

Fast image reconstruction for fluorescence microscopy

Mahesh Ravi Varma, K. Rajan, and Partha Pratim Mondal*

Citation: *AIP Advances* **2**, 032174 (2012); doi: 10.1063/1.4754604

View online: <http://dx.doi.org/10.1063/1.4754604>

View Table of Contents: <http://aip.scitation.org/toc/adv/2/3>

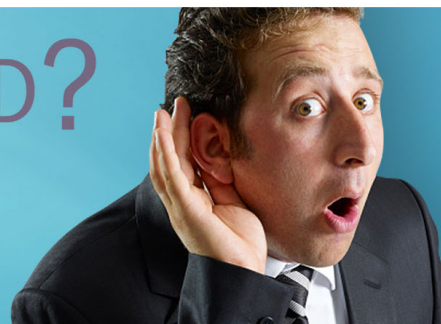
Published by the [American Institute of Physics](#)

HAVE YOU HEARD?

Employers hiring scientists and
engineers trust

PHYSICS TODAY | JOBS

www.physicstoday.org/jobs



Fast image reconstruction for fluorescence microscopy

Mahesh Ravi Varma,¹ K. Rajan,² and Partha Pratim Mondal^{1,3,a}

¹Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore 560012, India

²Department of Physics, Indian Institute of Science, Bangalore 560012, India

³Applied Photonics Initiative, Indian Institute of Science, Bangalore 560012, India

(Received 27 June 2012; accepted 10 September 2012; published online 17 September 2012)

Real-time image reconstruction is essential for improving the temporal resolution of fluorescence microscopy. A number of unavoidable processes such as, optical aberration, noise and scattering degrade image quality, thereby making image reconstruction an ill-posed problem. Maximum likelihood is an attractive technique for data reconstruction especially when the problem is ill-posed. Iterative nature of the maximum likelihood technique eludes real-time imaging. Here we propose and demonstrate a compute unified device architecture (CUDA) based fast computing engine for real-time 3D fluorescence imaging. A maximum performance boost of 210× is reported. Easy availability of powerful computing engines is a boon and may accelerate to realize real-time 3D fluorescence imaging. *Copyright 2012 Author(s). This article is distributed under a Creative Commons Attribution 3.0 Unported License.* [<http://dx.doi.org/10.1063/1.4754604>]

It's long being speculated the existence of crisp artifact-free 3D fluorescence image in real-time, which otherwise gets degraded by several physical processes including noise, blurring and scattering.¹ Iterative maximum likelihood estimation (MLE)^{2,3} is an attractive approach but requires a large number of iterations for convergence. This severely hampers the temporal resolution of the imaging system and eludes real-time fluorescence imaging.

Image reconstruction involves accurate modeling of the imaging process. In fluorescence imaging, the measured data g (fluorescence light recorded by photomultiplier tube or CCD camera), the object f (fluorescently-tagged) and the noise η are related by,

$$g(x, y) = Af(x, y) + \eta(x, y) \quad (1)$$

where, the operator A can be constructed from the relation $Af = k * f$ under space-invariance condition; $*$ is the convolution operator.

Due to the unknown noise distribution and data-loss during data acquisition, the image reconstruction problem does not have a unique solution.^{4,5} This calls for iterative image reconstruction procedures for tackling the ill-posed non-linear problem. The simplest way to estimate the distribution of fluorescent molecules in the object is to perform deconvolution.⁶⁻⁸ Deconvolution performs poorly because this does not incorporate statistical information in the image reconstruction process. A statistically accurate algorithm can be formulated using maximum likelihood / Richardson-Lucy (ML-RL) method.^{2,3} This statistical method utilizes the fact that, the fluorescence emission is a Poissonian process and the object estimate can be obtained by maximizing the cost function.^{1,5} The ML algorithm is quite successful for producing an approximate estimate of the fluorophore distribution. The flip side of iterative ML technique is it's poor temporal resolution. The algorithm requires long processing time for reaching to an approximate map of the fluorophore distribution. Some of the early work done in image reconstruction for optical microscopy include defocussing

^aCorrespondence: Partha Pratim Mondal PhD, Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore 560012, INDIA; E-mail: partha@isu.iisc.ernet.in.



techniques and its application in cellular imaging.^{9,10} Nevertheless, ML algorithm has been quite successful for image reconstruction in other imaging modalities such as, multiphoton microscopy and nuclear medicine imaging.¹¹⁻¹⁵

