

KWIK: Coulomb Energies in $O(N)$ Work

Jeremy P. Dombroski,[†] Stephen W. Taylor,[‡] and Peter M. W. Gill^{*,†}

Department of Chemistry, Massey University, Palmerston North, New Zealand, and
Department of Mathematics, University of Auckland, Tamaki, New Zealand

Received: September 22, 1995[⊗]

We introduce the KWIK algorithm for computing the Coulomb energy of N localized charge distributions. Asymptotically, like the Fast Multipole Method (FMM), the computational cost of the method scales linearly with N . This scaling can be traced to the Laws of Large Numbers and, in particular, to the statistics of the two-dimensional random walk. We have implemented the algorithm on a small workstation and applied it to systems with up to 10^6 charged particles.

1. Introduction

Many important problems in computational physics and chemistry can be treated by methods based on the total interaction energy E between N well-localized matter distributions. In many cases, the distributions interact through a two-particle inverse-square force law (gravitational or Coulombic, for example), and the interaction potential is therefore $1/r_{ij}$. Because this is a long-range potential, even widely separated distributions interact significantly, and straightforward cutoff approximations introduce substantial errors. On the other hand, if E is evaluated exactly by adding explicitly computed contributions from all pairs of distributions in the system, the cost of the resulting algorithm will necessarily scale *quadratically* with N .

The development of algorithms that treat all pairwise interactions in an N -particle system but whose computational costs grow only *linearly* with N has had a profound impact on a wide range of scientific activities. Many of these linear methods, which stem from seminal work^{1–4} in the mid-1980s, have been developed by Rokhlin and his collaborators^{4–11} and have since been applied to problems in molecular dynamics^{12,13} and quantum chemistry.^{14–17}

Although the FMM of Greengard and Rokhlin can be shown to be linear, its cost (in three dimensions) scales with the fourth power of the logarithm of the accuracy required.^{6,18} Furthermore, although emerging linear methods^{8–11} based on the Fast Wavelet Transform are likely to be faster than the FMM if implemented efficiently, their costs remain substantial. It thus remains important to search for still faster linear algorithms.

The breathtaking performance of the latest generation of computing machines, combined with the “fast” methods, allows us now to undertake calculations involving $N = O(10^6)$ particles, and future developments will doubtless soon render such heroic achievements routine. Given, however, that such large systems can now be contemplated, it is intriguing to consider whether very large values of N can actually become an asset, rather than a liability. More specifically, one could speculate that the Laws of Large Numbers (which underpin the success of statistical mechanics) ought somehow to be able to be brought to bear on the present problem. Despite the obvious appeal of such a proposition, however, we are not aware of any previous attempts along these lines.

In this paper, we describe the KWIK algorithm which, in that it partitions E into a short-range part (computed analytically)

and a long-range part (computed by Fourier summation), shares some of the features of the traditional Ewald approach and its variants.^{19–23} However, whereas the usefulness of Ewald schemes derives from their exploitation of periodic boundary conditions, the effectiveness of KWIK depends only on the Central Limit Theorem²⁴ and therefore increases with N . As a consequence, KWIK may be useful in the study of the complicated aperiodic systems which arise in astrophysical simulations, incompressible fluid dynamics, Brownian dynamics, and molecular quantum chemistry.

2. KWIK Theory

We define the energy of N distributions $Q_i(\mathbf{r})$ by

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \int \int Q_i(\mathbf{r}_1) \left(\frac{1}{r_{12}} \right) Q_j(\mathbf{r}_2) \, \mathbf{d}\mathbf{r}_1 \, \mathbf{d}\mathbf{r}_2 - E_{\text{SI}} \quad (1)$$

where $\mathbf{r}_{12} \equiv \mathbf{r}_1 - \mathbf{r}_2$ and the integrations are over all space. If included, the last term

$$E_{\text{SI}} = \frac{1}{2} \sum_{i=1}^N \int \int Q_i(\mathbf{r}_1) \left(\frac{1}{r_{12}} \right) Q_i(\mathbf{r}_2) \, \mathbf{d}\mathbf{r}_1 \, \mathbf{d}\mathbf{r}_2 \quad (2)$$

removes self-interactions. For convenience later, we assume that the system is confined to a d -dimensional cube with side length π . We seek to compute E within a relative accuracy ϵ .

