

## ESTIMATING PERIMETER USING GRAPH CUTS

NICOLÁS GARCÍA TRILLOS,\* *Brown University*

DEJAN SLEPČEV,\*\* *Carnegie Mellon University*

JAMES VON BRECHT,\*\*\* *California State University, Long Beach*

### Abstract

We investigate the estimation of the perimeter of a set by a graph cut of a random geometric graph. For  $\Omega \subseteq D = (0, 1)^d$  with  $d \geq 2$ , we are given  $n$  random independent and identically distributed points on  $D$  whose membership in  $\Omega$  is known. We consider the sample as a random geometric graph with connection distance  $\varepsilon > 0$ . We estimate the perimeter of  $\Omega$  (relative to  $D$ ) by the, appropriately rescaled, graph cut between the vertices in  $\Omega$  and the vertices in  $D \setminus \Omega$ . We obtain bias and variance estimates on the error, which are optimal in scaling with respect to  $n$  and  $\varepsilon$ . We consider two scaling regimes: the dense (when the average degree of the vertices goes to  $\infty$ ) and the sparse one (when the degree goes to 0). In the dense regime, there is a crossover in the nature of the approximation at dimension  $d = 5$ : we show that in low dimensions  $d = 2, 3, 4$  one can obtain confidence intervals for the approximation error, while in higher dimensions one can obtain only error estimates for testing the hypothesis that the perimeter is less than a given number.

*Keywords:* Perimeter; nonparametric estimation; graph cut; point cloud; random geometric graph; concentration inequality

2010 Mathematics Subject Classification: Primary 60D05

Secondary 62G20; 68R10

### 1. Introduction

In this paper we investigate the use of random graph cuts to obtain empirical estimates of the perimeter of a domain  $\Omega \subseteq D := (0, 1)^d$  for  $d \geq 2$ . Let  $x_1, \dots, x_n, \dots$  denote a sequence of independent random points uniformly distributed on the unit cube  $D$  and let  $V_n := \{x_1, \dots, x_n\}$ .

The problem of estimating the perimeter of  $\Omega$  based on knowing which points of  $V_n$  belong to  $\Omega$  is a classical question, see [4], [8], [9], [17], [23], and [24] for recent contributions, and see Subsection 1.2 below which contains a discussion about related work. Here we consider an estimator of the perimeter that is based on a geometric graph constructed from the point cloud  $V_n$ . More precisely, we select  $\varepsilon_n > 0$ , and connect two points in the cloud if they are within distance  $\varepsilon_n$  of each other; then we consider an appropriately scaled ‘cut’ determined by the number of edges in the graph that connect points in  $\Omega$  with points that belong to  $\Omega^c$ . This type of estimator is natural to consider, since graph cuts arise as a discretization of the perimeter in many applications such as clustering [3], [7], [14], [15], [18], [25]–[28]. Our choice of estimator is therefore based on its use in various statistical and machine learning applications.

Received 24 August 2016; revision received 15 June 2017.

\* Postal address: Division of Applied Mathematics, Brown University, Providence, RI 02912, USA.

Email address: nicolas\_garcia\_trillos@brown.edu

\*\* Postal address: Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

\*\*\* Postal address: Department of Mathematics and Statistics, California State University, Long Beach, CA 90840, USA.

We focus on estimating the approximation error of the perimeter of an arbitrary (but fixed) set  $\Omega$ ; the error estimates that we obtain are uniform on a class of sets where certain geometric quantities are controlled (see Remark 1.4 below). One of the important features of our estimator is that it has a small bias. Indeed, the expectation of our estimator provides a second-order approximation (in terms of the natural parameter  $\varepsilon_n$ ) of the true perimeter of  $\Omega$  under some regularity conditions on the boundary of the set (see (1.7) below); this turns out to be a sharp estimate for the bias. We also obtain precise estimates for the variance of the estimator; for these estimates to hold, we do not need any regularity assumptions on the boundary of the set  $\Omega$  except that it has a finite perimeter in the most general sense. Furthermore, we show that our estimator converges almost surely (a.s.) for remarkably sparse graphs (and indeed in settings which are sparser than for previously considered estimators). The almost sure convergence holds with no regularity assumptions on  $\Omega$  (other than the fact that it has a finite perimeter in the most general sense). Finally, assuming some smoothness on the boundary of  $\Omega$ , we are able to establish (in the dense graph regime  $1/n^{1/d} \ll \varepsilon \ll 1$ ) the asymptotic distribution of the error. These estimates lead to asymptotic confidence intervals (which we refer to simply as *confidence intervals*) and bounds on the type I and type II errors for hypothesis tests associated to the perimeter of a set.

**1.1. Set-up and main results**

Let us now be more precise about the setting we consider in this paper. We consider random geometric graphs with vertex set  $V_n$  and radius  $\varepsilon_n > 0$ . That is, graphs where  $x_i$  and  $x_j$  are connected by an edge if  $\|x_i - x_j\| \leq \varepsilon_n$ . The *graph cut* between  $A \subseteq V_n$  and  $A^c$  is given by

$$\text{cut}_{\varepsilon_n}(A, A^c) := \sum_{x_i \in A} \sum_{x_j \in V_n \setminus A} \mathbf{1}_{\{\|x_i - x_j\| \leq \varepsilon_n\}}.$$

We define the *graph perimeter* as a rescaling of the graph cut. For any  $\Omega \subseteq D$ ,

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) := \frac{2}{n(n-1)\varepsilon_n^{d+1}} \text{cut}_{\varepsilon_n}(V_n \cap \Omega, V_n \cap \Omega^c).$$

See Figure 1 for an illustration of this construction. The scaling is chosen so that  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  becomes a consistent estimator for the true (continuum) perimeter.

One of our interests lies in determining how well  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  performs as an estimate of the relative perimeter of  $\Omega$  in  $D$ . We first investigate for which scaling of  $\varepsilon_n$  on  $n$  does the convergence hold a.s. as  $n \rightarrow \infty$ . In other words, we want to understand the relation between edge sparsity of random geometric graphs and pointwise convergence of the graph perimeter to the continuum perimeter in the almost sure sense. We consider this question for a very broad family of sets  $\Omega \subseteq D$ , which are assumed to have finite relative perimeter only in  $D$  in the general sense of [2]. That is, we define the relative perimeter of  $\Omega$  with respect to  $D$  to be

$$\text{per}(\Omega) = \sup \left\{ \int_{\Omega} \text{div}(v) \, dx : \text{(for all } x \in D) \|v(x)\| \leq 1, v \in C_c^\infty(D, \mathbb{R}^d) \right\}. \tag{1.1}$$

If  $\Omega$  has a smooth relative boundary then  $\text{per}(\Omega)$  is nothing but the surface area of  $\partial\Omega \cap D$ . We remark that the notion of the perimeter we use is more general than the notion of the Hausdorff measure of the boundary  $\mathcal{H}^{d-1}(\partial\Omega \cap D)$  and the Minkowski content, which are the ones more typically used in the statistics literature [4], [8], [9]. In particular, as we see below, we work with consistent nonparametric estimators in the most general setting available.

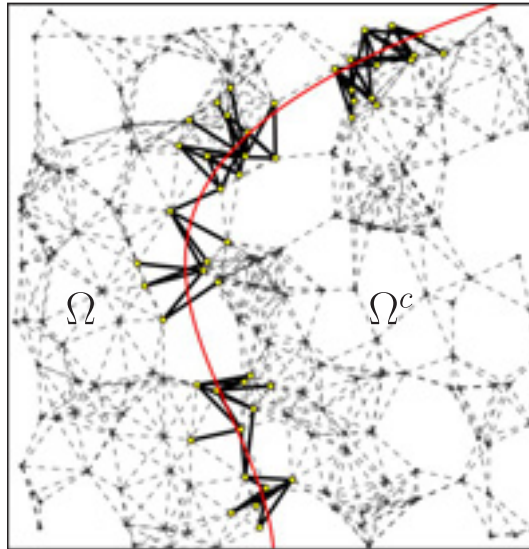


FIGURE 1: Edges of the cut between  $\Omega$  and  $\Omega^c$  are represented by bold lines, while other edges are dashed lines. Total number of vertices is  $n = 200$  and connectivity radius  $\varepsilon = 0.13$ .

As we recall below in (1.14)–(1.16), it is known that when  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  the bias of the estimator vanishes in the limit

$$\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \rightarrow \sigma_d \text{per}(\Omega) \quad \text{as } n \rightarrow \infty. \tag{1.2}$$

The scaling factor  $\sigma_d$  satisfies

$$\sigma_d := \int_{\|z\| \leq 1} |z_1| \, dz = \frac{2s_{d-2}}{(d+1)(d-1)}, \tag{1.3}$$

where  $z_1$  denotes the first component of the vector  $z \in \mathbb{R}^d$  and  $s_{d-2}$  is the area of the  $(d-2)$ -dimensional unit sphere (the boundary of the unit ball in  $\mathbb{R}^{d-1}$ ). We refer to the normalizing quantity  $\sigma_d$  as the *surface tension*.

We obtain the following estimates on the deviation of the graph perimeter  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  from its mean. Let

$$f(n, \varepsilon_n) := \begin{cases} \frac{1}{\sqrt{n\varepsilon_n}} & \text{if } 1/n^{1/d} \leq \varepsilon_n, \\ \frac{1}{n\varepsilon_n^{(d+1)/2}} & \text{if } 1/n^{2/(d+1)} \leq \varepsilon_n \leq 1/n^{1/d}. \end{cases} \tag{1.4}$$

**Theorem 1.1.** *Let  $p \geq 1$  and let  $\Omega \subseteq D$  be a set with finite perimeter. Assume that  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then*

$$\mathbb{E}(|G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega))|^p) \leq C_p (\max\{1, \text{per}(\Omega)\} f(n, \varepsilon_n))^p, \tag{1.5}$$

where  $C_p$  is a constant that depends only on  $p$ . In particular, if  $\varepsilon_n = n^{-\alpha}$  for  $\alpha \in (0, 2/(d+1))$  then

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) \rightarrow \sigma_d \text{per}(\Omega) \quad \text{a.s. as } n \rightarrow \infty.$$

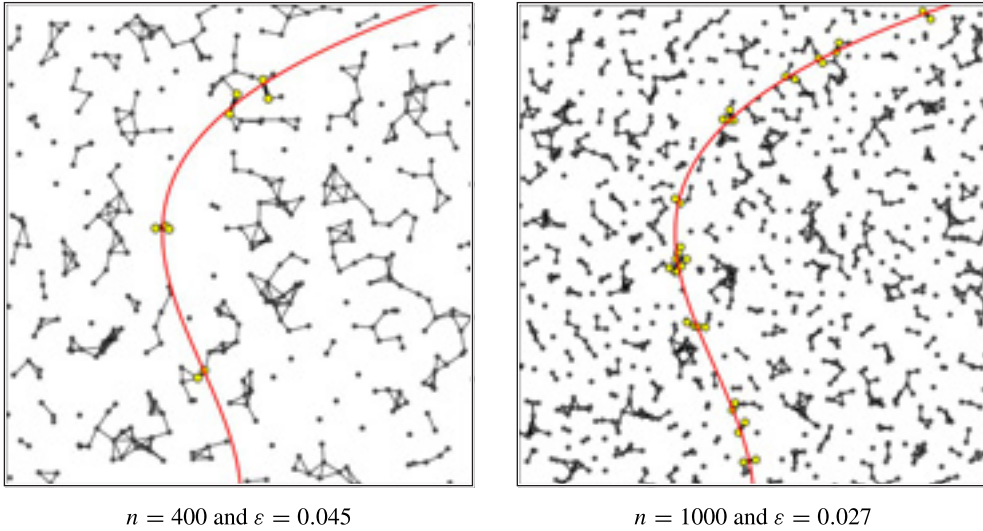


FIGURE 2: An illustrate of the ‘sparse’ regime when  $1/n^{1/d} \gg \varepsilon \gg (1/n)^{2/d+1}$  when the average degree goes to 0. Nevertheless, since the number of edges in the cut still increases as  $n$  does, the convergence of the cut to the perimeter still holds.

