Logarithmic Path-Length in Space-Filling Curves

Jens-Michael Wierum

Paderborn Center for Parallel Computing, PC[2]
Fürstenallee 11, 33102 Paderborn, Germany
jmwie@upb.de
http://www.upb.de/pc2/

Abstract. Data structures based on space-filling curves have shown to be a good approach in several applications. For the monitoring of moving objects, e. g. necessary for the contact detection in finite-element simulations, we need a special metrics to compare the quality of different curves.
This paper proposes the logarithmic index-range as a suitable quality metrics. It is based on the distance over the path of a space-filling curve to adjacent cells in the 2-dimensional grid. We present analytical results for the Hilbert, the Lebesgue curve, and the $\beta\Omega$-indexing and experimental results for the H-indexing.

1 Introduction

The performance of algorithms in computational geometry mainly depends on the efficiency of the underlying data structures. One problem is the representation of spatial moving objects in memory. Efficient sorting and searching is not easy because there is no unique sorting criterion in higher dimensions. There are two possibilities to circumvent this problem: Using complex data structures like trees which include several sorting criterions *or* mapping of the data into 1-dimensional structures to use efficient well known algorithms. Trees for higher dimensional spaces are much more sophisticated to implement if they should obtain given efficiency constraints.

Here we concentrate on the question how space-filling curves can be used to map 2-dimensional data into linear space. Space-filling curves are geometric representations of bijective mappings $M : \{1, \ldots, N^m\} \rightarrow \{1, \ldots, N\}^m$. The curve M traverses all N^m cells in the m-dimensional grid of size N. They define a sorting criterion for positions in the space. An (historic) overview on space-filling curves is given in [12].

Space-filling curves are used in many applications related to computational geometry like N-body simulations [11], image compression and browsing [10, 2], databases [1, 7], and contact search in finite element analysis [3]. The deployed indexing schemes have been examined with respect to several metrics. The most important are *locality* [4, 9], *clustering property* [1, 8], and *quality of partitions* [14, 6].

The costs for searching and sorting within 1-dimensional data structures often depend on the range to be examined. For example for searching we need $c \cdot \log m$ operations if the distance between the start and end point of the operation is lower than m. The same applies to the resorting of a single object in a tree-based data structure in average, if the distance between the position of deletion and insertion is less than m.

We concentrate our work on the *logarithmic index-range* on the path of the space-filling curve, which describes their efficiency in operations like searching and sorting:

Definition 1 (logarithmic index-range). *Let* curve *be an indexing scheme,* $(k,l) = curve(i)$ *the grid position of cell* i*. The* logarithmic index-ranges *to the direct neighbors of* i *are:*

$$
\begin{aligned}
r_1^{curve}(i) &= \log(|i - curve^{-1}(k-1,l)| + 1) \\
r_2^{curve}(i) &= \log(|i - curve^{-1}(k,l-1)| + 1) \\
r_3^{curve}(i) &= \log(|i - curve^{-1}(k+1,l)| + 1) \\
r_4^{curve}(i) &= \log(|i - curve^{-1}(k,l+1)| + 1)
\end{aligned}
$$

This formulation can be extended to a logarithmic index-range *of an indexing scheme:*

$$
\begin{aligned}
R_{\max}^{curve} &= \max_i\{r_j^{curve}(i)\}, \;\; j \in \{1,\ldots,4\} \\
R_{\text{avg}}^{curve} &= \text{avg}_i\{r_j^{curve}(i)\}, \;\; j \in \{1,\ldots,4\}
\end{aligned}
$$

2 Space-Filling Curves

This section introduces the production rules for the examined space-filling curves. For a more detailed description we refer to [9, 13] for H-indexing and $\beta\Omega$-indexing, respectively. Examples for all curves within a 32×32 grid are presented in Figure 6.

2.1 Hilbert Curve

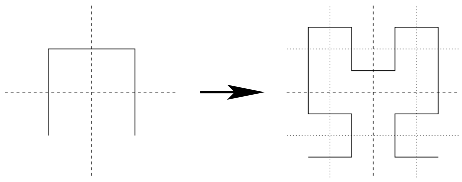

Fig. 1. Recursive definition of the Hilbert curve.

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 [5]. It is assumed to be highly local in terms of several metrics. The recursive definition of the curve is illustrated in Figure 1.

