An Exact Formula for the Expected Wire Length Between Two Randomly Chosen Terminals

David M. Lazoff[1] and Alan T. Sherman
Computer Science Department
University of Maryland Baltimore County
Baltimore, Maryland 21228-5398
email: lazoff@umbc8.umbc.edu, sherman@cs.umbc.edu

July 30, 1994

Keywords. Combinatorial problems, computational geometry, computer aided design, geometric probability, geometric Steiner tree problem, layout algorithms, minimum-cost spanning trees, two-terminal wire length problem, very large scale integration (VLSI).

1 Introduction

Motivated by geometric problems in VLSI design, we recently derived a formula for the expected Euclidean distance between two randomly-chosen points uniformly distributed in an arbitrary rectangle [10]. Although we subsequently discovered that Ghosh [6, 7] had published a solution in 1943 in the *Calcutta Statistical Association Bulletin*, we would like to make the formula conveniently available to modern practitioners in a simple, directly usable form. Oblivious to Ghosh's work, and overwhelmed by algebraic difficulties of the problem, previous computer science researchers have resorted to special cases, asymptotic bounds, and numerical approximations. Using elementary techniques, we give an exact closed-form solution involving square roots, natural logarithms, and rational functions of the two rectangle dimensions.

The *Two-Terminal Wire Length Problem (WLP_2)* is to compute the expected Euclidean distance between two randomly-chosen points uniformly distributed in an arbitrary rectangle. This problem characterizes the average wire length needed to connect two randomly chosen terminals on a VLSI chip. This problem also arises as the simplest ($n = 2$) case of the difficult problems of computing the expected cost of minimum-cost Euclidean spanning trees, minimum Euclidean Steiner trees, and traveling salesman paths, connecting n randomly chosen vertices in an arbitrary rectangle. Variations of WLP_2 also arise in the analysis of the ∞-version rectangle heuristic for Minimum Weighted Euclidean Matchings [14, p. 53] and in calculating the quantization error coefficient of a 2-dimensional lattice for the Euclidean Traveling Salesman Problem [5, pp. 72–74]. In addition, WLP_2 arises in cognitive science in the analysis of experiments assessing human memory of spatial relations,[2] and in military simulations of air defense suppression and aircraft survivability [8].

Despite the intuitive nature, apparent simplicity, and wide applicability of WLP_2, the general case of this problem has eluded exact analytical solution by many researchers. For example, in 1965, Gilbert [4] solved only the special case of WLP_2 for the unit square. Similarly, in 1983, Reingold and Supowit [14] solved a variation of WLP_2 only for the $\sqrt{2} \times 1$ rectangle, in which the two points

[1]Member, Applied Physics Laboratory, The Johns Hopkins University.
[2]David Uttal, Department of Psychology, Northwestern University, private correspondence (July 1994).

are required to lie on opposite sides of the rectangle. In 1990, Goddyn [5] analytically solved a much simpler variation of WLP_2 for which one point is fixed at the lower-left corner of a unit square. In his probability textbook, Pitman [12, Prob. 21, p. 356] considers WLP_2 on the unit square and challenges the reader to bound its solution from above and below, admonishing that even for the unit square, WLP_2 "is hard to do exactly by calculus." Several researchers have studied the more general problem of computing the expected costs of random minimum-cost spanning trees, but we are aware only of experimental or asymptotic bounds (for example, see [4, 9, 11]).

Our main result is the following closed-form solution to WLP_2

$$\frac{a^5+b^5-(a^4-3a^2b^2+b^4)\sqrt{a^2+b^2}}{15a^2b^2}+\frac{a^2}{6b}\ln\left(\frac{b+\sqrt{a^2+b^2}}{a}\right)+\frac{b^2}{6a}\ln\left(\frac{a+\sqrt{a^2+b^2}}{b}\right), \tag{1}$$

where a and b are the rectangle dimensions. By evaluating Equation 1, practitioners can solve instances of WLP_2 exactly, rather than resorting to upper and lower bounds or to numerical approximations. Only after deriving Equation 1 did we learn of the 1985 work of Sheng [15], who using the Borel overlap technique, solved WLP_2 and other related problems. In particular, Equation 1 can be produced by a page of routine algebra from Sheng's Equation 2.5.1, which also yields all higher moments of the distance.

