*Quantum Chemistry*

# Ab Initio NMR Spectra for Molecular Systems with a Thousand and More Atoms: A Linear-Scaling Method**

*Christian Ochsenfeld,* Jörg Kussmann, and Felix Koziol*

The importance of NMR spectroscopy for modern chemistry and biochemistry cannot be overestimated. Starting with the classical NMR experiments in 1946 by Purcell and Bloch, [1] contributions by numerous scientists have propelled NMR spectroscopy to be an extremely powerful tool in the investigation of structure and dynamics of molecular systems both in solution and in the solid state (e.g., reference [2]). Despite this progress, the understanding and reliable assignment of observed experimental spectra often remains a highly difficult task. Thus, theoretical methods can be extremely useful, which is the focus of this work.

The most reliable way to predict NMR spectra for a specific molecular system is to calculate the NMR chemical shieldings by using quantum-chemical methods. For this purpose an entire hierarchy of methods (and basis sets) exists, which allows the exact result to be systematically approached. The only drawback is, however, the dramatic growth of the computational effort in approaching the exact solution and in increasing the number of atoms in a molecular system. Nevertheless, the hierarchy of ab initio methods allows approximate solutions to be selected and validated, so that error bars can be estimated and the simplest, reliable approximation for studying a specific class of molecular systems can be found.

In recent years there has been much progress with respect to the size of molecular systems that can nowadays be treated by using Hartree–Fock (HF) and density-functional methods, due to a reduction of the scaling of the computational effort to linear. Progress has been made mostly for the calculation of energetics (e.g., see references [3–8]), the optimization of structures by using analytic gradients (e.g., references [9,10]), and for the calculation of molecular properties (e.g., references [6,11]). However, the linear-scaling computation of

[*] Prof. Dr. C. Ochsenfeld, J. Kussmann, Dr. F. Koziol
Institut für Physikalische und Theoretische Chemie
Abteilung für Theoretische Chemie
Universität Tübingen
Auf der Morgenstelle 8, 72076 Tübingen (Germany)
Fax: (+49) 7071-29-5490
E-mail: christian.ochsenfeld@uni-tuebingen.de

[**] We thank Prof. Dr. Reinhart Ahlrichs (Universität Karlsruhe, Germany) for the possibility of using the program TURBOMOLE for the initial comparison of NMR data obtained with our new NMR program. The authors also thank Prof. Dr. Frank-Gerrit Klärner (Universität Essen, Germany) for providing the experimental NMR spectrum of nicotinamide within a molecular clip displayed in the title figure. C.O. further acknowledges financial support by an Emmy Noether research grant of the DFG ("Deutsche Forschungsgemeinschaft"). In addition, the authors thank Sun Microsystems for their financial contribution towards the cost of color reproductions of the cover page and within this work.

 DOI: 10.1002/ange.200460336  

NMR chemical shieldings has not been possible so far, so that a particularly important link between experiment and theory for large molecular systems has been missing.

The cost for the calculation of NMR chemical shifts within even the simplest ab initio approximation, the Hartree–Fock (HF) approach, increases conventionally with the third power of the molecular size M ($\mathcal{O}(M^3)$, where $\mathcal{O}()$ stands for the scaling order). Therefore, the computation of NMR chemical shieldings has so far been limited to molecular systems in the order of 100 atoms without molecular symmetry. For larger systems it is crucial to reduce the increase of the computational effort, which is the central goal of this work. We have reduced the scaling of the computational effort to linear so that the computation of NMR shieldings of molecular systems with 1000 atoms becomes possible, while the accuracy and reliability of traditional methods is preserved. In addition, due to the linear-scaling computational cost, any increase of computer speed will now lead to the same increase in the treatable molecular size.