2.2 Lebesgue Curve

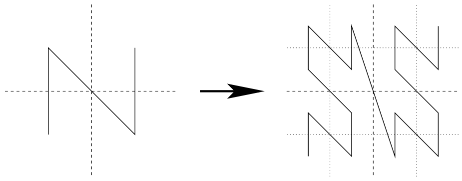

Fig. 2. Recursive definition of the Lebesgue curve.

Figure 2 illustrates the recursive definition of the Lebesgue-indexing. The resulting curve is also known as *bit interleaving* or *Z-code*. The large gaps in the curve indicate insufficient locality properties for some metrics. Nevertheless, it is adequate for other metrics. Furthermore, it is much easier to analyze than most of the other indexing schemes.

2.3 H-Indexing

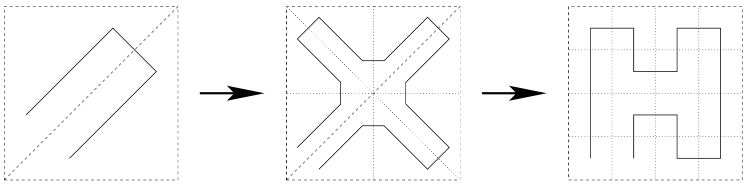

Fig. 3. Production rule for the H-indexing.

The H-indexing is related to the Sierpiński curve [12]. It was introduced by Niedermeier, Reinhardt, and Sanders [9]. Despite the other indexing schemes in this paper, the H-indexing is easier to describe by right-angled triangles. The left rule in Figure 3 describes the core production rule for the refinement of the triangles. The right rule shows the adjustment for regular grids as the last step of curve generation: The cells on the diagonals are assigned alternately to the left and right triangles of the diagonal. This step leads to the characteristic H-shaped forms within the curve.

2.4 $\beta\Omega$-Indexing

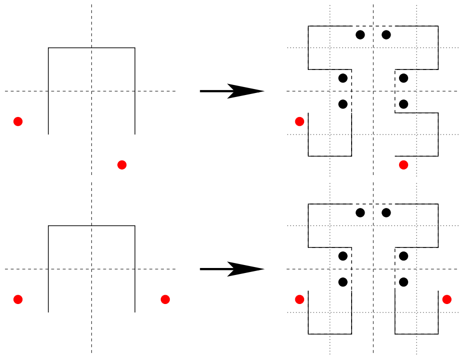

Fig. 4. Recursive definition of the $\beta\Omega$-indexing.

The $\beta\Omega$-indexing uses the same U-like base geometry as the Hilbert curve. There exists two sub types of this base geometry with two different refinement rules. The sub types illustrated in Figure 4 are labeled with small dots. They represent the two end points of the curve within this quarter. We named the new curve $\beta\Omega$-indexing due to the characteristic form of the two refinement rules [13].

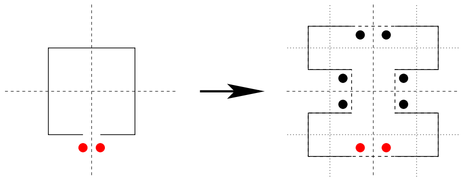

Fig. 5. Special rule for the first refinement step of the $\beta\Omega$-indexing.

It is obvious that any sub type chosen at the highest level leads to comparable asymptotic locality properties. To get a circular curve we use a special rule shown in Figure 5. The initial area is split into four objects of the first sub type.

A space-filling curve is called *recursive* if it can be recursively divided into four square space-filling curves of equal size [1]. Note, that the H-indexing does not belong to the class of recursive space-filling curves although it is also defined by a recursive refinement rule.

For recursive space-filling curves we distinguish the separators of the cells by their *level*. The level describes the phase in the recursive curve generation when the separator was introduced. The first (mean) separator is of level $\log n - 1$,

