How well do STARLAB and NBODY compare? II. Hardware and accuracy

P. Anders, H. Baumgardt, E. Gaburov and S. Portegies Zwart

ABSTRACT

Most recent progress in understanding the dynamical evolution of star clusters relies on direct N-body simulations. Owing to the computational demands, and the desire to model more complex and more massive star clusters, hardware calculational accelerators, such as Gravity Pipe (GRAPE) special-purpose hardware or, more recently, graphics processing units (GPUs) are generally utilized. In addition, simulations can be accelerated by adjusting parameters determining the calculation accuracy (i.e. changing the internal simulation time-step used for each star).

We extend our previous thorough comparison of basic quantities as derived from simulations performed either with STARLAB/KIRA or NBODY6. Here we focus on differences arising from using different hardware accelerations (including the increasingly popular graphic card accelerations/GPUs) and different calculation accuracy settings.

We use the large number of star cluster models (for a fixed stellar mass function, without stellar/binary evolution, primordial binaries, external tidal fields, etc.) already used in the previous paper, evolve them with STARLAB/KIRA (and NBODY6, where required), analyse them in a consistent way and compare the averaged results quantitatively. For this quantitative comparison, we apply the bootstrap algorithm for functional dependencies developed in our previous study.

In general, we find very high comparability of the simulation results, independent of the computer hardware (including the hardware accelerators) and the N-body code used. For the tested accuracy settings, we find that for reduced accuracy (i.e. time-step at least a factor of 2.5 larger than the standard setting) most simulation results deviate significantly from the results using standard settings. The remaining deviations are comprehensible and explicable.

Key words: methods: data analysis – methods: numerical – methods: statistical – open clusters and associations: general – galaxies: star clusters: general.

1 INTRODUCTION

In recent years, the modelling of star populations strongly advanced due to stellar dynamical studies. The fields covered comprehend a wide range of research questions, such as the modelling of individual star clusters (e.g. Baumgardt et al. 2003; Hurley et al. 2005; Heggie & Giersz 2009; Harfst, Portegies Zwart & Stolte 2010; Zonoozi et al. 2011), star cluster systems (e.g. Vesperini et al. 2003), populations of ‘exotic’ stellar objects (Portegies Zwart et al. 2004; massive black holes; Umbreit, Chatterjee & Rasio 2008; blue stragglers; Decressin et al. 2010: second populations of globular cluster stars; Gualandris, Portegies Zwart & Sipior 2005; Gvaramadze, Gualandris & Portegies Zwart 2008; Fujii & Portegies Zwart 2011: runaway stars) etc.

The major codes used in this field are the family of NODYX codes (Aarseth 1999) and the STARLAB environment with its N-body integrator KIRA (Portegies Zwart et al. 2001). Both codes are continuously expanded and improved, including the ability for the use of hardware dedicated to improve the simulation capabilities (in terms of simulation duration and available population size). This includes advances in software development (such as parallelization; see e.g. Harfst et al. 2007; Portegies Zwart et al. 2008; Spurzem et al. 2008) and the use of dedicated hardware, e.g. Gravity Pipe boards (GRAPEs; Ebisuzaki et al. 1993; Makino et al. 2003), more recently the use of GPUs (Portegies Zwart, Belleman & Geldof 2007; Gaburov, Harfst & Portegies Zwart 2009; Moore & Quillen 2011;
Nitadori & Aarseth, in preparation), supercomputers/computer clusters (Harfst et al. 2007) and large area networks, such as GRID (i.e. grid of special-purpose hardware, such as GRAPE modules, connected to form a joined computational network; Groen et al. 2011).

Additional N-body codes are emerging, such as MYRIAD (GRAPE-enabled; Konstantinidis & Kokkotas 2010), _NBSYMPLE_ (GRAPE-enabled; Capuzzo-Dolcetta, Mastrobuono-Battisti & Maschietti 2011) and _AMUSE_ (a compilation of various modules for gravitational dynamics, stellar evolution, radiative transfer and hydrodynamics, which can become independently combined; Portegies Zwart et al. 2009). GPUs are also applied in an additional wide range of astrophysical simulation environments (Ford 2009; Schive, Tsai & Chiuhe 2010; Thompson et al. 2010). Their general availability and usability for computationally intensive astrophysical modelling will continue the rise of GPU usage.

Despite their extensive use and computational specialties (see Sections 2 and 4.2), a detailed study concerning the reliability and comparability of simulations performed is missing. Important factors, which are studied in the present work, include the use of various computer hardware (including diverse special-purpose hardware to accelerate the simulations) and simulation accuracy settings. These factors determine the used calculation accuracy and therefore the accuracy to trace accurately fast important phases of star cluster evolution. Potentially, this can lead to incommensurable simulation results, although starting configuration and physical treatment are equivalent.

In Section 2, we describe the hardware and software used, including their limitations in calculation accuracy. Section 3 summarizes the simulation and analysis methods used. In Section 4, the simulation results are explained and analysed, and a summary of results and conclusions follow in Section 5.

2 UTILIZED HARDWARE AND SOFTWARE

We used three PCs for these simulations, whose specifications are summarized in Table 1. The use of these PCs was necessary to allow us to test the sensitivity of the results on the PC architecture and the availability of varying accelerator hardware. The labels ‘PC1’, ‘PC2’ and ‘PC3’ will be used to indicate the use of solely the CPU, without the available accelerator hardware. ‘PC1/GPU’, ‘PC2/GRAPE’ and ‘PC3/GPU’ indicate the usage of the available accelerator hardware.

For the majority of tests, we use the _STARLAB_ software package. Where we supplement these with simulations made with _NBODY_, we use _NBODY_6, contrary to our previous study (Anders et al. 2009, hereafter Paper I) where we used _NBODY_4. _NBODY_6 presents a significant enhancement of required computational time due to the AC neighbour scheme (see Ahmad & Cohen 1973 for the AC neighbour scheme and Aarseth 2001 for the _NBODY_ implementation).

2.1 Possible reasons for simulation differences

Numerical modelling of N-body systems, such as star clusters, usually involves long-term simulations. It is therefore necessary to understand the influence of different errors on the outcome. In a very basic case of a direct integration, there are two contributions to the overall error: one due to the time integration and the other due to the accuracy of force calculations. In the idealistic case, these two should be balanced, but in the majority of codes (e.g. _NBODY_, _STARLAB_, _PHGRAPE_; Harfst et al. 2008; _MYRIAD_; Konstantinidis & Kokkotas 2010) a simple direct summation method is employed to compute forces, whereas _NBODY_6 uses the Ahmad–Cohen neighbour scheme. While this in principle should result in exact accelerations, in practice, due to limited precision several factors affect the result. For example, the order in which partial forces are added affects the accuracy: if the partial forces are added from the weakest to strongest the total force on a particle is more accurate compared to the one if the summation is done in reverse or in arbitrary order. This is exacerbated if the precision of the force accumulation is below a certain threshold which depends on the system being modelled. On CPU, a generic double-precision force loop usually adds up partial forces in IEEE double-precision arithmetic, which, it is safe to say, results in the most accurate force. On special-purpose accelerations (such as GRAPE or GPU), for performance considerations, the summation is done in lower precision (e.g. Makino et al. 2003; Nitadori 2008; Gaburov et al. 2009). For example, Nitadori (2008) and Gaburov et al. (2009) use two single-precision numbers to emulate double precision, and the latter authors conduct force accumulation in single precision. The emulation of double precision is inherently less precise (2×23 bits mantissa) compared to IEEE double precision (53 bits mantissa), and its long-term errors are poorly understood. This may potentially inflict damage to long-term integrations.

