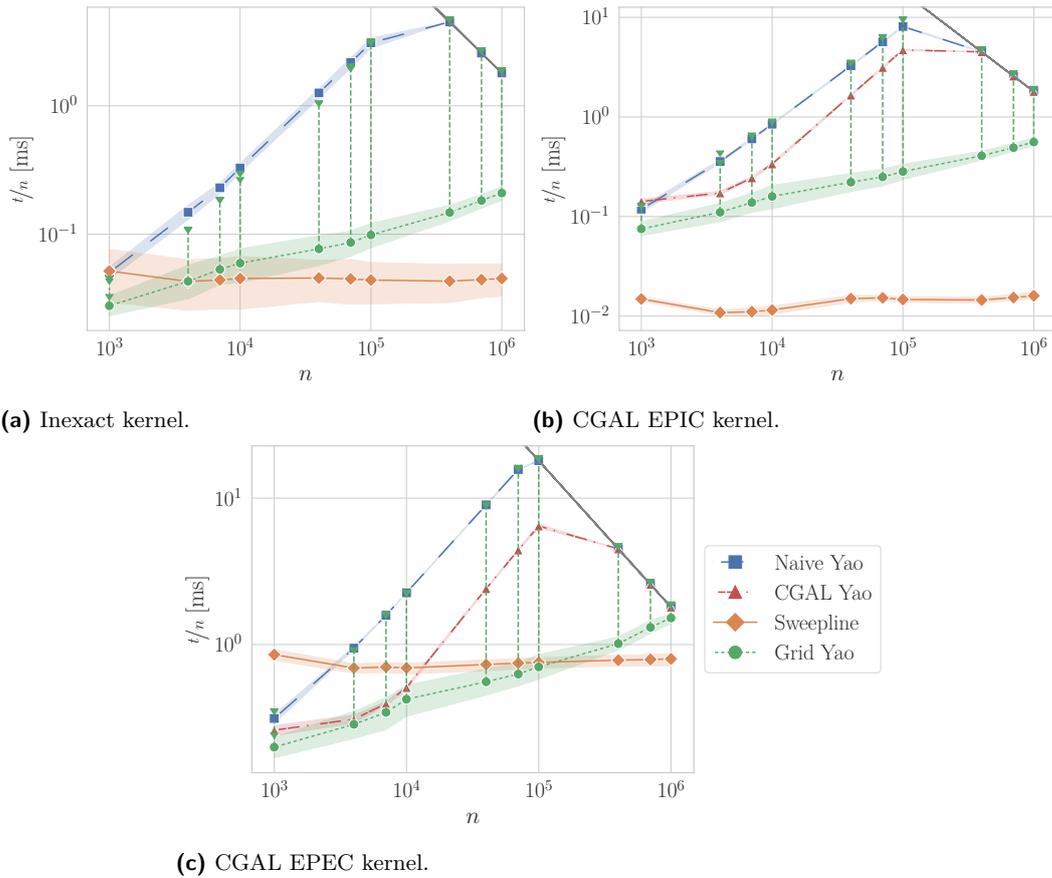


■ **Figure 7** Statistics for varying input point distribution and point set sizes for  $k = 6$  cones. Error bands give the variation over the different cones being calculated.

variance with regard to input size and distribution – except for the grid distribution and road graphs. Both exhibit larger variance, depending on whether the cone’s boundaries coincide with grid lines or not. Figure 7b shows the maximum number of intersection and deletion events that are in the priority queue at any given time during the algorithm execution. This number scales with  $\mathcal{O}(\sqrt{n})$  for most studied inputs, which motivates our choice of the two-part priority queue as discussed in Section 4.3. The behavior of the circle distribution requires further investigation. The maximum number of rays in the sweepline data structure at any point during algorithm execution shows no clear scaling behavior, refer to Figure 7c. It scales with  $\mathcal{O}(\sqrt{n})$  for most synthetic input sets, but approaches a constant fraction of the input size for the circle ( $\approx 10\%$ ), road ( $\approx 1\%$ ) and star ( $\approx 0.1\%$ ) datasets.