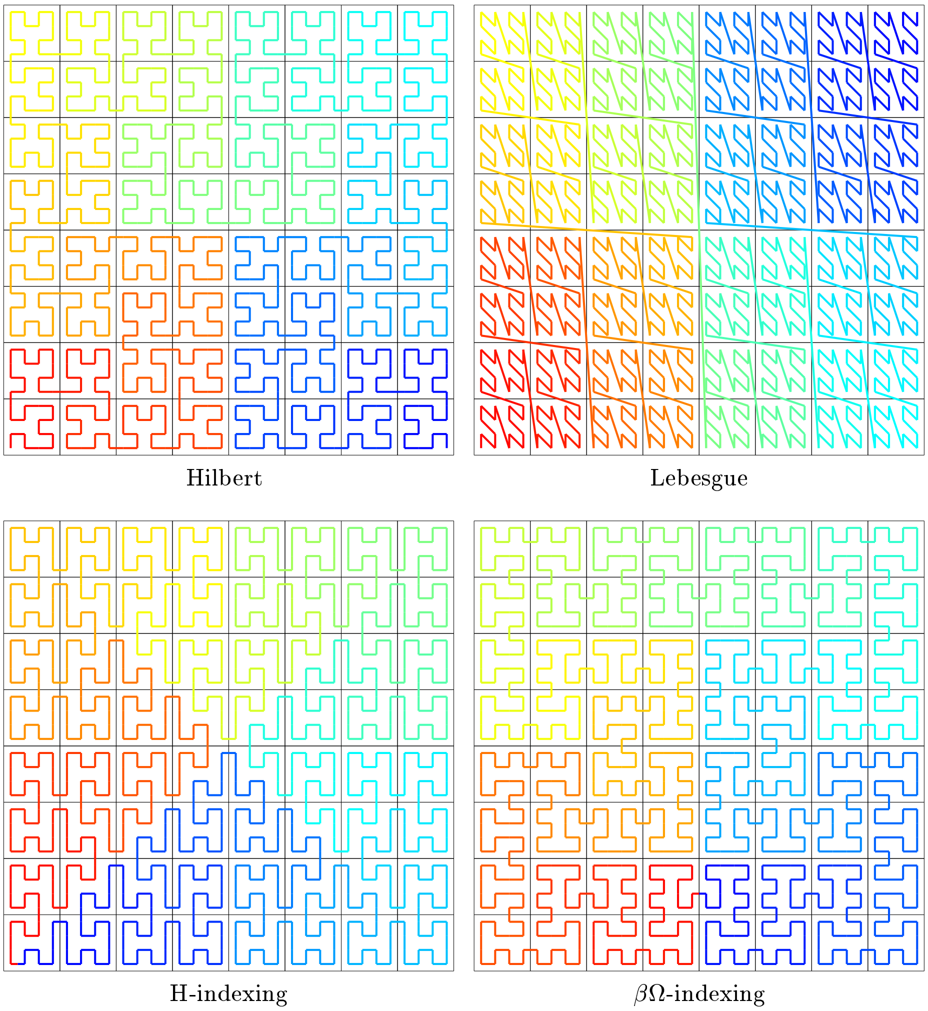

Fig. 6. The four evaluated curves in grids of size 32×32.

the last inserted separators are of level 0. For example, in Figure 6 only the separators of level 2, 3, and 4 are plotted while level 1 and 0 separators are omitted for clarity. It can be observed, that for large grids the probability for a separator being of level l converges to $p(l) = \frac{1}{2^{(l+1)}}$ [6].

3 Results

An overview on the results of this work is presented in Table 1. The logarithmic index-range R to the four adjacent cells is presented for the examined indexing schemes. The value for the H-indexing is obtained from extensive simulations (cf. Fig. 7), while the others are analytical results. In the average case the values of lower and upper bounds are presented.

R	worst case	average case lower	average case upper
Hilbert	$\log\left(\frac{5}{6}N\right)$	2.524	2.542
Lebesgue	$\log\left(\frac{N+2}{3}\right)$	2.745	2.746
H-Indexing	$\log N$	≈ 2.58	
$\beta\Omega$-Indexing	$\log N$	2.503	2.525

Table 1. Logarithmic index-range R to adjacent cells in a grid of size $N = n \times n$.

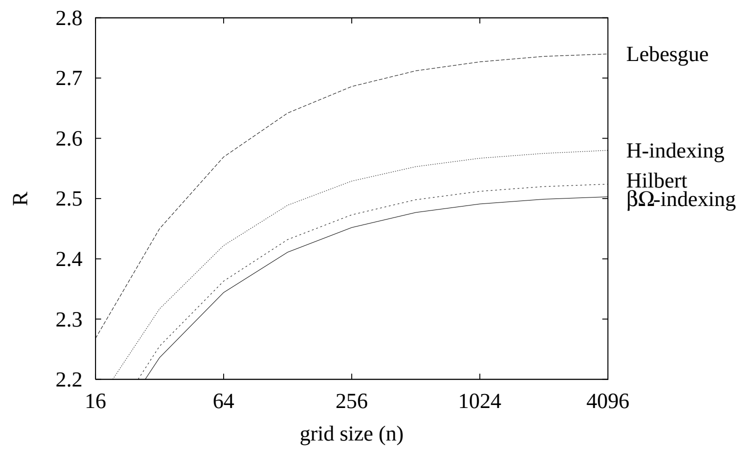

Fig. 7. Experimental results: logarithmic index-range for different grid sizes.

