

On the Quality of Partitions Based on Space-Filling Curves

Jan Hungershöfer and Jens-Michael Wierum

Paderborn Center for Parallel Computing, PC²
Fürstenallee 11, 33102 Paderborn, Germany
{hunger, jmwie}@upb.de
www.upb.de/pc2/

Abstract. This paper presents bounds on the quality of partitions induced by space-filling curves. We compare the surface that surrounds an arbitrary index range with the optimal partition in the grid, i. e. the square. It is shown that partitions induced by Lebesgue and Hilbert curves behave about 1.85 times worse with respect to the length of the surface. The Lebesgue indexing gives better results than the Hilbert indexing in worst case analysis. Furthermore, the surface of partitions based on the Lebesgue indexing are at most $\frac{5}{2\sqrt{3}}$ times larger than the optimal in average case.

1 Introduction

Data structures for maintaining sets of multidimensional points play an important role in many areas of computational geometry. While for example Voronoi diagrams have been established for efficient requests on neighborhood relationships, data structures based on space-filling curves are often used for requests on axis aligned bodies of arbitrary size. The aim of the requests is to find all points located in such multidimensional intervals. Those types of requests are needed in many applications like N-body simulations [12], image compression and browsing [10, 4], databases [2], and contact search in finite element analysis [5]. An overview on this and other techniques for range searching in computational geometry is given in [1]. Space-filling curves have other locality properties which are e. g. useful in parallel finite element simulations [3, 7].

Space-filling curves are geometric representations of bijective mappings $M : \{1, \dots, N^m\} \rightarrow \{1, \dots, N\}^m$. The curve M traverses all N^m cells in the m -dimensional grid of size N . An (historic) overview on space-filling curves is given in [11]. Experimental work and theoretical analysis have shown, that algorithms based on space-filling curves behave well on most inputs, while they are not well suited for some special inputs. Therefore, an analysis for the average case is often more important than for the worst case.

Due to the varying requirements on the *locality* properties, different metrics have been used to qualify, compare, and improve space-filling curves. A major metric for the analysis of the locality of space-filling curves is the ratio of index

interval to maximum distance within this index range. Results are published for different indexing schemes like Hilbert, Lebesgue, and H-indexing for Manhattan metric, Euclidean metric, and maximum metric [6, 9].

Other examinations concentrate on the number of index intervals which have to be determined for a given request region. Sharp results are given in [2] for squares in two-dimensional space. The costs for arbitrary shaped regions is discussed in [8].

Here we examine the surface of a partition which is induced by an interval of a space-filling curve. Practical results of this relationship for uniform grids and unstructured meshes can be found in [14]. We define a *quality coefficient* which represents a normed value for the quality of the induced partition in an uniform grid of size $N \times N$. We use the shape of an optimal partition, the square, as a reference:

Definition 1 (quality coefficient). *Let curve be an indexing scheme, p an index range, $S(p)$ the surface of a partition and $V(p) = |p|$ the size (volume) of it. $C^{curve}(p)$ defines the quality coefficient of the partition given by index range p :*

$$C^{curve}(p) = \frac{S^{curve}(p)}{4 \cdot \sqrt{V(p)}} \tag{1}$$

This formulation can be extended to a quality coefficient of an indexing scheme:

$$C_{\max}^{curve} = \max_p \{C^{curve}(p)\} \tag{2}$$

$$C_{\text{avg}}^{curve} = \text{avg}_p \{C^{curve}(p)\} \tag{3}$$

Definition 1 implies that $C(p) \geq 1$ for all indexing schemes.

2 Lebesgue Curves

Figure 1 illustrates the recursive definition of the Lebesgue indexing. The resulting curve is also known as *bit interleaving* or *Z-code*. In the following the edges of cells are assigned a level, depending on the step in which they were introduced during the recursive construction. The lines of the final step are of level 0. In the example shown dashed lines are of level 1, dotted lines of level 0. It is obvious that an arbitrary edge is of level l with an asymptotic probability of $2^{-(l+1)}$.

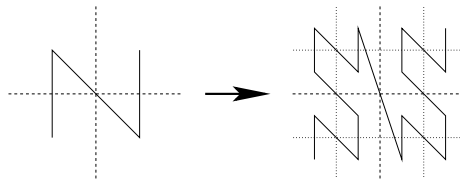


Fig. 1. Production rule for Lebesgue indexing.

