DDN® STORAGE

WHITE PAPER

Real-Time Deskew and Deconvolution for Improved Resolution of Microscopy Images

Marc Bruce
CEO and Co-Founder, Microvolution

George Vacek
Global Director, Life Sciences, Healthcare and Machine Learning, DDN

William Beaudin
Director of Solutions Engineering, DDN

Abstract

Lattice LightSheet™ microscopes and other high data rate instruments are revolutionizing biological research in areas such as neuroscience, developmental biology, cancer research and biomedical engineering. Storing and handling the large amounts of data generated by these instruments is often a concern for researchers looking to leverage the full potential of leading-edge imaging capture and analysis methods. We propose a technology platform that integrates a DDN® AI200® shared parallel storage appliance, a NVIDIA® DGX-1™ server and the Microvolution® advanced processing software to solve this challenge. The proposed solution enables and accelerates microscopy workflows for deconvolution and deskewing, and can deliver real-time image capture and processing capabilities with over 1,600-fold yield compared to traditional microscopy workflows.

Introduction to Deconvolution in Microscopy

In fluorescence microscopy, light comes into a sample and is collected at an objective focus. Improving signal to noise for increased resolution requires increasing the amount of light exposure of the sample. Since exposure to higher levels of illumination for longer periods of time compromises the sample, one goal is to maximize resolution with minimum light exposure to reduce photo damage. Other physical aspects can increase the resolution challenge, such as haze from out-of-plane fluorescence and diffraction, ultimately limiting the ability to resolve small objects in both 2D and 3D. Correcting for these aspects of blur in an image can help increase the resolution, which in turn helps reach the scientific goal of resolving each component of the sample so one can better understand the real fundamentals of biology that are otherwise hidden away.

Deconvolution is a widely recognized and sophisticated approach to removing the blur in an image that has evolved significantly over the last decades. By applying mathematics, an algorithm can compute the blurring (or spread) of any point in an image to be the result a point spread function (PSF). If the true image of a sample has been convolved by the PSF into the observed image stack, then to unblur the image, one simply needs to deconvolve the observed image stack back into the true object. Where the PSF is exactly known or can be measured, this is straightforward. In most real-world uses, the specific functional form of the contributions to blur and noise in the image are not exactly known, but nevertheless a close approximation for the PSF can be solved.

$$\text{Observed} = \text{True} * \text{Blur} + \text{Noise}$$

$$g = f * h + \varepsilon$$

$$f^{i+1} = f^{i} \cdot \left(\frac{g}{f^{i} * h} * \hat{h} \right)$$

Equation 1: Richardson-Lucy iterative algorithm for deconvolution, assuming Poisson-distributed noise.

Numerous deconvolution algorithms have been developed over the years, the most famous of which was developed by Richardson and Lucy and originally applied to the 2D case of astronomy. It is an iterative algorithm that assumes a Poisson distribution to the noise, which is reasonable especially for very dim samples that are limited more by photon noise than by camera readout noise, which follows a more Gaussian distribution. Equation 1 shows a somewhat simplified notation of the core Richardson - Lucy algorithm. The approach starts with the observed image and iteratively updates to new images that are successively more deblurred, eventually converging upon one that can be considered to be correct.

Deconvolution can be used for reanalysis of existing data collections, enabling researchers to discover new insights from their existing data. For instance, taking the exact same raw data, deconvolution produced a fourfold effective increase in contrast at Yeshiva University for a study of macrophages phagocytosing red blood cells. Using this post-processing reanalysis, researchers discovered insights into low-contrast areas because deconvolution essentially moved all the blurred light back to where it should have been to give a better contrast ratio*. More often, the contrast is fine, but the physics of diffraction is limiting resolution as the distance between two objects approaches the wavelength of the light source. Deconvolution can improve discrimination between objects much further, so that the exact same experimental set up yields better results. Even in cases where there is a good signal-to-noise ratio, there are benefits for deconvolution in 3D imaging to address out-of-plane fluorescence. In all of these cases, deconvolution enables renewed insight from data captured on older instruments and deliver better project results.