The last part of the previous theorem follows from (1.2), the moment estimates (1.5), Markov’s inequality, and the Borel–Cantelli lemma which imply that

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \rightarrow 0 \quad \text{a.s.}$$

**Remark 1.1.** We note that the almost sure convergence holds for rather sparse graphs (see Figure 2). Namely, the typical degree of a node is  $\omega_d n \varepsilon^d$ , where  $\omega_d$  is the volume of the unit ball in  $d$  dimensions. When  $n^{-2/(d+1)} \ll \varepsilon_n \ll n^{-1/d}$ , the almost sure convergence holds, while the average degree of a vertex converges to 0. The convergence is still possible because the expected number of edges crossing  $\partial\Omega$  is still a quantity converging to  $\infty$ .

**Remark 1.2.** Given that we show the almost sure consistency of our estimators for any arbitrary (but fixed) set for which the perimeter is finite, our construction provides a universal strongly convergent estimator which was the desired property listed as an open problem in [8] (for the estimator they considered).

We turn to estimating the bias of the empirical approximation:  $|\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) - \sigma_d \text{per}(\Omega)|$ . We first characterize the mean of the graph perimeter  $\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega))$  as the *nonlocal perimeter*  $\text{per}_{\varepsilon_n}(\Omega)$  of  $\Omega$ , defined as

$$\text{per}_{\varepsilon}(\Omega) := \frac{2}{\varepsilon^{d+1}} \int_{\Omega} \int_{D \setminus \Omega} \mathbf{1}_{\{\|x-y\| \leq \varepsilon\}} \, dx \, dy. \tag{1.6}$$

The nonlocal nature of the functional is essentially related to the fact that it involves averages of finite differences as opposed to a *local* approach where one considers derivatives. We then proceed to estimate  $|\text{per}_{\varepsilon_n}(\Omega) - \sigma_d \text{per}(\Omega)|$  explicitly. It is straightforward to check that  $|\text{per}_{\varepsilon_n}(\Omega) - \sigma_d \text{per}(\Omega)| = O(\varepsilon_n)$  for general subsets  $\Omega \subseteq D$  with a smooth relative boundary.

However, we show that the error is actually quadratic in  $\varepsilon_n$ , that is,

$$|\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) - \sigma_d \text{per}(\Omega)| = |\text{per}_{\varepsilon_n}(\Omega) - \sigma_d \text{per}(\Omega)| = O(\varepsilon_n^2) \tag{1.7}$$

under the extra condition that  $\text{dist}(\Omega, \partial D) > 0$ . This is the subject of the next lemma, whose proof may be found in Appendix A.

**Lemma 1.1.** *Let  $\Omega$  be a set with a smooth boundary such that  $\text{dist}(\Omega, \partial D) > 0$ . Let  $0 < \varepsilon < \text{dist}(\Omega, \partial D)$  and let  $\text{per}_{\varepsilon}(\Omega)$  be defined by (1.6). Then*

$$\text{per}_{\varepsilon}(\Omega) = \sigma_d \text{per}(\Omega) + O(\varepsilon^2). \tag{1.8}$$

**Remark 1.3.** The assumption that  $\text{dist}(\Omega, \partial D) > 0$  in the above lemma is needed in order to obtain a bias of order  $\varepsilon^2$ . If  $\Omega$  touches the boundary  $\partial D$ , the error of order  $\varepsilon^2$  is not expected, as can be seen, for example, by considering the rectangle  $\Omega := \{x = (x_1, \dots, x_d) \in D : x_1 \leq \frac{1}{2}\}$ , for which the error is of order  $\varepsilon$ ; in this situation the error is completely due to the region where  $\partial\Omega$  meets  $\partial D$  transversally. Thus, in general, for  $\Omega \subseteq D$  with a smooth relative boundary, the bias is of order  $\varepsilon$ . On the other hand, the smoothness of the boundary of  $\Omega$  is needed in the previous lemma only to guarantee that curvature and its derivatives are well defined. Finally, the constant involved in the term  $O(\varepsilon^2)$  depends on the reach of the set  $\Omega$ , and the intrinsic curvature of  $\partial\Omega$  together with its derivatives; this can be seen from our computations in Appendix A.

Combining the bias and variance of estimates allows us to obtain the rates of convergence for the error  $|G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \sigma_d \text{per}(\Omega)|$ . In particular, we estimate the ‘standard deviation’

$$\text{std}(n) := \mathbb{E}((G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \sigma_d \text{per}(\Omega))^2)^{1/2},$$

which we may quantify precisely by using the variance-bias decomposition

$$\text{std}^2(n) = \text{var}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) + (\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) - \sigma_d \text{per}(\Omega))^2.$$

Using the special case  $p = 2$  of Theorem 1.1 to estimate the variance and using Lemma 1.1 to estimate the bias we obtain the following.

**Theorem 1.2.** *Let  $\Omega \subseteq D$  be an open set with a smooth boundary. Assume that  $n^{-2/(d+1)} \ll \varepsilon_n \ll 1$  and consider  $f(n, \varepsilon_n)$  defined via (1.4). The error of approximating  $\sigma_d \text{per}(\Omega)$  by  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  satisfies*

$$\text{std}(n) = O(f(n, \varepsilon_n) + \varepsilon_n).$$

Furthermore, if we assume that  $\Omega$  does not touch the boundary of  $D$ , that is,  $\text{dist}(\Omega, \partial D) > 0$ , then a better estimate holds:

$$\text{std}(n) = O(f(n, \varepsilon_n) + \varepsilon_n^2).$$

A simple calculation using (1.4) allows us to choose a scaling of  $\varepsilon_n$  on  $n$  so that the error of the approximation is as small as possible. In particular, for a set  $\Omega$  which touches the boundary ( $\text{dist}(\Omega, \partial D) = 0$ ), the optimal scaling of  $\varepsilon_n$  is

$$\varepsilon_n \sim \begin{cases} n^{-1/3} & \text{if } d \leq 3, \\ n^{-2/(d+3)} & \text{if } d \geq 3, \end{cases} \quad \text{yielding} \quad \text{std}(n) \lesssim \begin{cases} n^{-1/3} & \text{if } d \leq 3, \\ n^{-2/(d+3)} & \text{if } d \geq 3. \end{cases}$$

We note that, for  $d < 3$ , the optimal  $\varepsilon_n$  is achieved in the regime  $n^{-1/d} \lesssim \varepsilon_n \ll 1$ , while if  $d > 3$  it is in the sparse regime. If we consider sets  $\Omega$  with a smooth boundary but such that  $\text{dist}(\Omega, \partial D) > 0$ , then the optimal scaling of  $\varepsilon_n$  on  $n$  is as follows:

$$\varepsilon_n \sim \begin{cases} n^{-1/5} & \text{if } d \leq 5, \\ n^{-2/(d+5)} & \text{if } d \geq 5, \end{cases} \quad \text{yielding} \quad \text{std}(n) \lesssim \begin{cases} n^{-2/5} & \text{if } d \leq 5, \\ n^{-4/(d+5)} & \text{if } d \geq 5. \end{cases}$$

Again we note that the optimal  $\varepsilon_n$  is in the sparse regime if  $d > 5$ . This has implications on how well the perimeter is estimated by graph cuts in the graphs considered in most machine learning applications. Namely, if  $d \geq 5$  and the graph has average degree bounded from below, that is, when  $n^{-1/d} \lesssim \varepsilon_n \ll 1$ , then, since the bias bound is sharp, most of the error is due to the bias term.

**Remark 1.4.** The constant appearing in the moment estimates of Theorem 1.1 depends exclusively on the power  $p$ , the dimension  $d$ , and the true perimeter of the set  $\Omega$ ; in particular, taking  $p = 2$  we see that the estimates for the variance of  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  are uniform on the class of sets  $\Omega$  whose perimeter is bounded above by some fixed constant. For general sets with large perimeter, we obtain uniform estimates for moments of relative error  $|G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega))|/\text{per}(\Omega)$  instead of the absolute error. Thus, by combining Theorem 1.1 and the proof of Lemma 1.1, we can derive that the corresponding error estimates for  $|G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \sigma_d \text{per}(\Omega)|$  (either relative or absolute) are uniform on the class of sets  $\Omega$  with a smooth boundary satisfying the following conditions: the reach of  $\Omega$  is bounded below by a fixed positive constant; the distance to the boundary  $\partial D$  is bounded below by a fixed positive constant; the curvature and first derivatives of curvature are bounded from above by a fixed constant. This last requirement comes from the  $O(\varepsilon^2)$  terms following Taylor’s theorem in the proof of our bias estimates.

We now consider how to obtain confidence intervals for the value of the true perimeter  $\text{per}(\Omega)$  based on the estimator  $(1/\sigma_d)G_{n,\varepsilon_n}^{\text{per}}(\Omega)$ . We focus on the dense regime (see Figures 3 and 4),  $1/n^{1/d} \ll \varepsilon_n \ll 1$ , and first obtain the asymptotic distribution of  $G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega))$ .

Since the term  $\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) = \text{per}_{\varepsilon_n}(\Omega)$  depends on the set  $\Omega$  itself, which is unknown (we assume that we have only an oracle access to it), we need the bias to be negligible compared to the standard deviation of our estimator. This allows us to construct confidence intervals for  $\text{per}(\Omega)$  without using any additional information about the bias (for example, upper bounds). From (1.5), the standard deviation of  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  scales as  $1/\sqrt{n\varepsilon_n}$ , while by (1.7) the bias scales as  $\varepsilon_n^2$ ; these estimates lead to restrictions on the dimensions for which the bias is negligible with respect to the standard deviation. Namely, this is possible for  $d = 2, 3$ , and 4.