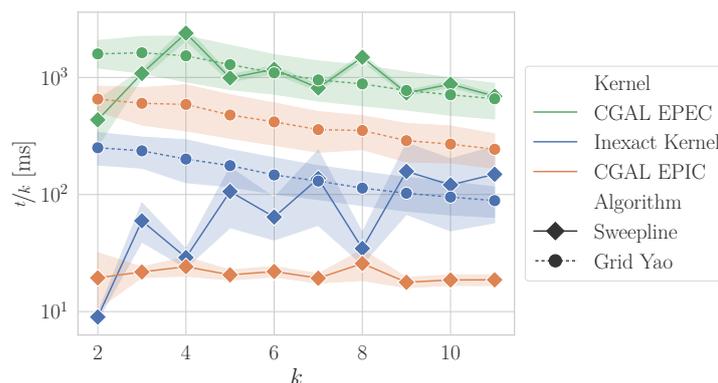
## 5.2 Runtime Evaluation

Figure 8 shows the results of our runtime experiments. Plots Figure 8a to Figure 8c show the (scaled) running time of the algorithms, displaying variations due to input distributions as error bands. Figure 12 in the appendix gives a more detailed picture of the runtime for the different distributions. Note that only the grid and the sweepline algorithm are sensitive



**Figure 8** Algorithm runtime experiments. Experiments over varying input sizes are performed with  $k = 6$  cones. Error bands give the runtime variation over the different input point distributions. For the grid algorithm, the circle distribution is plotted separately with triangular markers. The gray line represents the time limit of 30 min per algorithm. Experiments over varying number of cones are with  $n = 1 \times 10^5$  uniformly distributed input points.

to the input point distribution. As previously seen in Figure 7a, the number of processed events by the sweepline algorithm is relatively stable for all distributions. Therefore only little variation is seen in the runtime of the algorithm. This also shows, that the size of the sweepline data structure has only negligible influence on the algorithm runtime, as no higher runtime is observed for the road or circle datasets. Our inexact kernel shows more runtime variation than CGAL’s highly optimized kernels, mainly due to the grid distribution with its many points directly on cone boundaries. The sweepline algorithm clearly outperforms CGAL’s Yao graph implementation. Furthermore, even though the sweepline algorithm requires much more involved computations, it is superior to the simple grid algorithm for non-exact constructions. Only for exact constructions, large inputs are required to negate the more expensive operations of the sweepline algorithm. The exact construction kernel leads to runtime overhead of 100 compared to the EPIC kernel. However, if points lie directly on cone boundaries, exact constructions are necessary to obtain correct results, as seen in Figure 10c in the appendix. The data dependency is more pronounced for the grid algorithm, which performs well for most datasets but degenerates to the naive algorithm for the circle distribution, due to the many empty grid cells in the circle’s interior.



■ **Figure 9** Algorithm runtime experiments. Experiments over varying number of cones are with  $n = 1 \times 10^5$  uniformly distributed input points.

To compute a Yao graph with  $k$  cones, the sweepline algorithm requires  $k$  passes. This linear relationship can be seen in Figure 9. The grid algorithm has no dependency on  $k$  – except for the size of the neighborhood of a point. However, our experiments show that the runtime of the algorithm increases with increasing  $k$ . We attribute this to the fact that more grid cells need to be visited in order to settle all cones of a point  $p$ , since with narrower cones, chances are higher that no points lying in a specific cone of  $p$  are within a visited grid cell. We did not perform these experiments with the naive algorithm or the CGAL algorithm, due to their long runtimes. CGAL’s algorithm also requires one pass per cone, whereas the naive algorithm’s runtime dependency on  $k$  is negligible.