Deconvolution has benefits for newer microscope techniques as well. Light sheet microscopes reduce the amount of photo-induced damage to samples by illuminating only the imaged portion of the sample through a precise light source which is directed from the side, rather than an overhead light source that constantly illuminates the full sample during the entire capture. The sample resolution is directly tied to the thickness of the sheet of light used during imaging, and can further be enhanced through post-processing methods like deconvolution. Computational requirements for deconvolution grow with image size. Light sheet microscopy images are generally much bigger than traditional widefield microscopy images, and large 2K by 2K by 2K volumes are becoming quite common. The volumes have been growing quickly, not only for the reason that camera sensitivity allows for higher resolution, but also because the cameras are simultaneously getting faster and there are new sample preparation techniques, such as cleared tissue, which allow deeper imaging within the samples.

Lattice light sheet microscopy (LLSM), Nature method of the year in 2014, boosts image clarity while reducing phototoxicity and photobleaching, thus allowing researchers to image live cells at high resolution for extended periods. Researchers at such institutions as the University of Queensland, the University of New South Wales and The Department of Developmental Neurobiology, Advanced Neuroimaging Laboratory, Memphis, TN are using LLSM for applications in neuroscience, developmental biology, cancer research and biomedical engineering. Light sheet microscopes, notably Intelligent Imaging Innovations' (3i) Lattice LightSheet, can capture images at potentially much higher resolutions (in spatial, axial or temporal terms) than traditional instruments and may offer variable framerates depending on the field of view engaged. For instance, they might run at 100 frames per second with a full field of view, or at several hundred frames per second with a narrow field of view. Regardless of whether there are larger images, or more frames per second, the upshot is more data that must be deconvolved, greatly increasing computational requirements. To further complicate data interpretation with sample-scanning LLSM, the sample may be moved through the light sheet rather than the light sheet being scanned through the sample, which results in the raw images being skewed due to the geometry between the objective camera and the sample. This means the raw data needs to be deskewed before the deconvolution method is applied, which requires additional computation time and data storage resources.

* "Using Fluorescence Resonance Energy Transfer-Based Biosensors to Probe Rho GTPase Activation During Phagocytosis", V. Miskolci , L. Hodgson , and D. Cox, R. Botelho (ed.), *Phagocytosis and Phagosomes: Methods and Protocols, Methods in Molecular Biology*, **1519**, (2017) DOI: 10.1007/978-1-4939-6581-6_9, ©Springer Science+Business Media New York

GPU-Accelerated Deconvolution

Traditional deconvolution algorithms have been time-consuming and resource-intensive. CPU-based methods require significant compute resources running for long periods of time to process a single image. The lack of real-time feedback from the images reduces the efficiency of the microscopy pipeline and limits its usefulness. Sample cells may no longer be alive by the time that the image is processed, which severely limits the ability to adjust the experiment to acquire satisfactory results. The platform that we propose today, using Microvolution's GPU-based algorithms, enables real-time deconvolution and processing, providing immediate images and feedback to the microscope operator.

Microvolution uses the Richardson-Lucy deconvolution algorithm with some modifications and significant optimizations to efficiently correct images as quickly as possible. In particular, Microvolution has spent considerable effort parallelizing the approach to take full advantage of a GPU architecture and acceleration that it provides. For instance, calculating the PSF requires lots of integrals and other computationally-intensive operations. Instead of calculating once and relying on that result, parallelizing this operation allows recalculation of a new PSF for each new imaging condition, which improves results. In GPU programming, this makes additional sense because transferring a previously calculated PSF into GPU memory could take longer than calculating a new one from scratch. Microvolution has further considered memory block reuse to minimize transfer, locality such that two operations are as close as possible, and, access patterns such that, where not everything physically fits in one GPU's memory, the least accessed portion can reside on a different GPU leveraging peer-to-peer memory or in CPU RAM with pinned memory. Microvolution has also parallelized the deskew operation necessary for some datasets and optimized it to take full advantage of a GPU architecture and ensure fastest time to results.

The Microvolution software is fully-interoperable with standard microscopy image file formats and is already integrated within common applications such as Fiji and 3i's SlideBook. It also provides an extensive API for easy integration with other applications, or site-specific processing pipelines.