The routine calculation of NMR chemical shifts[12–14] by using quantum-chemical methods has become possible since the introduction of local gauge-origin methods,[15–19] which provide a solution to the gauge-origin problem within approximated schemes. In our formulation we use gauge-including atomic orbitals (GIAO),[17–19] which have proven to be particularly successful.[12] We constrain ourselves to the HF method (GIAO-HF),[18–20] which provides reasonably good results for many molecular systems. For example, in many systems the GIAO-HF method yields $^1$H NMR chemical shifts with an accuracy of typically 0.2–0.4 ppm.[21–24]

In the first part of our report, we outline briefly our new method for the linear-scaling calculation of NMR shieldings. In the second part, two first examples for applications of our new method are presented, in which we study two systems that are important in the area of molecular recognition and that contain 490 and 1003 atoms (no point-group symmetry). The first example illustrates the new possibilities in assigning solid-state NMR spectra by converging computed shieldings with the solid-state fragment size. The second example shows that our new method allows for the first time the study of solvent effects in an explicit and systematic manner.

NMR chemical shifts are calculated as second derivatives of the energy with respect to the external magnetic field $\boldsymbol{B}$ and the nuclear magnetic spin $m_{N_j}$ of a nucleus $N$ [Eq (1)]:

$$\sigma_{ij}^{N} = \frac{\partial^2 E}{\partial B_i \, \partial m_{N_j}}, \tag{1}$$

in which $i$, $j$ are $x$, $y$, $z$ coordinates. Within HF this leads to [Eq (2)]:

$$\sigma_{ij}^{N} = \sum_{\mu\nu} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial B_i \, \partial m_{N_j}} + \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial B_i} \frac{\partial h_{\mu\nu}}{\partial m_{N_j}}, \tag{2}$$

in which $P_{\mu\nu}$ represents the one-particle density matrix (with basis functions $\mu$, $\nu$) and $h_{\mu\nu}$ the one-electron matrix. For the calculation of NMR shieldings the perturbed one-particle density matrices $\frac{\partial P_{\mu\nu}}{\partial B_i}$ (short: $\mathbf{P}^{B_i}$) are required, which are obtained by solving the coupled perturbed HF (CPHF) equations. To solve these equations entirely in the molecular-orbital (MO) basis—also necessary for the calculation of vibrational frequencies for example—requires a transformation of the AO two-electron integrals, which leads to an effort scaling with $\mathcal{O}(M^5)$. In the context of NMR shieldings the effort is conventionally reduced to $\mathcal{O}(M^3)$.[19,20] This overall scaling of $\mathcal{O}(M^3)$ for the calculation of NMR chemical shifts hampers the investigation of large molecular systems. To reduce the overall scaling to linear, two key issues need to be tackled: first, the CPHF equations have to be solved with linear-scaling effort and, second, all integral contractions need to be done linearly.

To reduce the scaling of the CPHF equations, we avoid transformations that involve the nonlocal MO coefficient matrix $\mathbf{C}$ entirely and in contrast solve directly for the required (local) derivative density $\mathbf{P}^{B_i}$. We start from our reformulation of CPSCF (CP self-consistent field) theory for the computation of vibrational frequencies within a density matrix-based scheme.[6] The constraints of the idempotency and the number of electrons are preserved through first order by the use of the purification transformation of McWeeny.[25] In the present work, we introduce a new density-based coupled-perturbed method, which allows the density derivatives to be solved directly with respect to the magnetic field. In contrast to our formulation for vibrational frequencies, we exploit the projection properties of $\mathbf{P}$ to solve directly for the required occupied-virtual part of the perturbed density matrix $\mathbf{P}^{B_i}$ [Eq (3)]:

$$\begin{aligned}&\mathbf{FP}^{B_i}\mathbf{SPS} + \mathbf{SPSP}^{B_i}\mathbf{F} - \mathbf{FPSP}^{B_i}\mathbf{S} - \mathbf{SP}^{B_i}\mathbf{SPF} + \mathbf{G}[\mathbf{P}^{B_i}]\mathbf{PS}\\ &+\mathbf{SPG}[\mathbf{P}^{B_i}] - 2\mathbf{SPG}[\mathbf{P}^{B_i}]\mathbf{PS} = \mathbf{FPS}^{B_i} + \mathbf{S}^{B_i}\mathbf{PF}\\ &-\mathbf{FPS}^{B_i}\mathbf{PS} - \mathbf{SPS}^{B_i}\mathbf{PF} - \mathbf{YPS} - \mathbf{SPY} + 2\mathbf{SPYPS},\end{aligned} \tag{3}$$

with:

$$Y_{\mu\nu} = \frac{\partial h_{\mu\nu}}{\partial B_i} + \sum_{\lambda\sigma} P_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)]}{\partial B_i},$$

$$G_{\mu\nu}[P_{\lambda\sigma}^{B_i}] = -\frac{1}{2} \sum_{\lambda\sigma} \frac{\partial P_{\lambda\sigma}}{\partial B_i} (\mu\lambda|\nu\sigma).$$