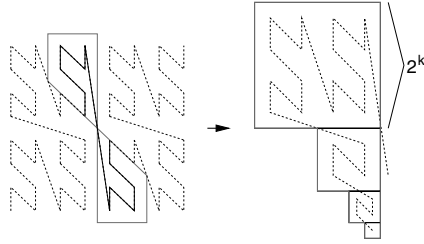


Fig. 2. Construction of a lower bound for Lebesgue indexing.

2.1 Lower Bound on Worst-Case Partitions

Theorem 1. *For the Lebesgue curve the quality coefficient is larger than or equal to $3 \cdot \sqrt{\frac{3}{8}} - \varepsilon$ with decreasing ε for increasing partition size.*

Proof. We construct a partition of size V and surface S which follows the Lebesgue curve and gives the stated bad quality: The symmetric partition is split by a high level border. Each half contains squares of size $4^k, 4^{k-1}, 4^{k-2}, \dots, 4^1, 4^0$. The first half of the partition is illustrated in Fig. 2. It follows:

$$V = 2 \cdot \frac{4^{k+1} - 1}{3} \text{ and } S = 2 \cdot 6 \cdot 2^k - 4 . \tag{4}$$

The quality coefficient of this partition is given by

$$\frac{S}{4 \cdot \sqrt{V}} > \frac{3}{2} \cdot \sqrt{\frac{3}{2}} \cdot \frac{2^k - 4}{2^k} = 3 \cdot \sqrt{\frac{3}{8}} - \varepsilon \approx 1.83 . \tag{5}$$

□

2.2 Upper Bound on Worst-Case Partitions

For the determination of an upper bound we examine partitions which start at the lower left corner of the grid. Due to the construction scheme the partition is always contiguous and its surface is equal to the surface of its bounding box.¹ We analyze the surface of those partitions with respect to a coarse granularity, to be able to examine a finite number of cases.

Lemma 1. *For each partition $p = [1, V]$ induced by the Lebesgue indexing $C_{\max}^{\text{Lebesgue}} \leq \frac{12}{4 \cdot \sqrt{5}}$.*

Proof. For a given partition size V chose $k \in \mathbb{N}$ that $4 \cdot 4^k < V \leq 16 \cdot 4^k$. For each V in the interval we can determine v with $v \cdot 4^k < V \leq (v + 1) \cdot 4^k$. The surface $S(p)$ of V is smaller than or equal to the surface of the partition $[1, (v+1) \cdot 4^k]$. The following table states upper bounds for surfaces of partitions v in granularity 2^k called s with $s \cdot 2^k \geq S$ (values for $v < 4$ are used in Theorem 2):

¹ This fact does not apply to all indexing schemes, e. g. Hilbert indexing.

v	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
s	4	6	8	8	10	12	12	14	14	16	16	16	16	16	16	16

$$\frac{S}{4 \cdot \sqrt{V}} \leq \frac{s}{4 \cdot \sqrt{v}} \leq \frac{12}{4 \cdot \sqrt{5}} \approx 1.34 \tag{6}$$

It is obvious that the equation holds for all unexamined partitions smaller than $4 \cdot 4^0$, too. □

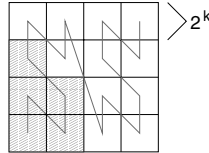


Fig. 3. Partition within a coarse structured Lebesgue indexing.

Example 1. Figure 3 shows the maximal partition induced by the Lebesgue indexing for $v = 4$. All partitions with size V , $4 \cdot 4^k < V \leq 5 \cdot 4^k$ are covered in this case. For all those partitions $S \leq 10 \cdot 2^k$ holds. It follows $s = 10$.

For the analysis of an arbitrary partition p within the Lebesgue indexing we use the fact that the curve is split at most into two sub-curves. The second part only has cells which lie in the same or a more right column *and* in the same or an upper row. It behaves like the partitions examined in Lemma 1. The same holds for the first part due to the symmetry of the curve. A partition p is examined as partitions p_1 and p_2 with $p = p_1 \circ p_2$ and $V = V_1 + V_2$. Again, the analysis is done on the coarse granularity used above.

Theorem 2. $C_{\max}^{\text{Lebesgue}} \leq \frac{7}{2 \cdot \sqrt{3}}$.