In this letter, we demonstrate a technique to accelerate the image reconstruction process for fluorescence microscopy. This will enable real-time image reconstruction and dynamic visualization of biological processes. We use CUDA architecture based fast computing engines (GTX-275 and Tesla-C2070) to execute the iterative data reconstruction process for realizing real-time imaging. Within the iterative image reconstruction process, the proposed technique requires taking forward and inverse Fourier transform(twice), point-to-point multiplication and division, and transpose of the 3D matrix which are computationally intensive, but parallelizable tasks. Depending on the data volume and the number of iterations, these computations can span up to several minutes. One major bottle-neck is the architecture(Intel x86) itself that does not support parallelization of these tasks. Here, we aim to take advantage of the general purpose graphics processor units (GPUs) for massively parallelizing these calculations. This work specifically aims to accelerate the maximum likelihood image reconstruction using a cost-effective programmable GPU, which is capable of accelerating the key computational operations (such as, 3D-FFT, 3D-IFFT and others) of the algorithm. Faster implementation of iterative image reconstruction technique can aid in low-light imaging and dynamic monitoring in real time. This is an essential requirement because the dye undergoes photobleaching and severely hampers the SNR. It may be noted that, noise level increases for long exposure times. In such a situation, ML-RL algorithm has proven to be effective in restoring the signal-to-noise ration (SNR) of the 3D image.^{5,6,8,16-18}

In fluorescence microscopy, we use fluorescent dyes to tag the protein of interest in the specimen. The dynamics of the protein is then followed with high spatial and temporal resolution. Molecular level spatial resolution is prohibited by the diffraction limit set by laws of physics.¹⁹ Even for super-resolution imaging, single molecule resolution is marred by random noise which is due to the source induced Poissonian noise, scattering and photobleaching. So, most of these high resolution techniques require real-time noise filtering. ML is an attractive algorithm but not suitable for real-time imaging which is due to slow convergence.⁵ This is because of the ill-posed nature of the inverse problem associated with fluorescence imaging.

We propose fast 3D image reconstruction using parallel computing engines for realizing temporal super-resolution. In graphics processing unit (GPU) based systems, this has been made possible using a large number of tiny compute nodes organized as grids and blocks. ML-algorithm based on the maximization of the likelihood function (eqn. (1)) for obtaining optimal estimate, is given by,⁵

$$f_n^{k+1} = f_n^k \left(A^T \frac{g}{Af_k} \right), \quad n = 1, 2, \dots, M \quad (2)$$

where A is the optical transfer function of the system, g is the raw image obtained from the microscope, f^k is the current estimate of the reconstructed image, and M is the number of pixels.

The parallelization with the CUDA architecture involves, single step computation of the following operations namely, Af or $K * f$ (3D FFT, point-to-point multiplication and 3D IFFT), A^T (3D matrix transpose) and $\frac{g}{Af}$ (point-to-point division). The algorithm demonstrating the fast computation of essential steps is given in supplementary 1 (Table 1).²⁰

The next step is the criterion for ceasing the iterative process which is taken as the I-divergence measure. I-divergence between the original object f and the reconstructed object \tilde{f} is defined as,

$$I_{f, \tilde{f}} = \frac{\sum_{n=1}^M \left[f_n \ln \frac{f_n}{\tilde{f}_n} - (f_n - \tilde{f}_n) \right]}{M} \quad (3)$$

But, since the original object is not known *a priori*, the I-divergence between the recorded image g and the object convolved with PSF, i.e., $f^k * K$ is calculated at each iteration k . When the difference between the I-divergence values of two consecutive estimates is smaller than a certain threshold, we stop the iteration. Beyond this threshold value, the change is found to be negligible. So for all practical purpose, this step gives the best possible estimate.

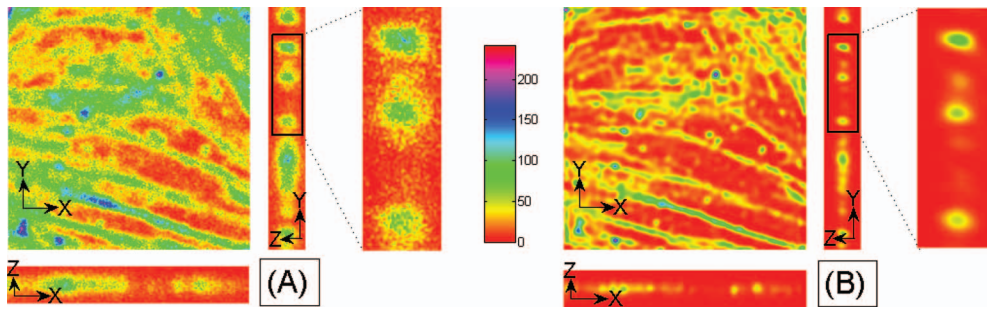


FIG. 1. 3D image (shown as, XY-, XZ- and YZ- planes) of F-Actin filaments in a BPAE cell. (A) Raw image, (B) Reconstructed image.