## 6 Conclusion

We present the – to the best of our knowledge – first implementation of Chang et al.’s optimal  $\mathcal{O}(n \log n)$ -time Yao graph algorithm. Our implementation uses carefully engineered data structures and algorithmic operations and outperforms current publicly available Yao graph implementations – particularly CGAL’s cone-based spanners package – by at least an order of magnitude. We furthermore present a very simple grid-based Yao graph algorithm that also outperforms CGAL’s implementation, but is inferior to Chang et al.’s algorithm for larger input. However, the algorithm could be further improved by using a precomputed mapping of the grid neighborhood to cones, in order to only visit grid cells that can contain points in hitherto unsettled cones. Moreover, the algorithm is trivially parallelizable over the input points, whereas Chang et al.’s algorithm can only be easily parallelized over the  $k$  cones. The parallelization within one sweepline pass remains for future work.

---

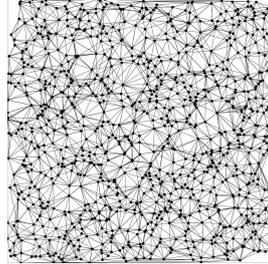
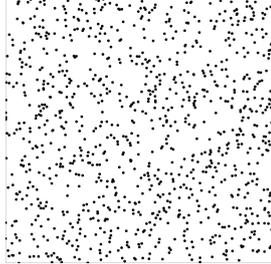
## References

- 1 Georgy Maksimovich Adelson-Velsky and Evgeny Mikhailovich Landis. An algorithm for organization of information. In *Doklady Akademii Nauk*, volume 146, pages 263–266. Russian Academy of Sciences, 1962.
- 2 V. Akman, W.R. Franklin, M. Kankanhalli, and C. Narayanaswami. Geometric computing and uniform grid technique. *Computer-Aided Design*, 21(7):410–420, 1989. doi:10.1016/0010-4485(89)90125-5.