Since $\text{erf}(x)$ is an odd function and $\text{erf}(x) \rightarrow 1$ rapidly as $x \rightarrow \infty$,²⁵ the identity

$$\frac{1}{r_{12}} \equiv \frac{\text{erfc}(\omega r_{12})}{r_{12}} + \frac{\text{erf}(\omega r_{12})}{r_{12}} \quad (3)$$

separates the interaction potential into a singular but rapidly decaying part and a nonsingular but slowly decaying part. Substituting (3) into (1) then leads to $E = E_{\text{short}} + E_{\text{long}}$ where

$$E_{\text{short}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \int \int Q_i(\mathbf{r}_1) \left(\frac{\text{erfc}(\omega r_{12})}{r_{12}} \right) Q_j(\mathbf{r}_2) \, \mathbf{d}\mathbf{r}_1 \, \mathbf{d}\mathbf{r}_2 - E_{\text{SI}} \quad (4)$$

$$E_{\text{long}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \int \int Q_i(\mathbf{r}_1) \left(\frac{\text{erf}(\omega r_{12})}{r_{12}} \right) Q_j(\mathbf{r}_2) \, \mathbf{d}\mathbf{r}_1 \, \mathbf{d}\mathbf{r}_2 \quad (5)$$

Note that, unlike the boxing schemes of the fast methods,^{1,2,5,6}

[†] Massey University.

[‡] University of Auckland.

[⊗] Abstract published in *Advance ACS Abstracts*, March 1, 1996.

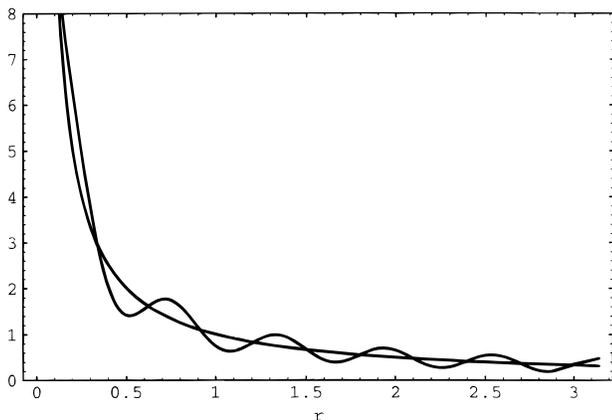


Figure 1. Function $1/r$ and its (10,10) approximation for $d = 1$.

this separation of short- and long-range interactions is infinitely differentiable. Replacing the long-range term in (3) by its truncated Fourier series expansion yields what we will call the (M,ω) approximation to $1/r_{12}$

$$\frac{1}{r_{12}} \approx \frac{\text{erfc}(\omega r_{12})}{r_{12}} + \sum_{|\mathbf{k}| \leq M} A(\mathbf{k}, \omega) \exp(i\mathbf{k} \cdot \mathbf{r}_{12}) \quad (6)$$

Figure 1 depicts the (10,10) approximation for $d = 1$. The sum in (6) then factorizes (5) to yield

$$E_{\text{long}} \approx \frac{1}{2} \sum_{|\mathbf{k}| \leq M} A(\mathbf{k}, \omega) I(\mathbf{k}) \quad (7)$$

where

$$I(\mathbf{k}) = \left| \sum_{j=1}^N \hat{Q}_j(\mathbf{k}) \right|^2 = \left| \sum_{j=1}^N \int Q_j(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \, d\mathbf{r} \right|^2 \quad (8)$$

is the intensity of the Fourier transform of the total distribution.

If the Fourier amplitudes $A(\mathbf{k}, \omega)$ and formulae for the transforms $\hat{Q}_j(\mathbf{k})$ are available, it requires $O(N)$ work to evaluate $A(\mathbf{k}, \omega) I(\mathbf{k})$. Thus, to the extent that M does not grow with N , E_{long} can be found in $O(N)$ work. The expression for E_{short} apparently involves N^2 terms. However, ω can be chosen large enough that any nominated fraction of these terms contributes negligibly to the sum and, in particular, ω can be employed to ensure that E_{short} can be found in $O(N)$ work. With optimistic reference to the final four letters in (7), we refer to our method as a KWIK algorithm.