In direct N-body simulations the error is usually due to time integration (among other errors coming from binary and multiple treatment, stellar evolution and close encounter treatment) rather than force summation. Generally in N-body codes, the time-step is defined by the Aarseth criterion (Aarseth 2003), which is usually scaled by an ‘accuracy parameter’ $\eta$. The time-step size monotonically depends on $\eta$ and changes in this parameter have proportional changes in the value of the time-step. In a generally employed fourth-order Hermite scheme, the error decreases as the fourth order of the time-step, and for a realistic time-step criterion which provides a good balance between the duration of simulations and accuracy, this error remains dominant. Furthermore, time-integration error has similar properties among different codes due to its simpler nature.

If $\eta$ is increased above a certain value, time-integration errors become intolerable, for example close encounters and dynamical binary integration become inaccurately simulated, but consistent among different codes. If $\eta$ decreases beyond a certain value, the dominant source of error becomes force computation, which varies among different implementations to a large degree. While this produces more accurate results, the outcome from different codes may show large scatter in sensitive quantities, such as energy conservation error, due to possibly significantly different force summation (or other) errors, rather than similar time-integration errors, influence long-term simulations. In this paper, we explore a range of $\eta$ parameters to identify their influence on the simulation results.

3 METHODS

In this section, we will summarize the methods developed in Paper I, which are applied in this paper. For additional details, see Paper I.

3.1 Standard set-up

We used start models from the previous study. They were created using _STARLAB_ (and, where necessary, transformed into the _NBODY_ file format) and have the following characteristics:

- (i) 1024 (1k) particles;
- (ii) Salpeter (1955) stellar mass function, with the highest mass $= 10 \times$ lowest mass;

© 2012 The Authors, _MNRAS_ 421, 3557–3569
Monthly Notices of the Royal Astronomical Society © 2012 RAS
Table 1. Specifications of PCs used.

| Specification          | PC1                  | PC2                  | PC3                  |
|------------------------|----------------------|----------------------|----------------------|
| Processor              | Quad Core Intel Q9300 | Dual Core Intel Xeon | Dual Core AMD Opteron|
| Operating system       | Debian 4.0           | Red Hat 2.6.9-67     | Debian 2.6.24.7       |
| C compiler             | gcc 4.2              | gcc 3.4.6            | gcc 4.3.2            |
| Special hardware       | GeForce 9800 GX2     | GRAPE6-BLX64         | GeForce 9800 GX2      |
| Special hardware libraries | Sapporo v1.5      | g6bx-0515            | GPU module by K. Nitadori |
| N-body software        | STARLAB 4.4.4        | STARLAB 4.4.4        | STARLAB6             |

(ii) no primordial binaries; and  
(iv) Plummer (1911) sphere initial density profile.

There are 50 start models, with identical parameters but varying statistical realisations.2

All simulations were performed with the following settings:
(i) no stellar/binary star evolution;  
(ii) no external tidal field;  
(iii) duration of simulation: 1000 N-body time units (i.e. until well after core collapse, which occurs around 60 N-body time units, see Paper I).

Unless otherwise specified, the simulations were performed using
(i) hardware acceleration: GeForce 9800GX2 GPU;  
(ii) calculation accuracy parameter $\eta = \eta_{\text{STARLAB}} = 0.1$ for STARLAB, equivalent to $\eta_{\text{NBODY}} = 0.01$ (due to a different definition of the accuracy parameter in STARLAB and NBODY; see Section 4.2). All quoted accuracy parameters are given as $\eta_{\text{STARLAB}}$.

3.2 Analysis of N-body simulation snapshots

As we want to trace differential effects between our settings (i.e. NBODY versus STARLAB, solely CPU versus CPU+GRAPE versus CPU+GPU, different accuracy parameters), we need to ensure that output from different simulations is treated identically. The main difficulty arises when comparing NBODY and STARLAB output, as the output content and format is very different. For example, NBODY output does not contain detailed information about binaries, while STARLAB output contains information on binary/triple/higher order hierarchies.

We therefore reduce the format of each simulation output to the most basic one possible, only retaining for each star its mass and vectors for its position and velocity. Binaries and higher order hierarchies are split up into their components. Other information possibly present is removed.

We then split the series of output into individual snapshots, pass them through KIRA, the N-body integrator of the STARLAB package, which evolves them for the 1/32th part of an N-body time unit. This procedure rebuilds the binary/higher order hierarchy structure, and stores it accessibly in the resulting snapshot. All snapshots with rebuilt hierarchy structure are then analysed using tools available within the STARLAB package. In Paper I we investigated the reliability of this procedure and did not find significant biases (except for the spurious rebuilding of multiple systems in the outer cluster regions, which are not subject of the present study). In addition, we calculate the change of total energy per N-body time unit as a measure of the energy conservation.

Results of this procedure are time series of cluster parameters (like core radius, virial ratio, etc.) and binary parameters (here only investigated at the end of the simulation; as shown in Paper I, testing at different times gives comparable results but poorer statistics). For the further analysis, results from the 50 individual runs per simulation setting are combined: the lists of binary parameters are merged, while for the overall cluster parameters a median time series and the 16/84 per cent quantiles are calculated.

3.3 Analyses of the simulation outcome

In this section, we summarize the methods used for the analysis of the N-body data. Many of the analysis results are given as probabilities. We will adopt the following (commonly used) nomenclature when discussing the significance level of statistical test results:
(i) highly significant: $p$-value < 1 per cent;  
(ii) significant: $p$-value < 5 per cent;  
(iii) weakly significant: $p$-value < 10 per cent.

‘Multiple testing’: if we perform 100 independent tests, a fraction of the order of 10 per cent of $p$-values below 0.1 will arise by chance even if none of the test null hypotheses is wrong. As we analyse various star cluster parameters of a large number of simulations, we need to take ‘multiple testing’ into account.

3.3.1 Binary and escaper parameters

We will analyse the parameters of dynamically created binaries (semimajor axis, eccentricity, mass ratio and binding energy) and the energy distributions of stars escaping from the cluster.

Differences between two such distributions for two different simulation set-ups are evaluated using a Kuiper test (Kuiper 1962), an advanced KS test (for the KS and Kuiper tests, see e.g. Numerical Recipes in FORTRAN by Press et al. 1992). A Kuiper test returns the probability that two distributions are drawn from the same parent distribution (or, more accurately, the probability that one is wrongly rejecting the null hypothesis ‘The two distributions are drawn from the same parent distribution.’, which in most cases is equivalent).

3.3.2 Time series

In Paper I we developed a bootstrap algorithm for functional dependencies and applied it to time series similar to the ones studied in this work. For example, we want to quantitatively compare the time evolution of the core radius (and of other parameters) for two different simulation set-ups.