Mulliken notation is used for two-electron integrals. Our new formulation of the density matrix-based D-CPSCF equations preserves their excellent convergence properties, while offering important advantages for sparse-matrix multiplications. Convergence can be typically achieved within four to six builds of $\mathbf{G}[\mathbf{P}^{B_i}]$, which is similar to the convergence behavior of MO-based theories. The main advantage is of course the reduced scaling of the computational effort of $\mathcal{O}(M)$, since for systems with a nonvanishing HOMO–LUMO gap all matrices scale linearly. For an overall linear-scaling behavior as well all Fock-type matrices ($\mathbf{G}[\mathbf{P}^{B_i}]$ or $\mathbf{Y}$) need to be calculated with linear-scaling effort. This can be done by using extensions of our LinK (linear exchange K) method[8,9] introduced for the formation of exchange contributions to energies and gradients. Since $\mathbf{P}^{B_i}$ is skew-symmetric, the Coulomb part vanishes for $\mathbf{G}$. For the Coulomb-type matrices in $\mathbf{Y}$ we adapt the CFMM (continuous fast multipole method) formulation of White and Head-Gordon.[3] We implemented

  

our new method into a development version of the program Q-Chem,[26] in which so far no NMR code existed.

Efficiencies of our new linear-scaling GIAO-HF and the standard method are compared for linear alkanes (6-31G* basis) in Figure 1 for the same accuracies: integral and sparse

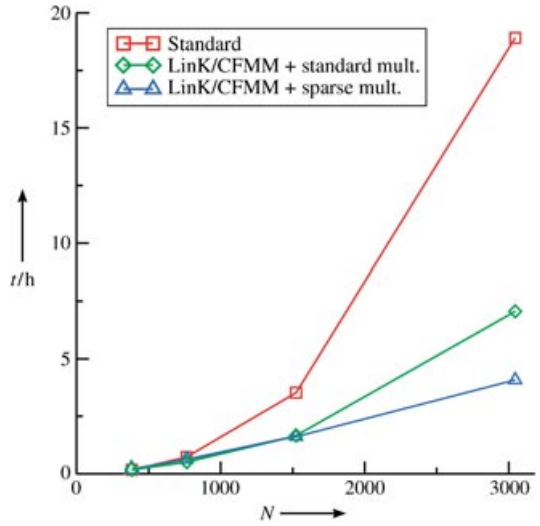


***Figure 1.*** Linear-scaling behavior of the density-matrix based GIAO-HF method with LinK/CFMM and sparse multiplications in comparison to the conventional $\mathcal{O}(M^3)$ scaling by using linear alkanes $C_nH_{2n+2}$ ($n = 20$, 40, 80, 160) within a 6-31G* basis (Itanium II 900 MHz, one processor). In addition, the D-GIAO-HF method with LinK/CFMM but with standard multiplications is plotted. $N$ is the number of basis functions, $t$ is the CPU time required to process the data.

multiplication thresholds are always $10^{-6}$. This is sufficient to obtain $^1$H NMR chemical shifts with accuracies of better than 0.1 ppm within HF/6-31G*. The accuracy can be rigorously controlled by numerical thresholds. Timings were calculated on one processor of an Itanium II 900 MHz; timings for a P4 2.8 GHz computer are listed in parentheses. It has to be stressed that in our present implementation the sparse-matrix multiplications are unusually slow on the Itanium II processor as compared to the P4. For $C_{80}H_{162}$ the overall gain of our preliminary implementation is a factor of 2.2 (2.5) for the total timing and 2.2–4.2 (2.9–4.0) for integral contractions within the D-CPHF scheme. For $C_{160}H_{322}$ the total win is a factor of 4.6; the wins for integral contractions are 3.9–9.1 for the Itanium II. In our preliminary implementation the $\mathcal{O}(M^3)$ matrix multiplications start to become dominant as compared to the integral contractions for alkanes with roughly 2100 (2850) basis functions. The crossover for sparse versus conventional matrix multiplications is in the region of 1400 basis functions for the Itanium II, whereas for the P4 computer the crossover occurs around 1150 basis functions. In addition, our new D-CPSCF scheme allows the reduction of the memory requirements to linear.