3.1 Worst Case

The logarithmic index-ranges in worst case are obvious. H-indexing and $\beta\Omega$-indexing are circular curves. Cell '0' is adjacent to cell '$N-1$', which results in an index-range of $N = n^2$.

The maximal index-range in the Lebesgue curve exists at the mean vertical separator. The distance on the path of the curve is $\frac{N+2}{3}$ for all cells at this separator [6].

The maximal distance in the Hilbert-indexing appears at the lowest cells close to the mean vertical separator. It includes $N/2$ cells in the upper half of the grid. For the lower quadrants we have

$$\sum_{i=1}^{\log n} \frac{1}{2} \cdot \frac{1}{4^i} N = \frac{1}{2} \cdot \frac{N-1}{3}$$

cells, due to the self-similarity of the curve[1]. The overall distance converges to $\frac{5}{6}$ for large grids.

3.2 Average Case

At first we show that the logarithmic index-range is monotonously increasing for increasing grid sizes for all recursive space-filling curves. We define two special terms for the analysis of those curves in grids of finite size. $R_{\text{avg}}^{curve}(k)$ denotes the average logarithmic index-range at separators of level k and $R_{\text{avg}}^{curve}(0..k)$ the average logarithmic index-range in a square of size $2^k \times 2^k$.

Lemma 1. *Let* curve *be a recursive space-filling curve, and $k \in \mathbb{N}$. It holds:*

$$R_{\text{avg}}^{curve}(0..k) > R_{\text{avg}}^{curve}(0..k-1)$$

Proof. It is

$$R_{\text{avg}}^{curve}(0..k) = \frac{R_{\text{avg}}^{curve}(0..k-1) \cdot (2^{k+1}-2) + R_{\text{avg}}^{curve}(k)}{2^{k+1}-1} \quad .$$

Therefore,

$$\begin{aligned} & R_{\text{avg}}^{curve}(0..k) > R_{\text{avg}}^{curve}(0..k-1) \\ \Leftrightarrow \quad & R_{\text{avg}}^{curve}(k) > R_{\text{avg}}^{curve}(k-1) \ , \end{aligned}$$

which is obviously true, because of the self-similarity of the curves. The distances at the mean separator of a $2^k \times 2^k$ grid are at least four times minus three larger than the distances in a $2^{k-1} \times 2^{k-1}$ grid (cf. index-ranges in Fig. 8 and 9). $\square$

In the following we focus on bounds for grids of infinite size.

[1] During the refinement, the quadrants which include the lowest cells close to the mean vertical separator are always of the same type: $\supset$ and $\subset$, resp.

3.2.1 Lebesgue Curve

Due to its regular structure without any reflections or rotations the Lebesgue curve is the easiest to analyze.

Theorem 1. *For the logarithmic index-range of the Lebesgue curve in average case holds:*

$$2.745 < R_{\text{avg}}^{\text{Lebesgue}} < 2.746 \ .$$

Proof. For each separator the distance on the path of the curve to adjacent cells only depends on the level of the separator and its direction. For a vertical separator of level q the distance is $R_q = 4 \cdot \frac{4^q-1}{3} + 2$ and for a horizontal separator of level p the distance is $U_p = 2 \cdot \frac{4^p-1}{3} + 1$. We get an average logarithmic index-range for a given level l of

$$R_{\text{avg}}^{\text{Lebesgue}}(l) = \frac{1}{2} \cdot (\log(U_l + 1) + \log(R_l + 1)) \ ,$$

and combined with the probability $p(l)$ for the level l of a separator

$$R_{\text{avg}}^{\text{Lebesgue}} = \sum_{l=0}^{\infty} R_{\text{avg}}^{\text{Lebesgue}}(l) \cdot p(l) \ , \tag{1}$$

for infinite grids.