**Theorem 1.3.** *Let  $\Omega \subseteq D$  be an open set with a smooth boundary such that  $\text{dist}(\Omega, \partial D) > 0$ . Let  $\varepsilon_n$  be such that*

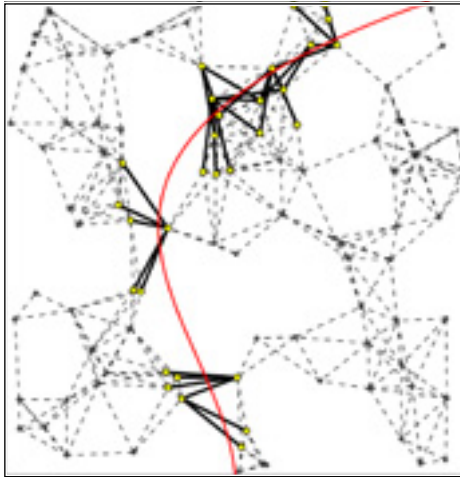
$$\frac{1}{n^{1/d}} \ll \varepsilon_n \ll 1.$$

Then

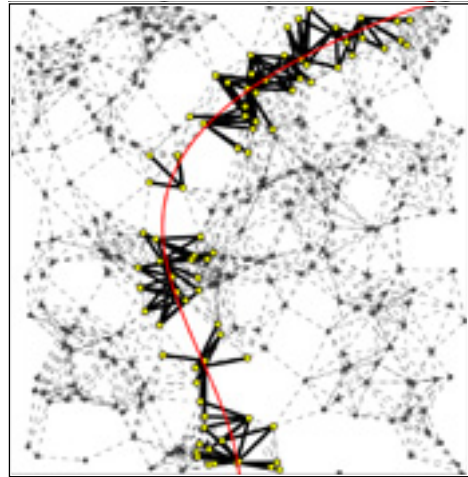
$$\sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}} (G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega)) \xrightarrow{w} N(0, 1),$$

where ‘ $\xrightarrow{w}$ ’ denotes weak convergence and  $C_d$  is given by

$$C_d := 2 \int_0^1 |B_d(0, 1) \cap \{x = (x_1, \dots, x_d) \in \mathbb{R}^d : x_d \geq t\}|^2 dt. \tag{1.9}$$

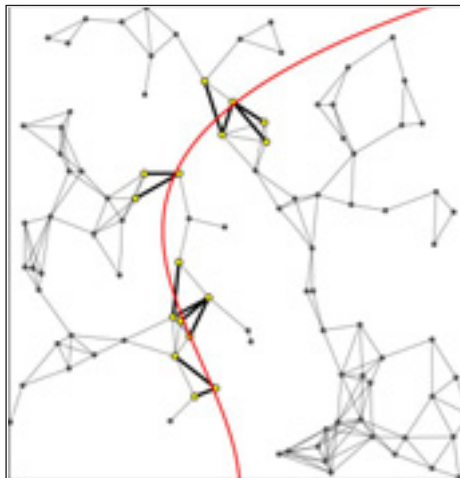


$n = 100$  and  $\epsilon = 0.17$

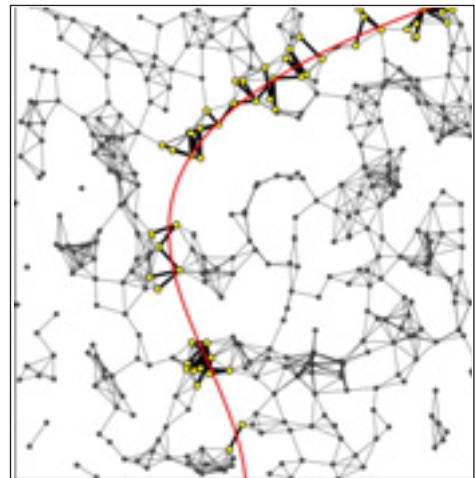


$n = 300$  and  $\epsilon = 0.12$

FIGURE 3: An illustrate of the ‘very dense’ regime,  $1 \gg \epsilon \gg (\ln(n)/n)^{1/d}$ , when the graphs are connected with high probability.



$n = 100$  and  $\epsilon = 0.12$



$n = 400$  and  $\epsilon = 0.07$

FIGURE 4: An illustrate of an intermediate ‘dense’ regime  $(\ln(n)/n)^{1/d} \gg \epsilon \gg (1/n)^{1/d}$ . The average degree still increases, but the graphs are disconnected with high probability.

If, in addition, the dimension  $d$  is either 2, 3, or 4, and if

$$\frac{1}{n^{1/d}} \ll \epsilon_n \ll \frac{1}{n^{1/5}}, \tag{1.10}$$

then

$$\sqrt{\frac{n\epsilon_n}{4C_d \text{per}(\Omega)}} (G_{n,\epsilon_n}^{\text{per}}(\Omega) - \sigma_d \text{per}(\Omega)) \xrightarrow{w} N(0, 1). \tag{1.11}$$

Naturally, the previous theorem implies that one can obtain confidence intervals for the value of  $\text{per}(\Omega)$  when the dimension  $d$  is 2, 3, or 4. Let us fix  $\alpha \in (0, 0.5)$  and let  $Z_\alpha$  be the  $1 - \alpha$  quantile of the standard normal distribution. That is,  $Z_\alpha$  is such that

$$\mathbb{P}(N(0, 1) \leq Z_\alpha) = 1 - \alpha.$$

Then provided that

$$\frac{1}{n^{1/d}} \ll \varepsilon_n \ll \frac{1}{n^{1/5}},$$

it follows that with probability converging to  $1 - \alpha$ ,  $\text{per}(\Omega) \in (a_n^-, a_n^+)$ , where

$$a_n^\pm := \frac{1}{\sigma_d} \left( G_{n,\varepsilon_n}^{\text{per}}(\Omega) \pm \sqrt{\frac{4C_d \text{per}(\Omega)}{n\varepsilon_n}} Z_{\alpha/2} \right).$$

We may equivalently write this inclusion as  $\sqrt{\text{per}(\Omega)} \in (\tilde{a}_n^-, \tilde{a}_n^+)$  for

$$\tilde{a}_n^\pm := \sqrt{\frac{C_d}{n\varepsilon_n} Z_{\alpha/2}^2 + \frac{G_{n,\varepsilon_n}^{\text{per}}(\Omega)}{\sigma_d}} \pm \sqrt{\frac{C_d}{n\varepsilon_n}} Z_{\alpha/2},$$

and thereby obtain an expression for  $\text{per}(\Omega)$  involving only observed quantities.

For  $d \geq 5$ , the confidence intervals cannot be obtained unless one has some extra quantitative information on the smoothness of  $\partial\Omega$ , such as curvature bounds or upper bounds on the bias. Nevertheless, since it is known that the nonlocal perimeter  $\text{per}_{\varepsilon_n}(\Omega) = \mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega))$  is less than  $\sigma_d \text{per}(\Omega)$  (see (2.3)), one can construct a test for the hypothesis that  $\text{per}(\Omega)$  is less than a certain number  $\rho$  without using any quantitative estimates on the smoothness of  $\partial\Omega$ . In this case, we have automatic upper bounds for the bias and may consider the hypothesis testing of  $H_0: \text{per}(\Omega) \leq \rho$ , rather than  $H_A: \text{per}(\Omega) > \rho$ , based on our estimator  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$ . We consider the statistic

$$l_n := \sqrt{\frac{n\varepsilon_n}{4C_d\rho}} (G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \sigma_d\rho).$$

The test consists of

$$\text{accept } H_0 \text{ if } l_n \leq Z_\alpha, \text{ reject otherwise.} \tag{1.12}$$

**Proposition 1.1.** *Assume that  $d \geq 2$ , that  $1/n^{1/d} \ll \varepsilon_n \ll 1$ , and that  $\Omega$  satisfies the regularity hypotheses from Theorem 1.3. Under these hypotheses, the type I error of (1.12) satisfies*

$$\limsup_{n \rightarrow \infty} \mathbb{P}_{H_0}(l_n > Z_\alpha) \leq \alpha,$$

that is, the type I error is asymptotically below  $\alpha$ . The type II error satisfies

$$\mathbb{P}_{H_A}(l_n \leq Z_\alpha) = O\left(\frac{1}{\sqrt{n\varepsilon_n}}\right).$$

**Remark 1.5.** (*Extensions.*) The estimates obtained in Theorems 1.1 and 1.2 are not exclusive to the case where the points are uniformly distributed in the unit cube and to geometric graphs. In fact, with slight modifications to the proofs of Theorems 1.1 and 1.2, we can extend these results to more general situations. For example, if the data points are distributed according to some smooth density  $p$  that is supported on a regular, bounded domain  $D \subset \mathbb{R}^d$  with  $p$  bounded below and above by positive constants, then the results still hold. In this case, the limiting value



of  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  is a *weighted* perimeter  $\text{per}(\Omega, p^2)$  (see, for example, [10]). Convergence is guaranteed for the same scaling for  $\varepsilon_n$  as in the uniform case. The fact that the weight is  $p^2$  and not  $p$  (as may be *a priori* expected) ultimately comes from the fact that a graph cut is a double sum.

Furthermore, if instead of weights  $\mathbf{1}_{\{\|x_i - x_j\| \leq \varepsilon_n\}}$  in the definition of the graph cut, one considers edge weights  $\eta(\|x_i - x_j\|/\varepsilon_n)$ , where  $\eta$  is nonnegative, integrable, and nonincreasing, the results still hold, provided we change the surface tension  $\sigma_d$  with a surface tension associated to  $\eta$  defined as in [10].

**Remark 1.6.** The previous results allow us to construct asymptotic confidence intervals for  $\text{per}(\Omega)$  for an arbitrary set  $\Omega$  with a smooth boundary and  $\text{dist}(\Omega, \partial D) > 0$ , using the estimator  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$ . Nevertheless, since very small sets (in the sense of volume) may have a very large perimeter, the error estimates are not uniform in  $\Omega$  and, thus, the asymptotic confidence intervals may not be of direct practical use. This issue is unavoidable without further assumptions on the set  $\Omega$  or a change of framework for testing. One possible way to restrict the class of  $\Omega$  considered is pointed out in Remark 1.4. In the class of sets  $\Omega$  to which Remark 1.4 applies, the error of approximation is uniformly controlled.

Alternatively, one may consider the *property testing* framework used in [21] and [23], where one could test the property ‘the set  $\Omega$  is close to a set whose perimeter is less than  $\rho$ ’ reliably. In that framework we would not be using the estimator  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  to test whether  $\text{per}(\Omega)$  is below  $\rho$  or not, but rather whether  $\Omega$  is *close* to a set whose perimeter is below  $\rho$  or not. We note that in this paper we have taken the classical hypothesis testing approach.

*Outline.* In Subsection 1.3 we establish (1.2) and provide an outline of the argument behind our main results. We present the proof of Theorem 1.1 in Section 2, while in Subsection 2.1 we show that the scaling is sharp (up to logarithmic corrections) in the sense that if  $n^2\varepsilon_n^{d+1} \rightarrow 0$  then  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  converges in probability to 0 and, hence, does not converge a.s. to the (rescaled) relative perimeter. In Section 3 we prove the results on the asymptotic distribution of the error stated in Theorem 1.3. In Subsection 3.1 we study the type I and type II errors of the hypothesis test of Proposition 1.1. Finally, in Appendix A we deal with the bias estimate from Lemma 1.1.

## 1.2. Discussion

Here we discuss the connections between our work and related works in the literature. First we relate it to other estimators of the perimeter based on a random sample. Then we contrast the type of the convergence and the scaling regimes considered in this paper with the ones needed for the convergence of graph-cut based machine learning algorithms for clustering and related tasks.

The problem of estimating the perimeter of a set  $\Omega$ , based on knowing which points of a random sample  $V_n$  belong to  $\Omega$ , has been considered by a number of authors. Cuevas *et al.* [9] considered estimators of the Minkowski content, which agrees with the perimeter for regular enough sets, but is a less general notion of the perimeter than the one we consider, see (1.1). Their estimator is based on counting the vertices near the boundary (relative to a parameter  $\varepsilon_n$ ), while we ‘count’ the edges of a graph. The error bound obtained was of order  $O(n^{-1/2d})$ . Pateiro-López and Rodríguez-Casal [24] considered a similar estimator and improved the bounds to  $O((\ln n/n)^{1/(d+1)})$  under qualitative regularity conditions (rolling ball conditions). Cuevas *et al.* [8] obtained the convergence of estimators similar to those of [9], when  $\varepsilon_n \gg 1/n^{1/d}$  and under weaker conditions on the regularity of  $\Omega$ , although still not in the full generality we consider in this paper.

