New Astronomy 15 (2010) 16–23

Contents lists available at ScienceDirect

New Astronomy

journal homepage: www.elsevier.com/locate/newast

Teraflop per second gravitational lensing ray-shooting using graphics processing units ☆

Alexander C. Thompson, Christopher J. Fluke *, David G. Barnes, Benjamin R. Barsdell

Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, Victoria 3122, Australia

ARTICLE INFO

Article history:
Received 5 March 2009
Received in revised form 15 May 2009
Accepted 18 May 2009
Available online 24 May 2009
Communication by G.F. Gilmore

PACS:
95.75.Pq
98.62.Sb
98.62.−g

Keywords:
Gravitational lensing
Methods: numerical

ABSTRACT

Gravitational lensing calculation using a direct inverse ray-shooting approach is a computationally expensive way to determine magnification maps, caustic patterns, and light-curves (e.g. as a function of source profile and size). However, as an easily parallelisable calculation, gravitational ray-shooting can be accelerated using programmable graphics processing units (GPUs). We present our implementation of inverse ray-shooting for the NVIDIA G80 generation of graphics processors using the NVIDIA Compute Unified Device Architecture (CUDA) software development kit. We also extend our code to multiple GPU systems, including a 4-GPU NVIDIA S1070 Tesla unit. We achieve sustained processing performance of 182 Gflop/s on a single GPU, and 1.28 Tflop/s using the Tesla unit. We demonstrate that billion-lens microlensing simulations can be run on a single computer with a Tesla unit in timescales of order a day without the use of a hierarchical tree-code.

1. Introduction

Gravitational microlensing is the study of the deflection of light by matter in a regime where high magnification and multiple-imaging occurs, but the individual micro-images are not resolvable. This includes high magnification events due to lenses in the Galactic bulge and halo (Alcock et al., 1993; Aubourg et al., 1993; Udalski et al., 1993) and microlensing by compact objects within macro-lenses at cosmological distances (Vanderriest et al., 1989; Irwin et al., 1989). While Galactic microlensing projects have focused on searches for dark matter and the detection of planets, cosmological microlensing has led to advances in the understanding of stellar mass functions, mean stellar masses, and the structure of quasars, including constraints on the physical size of the emission regions at different wavelengths. See Wambsganss (2006); Kochanek et al. (2007); Gould (2008) and Mao (2008) for recent reviews.

The standard signature of cosmological microlensing, especially when applied to observations of active galactic nuclei, is an uncorrelated change in brightness of a single macro-image within a multiply-imaged system (Schneider and Weiss, 1987). Intrinsic variation in source flux is seen as a correlated change in the brightness of all the images, separated by the (macro)lensing time-delay. Such observations require accurate light-curves to be obtained over long time periods, in many cases decades, as there is a wide variation in the time-delay: 2–30 h for the quadruple-lensed Q2237+0305 (Vakulik et al., 2006) and 423 days for Q0957+561 (Hjorth et al., 2002) – see Saha et al. (2006) and Oguri (2007) for further examples.

Determination of the source size, source intensity profile, and physical properties of the microlenses (mass function, mean mass), requires a statistical comparison between observed light-curves and microlensing models. This is achieved through the use of the gravitational lens equation:

$$\mathbf{y} = \mathbf{x} - \alpha(\mathbf{x}), \quad (1)$$

which relates the two-dimensional locations of a source, $\mathbf{y}$, and an image, $\mathbf{x}$, with the deflection angle term, $\alpha(\mathbf{x})$, dependent on the arrangement of lenses. A common choice for microlensing is the many-Schwarzschild lens model:

$$\alpha(\mathbf{x}) = \sum_{i=1}^{N_*} m_i \frac{(\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^2} \quad (2)$$

for N_* lenses with masses, m_i, at positions $\mathbf{x}_i$. The magnification, μ, due to a gravitational lens system is

$$\mu = 1/\det \mathbf{A} \quad (3)$$

☆ Research undertaken as part of the Commonwealth Cosmology Initiative (CCI: www.thecci.org), an International Collaboration supported by the Australian Research Council.

* Corresponding author. Tel.: +61 (0) 3 9214 5828; fax: +61 (0) 3 9214 8797.
E-mail address: cfluke@swin.edu.au (C.J. Fluke).

1384-1076/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.newast.2009.05.010

where $\mathbf{A} = \partial \mathbf{y}/\partial \mathbf{x}$ is the Jacobian matrix of Eq. (1), which measures the areal distortion between the image and source planes.

While an image position maps uniquely to a source location ($\mathbf{x} \rightarrow \mathbf{y}$ is a one-to-one mapping), the converse is not true ($\mathbf{y} \rightarrow \mathbf{x}$ is a one-to-many mapping). Except for a limited number of special cases [see Schneider et al. (1992) for examples], the lens equation is not invertible. In the cosmological microlensing case, where many millions of individual stars may contribute to the observed magnification of a macro-image, it is more common to use a numerical technique to solve for μ over a finite region of the source plane – a magnification map – rather than attempting to find all image locations from Eq. (1) for a given source position (e.g. Paczyński, 1986).

Inverse ray-shooting provides the most straightforward means to obtain magnification maps for an arbitrary lens distribution (see Kayser et al. (1986) and Schneider and Weiss (1986, 1987) for early versions of this technique). Inverse ray-shooting follows a large number (typically millions) of light rays backwards from the observer, through the lens plane to the source plane, which is represented as a pixellated grid. The number of light rays falling in each pixel, N_{ij}, compared to the (average) number if there was no lensing, $N_{\rm av}$, gives an estimate of the per-pixel magnification:

$$\mu_{ij} = N_{ij}/N_{\rm av} \tag{4}$$

A typical magnification map is shown in Fig. 1, with the characteristic pattern of caustics clearly visible. Caustics are regions of high magnification – formally those points where $\det \mathbf{A} = 0$. The relative motion of the observer, lens plane and source imparts an effective transverse velocity to the source, causing it to move across the caustic network, and resulting in a time-varying change in source brightness. Accordingly, a sample light curve is generated by moving a source profile across a simulated caustic network, and converting the magnification at each point to a magnitude change.