First, we introduce a measure of the difference between two time series (or, more generally, between two functional dependencies). Assume that we have two functional dependencies of one parameter from the independent variable $x$: $y_1(x)$ and $y_2(x)$. For each $x$
these dependencies have uncertainties $\sigma_1(x)$ and $\sigma_2(x)$. The relative difference between the functional dependencies at a given $x$ is then

$$
\delta_{12}(x) = \frac{y_1(x) - y_2(x)}{\sqrt{\sigma_1(x)^2 + \sigma_2(x)^2}}. 
$$

We then define the 'difference between functions 1 and 2' as

$$
\Delta_{12} = \frac{1}{N} \sum_x \delta_{12}(x),
$$

where $N$ is the number of data points used for the statistic. We consider only the absolute value, as we want to have a measure of the size of the difference, but not necessarily its direction. In addition, this ensures $\Delta_{12} \equiv \Delta_{21}$.

Equivalently, we define the 'absolute difference between functions 1 and 2' as

$$
\Gamma_{12} = \frac{1}{N} \sum_x|\delta_{12}(x)|.
$$

While $\Delta_{12}$ is more sensitive to systematic offsets, $\Gamma_{12}$ also traces statistical fluctuations.

We then utilize this measure of difference between the two time series to design a bootstrap-like algorithm to quantify the significance of this difference.

We calculate 300 test clusters with STARLAB using the standard settings (i.e. using PC1/GPU and with $\eta = 0.1$), and the same analysis routines as for the other clusters.

From these test clusters, we randomly select sets of 50 clusters each (i.e. the number of clusters in the main simulations) with replacement, and calculate for each parameter the median $y^T(x)$ and quantiles $\sigma^T(x)$.

We build 2000 such sets. Out of these we randomly select two sets (again with replacement) and derive the individual values of $\Delta^T_{12}$ and $\Gamma^T_{12}$. We repeat this procedure 10 000 times to estimate the $\Delta^T_{12}$ and $\Gamma^T_{12}$ test distributions for each parameter. As all test clusters are calculated with the same settings, the $\Delta^T_{12}$ and $\Gamma^T_{12}$ test distributions represent the null hypothesis 'functions 1 and 2 are drawn from the same parent distribution'. By comparing these test distributions with the values derived from the main simulations $\Delta^S_{12}$ and $\Gamma^S_{12}$, we can quantify the fraction of data in the test distribution with $\Delta^S_{12}$ or $\Gamma^S_{12}$ more deviating than the values derived from the main simulations $\Delta^S_{12}$ or $\Gamma^S_{12}$. This value serves as a measure of how similar the two main simulations are.

In order to avoid applying the test statistic to highly correlated data, which appear for the earliest time-steps (as all runs share the same type of start models, only different stochastic realizations) and which are beyond the area of application of the test statistic, we start the summation in equations (2) and (3) at 20 N-body time units. This is sufficiently well before core collapse, which appears at about 60 N-body time units (see Paper I), to capture the systems' behaviour during this important epoch. As we have shown in Paper I, the influence of varying this start time slightly is small and does not change the conclusions drawn.

4 RESULTS

In this section, we will compare simulations performed for a number of specifications.

The quantitative test results are tabulated in Appendix B and summarized in Section 5. Visual presentations of time series for selected results are provided in Appendix A.

4.1 Different hardware, using STARLAB

We have performed simulations on PC1 and PC2. For each PC, we realized one set of simulations with the accelerator hardware installed and one set without hardware acceleration.

The results are shown in Figs A1 and A2. The agreement appears to be very good, with possible exceptions being the evolution of the kinetic energy (though with large error bars, only the virial ratio is shown for consistency with the other sections) and the cumulative eccentricity and mass ratio distributions of the dynamically created binaries. The quantitative measures (our bootstrap algorithm for integrated cluster properties and the Kuiper test for properties of binaries and escaping stars), shown in Tables B1–B4, generally confirm the good agreement. The following issues are found:

(i) significant differences: kinetic energy of stars escaping from the cluster for #1.1 PC1 ⇔ PC1/GPU, and differences (visible in Fig. A2) in the mass ratio distributions for #1.1 PC1 ⇔ PC1/GPU #1.2 PC1/GPU ⇔ PC2/GRAPE;
(ii) weakly significant: the $\Delta_{12}$ test of King $W_0$ and the potential energy for #1.3 PC2 ⇔ PC2/GRAPE;
(iii) statistically insignificant: differences in the total kinetic energy of all stars and the binary eccentricity distributions (visible in Fig. A2).

Deviations in the mass ratio distributions can originate from the stochastical process of two-body interaction processes. The number and level of the remaining deviations is in agreement with ‘multiple testing’, i.e. the occurrence of statistical disagreement based on large numbers of tests. Therefore, N-body simulations based on equivalent start conditions and N-body code (here STARLAB) produce highly comparable results, independent of the computer hardware (including the accelerator hardware).

4.2 Accuracy settings

Using PC1/GPU, we test the impact of the calculation accuracy.

The 'calculation accuracy' is determined by a dimensionless constant $\eta$ in the formula determining the integration time-step (see Aarseth 2003, section 2.3):

$${\Delta t}_i = \sqrt{\frac{\eta_{\text{body}} \left( |F_i|^2 + |F_j|^2 \right)}{\left( |F_i|^2 + |F_j|^2 \right)}} = \eta_{\text{starlab}} \sqrt{\frac{\left( |F_i|^2 + |F_j|^2 \right)}{\left( |F_i|^2 + |F_j|^2 \right)}}.
$$

where $F$ is the force affecting particle $i$ and $F^{(j)}$ its $j$th time derivative.

Due to the slightly different definitions used by STARLAB and NBODY, we have $\eta_{\text{starlab}} = \sqrt{\eta_{\text{body}}}$. In the remainder of this paper, we will express all accuracy settings as $\eta = \eta_{\text{starlab}}$.

The results are shown in Figs A3 and A4. In most of the panels, simulations with the largest accuracy parameter (i.e. the ‘worst’ accuracy) studied are clearly offset from the other simulations. The other simulations seem to provide quite comparable results, except for the change in total energy during 1 N-body time unit (i.e. a measure for energy conservation). The change in total energy reduces with lowering the accuracy parameter $\eta$ both before and after the core collapse, i.e. total energy becomes better preserved when accuracy is improved.
These findings are confirmed by the quantitative measures from our tests:

(i) (Highly) significant: most results for the $\eta = 0.3$ and $\eta = 0.4$ runs strongly deviate from the results using the standard settings, independent of the hardware used and parameter studied. For the $\eta = 0.25$ runs, the results depend on the studied parameter: one part of the results follows the standard setting results, while for the other parameters the results deviate from the standard setting runs similar to the $\eta = 0.3$ run results (no clear correlation between the studied parameter and the deviation).

(ii) (Highly) significant: the change in total energy during 1 N-body time unit (i.e. our measure for energy conservation) deviates from the standard settings for almost all runs

(iii) (Highly) significant: additional individual deviations are found, without a specific correlation with simulation settings, used hardware or studied parameter.

The strongly deviating results for $\eta = 0.3$ and $\eta = 0.4$ runs (and partially the $\eta = 0.25$ runs) represent strong discouragement for such ‘inaccurate’ simulations.