The optimal choices of ω and M balance the work required to compute E_{short} and E_{long} . If ω and M are too small, excessive time will be spent finding E_{short} . Conversely, if they are too large, overmuch time will be expended computing E_{long} . The key to choosing ω and M well is to use accurate *a priori* estimates, as functions of ω and M , of the times T_{short} and T_{long} to evaluate E_{short} and E_{long} . Moreover, since KWIK becomes the quadratic method at $\omega = 0$, the strategy of selecting the ω and M that minimize $T_{\text{KWIK}} = T_{\text{short}} + T_{\text{long}}$ ensures a continuity between the KWIK and quadratic schemes.

3. Collinear Point Charges and a Random Walk in the Complex Plane

Our first application of KWIK has been to the one-dimensional (*i.e.*, $d = 1$) problem of N unit-charge particles at random points $0 < r_i < \pi$. To compute the short-range energy

$$E_{\text{short}} = \sum_{i < j}^N \frac{\text{erfc}(\omega r_{ij})}{r_{ij}} - \frac{N\omega}{\sqrt{\pi}} \quad (9)$$

efficiently, we utilize neighbor lists to include only those pairs of particles whose separation is less than the root of $\text{erfc}(\omega r_{12}) = \epsilon r_{12}$. Efficiently computing the long-range energy

$$E_{\text{long}} \approx \frac{1}{2} \sum_{k=0}^M A(k, \omega) I(k) \quad (10)$$

is a fascinating problem that depends delicately on the decay behavior of the $A(k, \omega) I(k)$.

If $\text{erfc}(\omega\pi)$ is negligible, it can be shown that

$$A(k, \omega) = \frac{1}{\pi} \left[E_1 \left(\frac{k^2}{4\omega^2} \right) + 2\text{Ci}(\pi k) \right] \quad (11)$$

where E_1 and Ci are the exponential integral and cosine integral functions,²⁵ respectively. For small k , the $A(k, \omega)$ are positive and, like $E_1(k^2/4\omega^2)$, decay approximately exponentially. However, asymptotic expansions²⁵ for E_1 and Ci can be used to show that, beyond

$$k_{\text{crit}} = 2\omega[\ln(2\pi^2\omega^2)]^{1/2} \quad (12)$$

the Ci term in (11) becomes dominant and the $A(k, \omega)$ begin to oscillate and decay only as k^{-2} . Figure 2 illustrates this for $\omega = 10$: whereas the $A(k, 10)$ decrease by 5 orders of magnitude between $k = 0$ and $k = 60$, it is not until $k = 10\,000$ that a second such reduction is achieved.

The Fourier transform of $Q_j(r) = \delta(r - r_j)$ is trivial, and one immediately obtains

$$I(k) = \left| \sum_{j=1}^N \exp(ikr_j) \right|^2 \quad (13)$$

Figure 3 shows $I(k)$ for a random (and typical) arrangement of $N = 10^6$ particles. The key feature is the rapid decay of the graph from 10^{12} to a limiting, albeit noisy, value of 10^6 . An illuminating way in which to understand this behavior is to realize that it is intimately connected to the two-dimensional random-walk problem. If each kr_j is assumed to be a random phase, $[I(k)]^{1/2}$ can be interpreted as the net progress of a drunkard who takes N unit steps, each in a different random direction, on the complex plane. This is known as the Pearson random walk in two dimensions,²⁴ and it is known^{26,27} that, for large N , $I(k)$ is an exponential random variable whose mean and standard deviation equal N .

Of course, the random-phase assumption is completely false for $k = 0$ and so $I(0) = N^2$. However, as k increases, the assumption improves and the $I(k)$ decay rapidly before eventually settling down and fluctuating about a mean value of N . Since the $A(k, \omega)$ drop smoothly but more slowly, it follows that the $A(k, \omega) I(k)$ decay rapidly to roughly $1/N$ of $A(0, \omega) I(0)$ and more slowly thereafter. A felicitous consequence of this is that the series for E_{long} rapidly converges to within a relative error which is $O(N^{-1})$, and it is this property that most strongly suggests that KWIK may be especially well suited to the investigation of very large systems. Indeed, in the sense that the KWIK solution to the Coulomb problem becomes *more* effective as N grows, KWIK may be viewed as complementary to the conventional quadratic method.