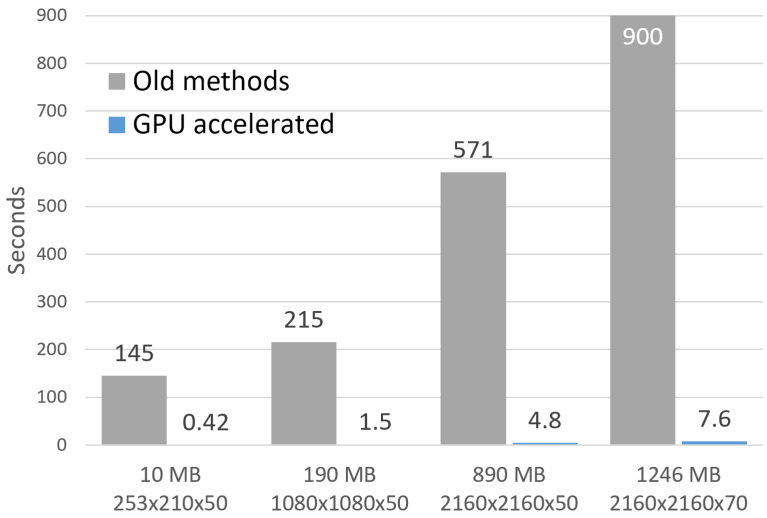

Figure 1: Microvolution's GPU optimized deconvolution is roughly 200x faster than CPU based approaches.

Deconvolution performance comparisons for traditional approaches and Microvolution GPU accelerated software appears in Figure 1*. Relative performance varies across the image sets, but for the largest, deconvolution run time drops from 15 minutes with CPUs to less than 8 seconds on NVIDIA Pascal GPU architecture. On average, Microvolution delivers deconvolution almost 200 times faster per thread than traditional methods, which lets users change how they leverage deconvolution.

Instead of limiting deconvolution to a post-processing step used to achieve high-quality results, Microvolution helps users discover what they can change within the experiment itself. They have started optimizing image capture, because they know that a lower signal-to-noise ratio is sufficient for quality results, so they can lower exposure doing less photo damage to their sample. Further, less exposure means one can acquire an image faster, so they might be able to take more time points, improving their experiment. Alternately, they would finish their run faster allowing the next person to get on an overbooked instrument sooner, allowing for more efficient utilization of shared instruments. Fast deconvolution, of course, also facilitates the reanalysis of old data to get better results; the faster deconvolution runs, the more past projects that can be reanalyzed.

High Performance Data Storage

The rapid evolution of microscopy techniques, camera technologies and processing applications has increased compute and data storage requirements for microscopy applications. Traditionally, a single workstation was used to capture data from the instrument on a local storage device, and for engaging post-processing operations. This approach imposes severe limitations on the performance, capability and scalability of the workflow. Moreover, even latest generation workstations are unable to keep up with real-time capture and processing of data generated by modern instruments used in LLSM.

Current generation microscopy workflows leverage a central data storage system. The shared architecture of the DDN AI200 provides tremendous benefits, simplicity, and flexibility to microscopy pipelines. It enables data to be captured directly from microscopes to a central storage system. The data is immediately available to GPU computing systems for processing and analysis, and to visualization workstations. The AI200 simultaneously serves all phases of the pipeline seamlessly and eliminates the need for time-consuming, error-inducing data management tasks like copying files from local storage on capture to analysis systems. Data ingest and access to a shared storage system also ensures maximum efficiency of specialized instruments and GPU computing resources, having no unnecessary downtime required for data movement. Virtually limitless data ingest and storage capacity allows for the creation of datasets of any size, removing barriers imposed by limitations of local storage on capture and analysis systems. This enables high-throughput analysis of many and large image volumes with maximum efficiency, whether to keep up with new data from multiple microscopes or for re-analysis of existing datasets. The AI200 can be configured for a small initial footprint, and scale easily in capacity, performance and capability to match evolving workflow requirements and

*"Real-time GPU-based 3D Deconvolution" M.A. Bruce, and M.J. Butte, *Optics Express* **21**, 4766-73 (2013) DOI:10.1364/OE.21.004766.

future investments in instruments, computing technology and analysis software. The DDN platform includes robust data protection mechanisms and can be coupled with DDN backup and archive solutions for enhanced data governance and regulatory compliance requirements.

The DDN AI200 provides breakthrough capabilities for GPU-accelerated applications. A single AI200 can support capture from multiple high data rate instruments and delivery of data to multiple GPUs for processing. The AI200 provides fast and efficient storage-to-GPU and GPU-to-GPU data delivery, even during peak utilization. The DDN shared parallel architecture allows linear scaling of microscopy workflows as more GPU compute resources are engaged.

The capabilities of the DDN shared parallel architecture and protocol are a cornerstone of accelerating data-intensive real-time 3D deconvolution with GPU computing. The optimal technology infrastructure for accelerating the Microvolution software integrates an AI200 all-flash system and a NVIDIA DGX-1 system connected with a high-speed, low-latency, RDMA-enabled EDR InfiniBand™ network. The DDN shared parallel architecture and protocol ensures that every GPU compute cycle is put to productive use for microscope image processing. DDN delivers over 8x higher application throughput than a cloud service with a similar number of GPUs.