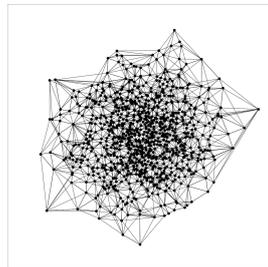
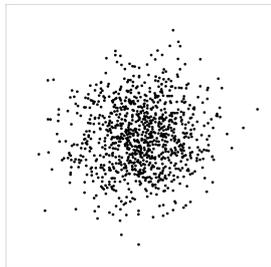
- 3 Luis Barba, Prosenjit Bose, Mirela Damian, Rolf Fagerberg, Wah Loon Keng, Joseph O'Rourke, André Van Renssen, Perouz Taslakian, Sander Verdonschot, and Ge Xia. New and improved spanning ratios for Yao graphs. *Journal of Computational Geometry*, 6(2):19–53, 2015. doi:10.20382/JOCG.V6I2A3.
- 4 J. L. Bentley and T. A. Ottmann. Algorithms for reporting and counting geometric intersections. *IEEE Transactions on Computers*, pages 643–647, 1979.
- 5 Hervé Brönnimann, Andreas Fabri, Geert-Jan Giezeman, Susan Hert, Michael Hoffmann, Lutz Kettner, Sylvain Pion, and Stefan Schirra. 2D and 3D Linear Geometry Kernel. In *CGAL User and Reference Manual*. CGAL Editorial Board, 5.5.1 edition, 2022. URL: <https://doc.cgal.org/5.5.1/Manual/packages.html#PkgKernel23>.
- 6 Maw Shang Chang, Nen-Fu Huang, and Chuan-Yi Tang. An optimal algorithm for constructing oriented Voronoi diagrams and geographic neighborhood graphs. *Information Processing Letters*, 35(5):255–260, 1990. doi:10.1016/0020-0190(90)90054-2.
- 7 Gaia Collaboration. Gaia Data Release 2 - Summary of the contents and survey properties. *Astronomy & Astrophysics*, 616:A1, 2018. doi:10.1051/0004-6361/201833051.
- 8 Mirela Damian and Naresh Nelavalli. Improved bounds on the stretch factor of Y4. *Computational Geometry*, 62:14–24, April 2017. doi:10.1016/j.comgeo.2016.12.001.
- 9 Camil Demetrescu, Andrew V Goldberg, and David S Johnson. *The shortest path problem: Ninth DIMACS implementation challenge*, volume 74. American Mathematical Soc., 2009.
- 10 Steven Fortune. A sweepline algorithm for Voronoi diagrams. *Algorithmica*, 2(1):153–174, 1987.
- 11 Daniel Funke and Peter Sanders. Efficient Yao Graph Construction, 2023. arXiv:2303.07858.
- 12 Daniel Funke, Peter Sanders, and Vincent Winkler. Load-Balancing for Parallel Delaunay Triangulations. In Ramin Yahyapour, editor, *Euro-Par 2019: Parallel Processing*, pages 156–169, Cham, 2019. Springer International Publishing.
- 13 Giri Narasimhan and Michiel Smid. *Geometric spanner networks*. Cambridge University Press, 2007.
- 14 Christian Schindelbauer, Klaus Volbert, and Martin Ziegler. Geometric spanners with applications in wireless networks. *Computational Geometry*, 36(3):197–214, 2007.
- 15 Weisheng Si, Quincy Tse, and Frédéric Paradis. Cone-Based Spanners. In *CGAL User and Reference Manual*. CGAL Editorial Board, 5.4 edition, 2022. URL: <https://doc.cgal.org/5.4/Manual/packages.html#PkgConeSpanners2>.
- 16 Svetlana Štrbac-Savić and Milo Tomašević. Comparative performance evaluation of the AVL and red-black trees. In *Proceedings of the Fifth Balkan Conference in Informatics*. ACM, 2012. doi:10.1145/2371316.2371320.
- 17 Andrew Chi-Chih Yao. On Constructing Minimum Spanning Trees in k-Dimensional Spaces and Related Problems. *SIAM Journal on Computing*, 11(4):721–736, 1982. doi:10.1137/0211059.
- 18 Xiujian Zhang, Jiguo Yu, Wei Li, Xiuzhen Cheng, Dongxiao Yu, and Feng Zhao. Localized Algorithms for Yao Graph-Based Spanner Construction in Wireless Networks Under SINR. *IEEE/ACM Transactions on Networking*, 25(4):2459–2472, 2017. doi:10.1109/TNET.2017.2688484.

**A** Evaluation

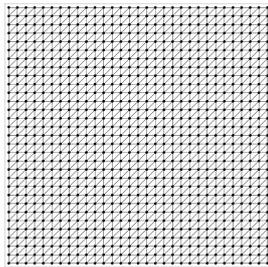
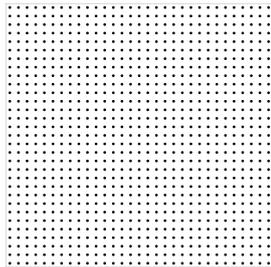
**A.1** Input Point Distributions



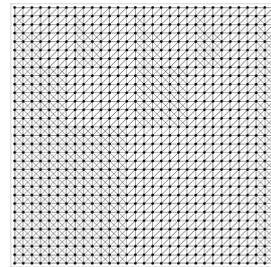
(a) Uniform distribution.



(b) Gaussian distribution.