To assess KWIK numerically, we have investigated its performance on several randomly chosen ensembles of collinear

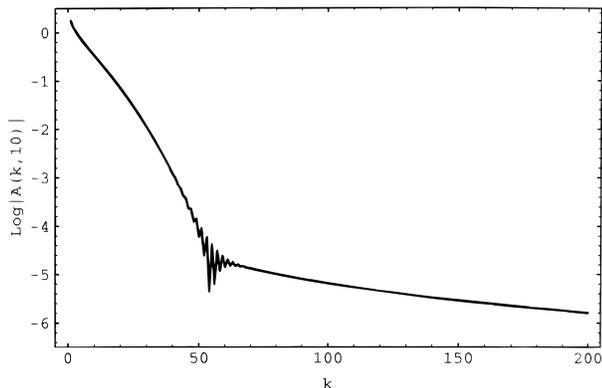


Figure 2. Function $\log |A(k,10)|$ for $d = 1$.

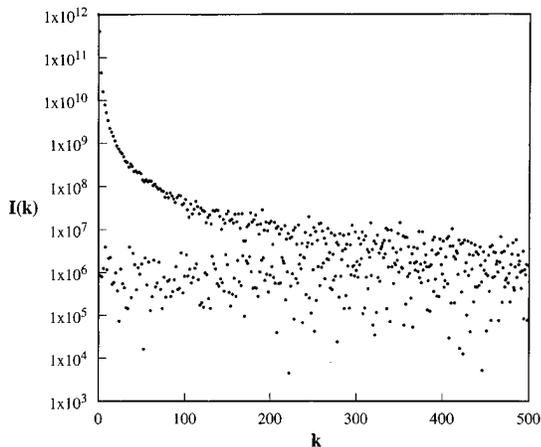


Figure 3. Typical Fourier intensity $I(k)$ for 10^6 random-point charges in $(0,\pi)$.

TABLE 1: CPU Timings for N Collinear Unit Charge Particles ($\epsilon = 10^{-3}$)

N	T_{Quad}	ω	M	T_{short}	T_{long}	T_{KWIK}
1 000	0.4	100	30	0.2	0.2	0.4
5 000	7.8	1 000	30	1.0	0.8	1.8
10 000	29.5	3 000	25	1.9	1.7	3.6
50 000	706	7 500	50	10.0	8.5	18.5
100 000	2 810	16 000	50	19.9	17.3	37.2
500 000	70 000	45 000	50	114	85.3	199
1 000 000	280 000	85 000	75	236	178	414

TABLE 2: CPU Timings for N Collinear Unit Charge Particles ($\epsilon = 10^{-5}$)

N	T_{Quad}	ω	M	T_{short}	T_{long}	T_{KWIK}
1 000	0.4	140	300	0.3	0.2	0.5
5 000	7.8	250	490	1.5	1.3	2.8
10 000	29.5	400	700	3.2	2.9	6.1
50 000	706	1 000	1600	23.7	22.5	46.2
100 000	2 810	2 300	1800	44.2	48.6	92.8
500 000	70 000	9 000	2300	267	294	561
1 000 000	280 000	24 000	1900	454	505	959

particles. We have explored $10^2 \leq N \leq 10^6$ and $10^{-3} \leq \epsilon \leq 10^{-10}$ and have adjusted ω and M manually to minimize $T_{\text{KWIK}} = T_{\text{short}} + T_{\text{long}}$. The resulting CPU timings on an IBM RS/6000 Model 355 workstation, together with the T_{Quad} obtained using an optimized quadratic program, are summarized in Tables 1–3 and Figures 4 and 5.

The raw data in Tables 1–3 show how the cost of KWIK increases with N for fixed ϵ . Figure 4 shows log–log plots of T_{KWIK} and T_{Quad} against N . As one would expect, the T_{Quad} graph is linear with a slope of two—as befits a quadratic method. In contrast, we find that the slopes of the T_{KWIK} curves are rather greater than one for small N but tend to one as N grows. Precise

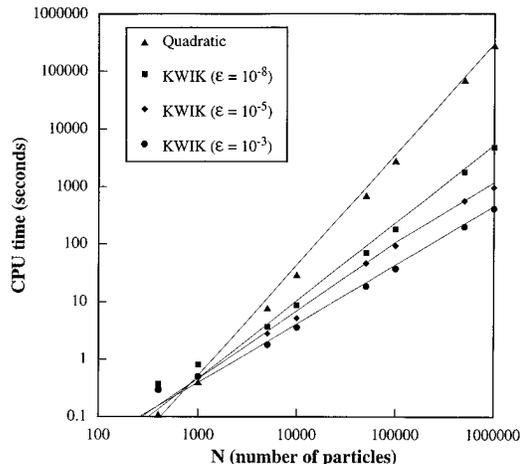


Figure 4. CPU time to compute the energy of N point charges ($d = 1$) using the quadratic method and the KWIK method with various accuracies.