The difficulty of WLP_2 is not conceptual—it is simple to understand the problem; it is straightforward to express the solution as a multiple integral; and it is easy to outline a high-level plan for solving the integral. Rather, the difficulty of WLP_2 lies solely in the excruciatingly toilsome algebra necessary to solve the integral to produce a usable formula. Simply stated, the extraordinary *grunge* of WLP_2 has thwarted many previous attempts to solve the problem. It would be natural to solve the integral with a symbolic math package, such as *Macsyma*, *Maple*, or *Mathematica*. Although we did use these packages to verify some of our intermediate calculations, none of these tools were able to solve the integral without extensive hints from our manual solution. Meticulously applying elementary techniques, we relentlessly overcame all algebraic difficulties.

In the rest of this correspondence, we sketch our solution; we verify it using computer simulations; and we reexpress it in terms of aspect ratio and area.

2 Exact Analytical Solution

Let the random variable D denote the Euclidean distance between two independently-chosen points, uniformly distributed over a rectangle of dimensions $a > 0$ and $b > 0$. Using elementary calculus, we compute exact closed-form expressions for the first two moments of D in terms of elementary functions of a and b.

By the definition of expected value, the expected value of D can be expressed as the quadruple integral

$$E[D]=\frac{1}{a^2b^2}\int_0^b\int_0^b\int_0^a\int_0^a\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}\,dx_1\,dx_2\,dy_1\,dy_2. \tag{2}$$

In Equation 2, each integral corresponds to one of the four coordinates needed to describe the two points. The difficulty of WLP_2 arises from the square root in the integrand.

After applying the variable substitutions $u = x_1 - x_2$ and $v = y_1 - y_2$, we solve the resulting simpler quadruple integral, one integral at a time. In this calculation, the six integrals $\int x^k\sqrt{x^2+a^2}$

with $0 \le k \le 3$ and $\int \ln(x + \sqrt{x^2 + a^2})$ and $\int x^2 \ln(a + \sqrt{x^2 + a^2})$ account for most of the calculus. Because the amount of algebra is extensive, we leave the detailed calculations to Appendix B of our technical report [10]. As we expected, the resulting formula for $E[D]$ given in Equation 1 is symmetrical in a and b.

Because squaring the integrand in Equation 2 eliminates the square root, an easy calculation determines the second moment of D to be

$$E[D^2] = \frac{a^2 + b^2}{6}. \tag{3}$$

3 Monte Carlo Simulations

To check our solution, we ran Monte Carlo simulations and compared the resulting sample mean and standard deviations of D with the corresponding exact values given by Equations 1 and 3. We implemented our simulations in a straightforward fashion and ran them on a Silicon Graphics workstation. For both simulations, we used L'Ecuyer's [13, p. 282] pseudorandom number generator with the Bays-Durham shuffle.

Table 1 summarizes the results of some of our simulations. In these simulations, we worked with rectangles of constant area 1 with selected aspect ratios $1 \le a/b \le 1024$. Using 10^6 trials per rectangle, the empirical values agree closely with their corresponding theoretical values; moreover, the extent of this agreement increased with the number of trials. Also, as we expected, the standard deviations in Table 1 become large for long, narrow rectangles. For additional simulations, see [10].

a/b	a	b	Exact D	Empirical D
1	1.000000	1.000000	0.521405 ± 0.247931	0.521598 ± 0.247950
2	1.414214	0.707107	0.569060 ± 0.304693	0.569337 ± 0.304542
4	2.000000	0.500000	0.713743 ± 0.445987	0.712539 ± 0.445795
8	2.828427	0.353553	0.964207 ± 0.651514	0.963945 ± 0.651531
16	4.000000	0.250000	1.342640 ± 0.935094	1.342688 ± 0.934585
32	5.656854	0.176777	1.889535 ± 1.329736	1.888086 ± 1.328148
64	8.000000	0.125000	2.668275 ± 1.884032	2.667324 ± 1.884120
128	11.313708	0.088388	3.771884 ± 2.665994	3.771421 ± 2.665872
256	16.000000	0.062500	5.333591 ± 3.770959	5.340682 ± 3.772555
512	22.627417	0.044194	7.542573 ± 5.333221	7.529635 ± 5.331104
1024	32.000000	0.031250	10.666706 ± 7.542428	10.646558 ± 7.537217

Table 1: Results of Monte Carlo simulations for unit-area rectangles of dimensions a and b, for selected aspect ratios $1 \le a/b \le 1024$. Values of D are listed as means $\pm$ standard deviations. Exact values are computed from Equations 1 and 3; empirical values are based on 10^6 trials for each rectangle.