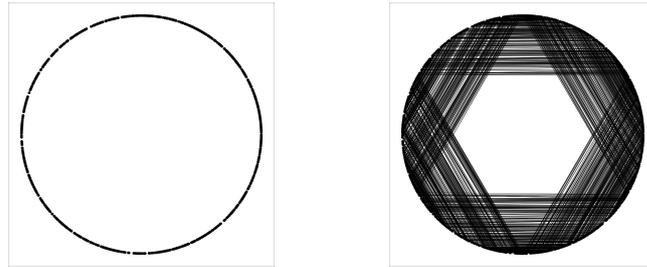


EPEC kernel

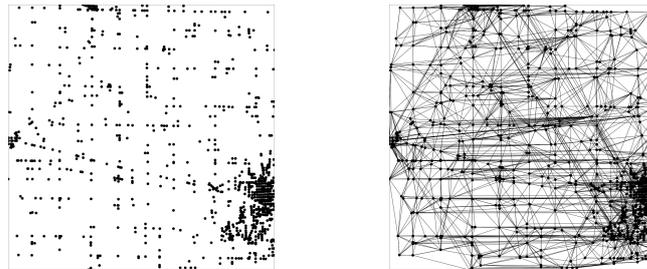


EPIC kernel

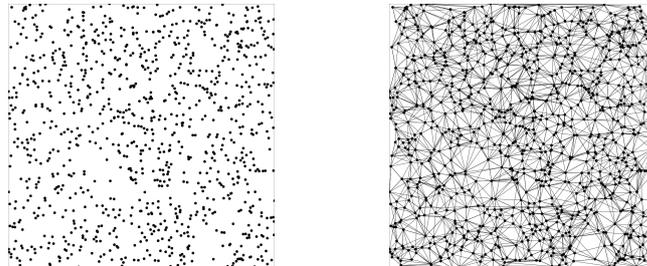
(c) Grid distribution.



(d) Circle distribution.



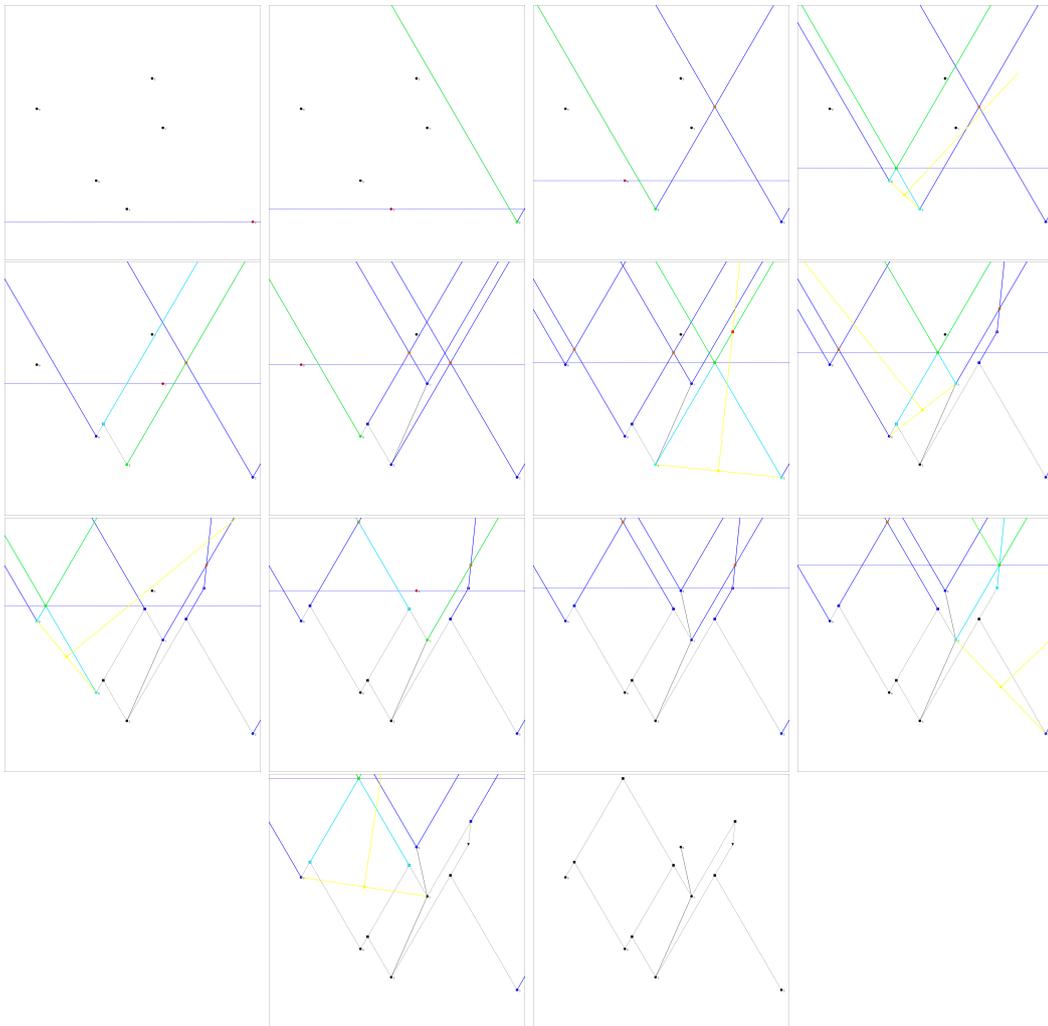
(e) Road dataset.