Armendáriz *et al.* [4] considered a similar set-up to that of [9] and [24] but with different sampling rates for  $\Omega$  and  $\Omega^c$ : let  $n$  denote the number of sample points in  $\Omega$  and  $k$  the number of sample points in  $\Omega^c$ . Under mild assumptions on the regularity of  $\partial\Omega$ , and under some conditions on  $n$  and  $k$  which include  $1 \gg \varepsilon_{n,k} \gg (1/n^{1/3})$  and  $k \gg (n/\varepsilon_{n,k})^{d/2}$  they obtained the asymptotic distribution of the error under the same scaling in  $n$  and  $\varepsilon$  that we consider in Theorem 1.3. We note that our estimator is different, and also that in [4] a very large number of points  $k$  in  $\Omega^c$  was needed for the consistency to hold. This allowed the authors to obtain the asymptotic distribution of the total error in any dimension, while in our setting we obtain it only in low dimensions. A further difference between their work and ours is that we allow for a wider range in  $\varepsilon$ , namely,  $1 \gg \varepsilon_n \gg (1/n)^{2/(d+1)}$ .

Jiménez and Yukich [17] stated a different use to the point cloud and instead of considering a parameter  $\varepsilon$  to count points close to the boundary of the set or to define a geometric graph, they considered a new estimator based on the Delaunay triangulation induced by the cloud. They obtained results not only on estimating the perimeter of the set, but also integrals of functions over  $\partial\Omega$ .

Kothari *et al.* [21] and Neeman [23] considered an estimator essentially based on the following procedure: pick  $n$  random points uniformly distributed on  $D$  and to each point associate a random direction for a ‘needle’ based at the point with length of order  $\sqrt{\varepsilon_n}$ ; then count how many of the needles touch the boundary of  $\Omega$ . Their main motivation was to consider the perimeter estimation from the viewpoint of *property testing* as introduced in [19]. In that setting, the idea is to produce an algorithm that requires a small number of samples (essentially independent of the dimension  $d$ ) in order to determine if a given set has a small perimeter or is ‘far away’ from a set that has a small perimeter. The authors demonstrated completeness and soundness of the test they designed (the notion of completeness and soundness is as in [19] and [21]). The notion of testing used in their work is one of the main differences with our work since we consider the perimeter testing in the more classical framework of hypothesis testing (Proposition 1.1). We note that the completeness of [21] is analogous to the type I error, but the soundness is fundamentally different from estimating the type II error.

It is also worth mentioning the work of Belkin *et al.* [5] where they considered an algorithm that requires as few samples as possible in order to estimate the perimeter of a convex body. Their results show that there is an algorithm that uses  $O(d^4\gamma^{-2})$  samples to obtain an estimator for the true perimeter of the convex set, with an error of approximation of  $\gamma$ ; this statement holds with high probability. We note the polynomial dependence on the dimension in their estimates.

Let us now contrast the type of the convergence and the scaling regimes that we consider in this paper with the ones needed for the convergence of graph-cut based machine learning algorithms for clustering and related tasks.

Graph-cut based algorithms for tasks such as clustering have played an important role in machine learning [3], [7], [14], [15], [18], [25]–[28]. Data clustering algorithms are called consistent if as the sample size  $n$  tends to  $\infty$  their outputs converge to a desired partitioning of the underlying measure being sampled. It is of interest to understand under what scaling of  $\varepsilon_n$  on  $n$  does the consistency hold. Here we showed that, for a *fixed* set  $\Omega$ , the value of graph perimeter (and indeed of the graph-cut based objective functionals such as Cheeger, ratio, or normalized cuts) converges to the perimeter of  $\Omega$ ; this result holds even for rather sparse graphs (Theorem 1.2). In particular, partitioning such sparse graphs (as in Figure 2) does not provide any meaningful information about the clusters present in the data. We conclude that the convergence of the graph perimeter of a fixed set towards the continuum perimeter

does not provide the needed information on the asymptotic properties of graph-cut based clustering algorithms. To obtain consistency of such algorithms we need a stronger notion of convergence of graph based functionals towards continuum functionals. Recently the authors of [10], together with Laurent and Bresson [11], have developed the appropriate notion of convergence (based on  $\Gamma$ -convergence from the calculus of variations), and have applied it to consistency of Cheeger, ratio, sparsest, and normalized cut based point cloud clustering. More specifically, in [11] it was shown that consistency holds if  $(\log(n))^{1/d}/n^{1/d} \ll \varepsilon \ll 1$  (for  $d \geq 3$ ), in that regime the graphs are connected with high probability as  $n \rightarrow \infty$ .

### 1.3. Outline of the argument

In order to understand the asymptotic behavior of the graph perimeter, following [3] we first define a symmetric kernel  $\phi_\varepsilon : D \times D \rightarrow (0, \infty)$  by

$$\phi_{\varepsilon_n}(x, y) = \frac{\mathbf{1}_{\{\|x-y\| \leq \varepsilon_n\}}}{\varepsilon_n^{d+1}} |\mathbf{1}_\Omega(x) - \mathbf{1}_\Omega(y)|.$$

Using the kernel  $\phi_{\varepsilon_n}$ , we can then write  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  as

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \phi_{\varepsilon_n}(x_i, x_j), \tag{1.13}$$

which is a U-statistic in the terminology of [16]. A simple computation shows that the mean of this U-statistic is the nonlocal perimeter  $\text{per}_\varepsilon(\Omega)$  defined in (1.6), that is,

$$\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) = \text{per}_{\varepsilon_n}(\Omega) = \int_D \int_D \phi_\varepsilon(x, y) \, dx \, dy. \tag{1.14}$$

Additionally, Remark 4.3 of [10] establishes that the nonlocal perimeter  $\text{per}_{\varepsilon_n}(\Omega)$  approaches a constant multiple of the relative perimeter of  $\Omega$  as the parameter  $\varepsilon_n$  goes to 0. More precisely, if  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  then

$$\text{per}_{\varepsilon_n}(\Omega) \rightarrow \sigma_d \text{per}(\Omega) \quad \text{as } n \rightarrow \infty \tag{1.15}$$

for  $\sigma_d$  the surface tension (1.3). This convergence also follows from the estimates in Appendix A in the special case that  $\Omega$  has a smooth boundary. Combining (1.15) with (1.14), we conclude that if  $\varepsilon_n$  converges to 0 as  $n \rightarrow \infty$  then

$$\mathbb{E}(G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \rightarrow \sigma_d \text{per}(\Omega) \quad \text{as } n \rightarrow \infty. \tag{1.16}$$

Since the graph perimeter  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  is a U-statistic of order two, we can use the general theory of U-statistics to obtain moment estimates for  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$ . Let us first note that Hoeffding’s decomposition theorem for U-statistics of order two (see [20]) implies that  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  can be written as

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega) = 2U_{n,1} + U_{n,2}, \tag{1.17}$$

where  $U_{n,1}$  is a U-statistic of order one (just a sum of centered independent random variables) and  $U_{n,2}$  is a U-statistic of order two, which is *canonical* or completely degenerate (see [20]). In order to define the variables  $U_{n,1}$  and  $U_{n,2}$ , let us introduce the functions

$$\begin{aligned} \bar{\phi}_{\varepsilon_n}(x) &:= \int_D \phi_{\varepsilon_n}(x, z) \, dz \quad \text{and} \quad g_{n,1}(x) := \bar{\phi}_{\varepsilon_n}(x) - \text{per}_{\varepsilon_n}(\Omega), \quad x \in D, \\ g_{n,2}(x, y) &:= \phi_{\varepsilon_n}(x, y) - \bar{\phi}_{\varepsilon_n}(x) - \bar{\phi}_{\varepsilon_n}(y) + \text{per}_{\varepsilon_n}(\Omega), \quad x, y \in D. \end{aligned} \tag{1.18}$$

With the previous definitions, we can now define

$$U_{n,1} = \frac{1}{n} \sum_{i=1}^n g_{n,1}(x_i), \quad U_{n,2} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{n,2}(x_i, x_j). \tag{1.19}$$

We remark that  $\int_D g_{n,1}(z) dz = 0$  and that  $\int_D g_{n,2}(x, z) dz = 0$  for all  $x \in D$ . Because of this,  $U_{n,1}$  and  $U_{n,2}$  are said to be canonical statistics of order one and two, respectively (see [20]). Bernstein’s inequality [6] applied to  $U_{n,1}$  yields

$$\mathbb{P}(|U_{n,1}| > t) \leq 2 \exp\left(\frac{-3n^2 t^2}{6B_{n-1}^2 + 2tA_{n,1}}\right),$$

where

$$A_{n,1} := \|g_{n,1}\|_\infty, \quad B_{n,1} := \sqrt{n}\|g_{n,1}\|_2. \tag{1.20}$$

In turn,

$$\mathbb{E}(|U_{n,1}|^p) = p \int_0^\infty \mathbb{P}(|U_{n,1}| > t) t^{p-1} dt \leq \frac{C_p}{n^p} \max(A_{n,1}^p, B_{n,1}^p) \tag{1.21}$$

for  $C_p$  a universal constant. See also [12] for a slight generalization of the previous moment estimates.

On the other hand,

$$\mathbb{E}(|U_{n,2}|^p) \leq \frac{C_p}{n^{2p}} \max(A_{n,2}^p, B_{n,2}^p, C_{n,2}^p), \tag{1.22}$$

where

$$A_{n,2} := \|g_{n,2}\|_\infty, \quad B_{n,2} := n\|g_{n,2}\|_2, \quad (C_{n,2})^2 := n \left\| \int_D g_{n,2}^2(\cdot, y) dy \right\|_\infty, \tag{1.23}$$

and  $C_p$  is a universal constant. This follows from Equation (3.3) of [12], which can be thought of as an analogue of (1.21) for U-statistics of order two.

**Remark 1.7.** The content of Theorem 3.2. of [12] provides better constants in the  $p$ -moment estimates for canonical U-statistics of order two than the estimates in Equation (3.3) of the same paper. As we do not keep track of the dependence of  $C_p$  on  $p$ , Equation (3.3) of [12] suffices for our purposes.

From the decomposition (1.17), it follows that, for  $p \geq 1$ ,

$$\mathbb{E}(|G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega)|^p) \leq C_p (\mathbb{E}(|U_{n,1}|^p) + \mathbb{E}(|U_{n,2}|^p))$$

for a universal constant  $C_p$ . Thus, in order to obtain the moment estimates for  $G_{n,\varepsilon_n}^{\text{per}}(\Omega)$  in Theorem 1.1, we focus on finding estimates for the quantities in (1.20) and (1.23).

**Remark 1.8.** The estimates on  $U_{n,1}$  and  $U_{n,2}$  exhibit a crossover in their behavior when the parameter  $\varepsilon_n$  transitions between the sparse ( $1/n^{2/(d+1)} \ll \varepsilon_n \ll 1/n^{1/d}$ ) and the dense regime ( $1/n^{1/d} \ll \varepsilon_n \ll 1$ ). From the theory of U-statistics, the crossover in the nature of the bounds is connected to the different nature of the two components in the canonical decomposition for U-statistics. Under the dense regime, the greatest source of error comes from the term  $U_{n,1}$ , while in the sparse regime the greatest source of error comes from the

term  $U_{n,2}$ . The two variables  $U_{n,1}$  and  $U_{n,2}$  exhibit different behaviors. In fact, one can think of  $U_{n,1}$  as a global quantity as it is the sum of averages, while  $U_{n,2}$  is simply the sum of pure interactions concentrating on the boundary of the set  $\Omega$ . We believe that there is a deeper geometric and analytic reason for the scaling of the error that appears in the sparse regime ( $1/n^{2/(d+1)} \ll \varepsilon_n \ll 1/n^{1/d}$ ). Moreover, we believe that such an understanding would allow us to complete the study of the asymptotic distribution in Theorem 1.3 for the sparse regime. We expect such a distribution to be of the Gaussian chaos type.

