SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 730

March 1987

LARGE TRAVELING SALESMAN PROBLEMS ARISING FROM EXPERIMENTS IN X-RAY CRYSTALLOGRAPHY: A PRELIMINARY REPORT ON COMPUTATION*

by

Robert G. Bland & David F. Shallcross

*Supported in part by NSF Grant ECS-8313853 to Cornell University.

Abstract

Determining efficient sequences of x-ray diffraction measurements on a four-circle diffractometer leads to extremely large traveling salesman problems. Simple TSP heuristics give substantial improvements in utilization at small computational expense. Furthermore, the Lin-Kernighan heuristic consistently produces sequences that are near-optimal.

1. Introduction

Experiments in x-ray crystallography often involve sequential collection of very large sets of data on a four-circle diffractometer. A small sample of the single crystal under study is mounted in the apparatus, which has computer-driven motors that can accurately position the crystal and a detector. When the diffractometer is positioned appropriately, the detector measures the intensity of monochromatic x-rays scattered by the crystal at a peak in the diffraction pattern. This gives, approximately, one coefficient in the electron density function of the crystal. The purpose of the experiment is to collect all of the coefficients in the electron density function, which is a Fourier series. Then, with substantial computational effort, the electron density function is calculated, and its contours are plotted, in order to attempt to discern the detailed structure of the crystal (see Figures 1 and 2). For information concerning x-ray analysis of crystals see, for example, Luger [8].

These experiments typically involve 5000 to 30000 readings on the diffractometer. The order in which the readings are taken is not relevant to the analysis of the data, but it can have a profound influence on the time to complete the experiment. Given a high intensity source of high energy x-rays, the measurements can be made relatively quickly, once the motors are set properly. However, the time to reposition the apparatus between readings is significant. The accumulation of these delays, the slewing time, may constitute a very substantial portion of the total time to complete an experiment. How large a portion depends on the energy and intensity of the x-ray source. Presently at the Cornell High Energy Synchrotron Source, slewing time is about 25% of measurement time; imminent improvements in intensity will result in a tenfold to hundredfold decrease in measurement time,