(f) Star dataset.

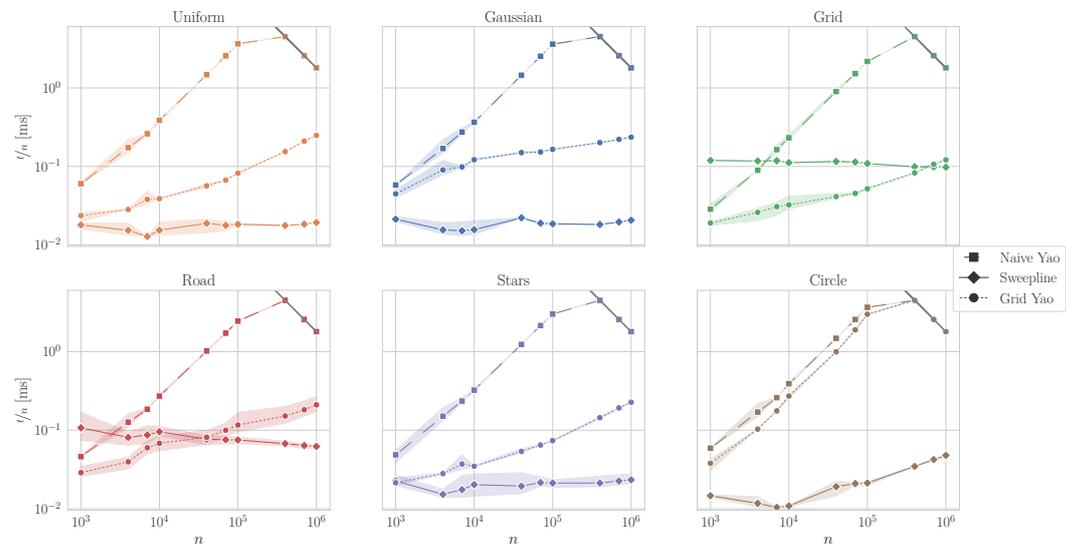
■ **Figure 10** Input distributions for  $n = 1000$  points and resulting Yao graph for  $k = 6$ . For the grid distribution, the resulting graphs from exact constructions and inexact constructions are shown.