Parallel computing is a form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently. In the present problem, we can observe a lot of *data parallelism* and *data independence* which facilitate parallelization. This validates the scope for using a parallel computing architecture like CUDA for speeding up the computation. The CUDA approach is based on partitioning the computational operations in to a large number of similar threads and these threads are spawned on to a large number of tiny compute nodes organized as multiprocessors. The reference CPU-based system is a 2.1GHz *Core2 Duo* system with 4GB RAM. We studied the performance of ML-based 3-D image reconstruction on two CUDA-based parallel architectures: (a) NVIDIA GTX-275 graphics card running at 1.49GHz, referred to as *GPU-A*^{21,22} and (b) NVIDIA Tesla-C2070 Fermi-architecture based graphics card, running at 1.15GHz, referred to as *GPU-B*.²³ The specifications of the two GPUs are given in the Supplementary 1 (Table 2).²⁰

To test the performance of these CUDA based methodology, 3D images were acquired using an Olympus confocal system. Images of dimension $256 \times 256 \times 40$ are acquired with a lateral (XY) and axial (Z) sampling of 60 nm and 120 nm respectively. BPAE cells are imaged for visualizing the structure of actin filaments tagged with green-fluorescent BODYPY FL phalloidin. The excitation and emission wavelengths are 488 nm and 520 nm respectively. The images are acquired using, Olympus FV1000 microscope, equipped with an immersion oil Apochromat 60X/NA = 1.42 objective.

Figure 1 shows 3D reconstructed images of actin filaments in BPAE cells using ML method. It is evident that noise is dominant in the raw image generated by the camera. Major improvement in the image quality is noticeable as far as noise is concerned (see inset). Along the z-axis, a reduction in point like structure is observed as shown in the zoomed inset of figure 1. This implies a straightway PSF reduction of almost 50% in the reconstructed image thereby implying an impressive improving in the resolution by a factor to 2. Another important indicator of image quality is intensity plot along a predefined line. This plot facilitates pixel fluctuations around a definite signal and in the background as well. Figure 2 show the intensity plots along a line passing through several Actin filaments. The algorithm is able to extract noise from minute features as well indication the noise cancelling ability of statistical ML algorithm.

Next, we investigate the stopping criterion for ceasing the iterative image reconstruction process. We use Ciszar's I-divergence test which evaluates the divergence between the raw and reconstructed image. This simultaneously ensures the statistical similarity (between the raw and reconstructed data set), and the best image quality. Two steps are followed, first is I-divergence value and second, the difference $\Delta I = I^{k+1} - I^k$ between two iterations. A threshold of 10^{-5} is found to be a reliable value for obtaining the best reconstruction. Figure 3 show the behavior of I-divergence and ΔI (see inset) with the number of iterations. This plot suggests that, the critical iteration is 30. This suggests that, these 30 iterations must be calculated at a rate of $\approx 30 \text{ frames/sec}$ for ensuring real time fluorescence imaging.

It may be noted that, the Tesla-C2070 has much more global memory than GTX-275 (see, Supplementary 1, Table 2).²⁰ With larger global memory area, more data can be transferred to the

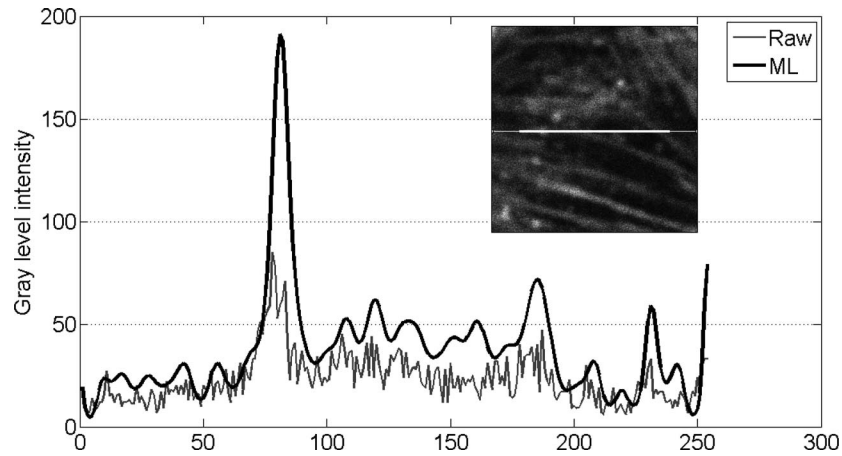


FIG. 2. Intensity plots along a predefined line passing through the image before and after reconstruction.