Demonstrating Real-Time 3D Deconvolution and Deskew

To achieve real-time analysis, the technology infrastructure must be capable of capturing data from an instrument while also delivering sufficient data to fully saturate GPU computing facilities. The following benchmarks demonstrate this idea of advanced architecture and workflow for real-time processing, using datasets from Daniel Stabley of The Department of Developmental Neurobiology, Advanced Neuroimaging Laboratory, Memphis, TN and James Springfield of the Institute for Molecular Bioscience at University of Queensland, both of whom use Microvolution to process and DDN storage to store data from their 3i LLS microscopes.

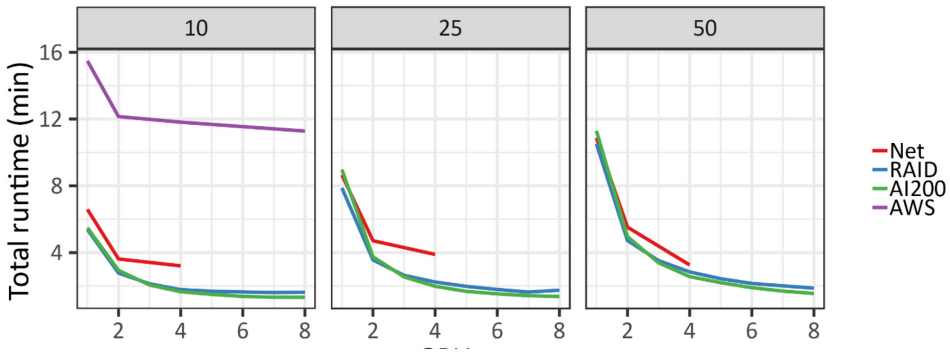

Figure 2: Runtime for Microvolution's deconvolution algorithm on lattice light sheet images for increasing numbers of GPUs across different storage architectures.

The first benchmark uses Microvolution software to deconvolve a LLS dataset with several processing architectures based on NVIDIA Tesla V100 GPUs. The LLS dataset is a total of 40 GB comprised of 540 image volumes in separated TIFs, each of which is 512x512x151 pixels in size. Each of the volumes run completely independently so there's no dependency across GPUs. Runs may deconvolve for 10, 25 or 50 iterations, with more iterations increasing the amount of compute vs IO, and in principle yielding better results. Standard practice is usually around 25 iterations, but may vary depending on the experimental particulars and availability of resources. No effort has been made to optimize performance for any of these platforms in terms of the file size, the number of concurrent IO operations, or anything else, but rather assumes a naïve user with plug-and-play default settings.

Figure 2 shows runtime performance, not including data transfer time to the local RAID or AWS. NET represents NVIDIA's internal system with 4 GPUs to which they generously provided access. AWS is an instance on Amazon Web Services comprised of 8 V100 GPUs. RAID and AI200 are run on the same DGX-1 system (8 GPU) using either the local RAID on the DGX-1 or AI200 as an (InfiniBand) network attached storage.

The most interesting aspect of these performance results is that the local RAID and the AI200 have similar profiles, and the AI200 performs better than a local RAID for more than 3 GPUs since it can deliver true parallel data streams from the high performance flash—all the way to the application running on GPUs. The runtime displayed ignores the amount of time it took to copy data from and results back to the local RAID, further lowering its effective performance compared to the AI200. This, plus the simplified workflows, reinforces the idea of skipping local storage entirely and instead working directly from a fast central data store.