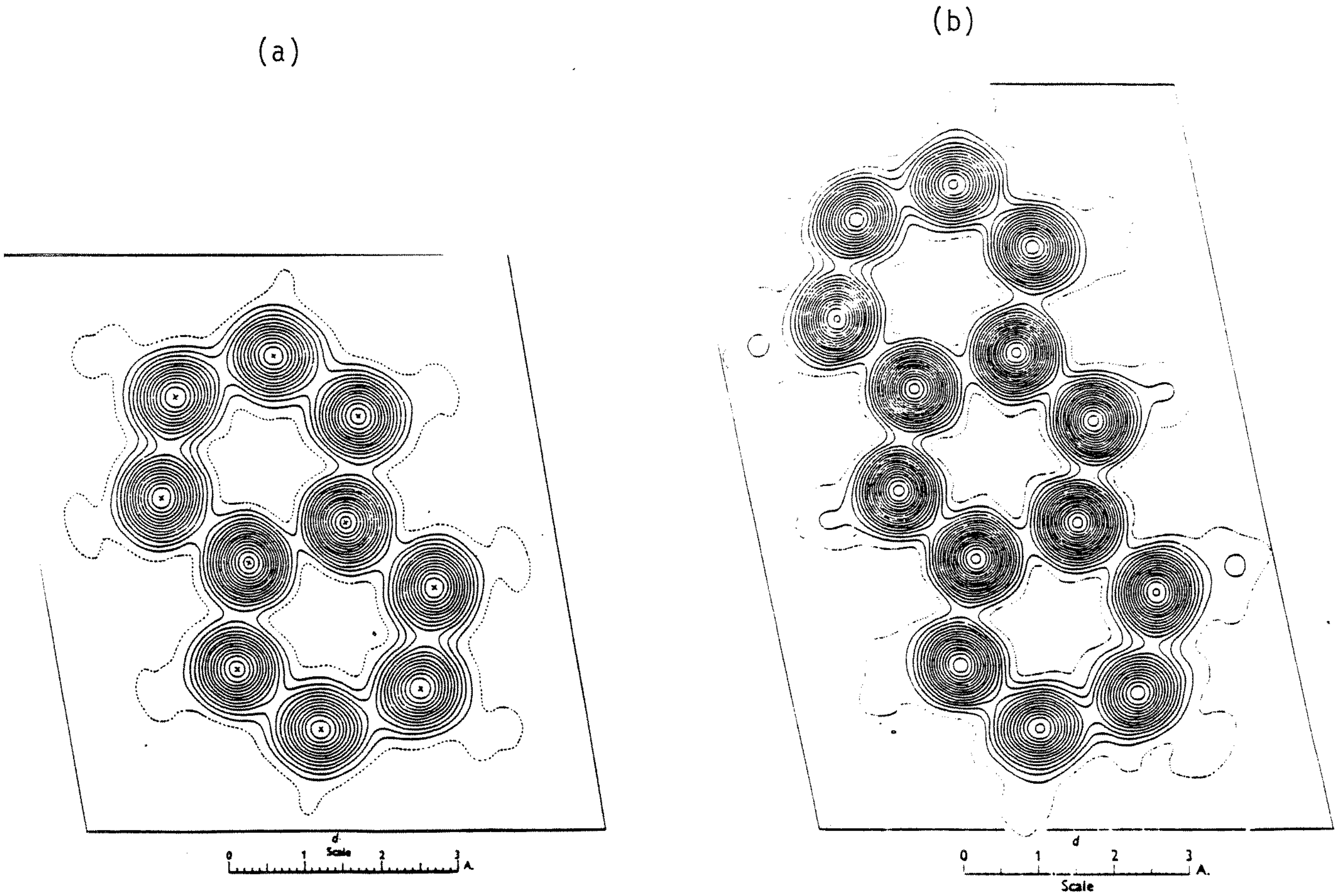

Figure 1. Electron density maps based on x-ray diffraction data: (a) napthalene ($C_{10}H_8$); (b) anthracene ($C_{14}H_{10}$). Reprinted with permission from: (a) *The crystal and molecular structure of napthalene. II. Structure investigation by the triple Fourier series method* by S.C. Abrahams, J.M. Robertson, and (in part) J.G. White, Acta Cryst. 2 (1949), copyright International Union of Crystallography; (b) *The crystal and molecular structure of anthracene. II. Structure investigation by the triple Fourier series method* by V.C. Sinclair, J.M. Robertson, and (in part) A.McL. Mathieson, Acta Cryst. 2 (1949), copyright International Union of Crystallography.