The results for the energy conservation are expected. The better the simulation accuracy the better even slight effects are traced. These slight effects, e.g. individual close encounters or binary evolution, have strong impact on the conservation of the system’s energy.

The number and level of additional deviations exceeds the expectations based on ‘multiple testing’, also without the $\eta = 0.3$ and $\eta = 0.4$ runs. The level of accuracy changes the simulation results slightly but cumulatively; therefore, fundamental agreement for all studied parameters from simulations using different accuracy settings is not assured.

### 4.3 STARLAB versus NBODY6

We have performed simulations on PC1 and PC3. For each PC, we realized one set of simulations with the GPUs installed and one set without hardware acceleration. The STARLAB simulations were performed with $\eta_{\text{STARLAB}} = 0.1$, while the NBODY6 simulations were performed with $\eta_{\text{STARLAB}} \approx 0.15$ ($\eta_{i} = \sqrt{0.02}$ and $\eta_{0} = \sqrt{0.03}$). As shown in Section 4.2, these settings give comparable results.

The results are shown in Figs A5 and A6. They indicate potential deviations in kinetic energy for the NBODY6 runs using PC3, in the change of total energy (during 1 N-body time unit) between NBODY6 and STARLAB runs and in binary eccentricity for the GPU runs using either NBODY6 or STARLAB.

Our quantitative tests confirm the deviations for the change in total energy during 1 N-body time unit. It is bigger (i.e. energy conservation is worse) for NBODY6 as compared to STARLAB. This agrees with our earlier finding for NBODY4 versus STARLAB in Paper I.

The remaining expectations, based on visual examination of Figs A5 and A6, are not confirmed. Our statistical tests indicate deviations in

(i) significant: the time evolution of $r_{\text{max}}$ (i.e. the distance of the furthest star from the cluster centre) when comparing simulations made with STARLAB and NBODY6, both without hardware acceleration;

(ii) weakly significant: the binary mass ratio distribution for #3.2 PC1 $\leftrightarrow$ PC3;

(iii) weakly significant: the distribution of the kinetic energy of stars escaping the cluster for #3.1 PC1/GPU $\leftrightarrow$ PC3/GPU.

These results are partially inconsistent with our study of STARLAB versus NBODY4 in Paper I. In Paper I, no significant deviations in $r_{\text{max}}$ were found, the binary mass ratio distributions were different (the probability of both being drawn from the same parent distribution was determined to be 14.9 per cent, slightly larger than required for a statistically slightly significant deviation), and for the kinetic energy of stars escaping the cluster, no significant deviation was found. The differences detected in this work are expected to originate from differences of codes between NBODY4 and NBODY6 (code enhancements, such as the AC neighbour scheme) of STARLAB (version 4.4.2) and STARLAB (version 4.4.4) (which included bug fixes).

The results for the time evolution of $r_{\text{max}}$ and the binary mass ratio distributions are based on processes with strong statistical effects, as these quantities are affected by rare events, e.g. direct two-body interaction processes. In general, the number and level of deviations agrees with that expected from ‘multiple testing’. These results allow reliable comparisons of NBODY6 and STARLAB simulations, based on comparable input parameters and independent of the computer hardware used.

### 5 SUMMARY OF RESULTS AND CONCLUSIONS

All of the quantitative test results are shown in Tables B1–B4. They are based on 50 input models and their consistently analysed output.

From these tables, the following picture emerges.

(i) In general, the results obtained from N-body simulations in various configurations are well comparable (with some exceptions, which will be discussed as follows), independent of the computer hardware, the accelerator type and N-body code used.

(ii) Results with too coarse accuracy (for our tests, $\eta_{\text{STARLAB}} = 0.3$ and $\eta_{\text{STARLAB}} = 0.4$, partially $\eta_{\text{STARLAB}} = 0.25$) give strongly significantly deviating results in a variety of quantities, hence should be avoided.

(iii) The quantity which shows the highest sensitivity to the configuration used is the energy conservation (here described by the change of total cluster energy during 1 N-body time unit). While it seems to be unaffected by the hardware used, it changes strongly with the accuracy parameter $\eta$, and between NBODY6 and STARLAB. The behaviour with the accuracy parameter is expected and desired. The differences between NBODY6 and STARLAB are comparable to our earlier finding when comparing NBODY4 and STARLAB in Paper I.

(iv) The significant deviation found for some configurations in the temporal evolution of $r_{\text{max}}$ (the distance of the furthest star from the cluster centre) and the mass ratio distribution of binaries could be due to stochastic effects, as these quantities are affected by rare events (e.g. close two-body encounters and binary evolution).

In our test setting, these quantities deviate statistically significantly, though larger scale simulations are required to fundamentally check the statistical significance.

(v) The remaining significant deviations can be real or a result of ‘multiple testing’. The number of tests performed and deviations observed is consistent with that result from ‘multiple testing’ for the study of hardware (#1.1–1.4 PC1 $\leftrightarrow$ PC2, utilizing or relinquishing the available accelerator hardware) and of the N-body code used (#3.1–3.3 STARLAB $\leftrightarrow$ NBODY6). For the study of accuracy settings, the number and strength of deviations is higher than that expected from ‘multiple testing’ for the energy-related cluster properties and the binary properties.

To conclude, in general, direct N-body simulations are assumed to be correct in a statistical sense rather than on the level of individual trajectories. Except for few well-understood cases (such as the...
results for various accuracy settings and the energy conservation), N-body simulations using different hardware or N-body codes give very comparable results, confirming their reliability in a statistical sense.

ACKNOWLEDGMENTS

PA acknowledges funding by the European Union (Marie Curie EIF grant MEIF-CT-2006-041108), the National Natural Science Foundation of China (NSFC, grant number 11073001) and wishes to thank N. Bissantz for helpful discussions. HB acknowledges support from the Australian Research Council through Future Fellowship grant FT0991052 and thanks K. Nitadori for providing the GPU module for NBODY6. SPZ thanks the Netherlands Research School for Astronomy (NOVA). Support for programme number HST-HF-51289.01-A was provided by NASA through a Hubble Fellowship grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