Proof. From $V = V_1 + V_2$ follows $3 \leq v_1 + v_2 \leq 15$ with $v_1, v_2 \in [0, 15]$. For the quality coefficient holds

$$C(p_1 \circ p_2) \leq \frac{s_1 + s_2}{4 \cdot \sqrt{v_1 + v_2}} . \tag{7}$$

The enumeration of all possible combinations for v_1 and v_2 shows that the maximum is achieved for $v_1 = 1$ and $v_2 = 2$. It follows $s_1 = 6$ and $s_2 = 8$ (compare table of Lemma 1) and

$$C(p_1 \circ p_2) \leq \frac{6 + 8}{4 \cdot \sqrt{1 + 2}} = \frac{7}{2 \cdot \sqrt{3}} \approx 2.02 . \tag{8}$$

□

The analysis of the upper bound is an enumeration of a finite number of cases with a maximum determination. We can shift the examined interval of

partition size V by a refinement of the underlying granularity with the help of computational evaluation. For the listed refinement steps the result improves towards the following values:

$$4 \cdot 4^k < V \leq 16 \cdot 4^k \Rightarrow C_{\max}^{\text{Lebesgue}} \leq \frac{6 + 8}{4 \cdot \sqrt{1 + 2}} < 2.021 \tag{9}$$

$$16 \cdot 4^k < V \leq 64 \cdot 4^k \Rightarrow C_{\max}^{\text{Lebesgue}} \leq \frac{24 + 24}{4 \cdot \sqrt{21 + 21}} < 1.852 \tag{10}$$

$$256 \cdot 4^k < V \leq 1024 \cdot 4^k \Rightarrow C_{\max}^{\text{Lebesgue}} \leq \frac{192 + 192}{4 \cdot \sqrt{1365 + 1365}} < 1.838 \tag{11}$$

Corollary 1. *For the quality coefficient of the Lebesgue indexing holds:*

$$1.837 < 3 \cdot \sqrt{\frac{3}{8}} - \varepsilon \leq C_{\max}^{\text{Lebesgue}} \leq \frac{96}{\sqrt{2730}} < 1.838 \tag{12}$$

2.3 Upper Bound in Average Case

In this section we will focus on the average case. As stated in the introduction, most algorithms based on space-filling curves profit from a good behavior in the average case of all performed operations. Due to space limitations we present an asymptotical estimation. An exact but rather complex solution is presented in [13].

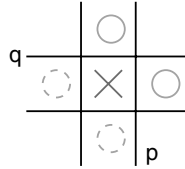


Fig. 4. Neighboring levels for an arbitrary cell within a grid structured by Lebesgue indexing.

For the evaluation of the surface the number of edges common to two cells is needed. It has to be subtracted twice from the number of all edges $4 \cdot V$. A cell has an inner edge on the right hand or upper side, if the index of the right or upper cell is small enough to be still a member of the partition. Given an arbitrary situation illustrated in Fig. 4 with levels p on the right hand and q on the upper side, the indices of the right and upper cell are $R_q = 4 \cdot \frac{4^q - 1}{3} + 2$ and $U_p = 2 \cdot \frac{4^p - 1}{3} + 1$.

Lemma 2. *For the surface of a partition induced by the Lebesgue curve holds in average case:*

$$S \leq \frac{3}{2^k} V + \frac{8}{3} \cdot 2^k - \frac{5}{3} \cdot \frac{1}{2^k} \quad \text{for } V \in \left[2 \frac{4^k - 1}{3} + 1, 4 \frac{4^k - 1}{3} + 2 \right] \quad \text{and}$$

$$S \leq \frac{2}{2^k} V + 4 \cdot 2^k - \frac{1}{2^k} \quad \text{for } V \in \left[4 \frac{4^k - 1}{3} + 2, 2 \frac{4^{k+1} - 1}{3} + 1 \right], \quad \text{with } k \in \mathbb{N}_0 .$$

Proof. The number of cells with a neighbor at level l is given by $\max\{V - R_l, 0\}$ and $\max\{V - U_l, 0\}$ for right and upper neighbor, resp. These terms can be used for a summation on all levels and its corresponding probabilities.

$$S \leq 4V - 2 \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} \max\{V - U_k, 0\} - 2 \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} \max\{V - R_k, 0\} \quad (13)$$