The bias estimates in Appendix A are obtained by a series of computations whose starting point is writing  $\text{per}_\varepsilon(\Omega)$  in terms of an iterated integral, the outer one taken over the manifold  $\partial\Omega$  and the inner one taken along the normal line to  $\partial\Omega$  at an arbitrary point  $x \in \partial\Omega$ . Such computations show that the first-order term of  $\text{per}_{\varepsilon_n}(\Omega)$  on  $\varepsilon_n$  vanishes.

Finally, Theorem 1.3 is obtained by using the canonical decomposition of U-statistics and by noting that in the dense regime  $1/n^{1/d} \ll \varepsilon_n \ll 1$ , the variable  $U_{n,2}$  is negligible in relation to  $U_{n,1}$ . We make use of the central limit theorem for triangular arrays after computing the variances of the involved variables.

### 2. Proof of Theorem 1.1

We first compute the moments of  $U_{n,1}$  and so we start by computing the quantities  $A_{n,1}$  and  $B_{n,1}$  from (1.20). Denote by  $T_\varepsilon$  the  $\varepsilon$ -tube around  $\partial\Omega$ , that is, consider the set

$$T_\varepsilon := \{x \in \mathbb{R}^d : \text{dist}(x, \partial\Omega) \leq \varepsilon\}.$$

We also consider the half-tubes  $T_\varepsilon^-$  and  $T_\varepsilon^+$ ,

$$T_\varepsilon^- := \{x \in \Omega : \text{dist}(x, \partial\Omega) \leq \varepsilon\}, \quad T_\varepsilon^+ := \{x \in \Omega^c : \text{dist}(x, \partial\Omega) \leq \varepsilon\}. \tag{2.1}$$

With these definitions it is straightforward to check that

$$\bar{\phi}_{\varepsilon_n}(x) = \begin{cases} \frac{|B_d(x, \varepsilon_n) \cap \Omega|}{\varepsilon_n^{d+1}} & \text{if } x \in T_{\varepsilon_n}^+, \\ \frac{|B_d(x, \varepsilon_n) \cap \Omega^c|}{\varepsilon_n^{d+1}} & \text{if } x \in T_{\varepsilon_n}^-, \\ 0 & \text{if } x \notin T_{\varepsilon_n}. \end{cases} \tag{2.2}$$

Since  $|B_d(x, \varepsilon_n) \cap \Omega|$  and  $|B_d(x, \varepsilon_n) \cap \Omega^c|$  are bounded by  $\omega_d \varepsilon_n^d$ , where  $\omega_d$  is the volume of the  $d$ -dimensional unit ball, we deduce that

$$A_{n,1} = O\left(\frac{1}{\varepsilon_n}\right).$$

In order to compute the quantity  $B_{n,1}$ , we use the following lemma, whose proof may be found in Appendix B.

**Lemma 2.1.** *Let  $p \geq 1$  and let  $\Omega \subseteq D$  be a set with finite perimeter. Then, for all  $\varepsilon > 0$ , we have*

$$\int_D \bar{\phi}_\varepsilon^p(x) \, dx \leq \frac{\omega_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \text{per}(\Omega).$$

In particular, taking  $p = 1$  in the previous expression, we obtain

$$\text{per}_\varepsilon(\Omega) \leq \sigma_d \text{per}(\Omega). \tag{2.3}$$

Using the previous lemma with  $p = 2$ , we deduce that  $\int_D \bar{\phi}_{\varepsilon_n}^2(x) dx = O(1/\varepsilon_n)$ , and since

$$\int_D g_{n,1}^2(x) dx = \int_D \bar{\phi}_{\varepsilon_n}^2(x) dx - (\text{per}_{\varepsilon_n}(\Omega))^2,$$

we conclude that

$$B_{n,1} = O\left(\sqrt{\frac{n}{\varepsilon_n}}\right).$$

From the previous computations, we deduce that

$$\mathbb{E}(|U_{n,1}|^p) \leq C_p \max\{1, \text{per}(\Omega)\}^p \max\left(\frac{1}{n^p \varepsilon_n^p}, \frac{1}{n^{p/2} \varepsilon_n^{p/2}}\right),$$

where  $C_p$  depends on  $p$ , but is independent of the set  $\Omega$ . If  $1/n^{2/(d+1)} \leq \varepsilon_n$ , so that, in particular,  $1/n\varepsilon_n$  is  $o(1)$ , then

$$\mathbb{E}(|U_{n,1}|^p) \leq \frac{C_p \max\{1, \text{per}(\Omega)\}^p}{n^{p/2} \varepsilon_n^{p/2}}. \tag{2.4}$$

**Remark 2.1.** Later on, in Lemma 3.1, we provide an explicit computation of  $\int_D \bar{\phi}_{\varepsilon_n}^2(x) dx$  up to order  $1/\varepsilon_n$ , which is useful when studying the asymptotic distribution of a rescaled version of  $U_{n,1}$ .

Now we turn to the task of obtaining moment estimates for  $U_{n,2}$ . We estimate the quantities  $A_{n,2}$ ,  $B_{n,2}$ , and  $C_{n,2}$  from (1.23). Let us start by estimating  $A_{n,2}$ . Note that, for any  $(x, y) \in D \times D$ ,  $\bar{\phi}_{\varepsilon_n}(x)$  and  $\bar{\phi}_{\varepsilon_n}(y)$  are of order  $1/\varepsilon_n$  and that  $\text{per}_{\varepsilon_n}(\Omega)$  is of order one. Thus, it is clear from the definition of  $g_{n,2}$  in (1.18) that

$$A_{n,2} = O\left(\frac{1}{\varepsilon_n^{d+1}}\right).$$

On the other hand, using  $\phi_{\varepsilon_n}^2(x, y) = (1/\varepsilon_n^{d+1})\bar{\phi}_{\varepsilon_n}(x, y)$ , we obtain, for every  $x \in D$ ,

$$\begin{aligned} & \int_D g_{n,2}^2(x, y) dy \\ &= \int_D \phi_{\varepsilon_n}^2(x, y) dy - \bar{\phi}_{\varepsilon_n}^2(x) + 2\theta_n \bar{\phi}_{\varepsilon_n}(x) - 2 \int_D \phi_{\varepsilon_n}(x, y) \bar{\phi}_{\varepsilon_n}(y) dy + \int_D \bar{\phi}_{\varepsilon_n}^2(y) dy - \theta_n^2 \\ &= \frac{1}{\varepsilon_n^{d+1}} \bar{\phi}_{\varepsilon_n}(x) - \bar{\phi}_{\varepsilon_n}^2(x) + 2\theta_n \bar{\phi}_{\varepsilon_n}(x) - 2 \int_D \phi_{\varepsilon_n}(x, y) \bar{\phi}_{\varepsilon_n}(y) dy + \int_D \bar{\phi}_{\varepsilon_n}^2(y) dy - \theta_n^2, \end{aligned} \tag{2.5}$$

where we are using  $\theta_n := \text{per}_{\varepsilon_n}(\Omega)$ . From this, it follows that

$$C_{n,2} = O\left(\sqrt{\frac{n}{\varepsilon_n^{d+2}}}\right).$$

Finally, upon integration of (2.5), and direct computations, we obtain

$$\|g_{n,2}\|_2^2 = \frac{\theta_n}{\varepsilon_n^{d+1}} - 2 \int_D \bar{\phi}_{\varepsilon_n}^2(y) dy + \theta_n^2,$$

which implies that

$$B_{n,2} = O\left(\frac{n}{\varepsilon_n^{(d+1)/2}}\right).$$

Thus, from (1.22), we deduce that

$$\mathbb{E}(|U_{n,2}|^p) \leq K_p \max\left(\frac{1}{n^{2p} \varepsilon_n^{p(d+1)}}, \frac{1}{n^p \varepsilon_n^{p(d+1)/2}}, \frac{1}{n^{3p/2} \varepsilon_n^{p(d+2)/2}}\right),$$

where  $K_p = C_p(\max\{1, \text{per}(\Omega)\})^p$  for  $C_p$  some constant that does not depend on the set  $\Omega$ . Hence, if  $1/n^{2/(d+1)} \leq \varepsilon_n$ , we have

$$\mathbb{E}(|U_{n,2}|^p) \leq \frac{K_p}{n^p \varepsilon_n^{p(d+1)/2}}. \tag{2.6}$$

Combining (2.4) and (2.6) and using the canonical decomposition (1.17), we obtain (1.5).

**2.1. Sharpness of the rate in Theorem 1.1**

A very simple argument shows that the rates for  $\varepsilon_n$  that guarantee the almost sure convergence of the graph perimeter to the actual perimeter in Theorem 1.1 are optimal in terms of scaling, up to logarithmic corrections.

In fact, suppose that  $n^2 \varepsilon_n^{d+1} = o(1)$  and let  $e_n$  denote the random variable that counts the number of edges that cross in the interface between  $\Omega$  and its complement. In other words, we define

$$e_n := \varepsilon_n^{d+1} \sum_{i=1}^n \sum_{j=i+1}^n \phi_{\varepsilon_n}(x_i, x_j).$$

As a consequence, if  $\Omega$  has a finite perimeter then we have

$$G_{n,\varepsilon_n}^{\text{per}}(\Omega) = \frac{2}{n(n-1)\varepsilon_n^{d+1}} e_n, \quad \mathbb{E}(e_n) = \frac{n(n-1)\varepsilon_n^{d+1}}{2} \text{per}_{\varepsilon_n}(\Omega). \tag{2.7}$$

Note that  $e_n$  takes integer values in the range  $\{0, 1, \dots, N\}$  for  $N = n(n-1)/2$ , so that

$$\mathbb{E}(e_n) = \sum_{k=1}^N k p_k^n, \quad p_k^n := \mathbb{P}(e_n = k).$$

The fact that  $p_0^n + \dots + p_N^n = 1$  implies that

$$\mathbb{E}(e_n) = \sum_{k=1}^N k p_k^n \geq \sum_{k=1}^N p_k^n = (1 - p_0^n).$$

In particular, from (2.7) and (1.15), we deduce that if  $n^2 \varepsilon_n^{d+1} \rightarrow 0$  and  $\Omega$  has a finite perimeter then

$$(1 - p_0^n) \leq \mathbb{E}(e_n) = o(1).$$

On the other hand, note that, for any given  $\gamma > 0$ , it holds that  $G_{n,\varepsilon_n}^{\text{pem}}(\Omega) > \gamma$  implies that  $e_n \neq 0$ . In turn,

$$\mathbb{P}(G_{n,\varepsilon_n}^{\text{pem}}(\Omega) > \gamma) \leq \mathbb{P}(e_n \neq 0) = 1 - p_0^n = o(1).$$

We conclude that if  $n^2 \varepsilon_n^{d+1} \rightarrow 0$  then  $G_{n,\varepsilon_n}^{\text{pem}}(\Omega)$  converges in probability to 0. Therefore, if  $\Omega$  has a nonzero, finite perimeter then  $G_{n,\varepsilon_n}^{\text{pem}}(\Omega)$  does not converge to  $\sigma_d \text{per}(\Omega)$  in probability (nor a.s., either).

This completes the proof of Theorem 1.1. □

### 3. Proof of Theorem 1.3

The proof of Theorem 1.3 relies on the following lemma, whose proof may be found in Appendix C.

**Lemma 3.1.** *Assume that  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , and that  $\Omega \subseteq D$  is an open set with a smooth boundary so that  $\text{dist}(\Omega, \partial D) > 0$ . Then*

$$\text{var}(g_{n,1}(X_1)) = \frac{C_d \text{per}(\Omega)}{\varepsilon_n} + O(1), \tag{3.1}$$

where  $C_d$  is given by (1.9).