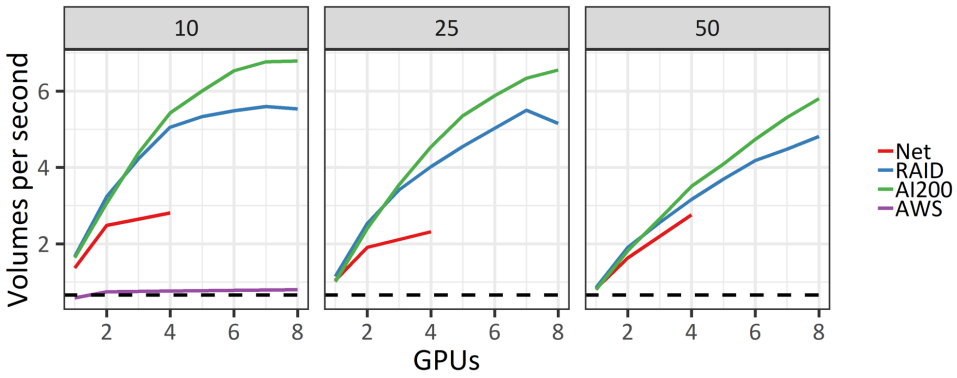

Figure 3: Throughput of Microvolution's deconvolution algorithm for increasing numbers of GPUs across different storage architectures, not including data transfer time for RAID or AWS. The breakeven points for real-time analysis is shown by the dashed line.

In considering real-time analysis, the breakeven point is the processing rate necessary to keep up with instrument data generation. Here we use a worst-case calculation based on the dataset's exposure time multiplied by the number of planes per volume. This is clearly an upper limit, since a) the instrument must move between Z positions, which takes time, and b) this doesn't account for transfer time from the camera's frame buffer to the system. Figure 3 shows the throughput for these analyses with the breakeven point for the dataset demarked. As this particular dataset is fairly small, most of the platforms can easily keep up. At 8x slower that the AI200 results, AWS can just barely keep up (assuming the transfer speed from the microscope to AWS is fast enough), so it could still be run there, but at current rates it is really no bargain. Further, as the effective bandwidth is limited to a paltry 160 MB/sec, AWS shows particularly poor scaling, being completely gated by IO for more than two GPUs, making it inappropriate for real-time analysis with larger image volumes. On the other platforms, scaling is better, with the AI200 delivering near linear performance to 8 GPUs with 50 iterations. Note that while leveraging a single GPU, Microvolution was about 200 times faster than traditional methods, so the combined platform here is nearly 1,600 times faster.

Dataset	Source	Modality	Total size (GB)	Total volumes	Size per volume (MB)	Acquisition time per frame (ms)	Volume dimensions
LLS1	Adv. Neuroimaging Laboratory	LLSM	40	540	74	10	512x512x151
LLS2	U. Queensland	LLSM + deskew	360	2494	144	10	768x768x131
LLS3	U. Queensland	LLSM + deskew	93	594	157	20	768x768x141
LLS4	U. Queensland	LLSM + deskew	174	553	315	20	1024x1024x161
SD	U. Queensland	Spinning disk	266	200	1330	10	2048x2048x162

Table 1: Description of datasets used for benchmarks

The true test of real-time image analysis comes with larger datasets, which naturally take longer to process. The next set of benchmarks use the datasets described in Table 1. For comparison, the dataset from the previous section is also in the table, labelled LLS1. Then, there are three more datasets labelled LLS2, LLS3 and LLS4, generated on 3i Lattice LightSheet microscopes. In addition to deconvolution, these also require deskewing, so the deconvolution input data size is about 50% larger than the raw data size indicated in Table 1. The final, labelled SD, is from a dual camera (each four megapixel) Andor Dragonfly spinning-disk microscope that generates data at about 1.6GB/s. The dataset has 200 volumes of large 16-bit images for 1.3GB of data per volume or about ¼ TB for the whole dataset.

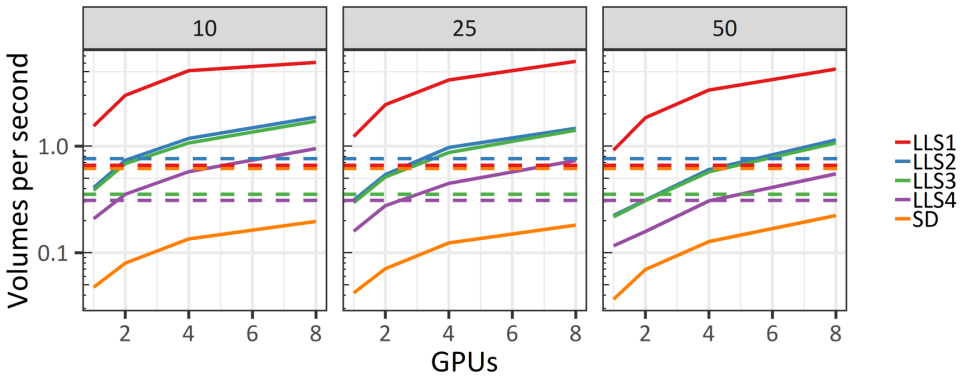

Figure 4: Throughput of Microvolution's deconvolution algorithm for increasing numbers of GPUs on the DGX-1 + AI200. Breakeven points for real-time analysis are shown by the dashed lines. Note the y-axis is in log scale to fit all benchmarks in one chart.