For further examinations of this formulation the evaluated space is split into two classes of intervals: $I_1 = [2\frac{4^k-1}{3} + 1, 4\frac{4^k-1}{3} + 2[$ and $I_2 = [4\frac{4^k-1}{3} + 2, 2\frac{4^{k+1}-1}{3} + 1[$. The size of the surface for all partitions p in intervals of class I_1 is given by:

$$\begin{aligned} S &\leq 4V - \sum_{i=0}^k \frac{1}{2^i} \left(V - 2\frac{4^i-1}{3} - 1 \right) - \sum_{i=0}^{k-1} \frac{1}{2^i} \left(V - 4\frac{4^i-1}{3} - 2 \right) \\ &= 4V - \sum_{i=0}^k \left(\left(V - \frac{1}{3} \right) \frac{1}{2^i} - \frac{2}{3} \cdot 2^i \right) - \sum_{i=0}^{k-1} \left(\left(V - \frac{2}{3} \right) \frac{1}{2^i} - \frac{4}{3} \cdot 2^i \right) \\ &= 4V - \left(V - \frac{1}{3} \right) \left(2 - \frac{1}{2^k} \right) + \frac{2}{3} (2^{k+1} - 1) \\ &\quad - \left(V - \frac{2}{3} \right) \left(2 - \frac{1}{2^{k-1}} \right) + \frac{4}{3} (2^k - 1) \\ &= \frac{3}{2^k} V + \frac{8}{3} 2^k - \frac{5}{3} \frac{1}{2^k} \end{aligned} \quad (14)$$

Using the same arithmetic technique, for interval class I_2 holds:

$$\begin{aligned} S &\leq 4V - \sum_{i=0}^k \frac{1}{2^i} \left(V - 2\frac{4^i-1}{3} - 1 \right) - \sum_{i=0}^k \frac{1}{2^i} \left(V - 4\frac{4^i-1}{3} - 2 \right) \\ &= \frac{2}{2^k} V + 4 \cdot 2^k - \frac{1}{2^k} \end{aligned} \quad (15)$$

□

It has to be kept in mind that the occurrence of the different edges of level l is not exactly $p(l) = 1/2^{l+1}$. For $l = 0$ the possibility is larger than $p(0)$, for all other levels it is smaller than $p(l)$. This results in an underestimation of inner edges and therefore in an overestimation for the size of the surface. For large grids the calculated values converge to the exact solutions (comp. [13]). However, the quality of the given estimation does *not* depend on the size of the partition.

Theorem 3. *The quality coefficient of the average case for the Lebesgue indexing scheme is less than or equal to $\frac{5}{2\sqrt{3}}$.*

Proof. For the determination of the upper bound for the average case the limits of the intervals of classes I_1 and I_2 has to be examined. For the upper limit of I_1 we get:

$$V = 4 \frac{4^{k-1} - 1}{3} + 2 \Leftrightarrow \sqrt{3V - 2} = 2^k \tag{16}$$

Using the result of Lemma 2 gives:

$$\begin{aligned} S &\leq \frac{4}{2^k} V + 2 \left(2^k - \frac{1}{2^k} \right) = \frac{4V - 2}{\sqrt{3V - 2}} + 2 \cdot \sqrt{3V - 2} = \frac{10V - 6}{\sqrt{3V - 2}} \\ &\leq \frac{10}{\sqrt{3}} \sqrt{V} \end{aligned} \tag{17}$$

The corresponding lower limit of I_1 (eq. to upper limit of I_2) results in:

$$S \leq \frac{7\sqrt{2}}{\sqrt{3}} \cdot \sqrt{V} \tag{18}$$

The surface size is obviously larger in the first case. The quality coefficient for the average case is:

$$C_{\text{avg}}^{\text{Lebesgue}} \leq \frac{10}{\sqrt{3}} \cdot \frac{1}{4} = \frac{5}{2 \cdot \sqrt{3}} \approx 1.44 \tag{19}$$

□

2.4 Summary

In Fig. 5 the analytical results are compared with computational results. Within a uniform 1024×1024 grid all possible partitions of size V (volume) are examined and the maximum, minimum, and average surface size is determined. The resulting values are plotted as solid lines while the analytical formulations for the worst case and average case are indicated by dashed lines. Two positions are tagged with an exclamation mark, where the computational results are very close to the analytical formulations.