REFERENCES

Aarseth S. J., 1999, PASP, 111, 1333
Aarseth S. J., 2001, New Astron., 6, 277
Aarseth S. J., 2003, Gravitational N-Body Simulations. Cambridge Univ. Press, Cambridge
Ahmad A., Cohen L., 1973, J. Comput. Phys., 12, 389
Anders P., Baumgardt H., Bissantz N., Portegies Zwart S., 2009, MNRAS, 395, 2304 (Paper I)
Baumgardt H., Makino J., Hut P., McMillan S., Portegies Zwart S., 2003, ApJ, 589, L25
Capuzzo-Dolcetta R., Mastrobuono-Battisti A., Maschietti D., 2011, New Astron., 16, 284
Decressin T., Baumgardt H., Charbonnel C., Kroupa P., 2010, A&A, 516, A73
Ebisuzaki T., Makino J., Fukushige T., Taijil M., Sugimoto D., Ito T., Okumura S. K., 1993, PASJ, 45, 269
Ford E. B., 2009, New Astron., 14, 406
Fujii M., Portegies Zwart S., 2011, Sci, 334, 1380
Gao B., Portegies Zwart S., Ishiyama T., Makino J., 2011, Comput. Sci. Disc., 4, 015001
Gualandris A., Portegies Zwart S., Sipior M. S., 2005, MNRAS, 363, 223
Gvaramadze V. V., Gualandris A., Portegies Zwart S., 2008, MNRAS, 385, 929
Harfst S., Gualandris A., Merritt D., Spurzem R., Portegies Zwart S., Berczik P., 2007, New Astron., 12, 357
Harfst S., Gualandris A., Merritt D., Mikkola S., 2008, MNRAS, 389, 2
Harfst S., Portegies Zwart S., Stolte A., 2010, MNRAS, 409, 628
Heggie D. C., Giersz M., 2009, MNRAS, 397, L46
Hurley J. R., Pols O. R., Aarseth S. J., Tout C. A., 2005, MNRAS, 363, 293
Konstantinidis S., Kokkotas K. D., 2010, A&A, 522, A70
Kotulla R., Fritze U., Weilbacher P., Anders P., 2009, MNRAS, 396, 462
Kuiper N. H., 1962, Proc. K. Ned. Akad. Wet. A, 63, 38
Makino J., Fukushige T., Koga M., Namura K., 2003, PASJ, 55, 1163
Moore A., Quillen A. C., 2011, New Astron., 16, 445
Nitadori K., 2008, PhD thesis, Natl Astron. Obser. Japan
Plummer H. C., 1911, MNRAS, 71, 460
Portegies Zwart S. F., McMillan S. L. W., Hut P., Makino J., 2001, MNRAS, 321, 199
Portegies Zwart S. F., Baumgardt H., Hut P., Makino J., McMillan S. L. W., 2004, Nat, 428, 724
Portegies Zwart S. F., Bellemann R. G., Geldof P. M., 2007, New Astron., 12, 641
Portegies Zwart S. F., McMillan S., Groen D., Gualandris A., Sipior M., Vermin W., 2008, New Astron., 13, 285
Portegies Zwart S. et al., 2009, New Astron., 14, 369
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in Fortran. The Art of Scientific Computing, 2nd edn. Cambridge Univ. Press, Cambridge
Salpeter E. E., 1955, ApJ, 121, 161
Schive H., Tsai Y., Chiu Y., 2010, ApJS, 186, 457
Spurzem R., Berentzen I., Berczik P., Merritt D., Amaro-Seoane P., Harfst S., Gualandris A., 2008, in Aarseth S. J., Tout C. A., Mardling R. A., eds., Lect. Notes Phys. Vol. 760, The Cambridge N-Body Lectures. Springer-Verlag, Berlin, p. 377
Thompson A. C., Fluke C. J., Barnes D. G., Barsdell B. R., 2010, New Astron., 15, 16
Umbricht S., Chatterjee S., Rasio F. A., 2008, ApJ, 680, L113
Vesperini E., Zepf S. E., Kundu A., Ashman K. M., 2003, ApJ, 593, 760
Zonoozi A. H., Küpper A. H. W., Baumgardt H., Haghi H., Kroupa P., Hilker M., 2011, MNRAS, 411, 1989

APPENDIX A: GRAPHICAL DISPLAY OF SIMULATION RESULTS

© 2012 The Authors, MNRAS 421, 3557–3569
Monthly Notices of the Royal Astronomical Society © 2012 RAS
Figure A1. Comparison of simulations using STARLAB and various hardware. The lines show the median values, the error bars give the uncertainty ranges from the 50 individual runs. Shown are the time evolutions of the core radius (top left, bottom lines), half-mass radius (top left, upper lines), the objects’ mean mass in the core = measure for mass segregation (top right), virial ratio (bottom left) and change in total energy during 1 N-body time unit (bottom right).

Figure A2. Comparison of binary parameters from simulations using STARLAB and various hardware after 1000 N-body time units (i.e. well after core collapse). Shown are the cumulative distributions of the semimajor axis (top left), the eccentricity (top right), mass ratio of secondary to primary (bottom left) and the binding energy (bottom right). The lines show the data, and the error bars give the uncertainty ranges from bootstrapping (at every 20th data point only, for clarity).
Figure A3. Comparison of simulations using STARLAB and various accuracy settings, all using PC1 and GPU. The lines show the median values, and the error bars give the uncertainty ranges from the 50 individual runs. Shown are the time evolutions of the core radius (top left, bottom lines), half-mass radius (top left, upper lines), the objects’ mean mass in the core = measure for mass segregation (top right), virial ratio (bottom left; $\eta = 0.4$ is not shown due to very large deviations and error bars) and change in total energy during 1 $N$-body time unit (bottom right).

Figure A4. Comparison of binary parameters from simulations using STARLAB and various accuracy settings after 1000 $N$-body time units (i.e. well after core collapse). All using PC1 and GPU. Shown are the cumulative distributions of the semimajor axis (top left), the eccentricity (top right), mass ratio of secondary to primary (bottom left) and the binding energy (bottom right). The lines show the data, and the error bars give the uncertainty ranges from bootstrapping (at every 20th data point only, for clarity).
Comparison of simulations using STARLAB or NBODY6, with either GPU acceleration or no hardware acceleration. The lines show the median values, and the error bars give the uncertainty ranges from the 50 individual runs. Shown are the time evolutions of the core radius (top left, bottom lines), half-mass radius (top left, upper lines), the objects’ mean mass in the core = measure for mass segregation (top right), virial ratio (bottom left) and change in total energy during 1 N-body time unit (bottom right).

Comparison of binary parameters from simulations using STARLAB or NBODY6, with either GPU acceleration or no hardware acceleration, after 1000 N-body time units (i.e. well after core collapse). Shown are the cumulative distributions of the semimajor axis (top left), the eccentricity (top right), mass ratio of secondary to primary (bottom left) and the binding energy (bottom right). The lines show the data, and the error bars give the uncertainty ranges from bootstrapping (at every 20th data point only, for clarity).
### APPENDIX B: ANALYSIS RESULTS

Table B1. Bootstrap results for structural cluster properties. Given are the fractions (in per cent) of the test cluster distributions (which represent the case that they are drawn from the same parent distribution) more deviating than the main simulations, i.e. the smaller this number the less alike are the distributions. The ‘standard’ set-up uses PC1/GPU and an accuracy setting of $\eta = 0.1$. Quantities are: $r_{\text{core}}$ = cluster core radius; $r_{\text{half}}$ = cluster half-mass radius; $r_{\text{max}}$ = distance of the furthest star from the cluster centre; (mass)$_{\text{core}}$ = mean mass of objects in the core (i.e. a measure for mass segregation); King $W_0$ = dimensionless potential $W_0$ of a King profile (fitted by the hyps_stars analysis task of starlab); $\langle W \rangle$ = density c. = absolute distance of the cluster density centre from origin. For each parameter, the probabilities from the $\Delta_{12}$ and $\Gamma_{12}$ distributions (see Section 3.3.2) are given. If no $\eta$ value is shown, $\eta = 0.1$ is used. For a more detailed description of the set-ups, see text. Results with (at least weakly) statistically significant deviations are marked with bold face in all Appendix tables.