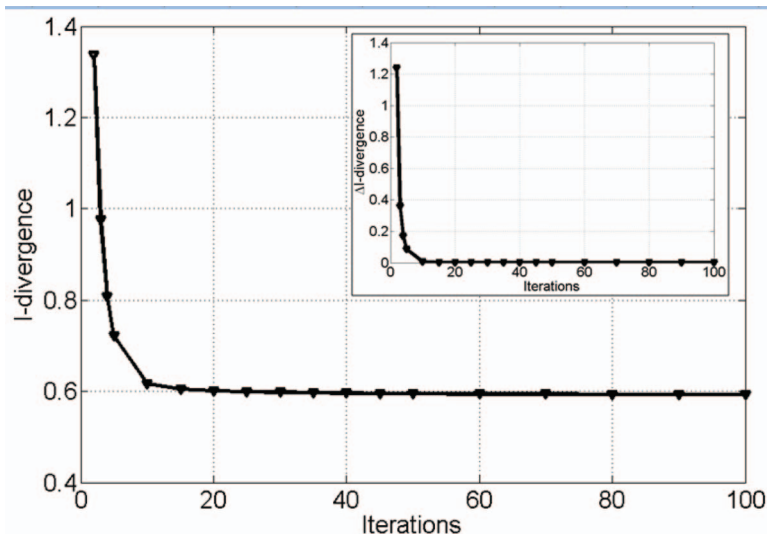


FIG. 3. I-divergence against the number of iteration showing convergence of the algorithm. Inset show the incremental change (ΔI) in the I-divergence value.

GPU side in a single step. The CUDA best practice guide²¹ recommends minimum number of data transfers between CPU and GPU. This is because the data transfer between GPU and CPU consumes large number of clock cycles. The total number of cores for Tesla-C2070 is 448 as against 240 for GTX-275. Each core is an independent processing unit. Higher the number of cores, higher degrees of parallelization can be achieved. Maximum threads per block for Tesla-C2070 is 1024 as against 512 for GTX-275. *Max. threads per block* figure tells the number of threads that can be launched simultaneously. This in-turn indicates the degree of parallelization that can be achieved. The *registers per block* and *shared memory per block* figures are higher for Tesla-C2070, which indicates that more computations can be carried out per block without consuming global memory. Confining the memory usage to the register memory and shared memory rather than global memory is a good way to increase the performance. Figure 4 shows the real-time performance of the computing engines. The most critical computational steps are 7–9 and 16–19, that consumes the maximum computational time.

Next, we compare the absolute time taken for executing the 3D image volume, as shown in figure 5. For small volume, both GTX and Tesla have similar time of execution, whereas, the difference increases with increasing the number of pixels. The data is purposefully indicated as,