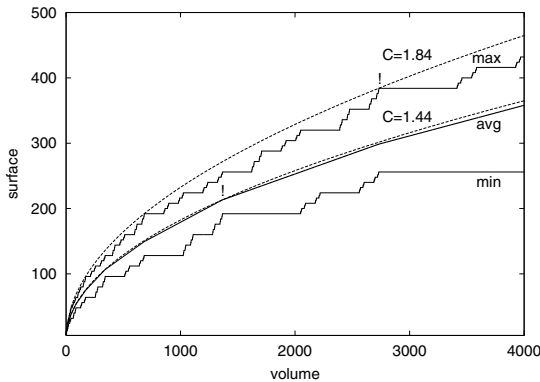


Fig. 5. Locality of partitions induced by the Lebesgue indexing.

3 Hilbert Curves

The Hilbert curve is presumably the most used and studied space-filling curve. It was introduced by Peano and Hilbert in the late 19th century [11]. It is known to be highly local in terms of several metrics mentioned in the introduction. The recursive definition of the curve is illustrated in Fig. 6. For the locality metric based on the quality coefficient this curve is much harder to analyze because the distance within the indexing for neighboring cells depends on the context during construction. An important result is that the lower bound on $C_{\max}^{\text{Hilbert}}$ is larger than the upper bound on $C_{\max}^{\text{Lebesgue}}$.

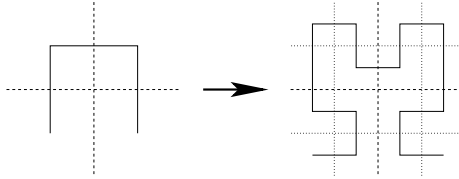


Fig. 6. Production rule for Hilbert indexing.

3.1 Lower Bound on Worst-Case Partitions

Theorem 4. *For the Hilbert curve the quality coefficient is larger than or equal to $3\sqrt{\frac{5}{13}}$.*

Proof. We construct a partition of size V and surface S which follows the Hilbert curve and gives the stated bad quality: Let k be an even number. The center of the partition is given by a square of size 4^{k+1} . On two sides of it 3 squares of sizes $4^k, 4^{k-2}, \dots, 4^2, 4^0$ are appended. Figure 7 shows the construction of the partition and its location within the Hilbert curve. It follows:

$$V = 4^{k+1} + 2 \sum_{i=0}^{k/2} 3 \cdot 4^{2i} = \frac{52}{5} 4^k - \frac{2}{5} \tag{20}$$

and

$$S = 8 \cdot 2^k + 12 \sum_{i=0}^{k/2} 2^{2i} = 24 \cdot 2^k - 4 \tag{21}$$

The quality coefficient of the partition is given by

$$\frac{S}{4 \cdot \sqrt{V}} = \frac{24 \cdot 2^k - 4}{4 \cdot \sqrt{\frac{52}{5} \cdot 4^k - \frac{2}{5}}} > \frac{24 \cdot 2^k}{4 \cdot \sqrt{\frac{52}{5} \cdot 2^k}} = 3 \cdot \sqrt{\frac{5}{13}} \approx 1.86 \tag{22}$$

□

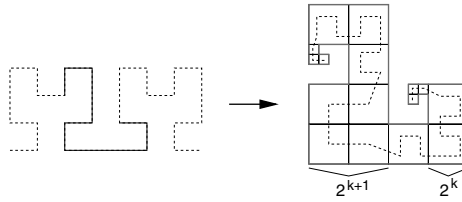


Fig. 7. Construction of lower bound for Hilbert indexing.

3.2 Summary

Figure 8 compares the lower bound on $C_{\max}^{\text{Hilbert}}$ with computational results in a 1024×1024 grid. It can be expected that the determined lower bound is close to the exact solution of $C_{\max}^{\text{Hilbert}}$. The dashed line for $C = 1.38$ seems to be an upper bound in the average case. This value would be lower than the corresponding for the Lebesgue indexing which proves the high locality of the Hilbert curve in another metric.