The performance results for deconvolving these datasets using Microvolution on a DGX-1 and AI200 are shown in Figure 4. For consideration of real-time processing, the break-even point for each dataset is also plotted. As shown in the previous benchmark, LLS1 is processed well ahead of its threshold. In general, the other LLS results also outperform their breakeven points. In the worst case, for LLS4 at 50 iterations, it requires four or more GPUs to keep up with the instrument. Across the datasets, it is clear that runtimes scale with the volume size. For instance, LLS2 and LLS3 have about the same size per volume, and they take about the same amount of time to process. The only dataset that presents any challenge from a real-time perspective is SD, in part because the very large volumes, at 1.3GB, simply require a lot of processing. As well, though it may be due to setting an artificially high breakeven bar, as discussed earlier, since the data likely was not actually acquired at the maximum frame rate. Ironically, although 10 millisecond exposure places a high-bar on real-time deconvolution, scanning at these rates is really is only tenable to begin with because deconvolution exists, due to the fact that 10 millisecond (or even 20 millisecond) exposure is so fast that it has a very low signal-to-noise ratio. Thus, the tested platform critically enables the increased temporal resolution that these instruments offer; efficient detection allows for imaging of dim and fast phenomenon that were not previously possible to resolve.

Conclusion

Rapid advancements in microscopy are revolutionizing biological research in areas such as neuroscience, developmental biology, cancer research and biomedical engineering. The proposed technology platform clearly demonstrates significant experimental, technical and economic benefits for microscopy workflows. By integrating a DDN AI200 shared parallel storage appliance, a NVIDIA DGX-1 server and the Microvolution advanced processing software, deconvolution and deskewing operations are massively accelerated. The solution delivers enhanced images 1,600X faster than with traditional microscopy workflows. It also provides results 8X faster than cloud-based infrastructure.

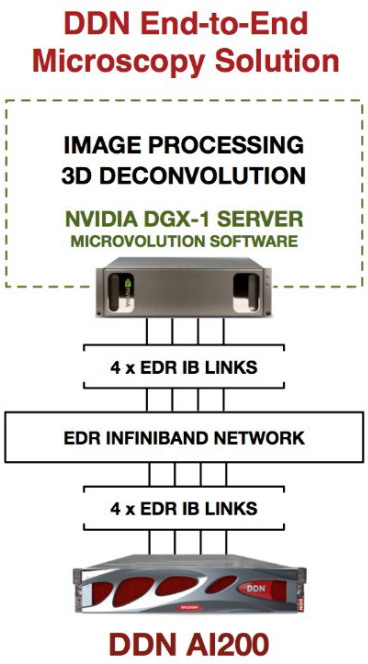

Figure 5: An end-to-end solution including for real-time deconvolution of even the largest microscopy volumes includes Microvolution software, a DDN AI200 all-flash system, a NVIDIA DGX-1 server for GPU computing, and a high-speed, low-latency, RDMA-enabled EDR InfiniBand network.

Real-time analysis and imaging provide unprecedented flexibility for experiments. They enable rapid modification of the capture and sample conditions to achieve best possible results. Live rendering of experiments eliminates ambiguities on exposure. This ensures that images are captured with the optimal amount of light for highest signal fidelity and minimal noise. GPU-accelerated deconvolution enables researchers and institutions to achieve higher yield from existing instruments and further enhances the imaging capabilities of newer instruments such as Lattice LightSheet microscopes.

The proposed technology platform enables more efficient utilization of shared instruments, provides unlimited data capture from multiple instruments and simplifies workflows by eliminating the need for time-consuming, error-inducing data management tasks. It can be configured for a small initial footprint, and scale easily in capacity, performance and capability to match evolving workflow requirements and future investments in instruments. It also includes robust data protection and access control mechanisms to meet data governance and regulatory compliance requirements.

Overall, the proposed technology platform helps researchers achieve the full potential of leading-edge instruments for faster, better results from microscopy workflows.

ABOUT DDN®

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. DDN has designed, developed, deployed, and optimized systems, software, and solutions that enable enterprises, service providers, research facilities, and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. For more information, visit our website www.ddn.com or call 1-800-837-2298.

SALES@DDN.COM
+1.800.837.2298

v1 (11/18)