4 Alternate Form and Special Cases

To interpret our solution further, we express $E[D]$ in terms of the *aspect ratio* $r = a/b$ and *area* $A = ab$ of the rectangle. In addition, we simplify our formula for the special cases when the points are uniformly distributed within a square or along a line segment.

4.1 Solution Expressed in Terms of Aspect Ratio and Area

Substituting $a = \sqrt{Ar}$ and $b = \sqrt{A/r}$ into Equations 1 and 3, and using the inverse hyperbolic substitution $\sinh^{-1} r = \ln(r + \sqrt{r^2+1})$, we obtain

$$E[D] = \sqrt{Ar}\left[\frac{3\sqrt{1+r^{-2}} + r(r - \sqrt{r^2+1}) + r^{-3}(1 - \sqrt{r^2+1})}{15} + \frac{r}{6}\sinh^{-1}(1/r) + \frac{r^{-2}}{6}\sinh^{-1} r\right]. \quad (4)$$

and

$$E[D^2] = \frac{A}{6}(r + r^{-1}). \quad (5)$$

Thus, for any fixed aspect ratio, $E[D]$ grows linearly with the square root of the area. Conversely, for any fixed area, $E[D] \in \Theta(\sqrt{r})$, because $\lim_{r\to\infty} E[D]/\sqrt{r} = \sqrt{A}[(3-0.5+0)/15+(1/6)+0] = \sqrt{A}/3$. Figure 1 shows a 3-dimensional graph of Equation 4 produced by *Maple* [2].

4.2 Special Case: Square

Let D_s denote D when the rectangle is a square. Substituting $b = a$ in Equations 1 and 3 yields

$$E[D_s] = a\left[\frac{2+\sqrt{2}}{15} + \frac{\ln(1+\sqrt{2})}{3}\right] \approx 0.521405a \qquad \text{and} \qquad E[D_s^2] = \frac{a^2}{3}, \quad (6)$$

which agrees with Gilbert's calculation [4, p. 387].

4.3 Special Case: Line Segment

We now consider the degenerate case when the two points are uniformly distributed along a line segment of length a. Let D_l denote the random variable D for this case. A simple integration of polynomials yields

$$E[D_l] = \frac{1}{a^2}\int_0^a\int_0^a |x_1 - x_2|\, dx_2\, dx_1 = \frac{a}{3} \qquad \text{and} \qquad E[D_l^2] = \frac{a^2}{6}. \quad (7)$$

As a final check of Equation 1, we verified that $\lim_{b\to 0} E[D] = E[D_l]$ and $\lim_{b\to 0} E[D^2] = E[D_l^2]$.

5 Conclusion and Open Problem

We have rediscovered an exact closed-form solution to the general case of WLP_2. Our conveniently stated solution has applications in a variety of geometric settings drawn from VLSI design, computational geometry, analysis of algorithms, and cognitive science.

We conclude by posing the following related open problem: Derive an exact, closed-form expression for the expected cost of minimum-cost spanning trees (respectively, minimum Steiner trees) connecting n randomly chosen points in an arbitrary rectangle, for small values of $n > 2$. For $n = 3$, we conjecture that this second problem is solvable by case analysis.

Acknowledgments

For pointing out the prior work of Ghosh [6] and Sheng [15], we thank Peter Ehlers, who also rediscovered the solution to WLP_2 as part of his 1972 thesis. In addition, we thank Konstantinos Kalpakis for helpful comments.

References

[1] Beyer, William H., ed., *CRC Standard Mathematical Tables and Formulae, 29th Edition*, CRC Press (Boca Raton, 1991).

[2] Char, Bruce W.; K. Geddes; G. Gonnet; B. Leong; M. Monagan; and S. Watt, *First Leaves: A Tutorial Introduction to Maple V*, Springer-Verlag (1992).

[3] Crofton, M., "Probability" in *Encyclopaedia Britannica*, 9th edition, **19** (1885), 768–788.

[4] Gilbert, E. N., "Random minimal trees," *J. SIAM*, **13**:2 (June 1965), 376–387.

[5] Goddyn, Luis A., "Quantizers and the worst-case Euclidean Traveling Salesman Problem," *Journal of Combinatorial Theory*, Series B, **50**:1 (October 1990), 65–81.

[6] Ghosh, B., "On the distribution of random distances in a rectangle," *Science and Culture*, **8** (1943), 388.

[7] Ghosh, B., "Random distances within a rectangle and between two rectangles," *Bulletin of the Calcutta Mathematical Society*, **43** (1951), 17–24.