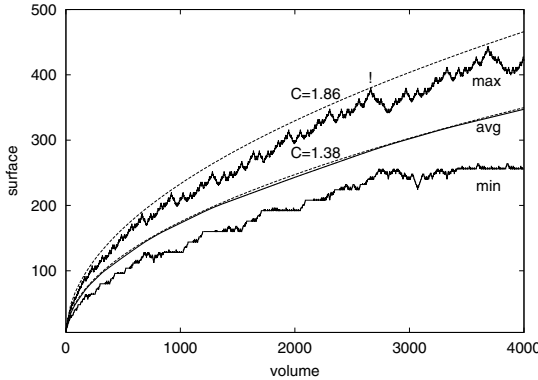


Fig. 8. Locality of partitions induced by the Hilbert indexing.

4 Concluding Remarks

The shown analytical results indicate that partitions based on the Lebesgue space-filling curve have good quality. We proved that they are slightly superior to Hilbert curves in the worst case. Computational results indicate that Hilbert curves behave better in average case. This is due to the fact that index intervals of the Hilbert curve are always connected.

It appears to be much harder to give sharp bounds for the Hilbert indexing than for Lebesgue indexing. We are near by an upper bound for $C_{\max}^{\text{Hilbert}} \leq \frac{26^{\frac{2}{3}}}{8 \cdot \sqrt{\frac{41}{15}}} \approx 2.02$, which still means a weaker result than for the Lebesgue curve.

Obviously, an open question is the lower bound on the quality coefficient for an arbitrary indexing scheme. It is hard to argue whether this bound is closer to the coefficients of the Lebesgue and Hilbert indexings or to 1, the optimal value given by the square. It is easy to generate bad cases for very small partitions, e. g. $V = 3$. This partition has a surface of at least 8. It follows $C = \frac{8}{4\sqrt{2}} = \sqrt{2}$. Excluding small volumes we can generate partitions with much lower quality coefficients. But it is an open question whether it is true for arbitrary partitions of an indexing scheme.

References

1. P. K. Agarwal and J. Erickson. Geometric range searching and its relatives. *Advances in Discrete and Computational Geometry*, 1998.
2. T. Asano, D. Ranjan, T. Roos, E. Welzl, and P. Widmayer. Space-filling curves and their use in the design of geometric data structures. *Theoretical Computer Science*, 181:3–15, 1997.
3. J. Behrens and J. Zimmermann. Parallelizing an unstructured grid generator with a space-filling curve approach. In A. Bode, T. Ludwig, W. Karl, and R. Wismüller, editors, *Euro-Par 2000*, LNCS 1900, pages 815–823. Springer, 2000.
4. S. Craver, B.-L. Yeo, and M. Yeung. Multilinearization data structure for image browsing. In *SPIE – The International Society for Optical Engineering*, pages 155–, 1998.
5. R. Diekmann, J. Hungershöfer, M. Lux, L. Taenzer, and J.-M. Wierum. Using space filling curves for efficient contact searching. In *Proc. IMACS*, 2000.
6. C. Gotsman and M. Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797, May 1996.
7. M. Griebel and G. Zumbusch. Parallel multigrid in an adaptive PDE solver based on hashing and space-filling curves. *Parallel Computing*, 25:827–843, 1999.
8. B. Moon, H. V. Jagadish, C. Faloutsos, and J. H. Saltz. Analysis of the clustering properties of the Hilbert space-filling curve. *IEEE Transaction on Knowledge and Data Engineering*, 13(1), Jan/Feb 2001.
9. R. Niedermeier, K. Reinhardt, and P. Sanders. Towards optimal locality in mesh-indexings. *LNCS 1279*, 1997.
10. R. Pajarola and P. Widmayer. An image compression method for spatial search. *IEEE Transactions on Image Processing*, 9(3):357–365, 2000.
11. H. Sagan. *Space Filling Curves*. Springer, 1994.
12. S.-H. Teng. Provably good partitioning and load balancing algorithms for parallel adaptive N-body simulation. *SIAM Journal on Scientific Computing*, 19(2):635–656, 1998.
13. J.-M. Wierum. Average case quality of partitions induced by the Lebesgue indexing. Technical Report TR-002-01, Paderborn Center for Parallel Computing, www.upb.de/pc2/, 2001.
14. G. Zumbusch. On the quality of space-filling curve induced partitions. *Zeitschrift für Angewandte Mathematik und Mechanik*, 81, SUPP/1:25–28, 2001.