| ID   | Set-up 1     | Set-up 2     | $r_{\text{core}}$ | $r_{\text{half}}$ | $r_{\text{max}}$ | $W_0$ | $\langle W \rangle$ | (mass)$_{\text{core}}$ |
|------|--------------|--------------|-------------------|-------------------|------------------|------|---------------------|------------------------|
| 1.1  | PC1/GPU      | PC1          | 69.84             | 40.34             | 32.25            | 29.99| 62.64               | 80.62                  |
| 1.2  | PC1/GPU      | PC2/GRAPE    | 53.04             | 71.99             | 27.32            | 32.04| 22.54               | 38.78                  |
| 1.3  | PC2/GRAPE    | PC2          | 31.78             | 35.53             | 18.52            | 5.76 | 33.24               | 47.83                  |
| 1.4  | PC1          | PC2          | 99.76             | 99.89             | 99.94            | 99.92| 99.94               | 99.89                  |
|      | Tests using different PCs and accelerator hardware   |              |                   |                   |                 |      |                     |                        |
| 2.1  | Standard     | PC1/GPU $\eta = 0.05$ | 90.53             | 43.26             | 45.83            | 80.62| 56.33               | 63.13                  |
| 2.2  | Standard     | PC1/GPU $\eta = 0.2$ | 24.23             | 84.23             | 59.08            | 31.21| 76.74               | 39.56                  |
| 2.3  | Standard     | PC1/GPU $\eta = 0.3$ | 16.76             | 51.91             | 6.67             | 3.68 | 3.74                | 4.93                   |
| 2.4  | Standard     | PC1/GPU $\eta = 0.4$ | 0.10              | 0.76              | 0.00             | 1.48 | 0.01                | 0.14                   |
| 2.5  | PC1/GPU $\eta = 0.05$ | 63.07          | 12.97             | 84.45             | 90.32            | 22.45| 16.88               | 71.13                  |
| 2.6  | PC1/GPU $\eta = 0.2$ | 40.35           | 78.80             | 73.41             | 23.97            | 84.90| 73.96               | 1.82                   |
| 2.7  | PC1/GPU $\eta = 0.4$ | 5.76            | 30.84             | 83.53             | 12.03            | 2.26 | 2.77               | 6.66                   |
| 2.8  | PC1 $\eta = 0.1$ | PC1 $\eta = 0.25$ | 38.52             | 62.60             | 6.54             | 17.27| 63.72               | 58.57                  |
| 2.9  | PC1 $\eta = 0.1$ | PC1 $\eta = 0.05$ | 83.51             | 16.69             | 45.93            | 35.25| 52.36               | 81.28                  |
| 2.10 | PC1 $\eta = 0.1$ | PC1 $\eta = 0.3$ | 7.73              | 63.79             | 10.86            | 1.58 | 1.84                | 2.82                   |
| 2.11 | PC1 $\eta = 0.1$ | PC1 $\eta = 0.4$ | 6.12              | 0.78              | 0.45             | 2.54 | 7.90               | 1.75                   |
| 2.12 | PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.05$ | 12.53            | 79.69             | 41.02            | 54.85| 57.90               | 84.51                  |
| 2.13 | PC2/GRAPE $\eta = 0.2$ | PC2/GRAPE $\eta = 0.2$ | 60.32             | 64.11             | 45.93            | 31.19| 55.00               | 82.72                  |
| 2.14 | PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.25$ | 36.41            | 52.11             | 2.40             | 12.73| 25.07               | 34.87                  |
| 2.15 | PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.3$ | 98.73             | 82.25             | 0.05             | 84.11| 54.87               | 29.73                  |
| 2.16 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.025$ | 35.15             | 64.30             | 3.94             | 13.60| 51.56               | 28.70                  |
| 2.17 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.05$ | 83.47             | 16.73             | 65.07            | 35.19| 56.32               | 81.35                  |
| 2.18 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.2$ | 36.16             | 43.41             | 6.45             | 41.31| 61.96               | 88.75                  |
| 2.19 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.25$ | 41.90             | 11.63             | 20.53            | 10.52| 34.10               | 72.97                  |
| 2.20 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.3$ | 7.65              | 62.93             | 0.18             | 10.53| 33.50               | 1.58                   |
| 2.21 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.4$ | 2.34              | 0.79              | 0.00             | 1.84 | 3.30                | 0.00                   |
| 2.22 | PC2 $\eta = 0.1$ | PC2 $\eta = 0.025$ | 99.85             | 99.89             | 99.94            | 99.91| 99.94               | 99.81                  |
| 2.23 | PC1 $\eta = 0.05$ | PC1 $\eta = 0.05$ | 99.82             | 99.94             | 99.94            | 99.94| 99.94               | 99.94                  |
| 2.24 | PC1 $\eta = 0.1$ | PC1 $\eta = 0.3$ | 99.69             | 99.94             | 99.94            | 99.94| 99.94               | 99.94                  |
| 2.25 | PC1 $\eta = 0.4$ | PC2 $\eta = 0.4$ | 55.01             | 62.69             | 13.21            | 65.91| 99.94               | 30.13                  |
| 2.26 | PC3/GPU      | PC3          | 87.41             | 62.78             | 88.98            | 65.80| 44.58               | 37.98                  |
| 2.27 | PC3          | PC3          | 94.11             | 81.33             | 12.73            | 41.68| 29.64               | 69.88                  |
| 2.28 | PC3/GPU      | PC3          | 88.41             | 89.97             | 12.73            | 82.22| 75.36               | 97.00                  |

*Data sets with ‘extreme’ uncertainties, i.e. ‘extreme run-to-run’ scatter (applies only to the $\eta = 0.4$ data sets).*

© 2012 The Authors, MNras 421, 3557–3569

Monthly Notices of the Royal Astronomical Society © 2012 RAS
Table B2. Bootstrap results for energy-related cluster properties. Given are the fractions (in per cent) of the test cluster distributions (which represent the case that they are drawn from the same parent distribution) more deviating than the main simulations, i.e. the smaller this number the less alike the distributions are. The ‘standard’ set-up uses PC1/GPU and an accuracy setting of $\eta = 0.1$. Quantities are: $E_{\text{pot}}$ = cluster potential energy; $E_{\text{kin}}$ = cluster kinetic energy; $Q_{\text{vir}}$ = cluster virial ratio; $E_{\text{tot}}$ = cluster total energy; $\Delta E_{\text{tot}}$ = change in the cluster’s total energy during 1 N-body time unit. For each parameter, probabilities from the $\Delta_{12}$ and $\Gamma_{12}$ distributions (see Section 3.3.2) are given.

| ID | Set-up 1          | Set-up 2          | $E_{\text{pot}}$ | $E_{\text{kin}}$ | $Q_{\text{vir}}$ | $E_{\text{tot}}$ | $\Delta E_{\text{tot}}$ |
|----|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-------------------|
|    |                  |                  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$     |
| 1.1 | PC1/GPU          | PC1              | 36.33 74.09     | 0.3 56.43       | 0.05            | 0.1 PC2         | 7.82 1.39         |
| 1.2 | PC1/GPU          | PC2/GrAPE       | 57.97 99.21     | 0.3 56.43       | 0.05            | 0.1 PC2         | 9.13 24.42        |
| 1.3 | PC2/GrAPE       | PC2              | 98.78 85.97     | 0.3 56.43       | 0.05            | 0.1 PC2         | 10.97 76.01       |
| 1.4 | PC1              | PC2              | 73.78 79.77     | 0.3 56.43       | 0.05            | 0.1 PC2         | 12.28 39.33       |