As another example, we selected a DNA fragment to investigate the performance of our new method. Here, we obtain for a DNA fragment with four base pairs (GIAO-HF/6-31G*; 280 atoms, 2862 basis functions) a performance win of 1.7–3.6 for integral contractions within our D-CPSCF scheme. In our present implementation we solve the D-CPSCF equations with a minimum number of integral contractions at the expense of more matrix multiplies, which have the smaller prefactor. Nevertheless, sparse matrix multiplies show a later onset of their linear-scaling behavior as compared to integral contractions, so that a good balance becomes important, which we are currently optimizing.

The linear-scaling behavior of our new density matrix-based GIAO-HF method allows for the first time the calculation of molecular systems with 1000 and more atoms without point-group symmetry on one-processor workstations or PCs. We present the first examples of an application below.

Solid-state NMR spectroscopy has evolved into a powerful tool in yielding important statical and dynamical information on molecules.[2] Recently, we demonstrated that some of the difficulties in assigning experimental solid-state NMR spectra can be overcome by using quantum-chemical methods, in which we studied examples of differently substituted hexabenzocoronenes[21,22] and host–guest systems.[23,24] In this way complementary information to X-ray spectroscopy can be obtained, which is particularly important for noncrystalline phases or positions of protons in hydrogen-bonded species. For the simulation of the spectra it is crucial to converge the NMR shieldings with respect to the size of the solid-state fragment to account for influences of neighboring molecules. However, so far both the size of the treatable system and that of the fragment was limited by the strong increase of the computational effort. Therefore one was constrained to relatively small systems and fragments, and it was not possible to avoid small basis sets and incremental approximations.[21,24] In contrast, with our new method we are now able to compute large systems and large fragments of the solid state, so that the size of treatable systems is presently at least increased by an order of magnitude.

As an example, we consider a host–guest complex with a tweezer-shaped host, binding a tetracyano-*p*-quinodimethane (TCNQ) guest used as model system for molecular recognition[27] (Figure 2). We studied similar systems earlier,[23,24] but were strongly limited with respect to the computation of NMR shieldings: for the largest computed system containing three complexes with a total of 276 atoms (guest: dicyanobenzene), we had to use a very small basis (3-21G)[24] at the GIAO-HF level. Since this basis set is too small to provide reliable $^1$H NMR predictions, an incremental approach was employed.[24] With our new linear-scaling NMR method these problems are solved. For the TCNQ host–guest complex it is now for the first time possible to include the nearest neighbors within a reasonable basis for $^1$H NMR shifts of the present system (GIAO-HF/6-31G*). The structure is shown in Figure 2 and consists of a total of five complexes (denoted as pentamer; 490 atoms; structure obtained by projecting the HF/6-31G* optimized monomer onto the X-ray solid-state ordering[24]). The calculations show that the accuracy of the $^1$H NMR chemical shifts for the central complex within the full pentamer calculation (GIAO-HF/6-31G*) is improved by 0.8 ppm as compared to an estimate of these shieldings for the central complex from GIAO-HF/3-21G trimer values, which was the largest system treatable before our new method. The latter estimate has been

  