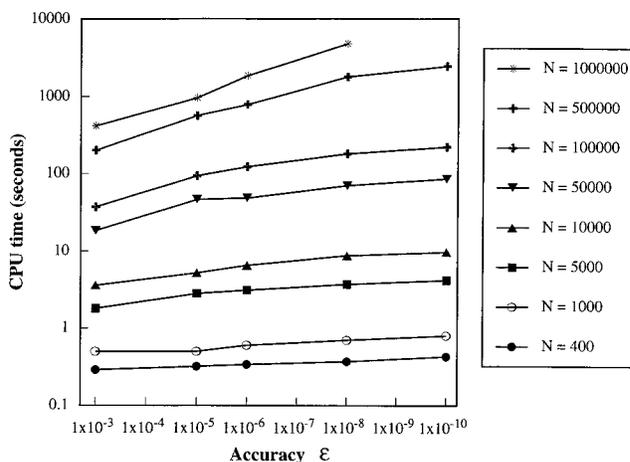


Figure 5. CPU time vs accuracy to compute the energy of N point charges ($d = 1$).

TABLE 3: CPU Timings for N Collinear Unit Charge Particles ($\epsilon = 10^{-8}$)

N	T_{Quad}	ω	M	T_{short}	T_{long}	T_{KWIK}
1 000	0.4	100	600	0.3	0.4	0.7
5 000	7.8	190	950	1.9	1.8	3.7
10 000	29.5	300	1 500	4.2	4.5	8.7
50 000	706	730	3 050	33.9	35.8	69.7
100 000	2 810	1100	4 300	85.9	93.2	179
500 000	70 000	2800	10 000	783	1000	1780
1 000 000	280 000	4150	14 000	2100	2660	4760

measurements reveal that the computational cost of KWIK grows roughly as $N^{1.3}$ when $N < 1/\epsilon$ but only as N when $N > 1/\epsilon$. On this basis, we may conclude that KWIK is an $O(N)$ method and, furthermore, that it performs best when both N and ϵ are large.

How can these empirical observations be rationalized? The key is the “random-walk” property of $I(k)$ discussed above. As k increases, the $I(k)$ decrease rapidly to approximately $1/N$ of $I(0)$ but then remain more or less constant. Because of this, it requires comparatively few terms in the E_{long} sum to achieve a relative accuracy of $1/N$. However, if greater accuracy than this is required, we must depend on the (much more slowly decaying) amplitudes $A(k,\omega)$ to converge the E_{long} sum.

The number of collinear particles beyond which the KWIK algorithm becomes faster than the quadratic method can be seen directly from Figure 4. Although the exact point depends on the value of ϵ used (and, of course, implementational details in

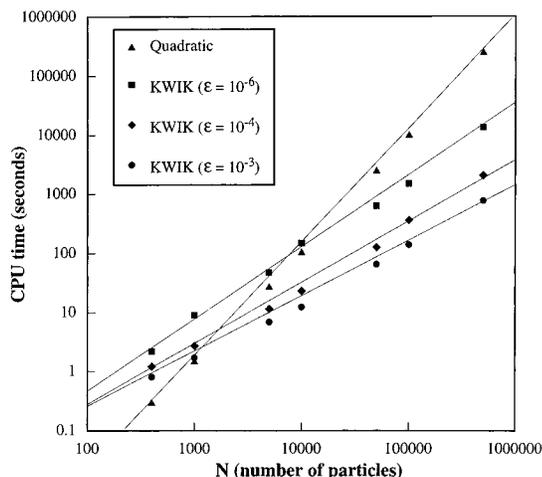


Figure 6. CPU time to compute the energy of N point charges ($d = 3$) using the quadratic method and the KWIK method with various accuracies.

both the KWIK and quadratic programs), it is clear that the crossover occurs at roughly 1000–3000 particles for ϵ values of interest to us. Similar crossover points have been found⁶ when comparing the quadratic and fast multipole methods. To put this in perspective, we note that the only problems for which the quadratic method is faster than KWIK are ones which, on a modern workstation, take less than 1 CPU s anyway. It may reasonably be inferred from this that the quadratic method is more or less obsolete.