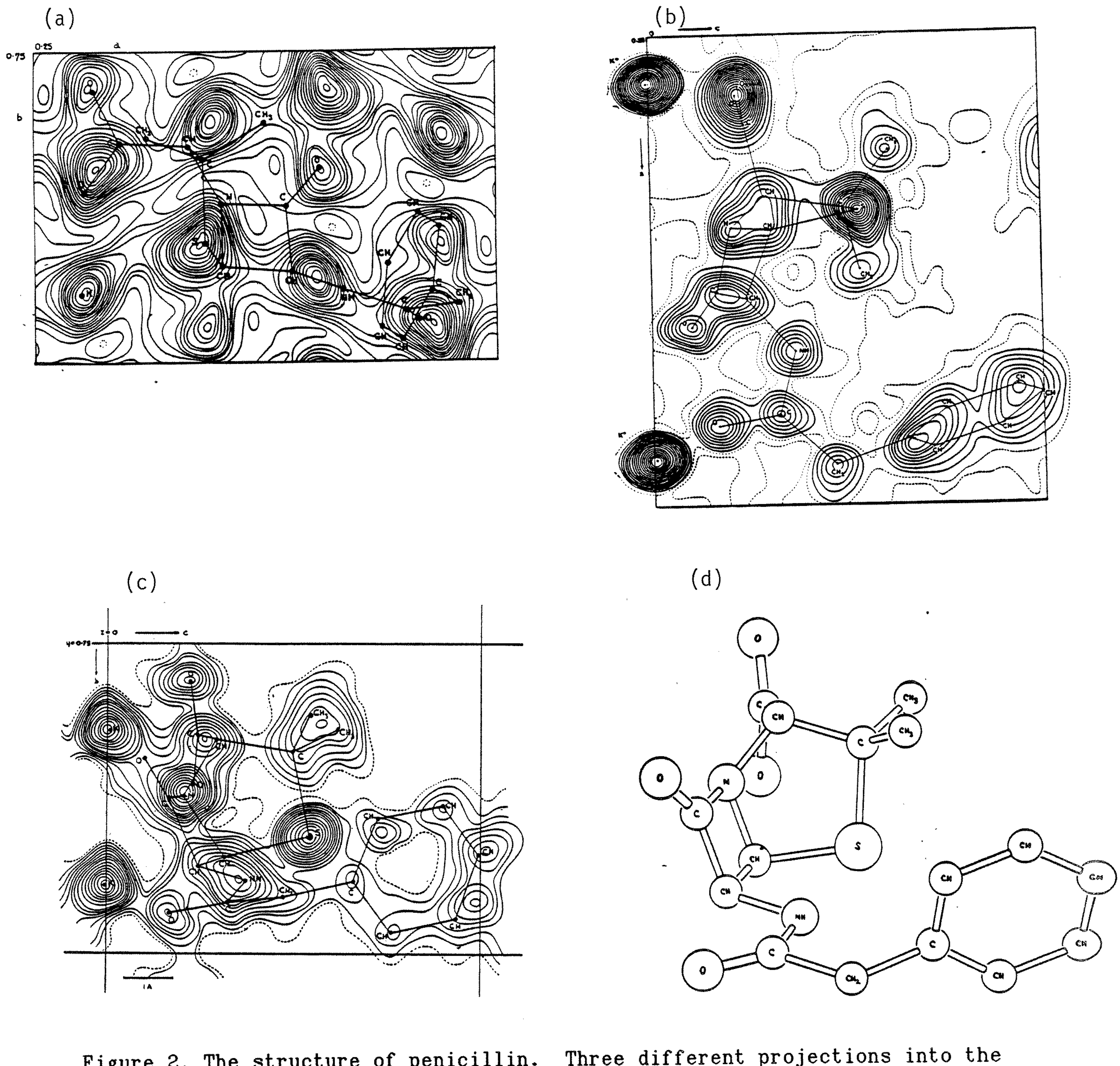

Figure 2. The structure of penicillin. Three different projections into the plane of electron density maps of potassium benzylpenicillin are presented as (a)-(c). These are from an early study (see reference below) that first revealed the exact chemical formula and fundamental stereochemical configuration of penicillin. (d) The arrangement of atoms in the benzylpenicillin ion as deduced from the x-ray crystallographic study. Reprinted with permission from: *The x-ray crystallographic investigation of the structure of penicillin* by D. Crowfoot, C.W. Bunn, B.W. Rogers-Low, and A. Turner-Jones, in: The Chemistry of Penicillin, eds: H.T. Clarke, J.R. Johnson, and R. Robinson, Princeton University Press, Princeton, New Jersey (1949).

resulting in slewing time becoming a very large fraction of the total time to completion. Even at the low end (5000) of the range of readings of interest, the slewing time is typically in the neighborhood of 15 - 24 hours.

There is a standard method used for sequencing the readings. It is based upon a simple lexicography on the lattice dual to the crystal lattice, the reciprocal lattice. The peaks in the diffraction pattern occur at points of the dual lattice. One chooses a basis of the dual lattice, represents each lattice point by a triple (h,k,l) with respect to this basis, and traces lexicographically through all observable triples. If the incident x-rays are of wavelength λ, then the observable triples are those within distance $2/\lambda$ of the origin.

It is natural to wonder whether one can improve significantly on the conventional lexicographic method of sequencing the readings, in order to improve the utilization of the diffractometer and x-ray source. Since the problem of sequencing the readings to minimize the total delay is a traveling salesman problem (TSP), standard TSP heuristics might work well here.(See the recent book of Lawler, Lenstra, Rinnooy Kan, and Shmoys [6] for a wealth of information on the TSP.) It should be noted, however, that the instances arising in this setting are: (1) enormously large by conventional TSP standards; and (2) non-Euclidean, though they do satisfy the triangle inequality. The readings correspond to nodes of the TSP. The length of an edge of the TSP is the delay time for the motors to reposition the apparatus from the settings corresponding to one of the two ends of the edge to the settings corresponding to the other. The lengths are symmetric and can be computed readily.

2. Computational Tests

We have tested three different heuristics on twelve large TSP instances. The instances have from 2762 to 14464 readings, with mean 8372. They come from five different crystals, with variation of the wavelength of the monochromatic x-rays resulting in more than one instance arising from a given crystal. Table 1 identifies the five crystals in the leftmost column, and indicates the number of readings in the next column. The instances are grouped according to wavelength. The heuristics were coded in Fortran (77), and executed on an IBM 3081 model K.