Statistical investigations of cosmological microlensing require the generation of many sample light-curves, however, the creation of magnification maps poses a significant computational challenge. The time to calculate a magnification map is directly proportional to the number of pixels in the source plane ($N_{\rm pix}$), the number of microlenses (N_*), and the number of floating point operations[1] ($N_{\rm flop}$) per deflection calculation. As a Monte Carlo technique, the computation time is extended in direct proportion to $N_{\rm av}$, which sets the accuracy of calculated magnifications, and the number of repeat (n) map generations. Long compute times – O(days-months) – limit the scope to vary the input parameters, such as the initial stellar mass function, mean stellar mass, and source grid resolution. To keep computation times feasible for a direct implementation of the inverse ray-shooting method, the product $\Phi = n \times N_{\rm flop} \times N_{\rm pix} \times N_* \times N_{\rm av}$ historically has been constrained to $\lesssim O(10^{16})$.

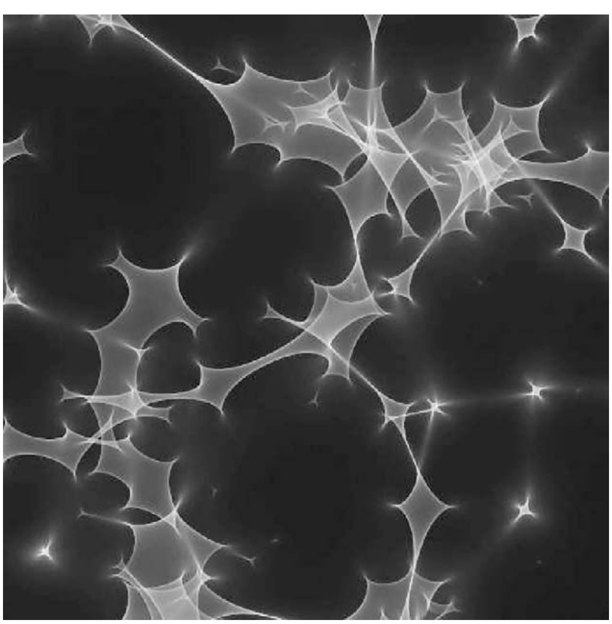

Fig. 1. A sample microlensing magnification map generated with a pair of NVIDIA GeForce 8800GT graphics cards. For model parameters: $N_* = 100$ lenses, $N_{\rm pix} = 1024^2$ pixels in the source plane, and $N_{\rm av} = 1000$ light rays per source pixel (on average), the processing time was 135 s. See Section 2 for details.

A number of approaches have been developed to overcome the processing time problem. Wambsganss (1990, 1999) used a hierarchical tree-code (Barnes and Hut, 1986), where lenses are treated differently depending on their distances from the light ray: lenses at a similar distance from a ray are grouped together and replaced with a single pseudo-lens of higher mass, effectively reducing the N_* factor. This introduces a slight error in the magnification map, which can be reduced by including higher order moments of the mass distribution. A parallel version of the tree-code, suitable for running billion-lens calculations on a parallel computing cluster – a region of parameter size previously unavailable to microlensing codes – has been implemented by Garsden and Lewis (submitted for publication).

Mediavilla et al. (2006) used a lattice of polygonal cells to map areas of the image plane to source plane pixels, rather than using a regular grid in the source plane. This greatly reduces $N_{\rm av}$, resulting in a $\sim 100\times$ speed-up to reach a given accuracy compared to standard inverse ray-shooting, however, preparing an appropriate polygonal lattice introduces a significant computation overhead.

A limitation of Monte Carlo-style methods is that many more magnification values, $N_{\rm pix}$, are constructed than may actually be required (e.g. to form a single light curve). A slightly larger (angular) size must be used for the image plane than for the source plane, as light rays at large impact parameters can be deflected into the source plane, contributing flux that would otherwise be lost. Consequently, more light rays must be generated than will actually fall within the source grid. Additionally, the finite source grid resolution means that true point source magnifications cannot be accurately calculated. To avoid these issues, Lewis et al. (1993) and Witt (1993) independently developed approaches based on imaging an infinite line in the source plane, which maps to a continuous, infinite line in the image plane, plus a number of closed loops – one for each microlens. Wyithe and Webster (1999) developed this technique further for extended sources.

In this work, we demonstrate that the redeployment of the direct inverse ray-shooting algorithm on modern, programmable graphics processing units (GPUs) can dramatically speed-up the calculation of microlensing magnification maps, without the programming overheads of implementing a more complex code. GPUs are macroscopic semiconductor arrays designed to accelerate the rendering of three-dimensional geometry for display on two-dimensional computer screens. Most modern computers contain a GPU, either on the system board or on a peripheral graphics card, which now regularly provide at least an order of magnitude greater raw computational power than the central processing unit (CPU). Rendering on-screen pixels is a highly parallel task, and this is reflected in the GPU architecture. Modern GPUs are primarily composed of stream processors, which are individual arithmetic logic units (ALUs) grouped in sets and controlled by an instruction scheduler with associated shared memory. Consequently, algorithms that lend themselves to the "stream processing" paradigm, where many individual data-streams all undergo identical opera-

[1] We use the notation: flop = floating point operation; 1 Gflop/s = 1 Gigaflop per second; and 1 Tflop/s = 1 Teraflop per second.