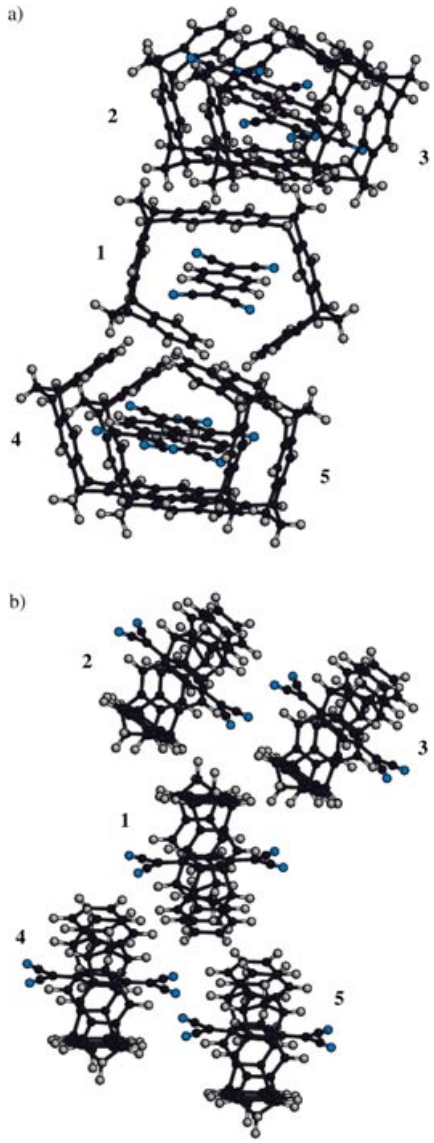


***Figure 2.*** Solid-state fragment of the molecular tweezer host-guest complex (TCNQ guest) with five neighbors (490 atoms): a) front and b) side-on view. C atoms are marked in dark grey, H atoms in light grey, N atoms in blue. N atoms occur only in the four cyano groups of the guest.

obtained by two trimer calculations, in which the influence of units **2**, **3** (Figure 2) and **4**, **5** on the central unit **1** have been added. This error with respect to the full calculation is slightly reduced to 0.7 ppm (the mean error from 0.5 to 0.2 ppm), if the trimer influences (GIAO-HF/3–21G) are added to the GIAO-HF/6-31G* monomer data in an incremental fashion. The calculation for the pentamer avoids the deficiencies of incremental approximations, which illustrates some of the new possibilities with our new linear-scaling method for predicting solid-state NMR spectra.

The second example is focused on an artificial receptor for $NAD^+$ (NAD is nicotinamide adenine dinucleotide) in water recently presented by the groups of Klärner and Schrader.[28] The form of the host is that of a molecular clip inside which $NAD^+$ is bound. However, the structure of the host–guest complex remained unclear based on experimental information only. Therefore a comparison of computed and measured NMR chemical shifts is helpful and allows some of the possible binding motifs to be discarded.[29] A complication arises as the NMR measurements have been performed in water. Here, our new method allows for the first time to estimate explicitly the influence of the solvent at the ab initio level, as we are able to treat both the host–guest complex (containing 88 atoms) and neighboring water molecules at the GIAO-HF level: the largest system computed contains 1003 atoms. In this work, our aim is only a proof of principle that we are now able to deal with such large systems and that we can converge the results with respect to the number of surrounding water molecules.

As a first step we performed a classical molecular dynamics simulation of *N*-methyl nicotinamide in a molecular clip[29] within a water surrounding (300 K; MMFF94; 1 ps equilibration; total 20 ps, 1.0 fs time-steps; 2575 water molecules in dielectricum). A snapshot of the MD simulation was selected, for which the water molecules were closest to the center of the guest ring within a radius of 800 pm. Around the host-guest system (88 atoms) different shells of surrounding water molecules were defined (total size: 169, 547, and 1003 atoms). These systems were used to calculate the NMR chemical shifts (GIAO-HF/6-31G**), with the largest system containing 1003 atoms and 8593 basis functions shown in Figure 3. The maximum change with respect to $^1$H NMR shieldings (protons at the phosphonate groups are not considered in the following) from the isolated host–guest to the 169 atoms system is 1.3 ppm, from 169 to 547 atoms 1.1 ppm, and from 547 to 1003 atoms 0.2 ppm. This result indicates that in the system that contains 1003 atoms, the proton chemical shifts of the host–guest complex are converged to 0.2 ppm. Overall, the maximum influence of the water molecules in the 1003 atom system on host–guest protons is 1.3 ppm and 0.7 ppm if only the aromatic guest protons are considered.