Figure 5 shows log–log plots of T_{KWIK} against ϵ for various values of N . It is evident that improving the error in a KWIK calculation, even by several orders of magnitude, can be achieved at a fairly modest additional cost. It is also apparent that KWIK calculations become very inexpensive when large errors can be tolerated.

4. Point Charges in Three Dimensions

Applying KWIK to three-dimensional ensembles of point charges is a straightforward extension of the linear case. In order to demonstrate this, we have studied collections of N unit charges at random points \mathbf{r}_i in the cubic box $0 < x, y, z < \pi$. The qualitative behavior of KWIK is similar for one-dimensional (1-D) and three-dimensional (3-D) systems, and to avoid an unilluminating repetition of the foregoing section, our discussion of the 3-D case will be terse.

A compact formula, analogous to (11), for the 3-D Fourier amplitudes $A(\mathbf{k}, \omega)$ is difficult to obtain, and we have therefore resorted to quadrature to evaluate these. This is not entirely satisfactory, and we will discuss more efficient generation of the amplitudes elsewhere.²⁸ As in the 1-D case, the $A(\mathbf{k}, \omega)$ decay rapidly at first with increasing $|\mathbf{k}|$ and more slowly thereafter. They are especially small when k_x , k_y , and k_z are all large. The 3-D intensity

$$I(\mathbf{k}) = \left| \sum_{j=1}^N \exp(i\mathbf{k} \cdot \mathbf{r}_j) \right|^2 \quad (14)$$

behaves similarly to its 1-D analogue (13), falling dramatically from $I(\mathbf{0}) = N^2$ to fluctuate around N as $|\mathbf{k}|$ becomes large. Again, this can be understood in terms of a 2-D Pearson walk.

Figure 6 shows plots of CPU times against N for various ϵ for this 3-D problem. The crossover points (at which KWIK becomes faster than the quadratic method) occur at higher N values than in the 1-D calculations, but this is mainly a consequence of the fact that our 3-D KWIK program is not yet

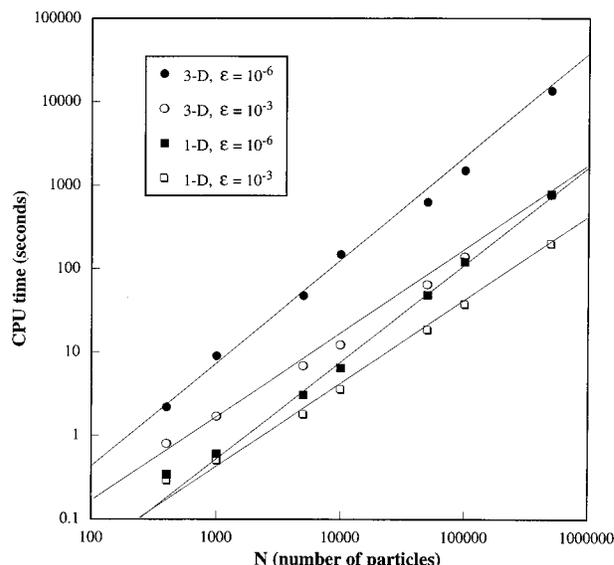


Figure 7. CPU time to compute the energy of N point charges showing the effects of varying accuracy and varying dimensionality.

optimized. The key observation, however, is that, as we found in the 1-D tests, the KWIK timings increase superlinearly for $N < 1/\epsilon$ and linearly thereafter.

Finally, in Figure 7, we directly compare our 1-D and (unoptimized) 3-D CPU timings. We observe that, unlike the FMM^{5,6} which is very much more expensive in 3-D than 1-D, the cost of KWIK appears to increase rather slowly with the system dimensionality d .

5. Concluding Remarks

In this paper, we have given a preliminary account of the KWIK algorithm, a novel solution to the Coulomb problem of N charge distributions. We have shown that, for a fixed relative accuracy ϵ , the computational cost of KWIK is $O(N)$. This asymptotic linear scaling behavior is observed in practice for $N > 1/\epsilon$. In this sense, large values of N actually become an asset, not a liability, for the same Laws of Large Numbers that underpin the success of statistical mechanics, used to advantage by KWIK. We find that KWIK is faster than the conventional quadratic approach for all but very small problems.