Now we turn our attention to the proof of Theorem 1.3.

*Proof of Theorem 1.3.* Note that, from (1.17)–(1.19), we obtain

$$\sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}} (G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega)) = \sqrt{\frac{\varepsilon_n}{nC_d \text{per}(\Omega)}} \sum_{i=1}^n g_{n,1}(x_i) + \sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}} U_{n,2}.$$

From the moment estimates (2.6), we deduce that

$$\sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}} U_{n,2} \xrightarrow{\mathbb{P}} 0, \tag{3.2}$$

where ‘ $\xrightarrow{\mathbb{P}}$ ’ denotes convergence in probability. On the other hand, we note that, from (3.1),

$$\frac{\text{std}(g_{n,1}(x_1))\sqrt{\varepsilon_n}}{\sqrt{C_d \text{per}(\Omega)}} \rightarrow 1 \quad \text{as } n \rightarrow \infty,$$

where  $\text{std}(g_{n,1}(x_1))$  is the standard deviation of  $g_{n,1}(x_1)$ . Lyapunov’s condition, which is sufficient to allow us to use the central limit theorem for triangular arrays, is easily checked from Lemma 2.1. We deduce that

$$\sqrt{\frac{\varepsilon_n}{nC_d \text{per}(\Omega)}} \sum_{i=1}^n g_{n,1}(X_i) \xrightarrow{w} N(0, 1).$$

Combining with (3.2) and the Slutsky’s theorem, we obtain the desired result. Finally, to obtain the last statement in the theorem, we note that from the bias estimates in Lemma 1.1,

$$\sqrt{n\varepsilon_n} |\text{per}_{\varepsilon_n}(A) - \sigma_d \text{per}(A)| = O(n^{1/2} \varepsilon_n^{5/2}).$$

Under condition (1.10), we conclude that  $\sqrt{n\varepsilon_n} |\text{per}_{\varepsilon_n}(A) - \sigma_\eta \text{per}(A)| \rightarrow 0$ . This implies (1.11). □

#### 3.1. Application to perimeter testing

Here we prove Proposition 1.1. We assume that  $\Omega \subseteq D$  is an open set with a smooth boundary such that  $\text{dist}(\Omega, \partial D) > 0$ . Note that under the null hypothesis, if  $l_n > Z_\alpha$  then

$$Z_\alpha < l_n \leq \sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}} (G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega)),$$



where we used the fact that  $\text{per}_{\varepsilon_n}(\Omega) \leq \sigma_d \text{per}(\Omega)$  by Lemma 2.1. Thus, using Theorem 1.3, we deduce that asymptotically, the type I error of our test is

$$\mathbb{P}_{H_0}(l_n > Z_\alpha) \leq \mathbb{P}\left(\sqrt{\frac{n\varepsilon_n}{4C_d \text{per}(\Omega)}}(G_{n,\varepsilon_n}^{\text{per}}(\Omega) - \text{per}_{\varepsilon_n}(\Omega)) > Z_\alpha\right) \rightarrow \mathbb{P}(N(0, 1) > Z_\alpha) = \alpha,$$

which establishes the first part of Proposition 1.1. In order to compute the type II error of our test, suppose that  $\text{per}(\Omega) = \rho'$ , where  $\rho' > \rho$ . In that case,

$$\begin{aligned} \mathbb{P}_{H_A}(l_n \leq Z_\alpha) &= \mathbb{P}_{H_A}\left(\sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\sigma_d\rho - G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \geq -Z_\alpha\right) \\ &= \mathbb{P}_{H_A}\left(\sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\text{per}_{\varepsilon_n}(\Omega) - G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \geq -Z_\alpha + \sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\text{per}_{\varepsilon_n}(\Omega) - \sigma_d\rho)\right). \end{aligned}$$

Now recall that  $\lim_{n \rightarrow \infty} \text{per}_{\varepsilon_n}(\Omega) = \sigma_d \text{per}(\Omega) = \sigma_d\rho' > \sigma_d\rho$ . In particular, we deduce that

$$\lim_{n \rightarrow \infty} \sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\text{per}_{\varepsilon_n}(\Omega) - \sigma_d\rho) = +\infty.$$

Thus, for large enough  $n$ ,

$$-Z_\alpha + \sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\text{per}_{\varepsilon_n}(\Omega) - \sigma_d\rho) \geq \frac{1}{2}\sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\sigma_d\rho' - \sigma_d\rho).$$

Hence, for large enough  $n$ ,

$$\mathbb{P}_{H_A}(l_n \leq Z_\alpha) \leq \mathbb{P}_{H_A}\left(\sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\text{per}_{\varepsilon_n}(\Omega) - G_{n,\varepsilon_n}^{\text{per}}(\Omega)) \geq \frac{1}{2}\sqrt{\frac{n\varepsilon_n}{4C_d\rho}}(\sigma_d\rho' - \sigma_d\rho)\right).$$

Using the moment estimates from Theorem 1.1 and Markov’s inequality, we deduce that

$$\mathbb{P}_{H_A}(l_n \leq Z_\alpha) = O\left(\frac{1}{(n\varepsilon_n)^{p/2}}\right) \text{ for any } p \geq 1,$$

where the constant in the above  $O$  expression depends on  $p$ .

### Appendix A. Proof of Lemma 1.1

Since  $\text{dist}(\Omega, \partial D) > 0$  and  $\Omega$  has smooth boundary, the relative perimeter of  $\Omega$  with respect to  $D$  in the generalized sense (1.1) simply corresponds to the usual perimeter of  $\partial\Omega$  in the sense that

$$\text{per}(\Omega) = \int_{\partial\Omega} d\mathcal{H}^{d-1} = \mathcal{H}^{d-1}(\partial\Omega).$$

Additionally, for all  $\varepsilon \leq \delta := \text{dist}(\Omega, \partial D)$ , we have

$$\text{per}_\varepsilon(\Omega) = \frac{2}{\varepsilon^{d+1}} \int_{\Omega} |B_d(x, \varepsilon) \cap \Omega^c| dx,$$

where  $B_d(x, r)$  denotes the ball of radius  $r$  in  $\mathbb{R}^d$  centered at  $x$  and  $\Omega^c$  denotes the complement of  $\Omega$  in all of space. Moreover, since  $\partial\Omega$  is a compact smooth manifold, we can assume without

the loss of generality (by taking  $\varepsilon$  small enough) that, for every  $x \in T_\varepsilon$ , there is a unique point  $P(x)$  in  $\partial\Omega$  closest to  $x$ . Furthermore, we can assume that the map  $P$  is smooth. We may further write

$$\text{per}_\varepsilon(\Omega) = \frac{2}{\varepsilon^{d+1}} \int_{T_\varepsilon^-} |B_d(x, \varepsilon) \cap \Omega^c| \, dx,$$

where  $T_\varepsilon^-$  is defined in (2.1). This reformulation makes it natural to write the previous integral as an iterated integral; the outer integral is taken over the manifold  $\partial\Omega$  and the inner integral is taken along the normal line to  $\partial\Omega$  at an arbitrary point  $x$  along the boundary.

To make this idea precise, we first let  $N(x)$  denote the outer unit normal to  $\partial\Omega$  at  $x \in \partial\Omega$  and then consider the transformation  $(x, t) \in \partial\Omega \times (0, 1) \mapsto x - t\varepsilon N(x)$  for all sufficiently small  $\varepsilon$ . The Jacobian of this transformation is equal to  $\varepsilon \det(I + t\varepsilon S_x)$ , where  $S_x$  denotes the shape operator (or second fundamental form) of  $\partial\Omega$  at  $x$ ; see, for example, [13]. For all sufficiently small  $\varepsilon$ , we may therefore conclude that

$$\frac{1}{\varepsilon} \int_{T_\varepsilon^-} |B_d(x, \varepsilon) \cap \Omega^c| \, dx = \int_{\partial\Omega} \left( \int_0^1 |B_d(x - t\varepsilon N(x), \varepsilon) \cap \Omega^c| \det(I + t\varepsilon S_x) \, dt \right) d\mathcal{H}^{d-1}(x).$$

As a consequence, we also have

$$\text{per}_\varepsilon(\Omega) = \frac{2}{\varepsilon^d} \int_{\partial\Omega} \left( \int_0^1 |B_d(x - t\varepsilon N(x), \varepsilon) \cap \Omega^c| \det(I + t\varepsilon S_x) \, dt \right) d\mathcal{H}^{d-1}(x). \tag{A.1}$$

With (A.1) in hand, we may now proceed to establish (1.8) by expanding  $\text{per}_\varepsilon(\Omega)$  in terms of  $\varepsilon$  and appealing to some elementary computations that show that the first-order term in  $\varepsilon$  vanishes.

For a fixed  $x \in \partial\Omega$ , we first wish to understand the behavior of the function

$$g_x(\varepsilon) := \frac{1}{\varepsilon^d} \left( \int_0^1 |B_d(x - t\varepsilon N(x), \varepsilon) \cap \Omega^c| \det(I + t\varepsilon S_x) \, dt \right)$$

for  $\varepsilon$  in the neighborhood of 0. Without loss of generality, we may assume that  $x = 0$ ,  $N(x) = e_d$ , and that around  $x$  the boundary  $\partial\Omega$  coincides with the graph  $\hat{x} = (x_1, \dots, x_{d-1}) \mapsto (\hat{x}, f(\hat{x})) \in \mathbb{R}^d$  of a smooth function  $f(\hat{x})$  that satisfies both  $f(0) = 0$  and  $\nabla f(0) = 0$  simultaneously. By symmetry of the shape operator  $S_x$ , there exists an orthonormal basis for  $\mathbb{R}^{d-1}$  (where we identify  $\mathbb{R}^{d-1}$  with the hyperplane  $\{(\hat{x}, x_d) : x_d = 0\}$ ) consisting of eigenvectors of the shape operator. We let  $v_1, \dots, v_{d-1}$  denote the eigenvectors of  $S_x$  and  $\kappa_1, \dots, \kappa_{d-1}$  the corresponding eigenvalues (also known as principal curvatures). In particular, whenever  $\|\hat{x}\| \leq \varepsilon$ , we have

$$f(\hat{x}) = \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \langle \hat{x}, v_i \rangle^2 + O(\varepsilon^3), \tag{A.2}$$

where curvatures  $\kappa_i = \kappa_i(x)$  and the  $O(\varepsilon^3)$  error term can be uniformly bounded.