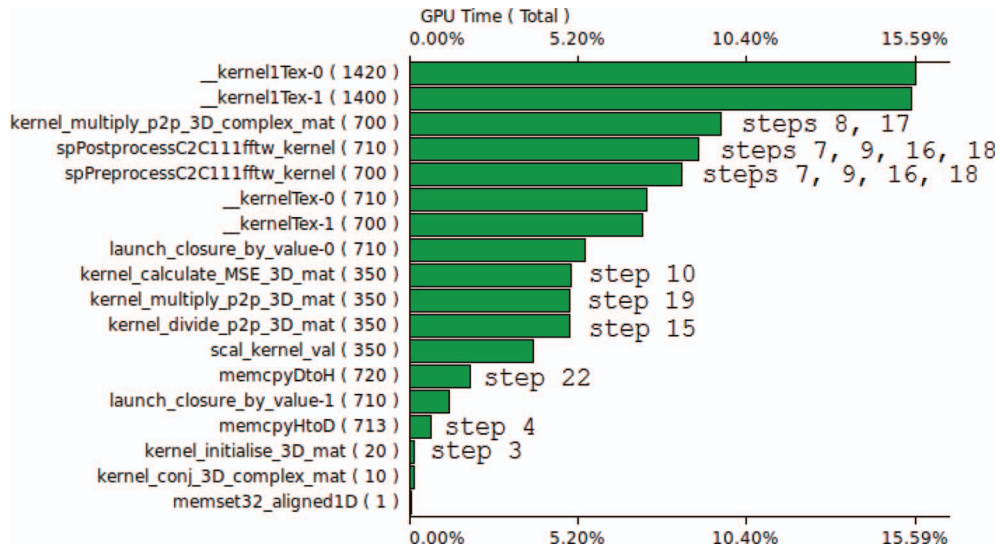


FIG. 4. CUDA profiler output that shows the percentage execution time taken by individual CUDA kernels. This shows that, the 3D forward and inverse FFT should take a significant amount of time. The 3D convolution operation, $A * f_k$ and $A^T * \frac{g}{(A * f_k)}$ is realized in frequency domain and corresponds to steps 7,9,16,18 and takes almost 20% of the total GPU time. Other Kernels corresponding to multiplication and division takes only 5% of the GPU time.

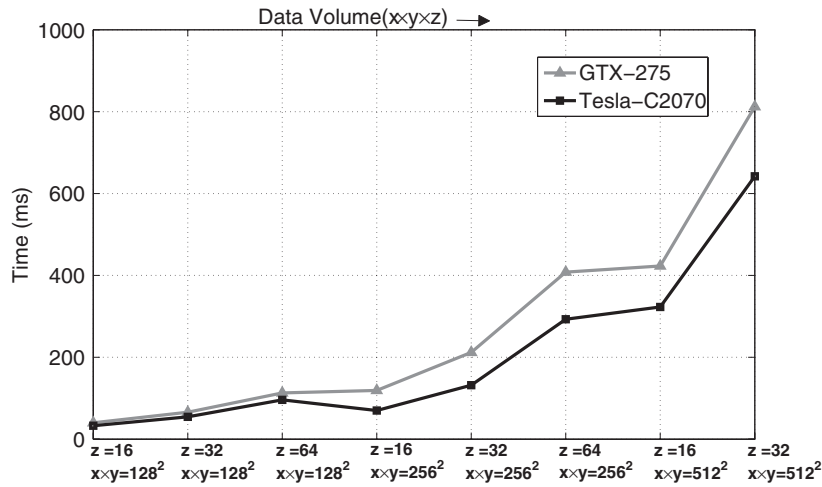


FIG. 5. Time taken versus data load for both GTX – 275 and Tesla – C2070.

image size and the number of planes for facilitating an intra-plane and plane-by-plane processing time. This is important when real time visualization of any oblique plane or plane-by-plane image is desired. Extensive study involving varying 3D image data set and number of iterations is provided in Supplementary 1 (Table 3).²⁰ Further, we define the performance for a particular computational operation on a CUDA device with respect to CPU device i.e., the ratio of the time taken for CUDA-based implementation to that for CPU-based implementation. The specification table for both GTX-275 and Tesla-C2070 can be found in supplementary 1 (Table 2).²⁰ Figure 6 shows the performance curve for GTX-275 and Tesla-C2070. It is evident that the performance of Tesla is superior than GTX as the computational task (total number of pixels) increases, which is because of the large global memory and double the number of cores of the Tesla-C2070 engine. Tesla-C2070 gives a performance boost of up to 210X, whereas, GTX – 275 has a maximum boost of 150X. This is critical because fluorescence imaging often involves processing of larger data size (image size involving 10.5 mega pixels) obtained from the CCD/CMOS camera. Future computing engines

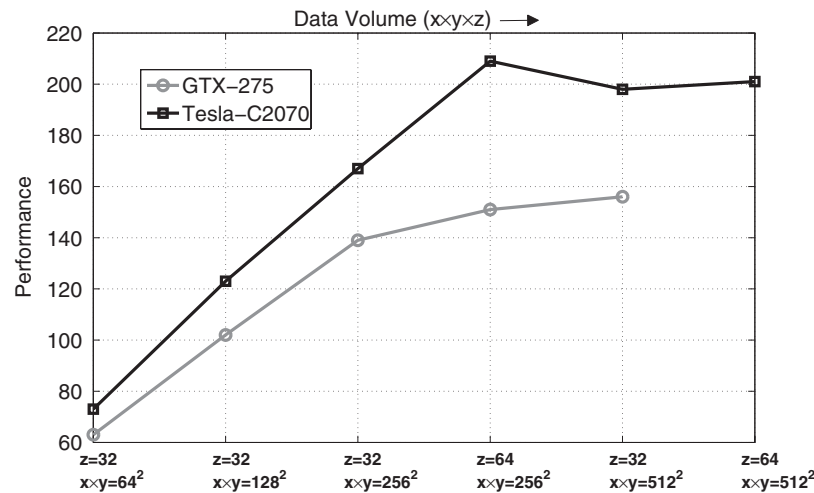


FIG. 6. Performance of GTX – 275 and Tesla – C2070.

with more global memory and large number of cores will be a great boon for real-time fluorescence imaging.