There is no further simplification for $R_{\text{avg}}^{\text{Lebesgue}}$. Therefore we evaluate the first k entries of the summation and specify an estimate for the missing entries. For levels l larger or equal to 3 holds

$$\begin{aligned} R_{\text{avg}}^{\text{Lebesgue}}(l) &= \frac{1}{2} \cdot \left[\log\left(2\frac{4^l-1}{3} + 2\right) + \log\left(4\frac{4^l-1}{3} + 3\right)\right] \\ &< \frac{1}{2} \cdot \log\left(4^{2 \cdot l}\right) = 2 \cdot l \ . \end{aligned}$$

For the summation of the missing entries we get

$$\sum_{l=k+1}^{\infty} R_{\text{avg}}^{\text{Lebesgue}}(l) \cdot \frac{1}{2^{l+1}} < \sum_{l=k+1}^{\infty} \frac{2 \cdot l}{2^{l+1}} = 2 \cdot \frac{k+2}{2^k} \ . \tag{2}$$

The evaluation of the entries 0 to 17 of the summation results in 2.74583. The error determined in Equation 2 is lower than $0.16 \cdot 10^{-3}$. □

3.2.2 Hilbert Curve

Theorem 2. *For the logarithmic index-range of the Hilbert curve in average case holds:*

$$R_{\text{avg}}^{\text{Hilbert}} < 3.25 \ .$$

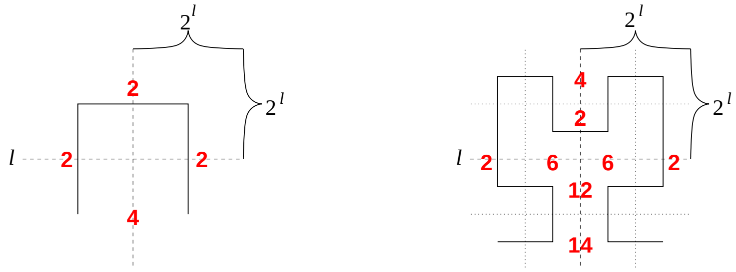

Fig. 8. Maximal index-ranges for the Hilbert curve at separator of level l as blocks of size $2^l \times 2^l$ (left) and as blocks of size $2^{l-1} \times 2^{l-1}$ (right).

Proof. For the upper bound on the Hilbert curve we have to consider a coarser structure shown in Figure 8 (left). In a given square of size $4 \cdot 4^l$ with separators of level l we always have the 'U'-like base structure. For three pairs of the quarters we have an index-range of at most $2 \cdot 4^l$ while the other index-range may include the whole square of size $4 \cdot 4^l$. Together with the probability of a level l separator we get

$$\begin{aligned} R_{\text{avg}}^{\text{Hilbert}} &< \sum_{l=0}^{\infty} \left(\frac{1}{4} \left[3 \cdot \log(2 \cdot 4^l) + 1 \cdot \log(4 \cdot 4^l) \right] \cdot \frac{1}{2^{l+1}} \right) \\ &= \frac{1}{4}(3 \cdot 3 + 4) = 3\frac{1}{4} \end{aligned}$$

□

We can extend this technique for finer grids. In Figure 8 (right) the square of size $4 \cdot 4^l$ is split into 4×4 blocks. The evaluation of the shown index-ranges like above leads to 2.0976. It neglects the index-ranges at the separators of the lowest level. For the four pairs of quarters on this level we have an index-range of $R_{\text{avg}}^{\text{Hilbert}}(0) = \frac{5}{4}$ in average and the probability for this level is $\frac{1}{2}$. Adding the resulting $\frac{5}{8}$ for the lowest level to the determined index-range for all higher levels results in 2.7226. The same technique can be applied to finer structures resulting in the following values:

$$\begin{aligned} 2 \times 2 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\text{Hilbert}} < 3.25 \\ 4 \times 4 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\text{Hilbert}} < 2.723 \\ 8 \times 8 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\text{Hilbert}} < 2.580 \\ 16 \times 16 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\text{Hilbert}} < 2.542 \end{aligned}$$

An evaluation of the Hilbert curve in the 4096×4096 grid results in an logarithmic index-range of 2.524 in average. From Lemma 1 follows:

Corollary 1. *For the logarithmic index-range of the Hilbert curve holds:*

$$2.524 < R_{\text{avg}}^{\text{Hilbert}} < 2.542$$

3.2.3 $\beta\Omega$-Indexing

Theorem 3. *For the logarithmic index-range of the $\beta\Omega$-indexing holds:*

$$2.503 < R_{\text{avg}}^{\beta\Omega-\text{indexing}} < 2.525$$

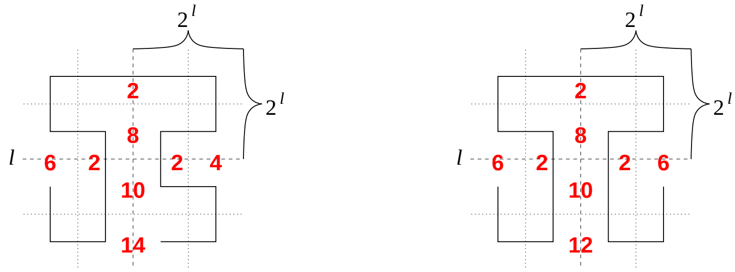

Fig. 9. Maximal index-ranges as blocks of size $2^{l-1} \times 2^{l-1}$ for the two refinement rules of the $\beta\Omega$-indexing.