The lexicographic method can be applied with any of the six orderings of the indices h, k, and l. In order to be conservative in comparing the heuristics to lexicography, all comparisons will be made with the best of the six lexicographic tours.

Before applying any of the heuristics, some preprocessing was done. As suggested in Lin and Kernighan [7], the complete graph on which the problem is defined was replaced by a sparse graph in which only the d shortest edges at each node were included, for an appropriately small d. We set d=10 and restricted the choice of edges at a given vertex (h,k,l) to those joining it to (h',k',l') with h',k' and l' differing from h,k, and l, respectively, by at most one. The preprocessing times are included in our reports of the execution times of the first heuristic, an approximation to Christofides' heuristic.

Christofides' approach (see Golden and Stewart [1], or Johnson and Papadimitriou[5]) begins with a minimum length spanning tree, adds edges to make it Eulerian, traces an Euler tour, and compresses it to a traveling salesman tour. In order to to get even degrees at every vertex, Christofides computes a minimum length perfect matching on the set of odd-degree vertices of the spanning tree. The difficulty here is that the problems of interest will have many thousands of odd-degree vertices in the minimum length spannning trees. Thus we are unwilling to apply exact algorithms for the matching problem, for which running time would grow faster than a quadratic function of the number of odd vertices. Since the matching problem arises as a subproblem within a heuristic, it is reasonable to do the matching heuristically. We employed a greedy approach. It selects an odd vertex that has not yet been matched and matches it with the nearest unmatched odd vertex, continuing until all odd vertices are matched. It should be noted that the simplest approach to getting even degrees, doubling the spanning tree edges, generally led to traveling salesman tours longer than those produced by lexicography.

The approximation to Christofides' heuristic was used to construct an initial traveling salesman tour. In all twelve problem instances this tour improved on the best lexicographic tour by 25% to 36% and used 34 to 239 CPU seconds (including the preprocessing). Then two-opt (see [1]), a simple local improvement heuristic, was applied to the previously constructed tour. Two-opt iteratively interchanges a pair of edges in the tour with the pair of missing edges that will re-connect it appropriately, if the tour length will be reduced by the interchange. This is continued until the current tour is two-optimal, i.e. no more improving interchanges can be made. This took no more than 67 seconds on any of the twelve instances, and resulted in improvements over the best lexicographic tour of 30% to 42%. The more sophisticated variable depth local improvement heuristic of Lin and Kernighan [7] was then applied to the two-optimal tour. The Lin-Kernighan heuristic typically works extremely well in practice, but is more intensive computationally than the first two heuristics. Given the expectation that

Lin-Kernighan would use much more CPU-time, it seemed reasonable to invest more effort in the preprocessing. We searched the entire complete graph for edges among the ten shortest at some vertex, but absent from the template used in the first preprocessor. The Lin-Kernighan CPU-times, including the times for the additional preprocessing, range up to 10279 seconds. Lin-Kernighan resulted in improvement over the best lexicographic tour of 34% to 46% in each of the twelve instances. What is even more impressive is that Lin-Kernighan got very close to optimality in each case. We used the Held and Karp [2,3] method of producing a lower bound on the optimal length of a traveling salesman tour by one-tree relaxations and subgradient optimization. Table 2 shows that the Lin-Kernighan tours are within 1.7% of the lower bound in all twelve instances, and within 1.5% in all seven of the instances with more than 6000 readings. Table 2 also lists the deviations from the lower bound for the other heuristics, and for the best lexicography. Table 3 gives the execution times of the heuristics and of the lower bound calculations. Note that in practice the lower bounds need not be calculated.

These tests indicate that one can expect to be able to make substantial improvements over lexicography in sequencing the four-circle diffractometer. In further work we will examine whether one can approach the excellent performance of the Lin-Kernighan heuristic at less expense by partitioning. In the instances tested here, it appears that the rate of growth of computation in the Lin-Kernighan heuristic is approximately $n^{1.85}$, as opposed to $n^{2.2}$ reported in [7]. Johnson, McGeoch, and Rothberg [4] report $n^{1.6}$ in testing of a somewhat more sophisticated implementation on other large TSP instances. However, given the problem sizes of interest here, one would like to be able to employ linear or near-linear-time heuristics.