Tests using different PCs and accelerator hardware

| ID | Set-up 1          | Set-up 2          | $E_{\text{pot}}$ | $E_{\text{kin}}$ | $Q_{\text{vir}}$ | $E_{\text{tot}}$ | $\Delta E_{\text{tot}}$ |
|----|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-------------------|
|    |                  |                  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$     |
| 2.1 | Standard         | PC1/GPU $\eta = 0.05$ | 7.70 0.80       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.2 | Standard         | PC1/GPU $\eta = 0.2$ | 77.95 83.90     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.3 | Standard         | PC1/GPU $\eta = 0.3$ | 19.04 7.42      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.4 | Standard         | PC1/GPU $\eta = 0.4$ | 0.83 0.90       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.5 | PC1/GPU $\eta = 0.05$ | PC1 $\eta = 0.05$ | 15.92 0.80      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.6 | PC1/GPU $\eta = 0.2$ | PC1 $\eta = 0.2$  | 33.84 24.99     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.7 | PC1/GPU $\eta = 0.3$ | PC1 $\eta = 0.3$  | 19.04 7.42      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.8 | PC1/GPU $\eta = 0.4$ | PC1 $\eta = 0.4$  | 0.83 0.90       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |

Tests using different accuracy settings $\eta$

| ID | Set-up 1          | Set-up 2          | $E_{\text{pot}}$ | $E_{\text{kin}}$ | $Q_{\text{vir}}$ | $E_{\text{tot}}$ | $\Delta E_{\text{tot}}$ |
|----|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-------------------|
|    |                  |                  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$     |
| 2.1 | Standard         | PC1/GPU $\eta = 0.05$ | 7.70 0.80       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.2 | Standard         | PC1/GPU $\eta = 0.2$ | 77.95 83.90     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.3 | Standard         | PC1/GPU $\eta = 0.3$ | 19.04 7.42      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.4 | Standard         | PC1/GPU $\eta = 0.4$ | 0.83 0.90       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.80 0.80         |
| 2.5 | PC1/GPU $\eta = 0.05$ | PC1 $\eta = 0.05$ | 15.92 0.80      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.6 | PC1/GPU $\eta = 0.2$ | PC1 $\eta = 0.2$  | 33.84 24.99     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.7 | PC1/GPU $\eta = 0.3$ | PC1 $\eta = 0.3$  | 19.04 7.42      | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 2.8 | PC1/GPU $\eta = 0.4$ | PC1 $\eta = 0.4$  | 0.83 0.90       | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |

Tests using STARLAB and NBODY6

| ID | Set-up 1          | Set-up 2          | $E_{\text{pot}}$ | $E_{\text{kin}}$ | $Q_{\text{vir}}$ | $E_{\text{tot}}$ | $\Delta E_{\text{tot}}$ |
|----|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-------------------|
|    |                  |                  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$  | $\Delta_{12}$     |
| 3.1 | Standard         | PC3/GPU          | 57.81 93.14     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 3.2 | Standard         | PC3              | 33.04 62.02     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |
| 3.3 | Standard         | PC3/GPU          | 53.66 54.56     | 0.3 56.43       | 0.05            | 0.1 PC2         | 0.79 0.71         |

*Data sets with ‘extreme’ uncertainties, i.e. ‘extreme run-to-run’ scatter (applies only to the $\eta = 0.4$ data sets).*
Table B3. Kuiper test results for binary properties. Given is the probability (in per cent) that for the given set-ups the binary properties are drawn from the same distribution. The ‘standard’ set-up uses PC1/GPU and an accuracy setting of $\eta = 0.1$. Quantities are: axis = semimajor axis; ecc. = eccentricity; $E/kT = $ binding energy in $E/kT$, $m2/m1 = $ mass ratio secondary mass/primary mass. #X = number of binaries in simulation X.

| ID | Set-up 1 | Set-up 2 | #1 | #2 | Axis | Ecc. | $E/kT$ | $m2/m1$ |
|----|----------|----------|----|----|------|------|-------|--------|
| 1.1 PC1/GPU | PC1 | 201 | 202 | 83.57 | 83.99 | 60.39 | 4.99 |
| 1.2 PC1/GPU | PC2/GRAPE | 201 | 206 | 80.18 | 32.28 | 69.14 | 1.63 |
| 1.3 PC2/GRAPE | PC2 | 206 | 202 | 99.97 | 13.67 | 49.19 | 76.06 |
| 1.4 PC1 | PC2 | 202 | 202 | 100.00 | 100.00 | 100.00 | 100.00 |
| 2.1 Standard PC1/GPU $\eta = 0.05$ | 201 | 190 | 88.96 | 38.18 | 42.01 | 97.89 |
| 2.2 Standard PC1/GPU $\eta = 0.2$ | 201 | 213 | 97.26 | 72.84 | 90.72 | 44.90 |
| 2.3 Standard PC1/GPU $\eta = 0.3$ | 201 | 213 | 9.20 | 51.32 | 0.81 | 28.47 |
| 2.4 Standard PC1/GPU $\eta = 0.4$ | 201 | 253 | 0.00 | 65.53 | 0.00 | 51.58 |
| 2.5 PC1/GPU $\eta = 0.05$ | PC1 | 190 | 202 | 29.28 | 77.83 | 53.79 | 33.28 |
| 2.6 PC1/GPU $\eta = 0.3$ | PC1 | 213 | 220 | 58.73 | 96.36 | 56.65 | 32.64 |
| 2.7 PC1/GPU $\eta = 0.4$ | PC1 | 253 | 260 | 67.64 | 85.16 | 0.12 | 79.14 |
| 2.8 PC1 $\eta = 0.1$ | PC1 $\eta = 0.025$ | 202 | 204 | 94.15 | 6.24 | 82.66 | 28.19 |
| 2.9 PC1 $\eta = 0.1$ | PC1 $\eta = 0.05$ | 202 | 202 | 20.43 | 37.05 | 58.81 | 2.33 |
| 2.10 PC1 $\eta = 0.1$ | PC1 $\eta = 0.4$ | 202 | 260 | 64.77 | 3.43 | 41.88 | 33.95 |
| 2.11 PC1 $\eta = 0.1$ | PC1 $\eta = 0.25$ | 202 | 202 | 66.44 | 30.82 | 73.82 | 0.27 |
| 2.12 PC1 $\eta = 0.1$ | PC1 $\eta = 0.3$ | 202 | 220 | 20.24 | 18.05 | 1.10 | 3.38 |
| 2.13 PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.05$ | 206 | 191 | 35.85 | 92.07 | 47.82 | 27.43 |
| 2.14 PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.2$ | 206 | 204 | 43.81 | 64.46 | 42.24 | 20.01 |
| 2.15 PC2/GRAPE $\eta = 0.1$ | PC2/GRAPE $\eta = 0.25$ | 206 | 198 | 28.62 | 75.40 | 56.65 | 32.64 |
| 2.16 PC2 $\eta = 0.1$ | PC2 $\eta = 0.025$ | 202 | 204 | 93.99 | 8.03 | 82.66 | 28.19 |
| 2.17 PC2 $\eta = 0.1$ | PC2 $\eta = 0.05$ | 202 | 201 | 20.43 | 37.05 | 58.81 | 2.33 |
| 2.18 PC2 $\eta = 0.1$ | PC2 $\eta = 0.2$ | 202 | 204 | 64.77 | 3.43 | 41.88 | 33.95 |
| 2.19 PC2 $\eta = 0.1$ | PC2 $\eta = 0.25$ | 202 | 202 | 66.44 | 30.82 | 73.82 | 0.27 |
| 2.20 PC2 $\eta = 0.1$ | PC2 $\eta = 0.3$ | 202 | 220 | 20.24 | 18.05 | 1.10 | 3.38 |
| 2.21 PC2 $\eta = 0.1$ | PC2 $\eta = 0.4$ | 202 | 248 | 0.03 | 54.16 | 0.00 | 19.66 |
| 2.22 PC2 $\eta = 0.1$ | PC2 $\eta = 0.025$ | 202 | 204 | 93.99 | 8.03 | 82.66 | 28.19 |
| 2.23 PC1 $\eta = 0.025$ | PC1 $\eta = 0.05$ | 204 | 204 | 100.00 | 100.00 | 100.00 | 100.00 |
| 2.24 PC1 $\eta = 0.05$ | PC1 $\eta = 0.05$ | 202 | 202 | 100.00 | 100.00 | 100.00 | 100.00 |
| 2.25 PC1 $\eta = 0.3$ | PC1 $\eta = 0.3$ | 220 | 220 | 100.00 | 100.00 | 100.00 | 100.00 |
| 2.26 PC1 $\eta = 0.4$ | PC1 $\eta = 0.4$ | 260 | 248 | 93.34 | 99.88 | 0.22 | 99.97 |
| 3.1 standard | PC3/GPU | 201 | 189 | 48.81 | 12.63 | 54.46 | 76.14 |
| 3.2 PC1 | PC3 | 202 | 209 | 46.91 | 34.27 | 74.56 | 7.80 |
| 3.3 PC3/GPU | PC3 | 189 | 209 | 22.69 | 43.63 | 80.84 | 49.64 |