In conclusion, we propose a multi-core computing engine for real-time high resolution fluorescence imaging. Massive parallelization of key operations of the ML algorithm comprising of FFT, IFFT, point-by-point multiplication and transpose operation, is responsible for temporal super-resolution. The very architecture of computing engines such as, GTX and Tesla, are tailored to carry out the key operation of the ML algorithm. ML implementation on these fast machines shows an improvement of about 210X, thereby achieving real-time imaging. This technique may find immediate application in-almost all forms of fluorescence imaging and 3D data visualization.

The authors acknowledge DAE and DST for research funding. They would like to thank SERC at IISc for giving access to the Tesla computational cluster, the NVIDIA corporation for providing Tesla hardware and research support.

¹ M. Bertero and P. Boccacci, *Introduction of Inverse Problems in Imaging* (IOP, London, 1998).

² W. H. Richardson, *Jl. Opt. Soc. Amer.* **62**, 55 (1972).

³ L. B. Lucy, *Astron. Jl.* **79**, 745 (1974).

⁴ A. Dempster, N. Laird, and D. Rubin, *J. R. Stat. Soc. Ser. B (Methodol.)* **39**, 1 (1977).

⁵ P. P. Mondal, G. Vicidomini, and A. Diaspro, *Jl. Appl. Phys.* **102**, 044701 (2007).

⁶ P. J. Verwee, M. J. Gemkow, and T. M. Jovin, *J. Microsc.* **193**, 50 (1999).

⁷ G. M. P. van Kempen, L. J. van Vliet, P. J. Verwee, and H. T. M. van der Voort, *J. Microsc.* **185**, 354 (1997).

⁸ M. Bertero, P. Boccacci, G. J. Brakenhoff, F. Malfanti, and H. T. M. van der Voort, *J. Microsc.* **157**, 3 (1990).

⁹ A. Diaspro, M. Sartore, and C. Nicolini, *Img. Vis. Comp.* **8**, 130 (1990).

¹⁰ B. Bianco and A. Diaspro, *Cell Biophys.* **15**, 189 (1989).

¹¹ A. Periasamy and A. Diaspro, *J. Biomed. Optics* **8**, 327 (2003).

¹² K. Rajan, L. M. Patnaik, and J. Ramakrishna, *IEEE Nucl. Sci.* **41**, 1721 (1994).

¹³ K. Rajan and L. M. Patnaik, *Microprocessors and Microsystems* **25**, 233 (2001).

¹⁴ P. P. Mondal and K. Rajan, *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, (Ontario, Canada, May 17–19, 2005).

¹⁵ P. P. Mondal and K. Rajan, *Proceedings of IEEE Nuclear Science Symposium and Medical Imaging Conference*, (Rome, Italy, October 16–22, 2004).

¹⁶ G. Vicidomini, P. P. Mondal, and A. Diaspro, *Opt. Lett.* **31**, 3582 (2006).

¹⁷ P. P. Mondal, G. Vicidomini, and A. Diaspro, *Appl. Phys. Lett.* **92**, 103902 (2008).

¹⁸ J. A. Conchello and J. G. McNally, in *Three Dimensional Microscopy: Image Acquisition and Processing III*, edited by C. J. Cogswell, G. Kino, and T. Wilson, (Proceedings of SPIE, 1996), Vol. 2655, p. 199.

¹⁹ E. Abbe, "Note on the proper definition of the amplifying power of a lens or a lens system," *J. Roy. Microsc. Soc.* **4**, 348–351 (1884).

²⁰ See supplementary materials <http://dx.doi.org/10.1063/1.4754604> for the algorithm for fast computation, GPU specifications and extensive data characterization.

²¹ <http://www.geforce.com/hardware/desktop-gpus/geforce-gtx-275>.

²² <http://www.nvidia.com/object/personal-Supercomputing.html>.

²³ <http://www.nvidia.com/object/fermi-architecture.html>.