Proof. (sketch) The same technique as for Theorem 2 is used. For a grid of 2×2 blocks we get the same result as for the Hilbert curve. Due to the two different refinement rules of the $\beta\Omega$-indexing two different grids have to be examined for finer grids. The curves in a grid of 4×4 blocks are presented in Figure 9. The results are combined according to the asymptotic probability of the corresponding structure which is $4:1$ [13]. For the different examined granularities we get the following results:

$$\begin{aligned}
2 \times 2 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\beta\Omega-indexing} < 3.25 \\
4 \times 4 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\beta\Omega-indexing} < 2.737 \\
8 \times 8 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\beta\Omega-indexing} < 2.570 \\
16 \times 16 \text{ blocks} &\Rightarrow R_{\text{avg}}^{\beta\Omega-indexing} < 2.525
\end{aligned}$$

The lower bound for the logarithmic index-range is derived from the evaluation of a 4096×4096 grid, combined with the result of Lemma 1. $\square$

4 Conclusions and Future Work

This paper presents the metrics of *logarithmic index-range* as a quality measure for mappings from space into linear data structures. This measure is important for applications with moving objects in space based on space-filling curves, like contact detection in finite-element simulation [3].

In worst case analysis the Lebesgue curve performs best in terms of logarithmic index-range, while the circular curves H-indexing and $\beta\Omega$-indexing perform worst. In average case, the $\beta\Omega$-indexing seems to have the best quality, although its upper bound is slightly higher than the lower bound of the Hilbert curve (cf. Tab. 1 and Fig. 7).

For the Lebesgue curve we are able to present a term describing the average case exactly. A partial evaluation of the summation gives strong bounds due to a strongly declining error-term.

Future work will concentrate on the improvement of the lower bounds for Hilbert curve and $\beta\Omega$-Indexing. A summation like for the Lebesgue curve fails because there is no term describing the distances within the curves. One approach could be the refinement of Lemma 1 by an estimation of the acceleration of the logarithmic index-range for arbitrary recursive space-filling curves. Furthermore, we are interested to extend the shown results to higher dimensional spaces.

References

1. 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.
2. S. Craver, B.-L. Yeo, and M. Yeung. Multilinearization data structure for image browsing. In *SPIE – The International Society for Optical Engineering*, 1998.
3. R. Diekmann, J. Hungershöfer, M. Lux, L. Taenzer, and J.-M. Wierum. Efficient contact search for finite element analysis. In *Proc. ECCOMAS*, Barcelona, 2000.
4. 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.
5. D. Hilbert. Über die stetige Abbildung einer Linie auf ein Flächenstück. *Mathematische Annalen*, 38:459–460, 1891.
6. J. Hungershöfer and J.-M. Wierum. On the quality of partitions based on space-filling curves. In *International Conference on Computational Science ICCS*, volume 2331 of *LNCS*, pages 36–45. Springer, 2002.
7. H. Jagadish. Analysis of the Hilbert curve for representing two-dimensional space. *Information Processing Letters*, 62:17–22, 1997.
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):124–141, Jan/Feb 2001.
9. R. Niedermeier, K. Reinhardt, and P. Sanders. Towards optimal locality in mesh-indexings. In B. S. Chlebus and L. Czaja, editors, *Fundamentals of Computation Theory*, volume 1279 of *LNCS*, pages 364–375. Springer, 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. W. Rankin, J. Board, and V. Henderson. The impact of data ordering strategies on a distributed hierarchical multipole algorithm. In *Proceedings of Ninth SIAM Conference on Parallel Processing for Scientific Computing*, March 1999.
12. H. Sagan. *Space Filling Curves*. Springer, 1994.
13. J.-M. Wierum. Definition of a new circular space-filling curve – $\beta\Omega$-indexing. Technical Report TR-001-02, Paderborn Center for Parallel Computing, http://www.upb.de/pc2/, 2002.
14. G. Zumbusch. On the quality of space-filling curve induced partitions. *Zeitschrift für Angewandte Mathematik und Mechanik*, 81, SUPP/1:25–28, 2001.