With these reductions in place, we first define  $u(\hat{y}) := \sqrt{\varepsilon^2 - \|\hat{y}\|^2}$  and then let

$$h(\hat{y}, t; \varepsilon) := \begin{cases} 2u(\hat{y}) & \text{if } f(\hat{y}) + \varepsilon t < -u(\hat{y}), \\ u(\hat{y}) - \varepsilon t - f(\hat{y}) & \text{if } -u(\hat{y}) \leq f(\hat{y}) + \varepsilon t \leq u(\hat{y}), \\ 0 & \text{otherwise.} \end{cases}$$

A direct calculation then shows that

$$|B_d(x - t\varepsilon N(x), \varepsilon) \cap \Omega^\varepsilon| = \int_{B_{d-1}(0, \varepsilon)} h(\hat{y}, t; \varepsilon) d\hat{y}, \tag{A.3}$$

and an application of (A.2) shows that  $h(\hat{y}, t; \varepsilon) = 2\sqrt{\varepsilon^2 - \|\hat{y}\|^2}$  only if

$$\|\hat{y}\|^2 = \varepsilon^2 - O(\varepsilon^4) \quad \text{and} \quad u(\hat{y}) = O(\varepsilon^2).$$

Therefore, it follows that

$$\int_{B_{d-1}(0, \varepsilon) \cap \{f(\hat{y}) + \varepsilon t < -u(\hat{y})\}} h(\hat{y}, t; \varepsilon) d\hat{y} \leq O(\varepsilon^2) \int_{B_{d-1}(0, \varepsilon) \cap \{\|\hat{y}\| \geq \sqrt{\varepsilon^2 - O(\varepsilon^4)}\}} d\hat{y} = O(\varepsilon^{d+3}).$$

We then let  $A_t^\varepsilon$  denote the set  $A_t^\varepsilon := \{\hat{y} \in B_{d-1}(0, \varepsilon) : -u(\hat{y}) \leq f(\hat{y}) + \varepsilon t \leq u(\hat{y})\}$  and use the previous estimate in (A.3) to uncover

$$|B_d(x - t\varepsilon N(x), \varepsilon) \cap \Omega^\varepsilon| = \int_{B_{d-1}(0, \varepsilon) \cap A_t^\varepsilon} (u(\hat{y}) - \varepsilon t - f(\hat{y})) d\hat{y} + O(\varepsilon^{d+3}). \tag{A.4}$$

We may then note that

$$\det(I + \varepsilon t S_x) = (1 + t\varepsilon \kappa_1) \cdots (1 + t\varepsilon \kappa_{d-1}) = 1 + t\varepsilon H_x + O(\varepsilon^2),$$

where  $H_x := \sum_{i=1}^{d-1} \kappa_i$  represents the mean curvature. Using this fact in (A.4) then yields

$$g_x(\varepsilon) = \frac{1}{\varepsilon^d} \int_0^1 \left( \int_{B_{d-1}(0, \varepsilon) \cap A_t^\varepsilon} u(\hat{y}) - \varepsilon t - f(\hat{y}) d\hat{y} \right) (1 + t\varepsilon H_x) dt + O(\varepsilon^2).$$

Now let  $f^\varepsilon(z) := (1/\varepsilon)f(\varepsilon z)$  and define the corresponding subset  $C_t^\varepsilon$  of  $(0, 1) \times B_{d-1}(0, 1)$  as

$$C_t^\varepsilon := \{(t, z) \in (0, 1) \times B_{d-1}(0, 1) : -\sqrt{1 - \|z\|^2} \leq f^\varepsilon(z) + t \leq \sqrt{1 - \|z\|^2}\},$$

then make the change of variables  $\hat{y} = \varepsilon z$  to yield

$$g_x(\varepsilon) = \int_{C_t^\varepsilon} (\sqrt{1 - \|z\|^2} - t - f^\varepsilon(z))(1 + t\varepsilon H_x) dz dt + O(\varepsilon^2).$$

Recalling (A.2), it follows that

$$f^\varepsilon(z) = \frac{\varepsilon}{2} \sum_{i=1}^{d-1} \kappa_i \langle z, v_i \rangle^2 + O(\varepsilon^2),$$

which then allows us to obtain an expansion of  $g_x(\varepsilon)$  in terms of  $\varepsilon$  according to the relation

$$g_x(\varepsilon) = \int_{C_t^\varepsilon} (\sqrt{1 - \|z\|^2} - t) dt dz + \varepsilon \int_{C_t^\varepsilon} \left( t H_x (\sqrt{1 - \|z\|^2} - t) - \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \langle z, v_i \rangle^2 \right) dt dz + O(\varepsilon^2).$$

The bias estimate (1.8) then directly follows after computing each of these terms individually.

We begin by considering the first term in the expansion, that is,

$$I := \int_{C_t^\varepsilon} (\sqrt{1 - \|z\|^2} - t) dt dz.$$

Given  $\varepsilon > 0$  and  $z \in B_{d-1}(0, 1)$ , define  $c(z) := \max\{-\sqrt{1 - \|z\|^2} - f^\varepsilon(z), 0\}$  and  $C(z) := \min\{\sqrt{1 - \|z\|^2} - f^\varepsilon(z), 1\}$ , so that we may easily write

$$I = \int_{B_{d-1}(0,1)} (C(z) - c(z)) \left( \sqrt{1 - \|z\|^2} - \frac{C(z) + c(z)}{2} \right) dz.$$

On the set where  $c(z) \neq 0$  has measure at most  $O(\varepsilon^2)$ , we easily conclude that

$$I = \int_{B_{d-1}(0,1)} C(z) \left( \sqrt{1 - \|z\|^2} - \frac{C(z)}{2} \right) dz + O(\varepsilon^2).$$

If  $C(z) = 1$  then  $\sqrt{1 - \|z\|^2} - C(z)/2 = \frac{1}{2}(1 - \|z\|^2) + O(\varepsilon^2)$  as well. In any case, it follows that

$$I = \frac{1}{2} \int_{B_{d-1}(0,1)} (1 - \|z\|^2) dz + O(\varepsilon^2) = \frac{\sigma_d}{2} + O(\varepsilon^2).$$

We now proceed to compute the second term in the expansion

$$II := H_x \int_{C_t^\varepsilon} (t\sqrt{1 - \|z\|^2} - t^2) dt dz = H_x \int_{B_{d-1}(0,1)} C^2(z) \left( \frac{\sqrt{1 - \|z\|^2}}{2} - \frac{C(z)}{3} \right) dz + O(\varepsilon^2)$$

and the third term in the expansion

$$III := \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \int_{C_t^\varepsilon} \langle z, v_i \rangle^2 dt dz = \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \int_{B_{d-1}(0,1)} \langle z, v_i \rangle^2 C(z) dz + O(\varepsilon^2)$$

in a similar fashion. We always have  $C(z) = \sqrt{1 - \|z\|^2} + O(\varepsilon)$ , so that

$$II = \frac{H_x}{6} \int_{B_{d-1}(0,1)} (1 - \|z\|^2)^{3/2} dz + O(\varepsilon) = \frac{H_x \text{vol}(\mathcal{S}^{d-2})}{6} \int_0^1 (1 - r^2)^{3/2} r^{d-2} dr + O(\varepsilon).$$

The third term follows similarly by appealing to spherical coordinates, in that we have

$$\begin{aligned} III &= \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \int_{B_{d-1}(0,1)} \sqrt{1 - \|z\|^2} \langle z, v_i \rangle^2 dz + O(\varepsilon) \\ &= \frac{H_x \text{vol}(\mathcal{S}^{d-2})}{2(d-1)} \int_0^1 \sqrt{1 - r^2} r^d dr + O(\varepsilon) \\ &= II + O(\varepsilon) \end{aligned}$$

using integration by parts in the final term. Therefore, we have  $I = \sigma_d/2 + O(\varepsilon^2)$ , and  $II$  and  $III = O(\varepsilon)$ , so that  $g_x(\varepsilon) = \sigma_d/2 + O(\varepsilon^2)$  and

$$\text{per}_\varepsilon(\Omega) = 2 \int_{\partial\Omega} g_x(\varepsilon) d\mathcal{H}^{d-1} = \sigma_d \text{per}(\Omega) + O(\varepsilon^2)$$

as desired.

We may also show that when  $\Omega$  is a fixed ball, say  $\Omega = B_d(x_c, \frac{1}{3})$  for  $x_c \in \mathbb{R}^d$ , the center point of  $D$ , the absolute value of the difference between  $\text{per}_\varepsilon(\Omega)$  and  $\sigma_d \text{per}(\Omega)$  remains bounded from below by  $c\varepsilon^2$  for  $c > 0$  some positive constant. The proof proceeds similarly to the proof of the bias estimate above. In particular, this shows that the bound in Lemma 1.1 is optimal in terms of scaling for general sets with a smooth boundary.  $\square$

### Appendix B. Proof of Lemma 2.1

The proof follows the same argument used in the proof of Theorem 4.1 of [10] or Theorem 6.2 of [1]. We assume that  $\text{dist}(\Omega, \partial D) > 0$ . Such an assumption implies that the perimeter of  $\Omega$  with respect to  $D$ , that is,  $\text{per}(\Omega)$  defined in (1.1) is equal to the perimeter of  $\Omega$  with respect to  $\mathbb{R}^d$ . We remark that a slight modification of the argument we present below proves the result in the general case and, hence, we omit the details (see, for example, the proof of Theorem 4.1 of [10]).

First we prove that, for any function  $u : \mathbb{R}^d \rightarrow [0, 1]$  with  $u \in W^{1,1}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$  and for all  $\varepsilon > 0$ , we have

$$\int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\|x-y\| \leq \varepsilon\}}}{\varepsilon^{d+1}} |u(y) - u(x)| \, dy \right)^p dx \leq \frac{\omega_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \int_{\mathbb{R}^d} \|\nabla u(x)\| \, dx. \tag{B.1}$$

Inequality (B.1) follows from

$$\begin{aligned} & \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\|x-y\| \leq \varepsilon\}}}{\varepsilon^{d+1}} |u(y) - u(x)| \, dy \right)^p dx \\ &= \frac{1}{\varepsilon^p} \int_{\mathbb{R}^d} \left( \int_{B_d(0,1)} |u(x + \varepsilon h) - u(x)| \, dh \right)^p dx \\ &\leq \frac{\omega_d^{p-1}}{\varepsilon^p} \int_{\mathbb{R}^d} \int_{B_d(0,1)} |u(x + \varepsilon h) - u(x)|^p \, dh \, dx \\ &\leq \frac{\omega_d^{p-1}}{\varepsilon^p} \int_{\mathbb{R}^d} \int_{B_d(0,1)} |u(x + \varepsilon h) - u(x)| \, dh \, dx \\ &= \frac{\omega_d^{p-1}}{\varepsilon^{p-1}} \int_{\mathbb{R}^d} \int_{B_d(0,1)} \left| \int_0^1 \nabla u(x + t\varepsilon h) \, dt \right| \, dh \, dx \\ &\leq \frac{\omega_d^{p-1}}{\varepsilon^{p-1}} \int_{\mathbb{R}^d} \int_{B_d(0,1)} \int_0^1 |\nabla u(x + t\varepsilon h)| \, dt \, dh \, dx \\ &= \frac{\omega_d^{p-1}}{\varepsilon^{p-1}} \int_0^1 \int_{B_d(0,1)} \int_{\mathbb{R}^d} |\nabla u(x)| \, dx \, dh \, dt \\ &= \frac{\omega_d^{p-1}}{\varepsilon^{p-1}} \int_0^1 \int_{\mathbb{R}^d} \|\nabla u(x)\| \int_{B_d(0,1)} \left| \frac{\nabla u(x)}{\|\nabla u(x)\|} \cdot h \right| \, dh \, dx \, dt \\ &= \frac{\omega_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \int_{\mathbb{R}^d} \|\nabla u(x)\| \, dx, \end{aligned}$$

where in the first equation we used the change of variables  $h = (x - y)/\varepsilon$ , in the first inequality we used Jensen’s inequality, and in the second inequality the fact that  $u$  takes values in  $[0, 1]$ .