It is clear that for a reliable description of solvent effects not only more extensive MD simulations are necessary but also the NMR shieldings for a multitude of snapshots need to be calculated. Our example application shows that we are now for the first time able to converge the NMR shieldings with respect to the number of surrounding water molecules explicitly, so that the minimum number of water molecules required for an adequate description can be obtained. In addition, the maximum influence of 0.7 ppm calculated for the water-surrounded complex with 1003 atoms presented above is a first hint that complex-induced changes of 6 ppm for guest-proton shieldings calculated in the vacuum for some binding motifs of $NAD^+$ within a clip,[29] could also be significant in water.

Our new linear-scaling GIAO-HF method allows for the first time the ab initio calculation of NMR chemical shifts for molecular systems with 1000 and more atoms. The accuracy and reliability of conventional methods is fully preserved. Since it is difficult to obtain highly accurate energetics or structural parameters for such large systems due to the large number of energetically close configurations, our new linear-scaling method for the computation of NMR chemical shifts

  

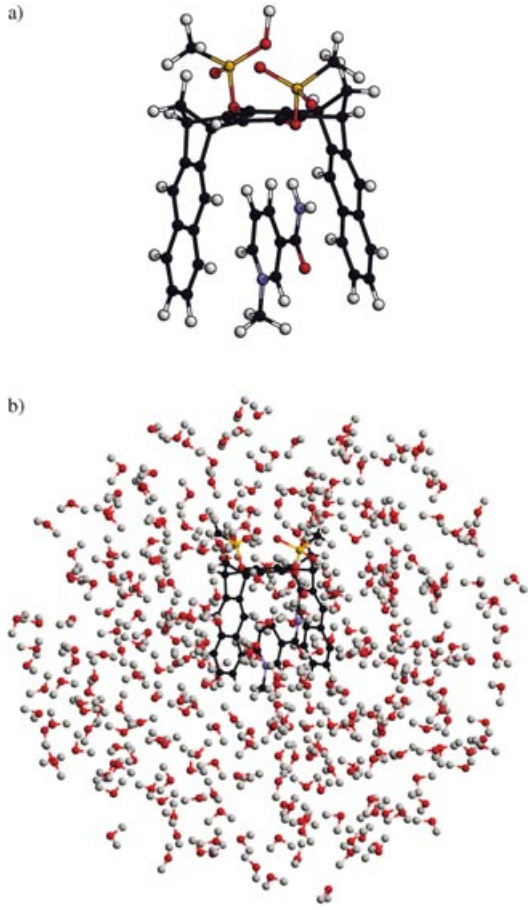


***Figure 3.*** a) Snapshot of *N*-methyl nicotinamide within a molecular clip and b) the same guest-clip system, including the surrounding water molecules comprising 1003 atoms, used to estimate solvent effects on the NMR chemical shifts (GIAO-HF/6-31G*). C atoms are marked in dark grey, H atoms light grey, N atoms blue, O atoms red, and P atoms orange.

opens up the way to a molecular property that allows for a direct interplay with experimental studies as well for large molecular systems. Besides our first two example applications that investigate the convergence of solid-state NMR shieldings and explicit solvent effects, a multitude of new possibilities arise for the study of large molecular systems such as proteins, enzymes, or active centers.

Received: April 15, 2004
Revised: June 1, 2004 [Z460336]