© 2012 The Authors, MNRAS 421, 3557–3569
Monthly Notices of the Royal Astronomical Society © 2012 RAS
Table B4. Kuiper test results for energy-related properties of unbound stars at 1000 N-body time units (i.e. well after core collapse). Results with a ‘B’ superscript denote results for unbound binary systems. Given is the probability (in per cent) that for the given set-ups the properties are drawn from the same distribution. The ‘standard’ set-up uses PC1/GPU acceleration and an accuracy setting of $\eta = 0.1$. Quantities are: $E_{\text{kin}}$ = reduced kinetic energy; $E_{\text{pot}}$ = reduced potential energy; $E_{\text{tot}}$ = reduced total energy. $X$ = number of systems in simulation $X$.

| ID | Set-up 1 | Set-up 2 | $E_{\text{kin}}$ | $E_{\text{pot}}$ | $E_{\text{tot}}$ | $E_{\text{kin}}^{B}$ | $E_{\text{pot}}^{B}$ | $E_{\text{tot}}^{B}$ |
|----|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|
| 1.1 | PC1/GPU  | PC1      | 7554             | 7441             | 4.71             | 31.42            | 13.47            | 144              | 150              | 96.68            | 94.73            | 96.46            |
| 1.2 | PC1/GPU  | PC2/GRAPE| 7554             | 7518             | 11.47            | 11.08            | 34.37            | 144              | 142              | 86.53            | 87.59            | 86.53            |
| 1.3 | PC2/GRAPE| PC2      | 7518             | 7441             | 49.20            | 73.55            | 59.18            | 142              | 150              | 84.42            | 62.13            | 91.79            |
| 1.4 | PC1      | PC2      | 7441             | 7441             | 100.00           | 100.00           | 100.00           | 150              | 150              | 100.00           | 100.00           | 100.00           |

Tests using different PCs and accelerator hardware

| ID | Set-up 1 | Set-up 2 | $E_{\text{kin}}$ | $E_{\text{pot}}$ | $E_{\text{tot}}$ | $E_{\text{kin}}^{B}$ | $E_{\text{pot}}^{B}$ | $E_{\text{tot}}^{B}$ |
|----|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|
| 2.1 | Standard | PC1/GPU  | 7554             | 7592             | 30.86            | 16.44            | 62.53            | 144              | 144              | 86.44            | 92.10            | 86.44            |
| 2.2 | Standard | PC1/GPU  | 7554             | 7640             | 21.75            | 16.36            | 34.70            | 144              | 154              | 86.56            | 36.04            | 74.98            |
| 2.3 | Standard | PC1/GPU  | 7554             | 8199             | 0.00             | 0.00             | 0.00             | 144              | 160              | 88.99            | 67.80            | 91.37            |
| 2.4 | Standard | PC1/GPU  | 7554             | 10052            | 0.00             | 0.00             | 0.00             | 144              | 209              | 15.77            | 35.88            | 30.41            |

Tests using different accuracy settings $\eta$

| ID | Set-up 1 | Set-up 2 | $E_{\text{kin}}$ | $E_{\text{pot}}$ | $E_{\text{tot}}$ | $E_{\text{kin}}^{B}$ | $E_{\text{pot}}^{B}$ | $E_{\text{tot}}^{B}$ |
|----|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|
| 2.1 | Standard | PC1/GPU  | 7554             | 7592             | 30.86            | 16.44            | 62.53            | 144              | 144              | 86.44            | 92.10            | 86.44            |
| 2.2 | Standard | PC1/GPU  | 7554             | 7640             | 21.75            | 16.36            | 34.70            | 144              | 154              | 86.56            | 36.04            | 74.98            |
| 2.3 | Standard | PC1/GPU  | 7554             | 8199             | 0.00             | 0.00             | 0.00             | 144              | 160              | 88.99            | 67.80            | 91.37            |
| 2.4 | Standard | PC1/GPU  | 7554             | 10052            | 0.00             | 0.00             | 0.00             | 144              | 209              | 15.77            | 35.88            | 30.41            |

Tests using STARLAB and NBODY6

| ID | Set-up 1 | Set-up 2 | $E_{\text{kin}}$ | $E_{\text{pot}}$ | $E_{\text{tot}}$ | $E_{\text{kin}}^{B}$ | $E_{\text{pot}}^{B}$ | $E_{\text{tot}}^{B}$ |
|----|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|
| 3.1 | Standard | PC3/GPU  | 7554             | 7501             | 9.59             | 48.69            | 21.44            | 144              | 135              | 99.84            | 99.22            | 99.94            |
| 3.2 | PC1      | PC3      | 7441             | 7507             | 81.64            | 17.46            | 63.79            | 150              | 146              | 92.75            | 71.24            | 87.68            |
| 3.3 | PC3/GPU  | PC3      | 7501             | 7507             | 72.60            | 37.09            | 82.08            | 135              | 146              | 79.35            | 29.11            | 85.84            |

This paper has been typeset from a TeX/LaTeX file prepared by the author.