Now, for any set  $\Omega \subseteq D$  as in the statement, we can find a sequence of functions  $\{u_k\}_{k \in \mathbb{N}}$  with  $u_k: \mathbb{R}^d \rightarrow [0, 1]$ ,  $u_k \in W^{1,1}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$  and such that

$$u_k \rightarrow \mathbf{1}_\Omega \quad \text{in } L^1(\mathbb{R}^d), \quad \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \|\nabla u_k(x)\| \, dx = \text{per}(\Omega). \tag{B.2}$$

Such a sequence can be obtained, for example, with the aid of standard mollifiers (see Theorem 13.9 of [22]). It follows from (B.1) and (B.2) that

$$\int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\|x-y\| \leq \varepsilon\}}}{\varepsilon^{d+1}} |\mathbf{1}_\Omega(y) - \mathbf{1}_\Omega(x)| \, dy \right)^p \, dx \leq \frac{\omega_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \text{per}(\Omega).$$

Finally, note that

$$\int_D \bar{\phi}_{\varepsilon_n}^p(x) \, dx \leq \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\|x-y\| \leq \varepsilon\}}}{\varepsilon^{d+1}} |\mathbf{1}_\Omega(y) - \mathbf{1}_\Omega(x)| \, dy \right)^p \, dx \leq \frac{\omega_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \text{per}(\Omega). \quad \square$$

### Appendix C. Proof of Lemma 3.1

The proof is based on similar computations to those used in Appendix A and, thus, we simply highlight the main ideas. First of all note that

$$\text{var}(g_{n,1}(X_1)) = \int_D \bar{\phi}_{\varepsilon_n}^2(x) \, dx - (\text{per}_{\varepsilon_n}(\Omega))^2.$$

Since  $\text{per}_{\varepsilon_n}(\Omega) = O(1)$ , our task reduces to computing the integral in the above expression. It follows from (2.2) that, for all small enough  $\varepsilon_n$  (so that  $T_{\varepsilon_n} \subseteq D$ ),

$$\int_D \bar{\phi}_{\varepsilon_n}^2(x) \, dx = \frac{1}{\varepsilon_n^{2(d+1)}} \int_{T_{\varepsilon_n}^-} |B_d(x, \varepsilon_n) \cap \Omega^c|^2 \, dx + \frac{1}{\varepsilon_n^{2(d+1)}} \int_{T_{\varepsilon_n}^+} |B_d(x, \varepsilon_n) \cap \Omega|^2 \, dx.$$

We compute the first of the integrals from the above expression. As in the proof of Lemma 1.1, we write

$$\begin{aligned} & \frac{1}{\varepsilon_n^{2(d+1)}} \int_{T_{\varepsilon_n}^-} |B_d(x, \varepsilon_n) \cap \Omega^c|^2 \, dx \\ &= \frac{1}{\varepsilon_n} \int_{\partial\Omega} \left( \frac{1}{\varepsilon_n^{2d}} \int_0^1 |B_d(x - t\varepsilon_n \mathbf{N}(x), \varepsilon_n) \cap \Omega^c|^2 \det(I + t\varepsilon_n \mathbf{S}_x) \, dt \right) \, d\mathcal{H}^{d-1}(x). \end{aligned}$$

For  $x \in \partial\Omega$ , we study the expression

$$h_{\varepsilon_n}(x) := \int_0^1 \frac{|B_d(x - t\varepsilon_n \mathbf{N}(x), \varepsilon_n) \cap \Omega^c|^2}{\varepsilon_n^{2d}} \det(I + t\varepsilon_n \mathbf{S}_x) \, dt.$$

Note that  $\det(I + t\varepsilon_n \mathbf{S}_x) = (1 + t\varepsilon_n \kappa_1) \cdots (1 + t\varepsilon_n \kappa_{d-1}) = 1 + O(\varepsilon_n)$ , where  $\kappa_1, \dots, \kappa_{d-1}$  are the principal curvatures (eigenvalues of the shape operator  $\mathbf{S}_x$ ). Hence,

$$h_{\varepsilon_n}(x) = \int_0^1 \frac{|B_d(x - t\varepsilon_n \mathbf{N}(x), \varepsilon_n) \cap \Omega^c|^2}{\varepsilon_n^{2d}} \, dt + O(\varepsilon_n).$$

Without loss of generality, we may assume  $x = 0$ ,  $\mathbf{N}(x) = e_d$ , and that around  $x$  the boundary  $\partial\Omega$  coincides with the graph  $\hat{x} = (x_1, \dots, x_{d-1}) \mapsto (\hat{x}, f(\hat{x})) \in \mathbb{R}^d$  of a smooth function  $f(\hat{x})$

that satisfies both  $f(0) = 0$  and  $\nabla f(0) = 0$  simultaneously, and we denote by  $v_1, \dots, v_{d-1}$  the eigenvectors of  $S_x$  (just as in Appendix A). Then  $f: B_{d-1}(0, \varepsilon_n) \rightarrow \mathbb{R}$  satisfies  $f(\hat{x}) = \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \langle \hat{x}, v_i \rangle^2 + O(\varepsilon_n^3)$ .

Now, for fixed  $t \in [0, 1]$ , we define  $H_t$  to be the half-plane

$$H_t := \{x = (\hat{x}, x_d) : x_d \geq t\},$$

and we let

$$A_t := |B_d(0, 1) \cap H_t|.$$

With these definitions, we can write

$$\begin{aligned} |B_d(x - t\varepsilon_n N(x), \varepsilon_n) \cap \Omega^c| &= \varepsilon_n^d A_t - \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \int_{B_{d-1}(0, \varepsilon_n \sqrt{1-t})} \langle \hat{x}, v_i \rangle^2 d\hat{x} + O(\varepsilon_n^{d+2}) \\ &= \varepsilon_n^d A_t - \frac{\sum_{i=1}^{d-1} \kappa_i}{2(d-1)} \int_{B_{d-1}(0, \varepsilon_n \sqrt{1-t})} \|\hat{x}\|^2 d\hat{x} + O(\varepsilon_n^{d+2}) \\ &= \varepsilon_n^d A_t + O(\varepsilon_n^{d+1}), \end{aligned}$$

where the second equality holds due to symmetry, and where the last equality follows after computing  $\int_{B_{d-1}(0, \varepsilon_n \sqrt{1-t})} \|\hat{x}\|^2 d\hat{x}$  using polar coordinates. From the above, it follows that

$$\frac{|B_d(x - t\varepsilon_n N(x), \varepsilon_n) \cap \Omega^c|^2}{\varepsilon_n^{2d}} = A_t^2 + O(\varepsilon_n).$$

We conclude that

$$h_{\varepsilon_n}(x) = \int_0^1 A_t^2 dt + O(\varepsilon_n).$$

Therefore,

$$\frac{1}{\varepsilon_n^{2(d+1)}} \int_{T_{\varepsilon_n}^-} |B_d(x, \varepsilon_n) \cap \Omega^c|^2 dx = \frac{1}{\varepsilon_n} \int_0^1 A_t^2 dt \text{ per}(\Omega) + O(1).$$

Analogously, we can obtain a similar expression for  $(1/\varepsilon_n^{2(d+1)}) \int_{T_{\varepsilon_n}^+} |B_d(x, \varepsilon_n) \cap \Omega|^2 dx$  and from this we deduce (3.1).

### Acknowledgements

D.S. and N.G.T. are grateful to the National Science Foundation (NSF) for its support (grant numbers DMS-1211760 and DMS-1516677). J.v.B. was supported by NSF grant DMS 1312344/DMS 1521138. The authors are grateful to ICERM (supported by NSF grant 0931908), where part of the research was done during the research cluster: *Geometric analysis methods for graph algorithms*. Furthermore, they are grateful to Ki-Net (NSF Research Network Grant RNMS11-07444) for opportunities provided. The authors would like to thank the Center for Nonlinear Analysis of the Carnegie Mellon University for its support.

### References

- [1] ALBERTI, G. AND BELLETTINI, G. (1998). A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *Europ. J. Appl. Math.* **9**, 261–284.
- [2] AMBROSIO, L., FUSCO, N. AND PALLARA, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press.

- [3] ARIAS-CASTRO, E., PELLETIER, B. AND PUDLO, P. (2012). The normalized graph cut and Cheeger constant: from discrete to continuous. *Adv. Appl. Prob.* **44**, 907–937.
- [4] ARMENDÁRIZ, I., CUEVAS, A. AND FRAIMAN, R. (2009). Nonparametric estimation of boundary measures and related functionals: asymptotic results. *Adv. Appl. Prob.* **41**, 311–322.
- [5] BELKIN, M., NARAYANAN, H. AND NIYOGI, P. (2006). Heat flow and a faster algorithm to compute the surface area of a convex body. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, IEEE, New York, pp. 47–56.
- [6] BERNSTEIN, S. (1924). On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine Sect. Math.* **1**, 38–49 (in Russian).
- [7] BRESSON, X., LAURENT, T., UMINSKY, D. AND VON BRECHT, J. H. (2012). Convergence and energy landscape for Cheeger cut clustering. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1394–1402.
- [8] CUEVAS, A., FRAIMAN, R. AND GYÖRFI, L. (2013). Towards a universally consistent estimator of the Minkowski content. *ESAIM Prob. Statist.* **17**, 359–369.
- [9] CUEVAS, A., FRAIMAN, R. AND RODRÍGUEZ-CASAL, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35**, 1031–1051.
- [10] GARCÍA TRILLOS, N. AND SLEPČEV, D. (2016). Continuum limit of total variation on point clouds. *Arch. Rational Mech. Anal.* **220**, 193–241.
- [11] GARCÍA TRILLOS, N. *et al.* (2016). Consistency of Cheeger and ratio graph cuts. *J. Machine Learning Res.* **17**, 181.
- [12] GINÉ, E., LATAŁA, R. AND ZINN, J. (2000). Exponential and moment inequalities for  $U$ -statistics. In *High Dimensional Probability, II* (Seattle, WA, 1999; Progr. Prob. **47**), Birkhäuser, Boston, MA, pp. 13–38.
- [13] GRAY, A. (2004). *Tubes*, 2nd edn. Birkhäuser, Basel.
- [14] HAGEN, L. AND KAHNG, A. B. (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. Comput. Aided Design* **11**, 1074–1085.
- [15] HEIN, M. AND SETZER, S. (2011). Beyond spectral clustering – tight relaxations of balanced graph cuts. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 2366–2374.
- [16] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.
- [17] JIMÉNEZ, R. AND YUKICH, J. E. (2011). Nonparametric estimation of surface integrals. *Ann. Statist.* **39**, 232–260.
- [18] KANNAN, R., VEMPALA, S. AND VETTA, A. (2004). On clusterings: good, bad and spectral. *J. ACM* **51**, 497–515.
- [19] KEARNS, M. AND RON, D. (2000). Testing problems with sublearning sample complexity. *J. Comput. System Sci.* **61**, 428–456.
- [20] KOROLJUK, V. S. AND BOROVSKICH, Y. V. (1994). *Theory of U-Statistics* (Math. Appl. **273**). Kluwer, Dordrecht.
- [21] KOTHARI, P., NAYYERI, A., O'DONNELL, R. AND WU, C. (2014). Testing surface area. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, pp. 1204–1214.
- [22] LEONI, G. (2009). *A First Course in Sobolev Spaces* (Graduate Studies Math. **105**). American Mathematical Society, Providence, RI.
- [23] NEEMAN, J. (2014). Testing surface area with arbitrary accuracy. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, ACM, New York, pp. 393–397.
- [24] PATEIRO-LÓPEZ, B. AND RODRÍGUEZ-CASAL, A. (2008). Length and surface area estimation under smoothness restrictions. *Adv. Appl. Prob.* **40**, 348–358.
- [25] SHI, J. AND MALIK, J. (1997). Normalized cuts and image segmentation. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, New York, pp. 731–737.
- [26] SZLAM, A. AND BRESSON, X. (2010). Total variation and Cheeger cuts. In *ICML'10*, ACM, New York, pp. 1039–1046.
- [27] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Statist. Comput.* **17**, 395–416.
- [28] WEI, Y.-C. AND CHENG, C.-K. (1989). Towards efficient hierarchical designs by ratio cut partitioning. In *1989 IEEE International Conference on Computer-Aided Design*, IEEE, New York, pp. 298–301.