[8] Hoeber, Francis P., *Military Applications of Modeling: Selected Case Studies*, Gordon and Breach Science Publishers (New York, 1982).

[9] Jain, Anjani; and John W. Mamer, "Approximations for the random minimal spanning tree with application to network provisioning," *Operations Research*, **36**:4 (July-August 1988), 575–584.

[10] Lazoff, David M.; and Alan T. Sherman, "An exact formula for the expected wire length between two randomly chosen terminals," Technical Report TR CS-94-08, Computer Science Department, University of Maryland Baltimore County (July 20, 1994). 14 pages.

[11] Lueker, George S., "Optimization problems on graphs with independent random edge weights," *SIAM Journal on Computing*, **10**:2 (May 1981), 338–351.

[12] Pitman, Jim, *Probability*, Springer-Verlag (New York, 1993).

[13] Press, William H.; Saul A. Teukolsky; William T. Vetterling; and Brian P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition, Cambridge University Press (Cambridge, 1992).

[14] Reingold, Edward M.; and Kenneth P. Supowit, "Probabilistic analysis of divide-and-conquer heuristics for minimum weighted Euclidean matching," *Networks*, **13**:1 (spring 1983), 49–66.

[15] Sheng, T. K., "The distance between two random points in plane regions," *Advances in Applied Probability*, **17**:4 (December 1985), 748–773.

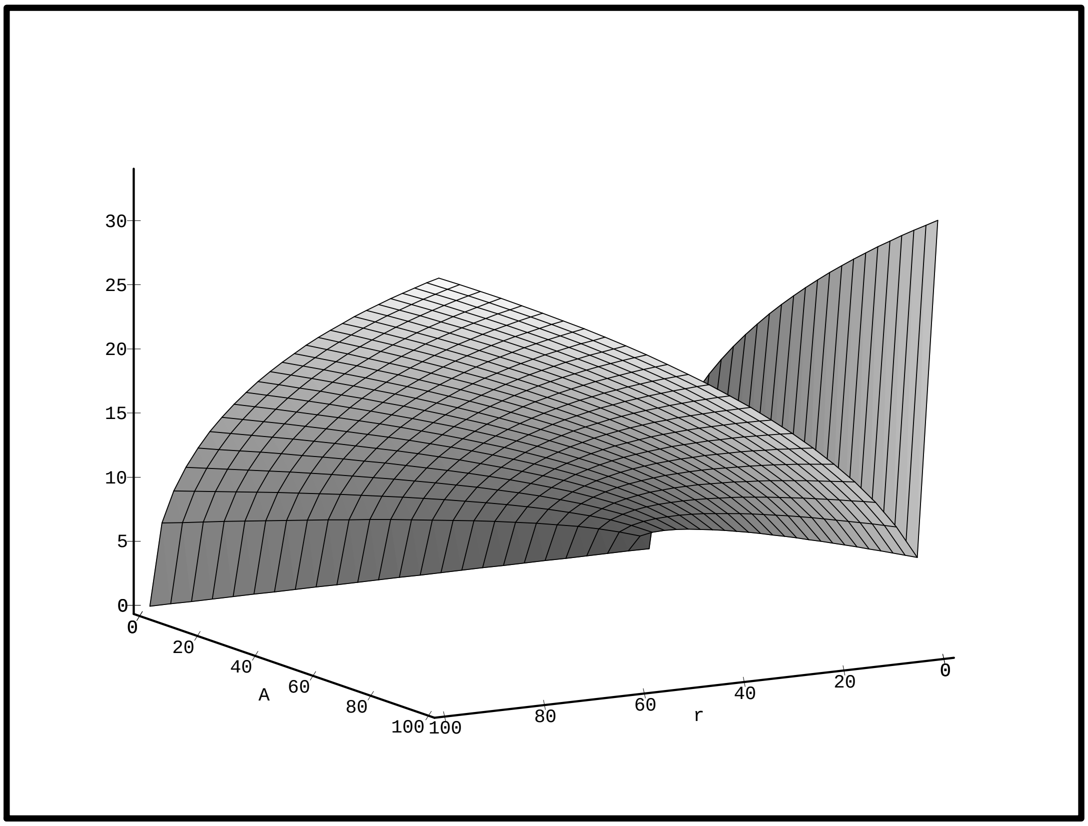

Figure 1: Expected value of D for rectangles of dimensions a and b, as function of aspect ratio $r = a/b$ and rectangle area $A = ab$.