Beyond the immediate relevance to the x-ray diffraction experiments, it is interesting that in real problem instances of such large size, Lin-Kernighan and Held-Karp consistently produced very good upper and lower bounds, respectively, on the optimal length traveling salesman tour. This is in keeping with previous wisdom on smaller problem instances, and with the recent experimental work of Johnson, McGeoch, and Rothberg [4]. In spite of the TSP being NP-hard, it seems that oftentimes it is not so difficult to get close to an optimal solution, even for big instances. Indeed, very recently Padberg and Rinaldi have used polyhedral techniques to solve to optimality on a supercomputer a TSP instance with 2392 nodes (see the note at the end of [9]).

Acknowledgment

This work has benefited from the advice of Philip Coppens, who is collaborating with us on a more extensive study, and from the participation of Jonathan Lee and Walter Morris during its very early stages.

APPENDIX

Table 1

The Twelve Test Problems

	crystal (see key)	wavelength (angstroms)	number of readings
1	A	1.70	4472
2	B	1.70	2950
3	D	1.70	7008
4	E	1.70	2762
5	F	1.70	6922
6	A	1.35	9070
7	B	1.35	5888
8	D	1.35	14012
9	E	1.35	5520
10	F	1.35	13804
11	B	1.00	14464
12	E	1.00	13590

Key
- A: ammonium tartrate
- B: biphenyl
- D: dinitrodiphenyltetrathiofulvalene
- E: bis-2-imidazole iron (octaethylporphyrin)
- F: iron dipyridyltetraphenylporphyrin

Table 2

% Over Lower Bound

	Best Lexicography	Approximate Christofides	Two-opt	Lin & Kernighan
1	56.0	16.0	8.4	1.7
2	59.5	15.2	9.0	1.7
3	55.9	13.7	7.3	1.2
4	78.3	15.9	9.6	1.4
5	79.0	16.2	8.3	1.2
6	60.1	14.0	7.4	1.3
7	63.1	13.4	7.3	1.7
8	59.3	14.4	7.8	1.5
9	82.5	15.6	8.6	1.5
10	82.3	15.5	8.5	1.4
11	66.4	12.5	7.0	1.5
12	88.4	13.6	7.8	1.4

Table 3

Execution Times (IBM 3081 seconds)

	Approximate Christofides	Two-opt	Lin & Kernighan	Held & Karp
1	51	18	1061	547
2	37	10	417	259
3	93	26	2071	1147
4	34	9	390	273
5	102	28	2947	1169
6	127	43	2503	1562
7	80	24	1462	774
8	213	72	6990	?
9	72	21	1271	708
10	239	67	10279	3079
11	201	60	7797	3274
12	204	65	7077	3035

REFERENCES

1. N. Christofides, "Worst–case analysis of a new heuristic for the traveling salesman problem," Report 388, Graduate School of Industrial Administration, Carnegie–Mellon University (1975).

2. G. Cornuejols and G.L. Nemhauser, "Tight bounds for Christofides' traveling salesman heuristic," Mathematical Programming 14 (1978) 116–121.

3. B.L. Golden and W.R. Stewart, "Empirical analysis of heuristics," in The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (eds.), Wiley & Sons, 1985.

4. M. Held and R.M. Karp, "The traveling–salesman problem and minimum spanning trees," Operations Research 18 (1970) 1138–1162.

5. M. Held and R.M. Karp, "The traveling–salesman problem and minimum spanning trees: part II," Mathematical Programming 1 (1971) 6–25.

6. D.S. Johnson, L.A. McGeoch and E.E. Rothberg, "Near optimal solutions to very large traveling salesman problems," in preparation.

7. D.S. Johnson and C.H. Papadimitriou, "Performance guarantees for heuristics," in The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (eds.), Wiley & Sons, 1985.

8. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, Wiley & Sons, 1985.

9. S. Lin and B. Kernighan, "An effective heuristic algorithm for the traveling salesman problem," Operations Research 21 (1973) 498–516.

10. P. Luger, Modern X–ray Analysis on Single Crystals, W. de Gruyter & Co., Berlin, 1980.

11. M. Padberg and G. Rinaldi, "Optimization of a 532–city symmetric traveling salesman problem by branch–and–cut," Operations Research Letters 6 (1987) 1–8.

12. M. Padberg and G. Rinaldi, "Facet identification for the symmetric traveling salesman polytope," preprint (1987).