**Keywords:** ab initio calculations · molecular recognition · NMR spectroscopy · quantum chemistry

[1] a) E. M. Purcell, H. G. Torrey, R. V. Pound, *Phys. Rev.* **1946**, *69*, 37; b) F. Bloch, W. Hansen, M. E. Packard, *Phys. Rev.* **1946**, *69*, 127.
[2] a) K. Wüthrich, *Angew. Chem.* **2003**, *115*, 3462; *Angew. Chem. Int. Ed.* **2003**, *42*, 3340; b) S. P. Brown, H. W. Spiess, *Chem. Rev.* **2001**, *101*, 4125.
[3] C. A. White, M. Head-Gordon, *Chem. Phys. Lett.* **1994**, *230*, 8.
[4] M. Challacombe, E. Schwegler, J. Almlöf, *J. Chem. Phys.* **1996**, *104*, 4685.
[5] M. C. Strain, G. E. Scuseria, M. J. Frisch, *Science* **1996**, *271*, 51.
[6] C. Ochsenfeld, M. Head-Gordon, *Chem. Phys. Lett.* **1997**, *270*, 399.
[7] G. E. Scuseria, *J. Phys. Chem. A* **1999**, *103*, 4782.
[8] C. Ochsenfeld, C. A. White, M. Head-Gordon, *J. Chem. Phys.* **1998**, *109*, 1663.
[9] C. Ochsenfeld, *Chem. Phys. Lett.* **2000**, *327*, 216.
[10] Y. H. Shao, C. A. White, M. Head-Gordon, *J. Chem. Phys.* **2001**, *114*, 6572.
[11] H. Larsen, T. Helgaker, J. Olsen, P. Jorgensen, *J. Chem. Phys.* **2001**, *115*, 10344.
[12] J. Gauss, *Ber. Bunsen-Ges. Phys. Chem.* **1995**, *99*, 1001.
[13] T. Helgaker, M. Jaszunski, K. Ruud, *Chem. Rev.* **1999**, *99*, 293.
[14] "Ab initio NMR Chemical Shift Computation": U. Fleischer, W. Kutzelnigg, C. van Wüllen in *Encyclopedia of Computational Chemistry* (Ed.: P. v. R. Schleyer, N.-L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, P. R. Schreiner), Wiley, Chichester, **1998**, pp. 1827–1835.
[15] W. Kutzelnigg, *Isr. J. Chem.* **1980**, *19*, 193.
[16] A. E. Hansen, T. D. Bouman, *J. Chem. Phys.* **1985**, *82*, 5035.
[17] F. London, *J. Phys. Radium* **1937**, *8*, 397.
[18] R. Ditchfield, *Mol. Phys.* **1974**, *27*, 789.
[19] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251.
[20] M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, H. Horn, *Theor. Chim. Acta* **1992**, *83*, 455.
[21] C. Ochsenfeld, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2153.
[22] C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss, H. W. Spiess, *J. Am. Chem. Soc.* **2001**, *123*, 2597.
[23] S. P. Brown, T. Schaller, U. P. Seelbach, F. Koziol, C. Ochsenfeld, F.-G. Klärner, H. W. Spiess, *Angew. Chem.* **2001**, *113*, 740; *Angew. Chem. Int. Ed.* **2001**, *40*, 717.
[24] C. Ochsenfeld, F. Koziol, S. P. Brown, T. Schaller, U. P. Seelbach, F.-G. Klärner, *Solid State Nucl. Magn. Reson.* **2002**, *22*, 128.
[25] R. McWeeny, *Rev. Mod. Phys.* **1960**, *32*, 335.
[26] J. Kong, C. A. White, A. I. Krylov, C. D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, S. R. Gwaltney, T. R. Adams, H. Daschel, W. Zhang, P. P. Korambath, C. Ochsenfeld, A. T. B. Gilbert, G. S. Keziora, D. R. Maurice, N. Nair, Y. Shao, N. A. Besley, P. E. Maslen, J. P. Dombroski, J. Baker, E. F. C. Byrd, T. v. Voorhis, M. Oumi, S. Hirata, C.-P. Hsu, N. Ishikawa, J. Florian, A. Warshel, B. G. Johnson, P. M. W. Gill, M. Head-Gordon, J. A. Pople, *J. Comput. Chem.* **2000**, *21*, 1532.
[27] F.-G. Klärner, U. Burkert, M. Kamieth, R. Boese, J. Benet-Buchholz, *Chem. Eur. J.* **1999**, *5*, 1700.
[28] C. Jasper, T. Schrader, J. Panitzky, F.-G. Klärner, *Angew. Chem.* **2002**, *114*, 1411; *Angew. Chem. Int. Ed.* **2002**, *41*, 1355.
[29] M. Fokkens, C. Jasper, T. Schrader, F. Koziol, C. Ochsenfeld, J. Polkowska, F.-G. Klärner, submitted.

  