## A.2 Example Execution

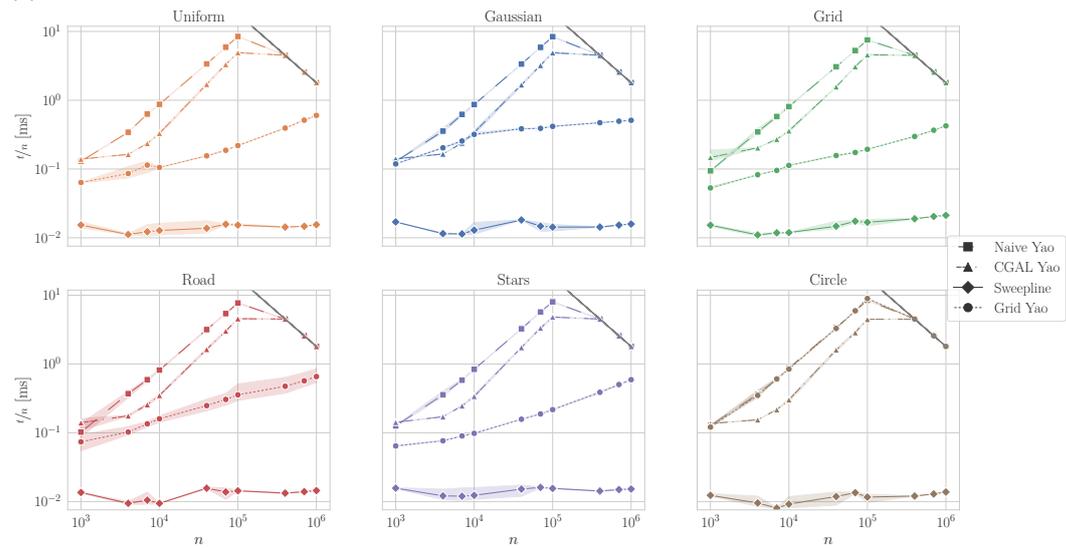


■ **Figure 11** Sample execution for  $n = 6$  points and cone  $\mathcal{C} = (4\pi/3, 5\pi/3)$ , resulting in  $\tau = \pi/2$  (upward). The currently processed point is marked in red, the sweepline is a dashed, blue line. All rays currently intersecting the sweepline are blue, except for the left boundary ray  $B_L$  (cyan) and right boundary ray  $B_R$  (green). Intersection points are marked as squares, deletion points as triangles. For intersection events, the intersecting rays are cyan and the bisector line is yellow. Edges of the Yao graph are solid black lines and settled cone boundaries are dashed.

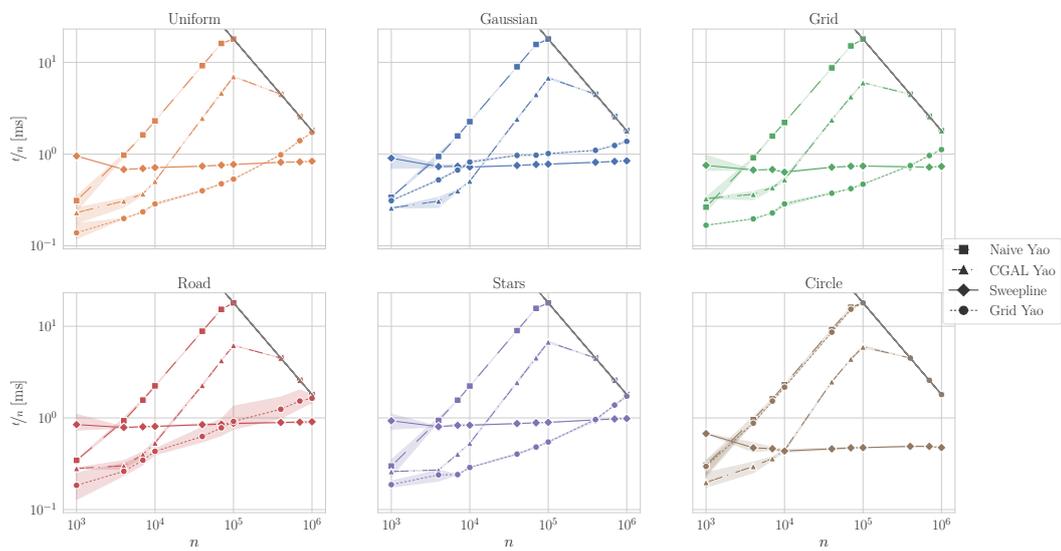
A.3 Results



(a) Inexact kernel.



(b) CGAL EPIC kernel.



(c) CGAL EPEC kernel.

■ **Figure 12** Algorithm runtime experiments. Experiments over varying input sizes are performed with  $k = 6$  cones. The gray line represents the time limit of 30 min per algorithm.