KWIK can be applied to localized continuous charge distributions (such as arise in ab initio molecular orbital theory), and we are currently examining its usefulness in computing the \mathbf{J} matrix, the infamous bottleneck in Hartree–Fock or Kohn–Sham self-consistent field (SCF) calculations. There are strong indications that the quadratic approaches—preeminent in quantum chemistry for half a century—may soon be displaced. We will report our findings and comparisons with the CFMM¹⁷ in the near future.²⁸

Acknowledgment. It is a pleasure to cite the landmark work on the quantum chemical Coulomb problem by S. F. Boys and I. Shavitt²⁹ and to acknowledge its influence on the development of KWIK. The latter was presented at the Molecular Quantum Mechanics conference held in Boys’ and Shavitt’s honor at Cambridge University (Sept 3–7, 1995). This research was supported by the National Science Foundation (Grant DMI-9460396), the NZ Lottery Grants Board (Grant AP-32002), and Q-Chem Inc. J.P.D. thanks Massey University, the NZ Vice Chancellors’ Committee, and the NZ Institute of Chemistry for financial support.

References and Notes

- (1) Appel, A. W. *SIAM J. Sci. Stat. Comput.* **1985**, *6*, 85.
- (2) Barnes, J.; Hut, P. *Nature* **1986**, *324*, 446.

- (3) Daubechies, I. *Comm. Pure Appl. Math.* **1988**, *41*, 909.
- (4) Rokhlin, V. J. *Comput. Phys.* **1985**, *60*, 187.
- (5) Greengard, L.; Rokhlin, V. J. *Comput. Phys.* **1987**, *73*, 325.
- (6) Greengard, L. F. *The Rapid Evaluation of Potential Fields in Particle Systems*; MIT: Cambridge, MA, 1988.
- (7) Greengard, L.; Rokhlin, V. *Comm. Pure Appl. Math.* **1991**, *44*, 419.
- (8) Alpert, B. K.; Rokhlin, V. *SIAM J. Sci. Stat. Comput.* **1991**, *12*, 158.
- (9) Beylkin, G.; Coifman, R.; Rokhlin, V. *Comm. Pure Appl. Math.* **1991**, *44*, 141.
- (10) Alpert, B. K. In *Wavelets: A Tutorial in Theory and Applications*; Chui, C. K., Ed.; Academic: New York, 1992.
- (11) Beylkin, G. *SIAM J. Numer. Anal.* **1992**, *6*, 1716.
- (12) Board, J. A.; Causay, J. W.; Leathrum, J. F.; Windemuth, A.; Schulten, K. *Chem. Phys. Lett.* **1992**, *198*, 89.
- (13) Ding, H.-Q.; Karasawa, N.; Goddard, W. A. *J. Chem. Phys.* **1992**, *97*, 4309.
- (14) Cho, K.; Arias, T. A.; Joannopoulos, J. D. *Phys. Rev. Lett.* **1993**, *71*, 1808.
- (15) Li, Z.; Borrmann, A.; Martens, C. C. *Chem. Phys. Lett.* **1993**, *214*, 362.
- (16) White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593.
- (17) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.
- (18) Petersen, H. G.; Smith, E. R.; Soelvason, D. *Proc. R. Soc. London, Ser. A* **1995**, *448*, 401.
- (19) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253.
- (20) Perram, J. W.; Petersen, H. G.; de Leeuw, S. W. *Mol. Phys.* **1988**, *65*, 875.
- (21) Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Oxford Science: Oxford, 1990.
- (22) Natoli, V.; Ceperley, D. M. *J. Comput. Phys.* **1995**, *117*, 171.
- (23) Hummer, G. *Chem. Phys. Lett.* **1995**, *235*, 297.
- (24) Weiss, G. H. *Aspects and Applications of the Random Walk*; North Holland: New York, 1994.
- (25) Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*; Dover: New York, 1965.
- (26) Pearson, K. *Nature* **1905**, *72*, 294.
- (27) Kluyver, J. C. *Konink. Acad. Wetenschap Amst.* **1906**, *14*, 325.
- (28) Dombroski, J. P.; Taylor, S. W.; Gill, P. M. W. In preparation.
- (29) Boys, S. F.; Cook, G. B.; Reeves, C. M.; Shavitt, I. *Nature* **1956**, *178*, 1207.

JP952841B