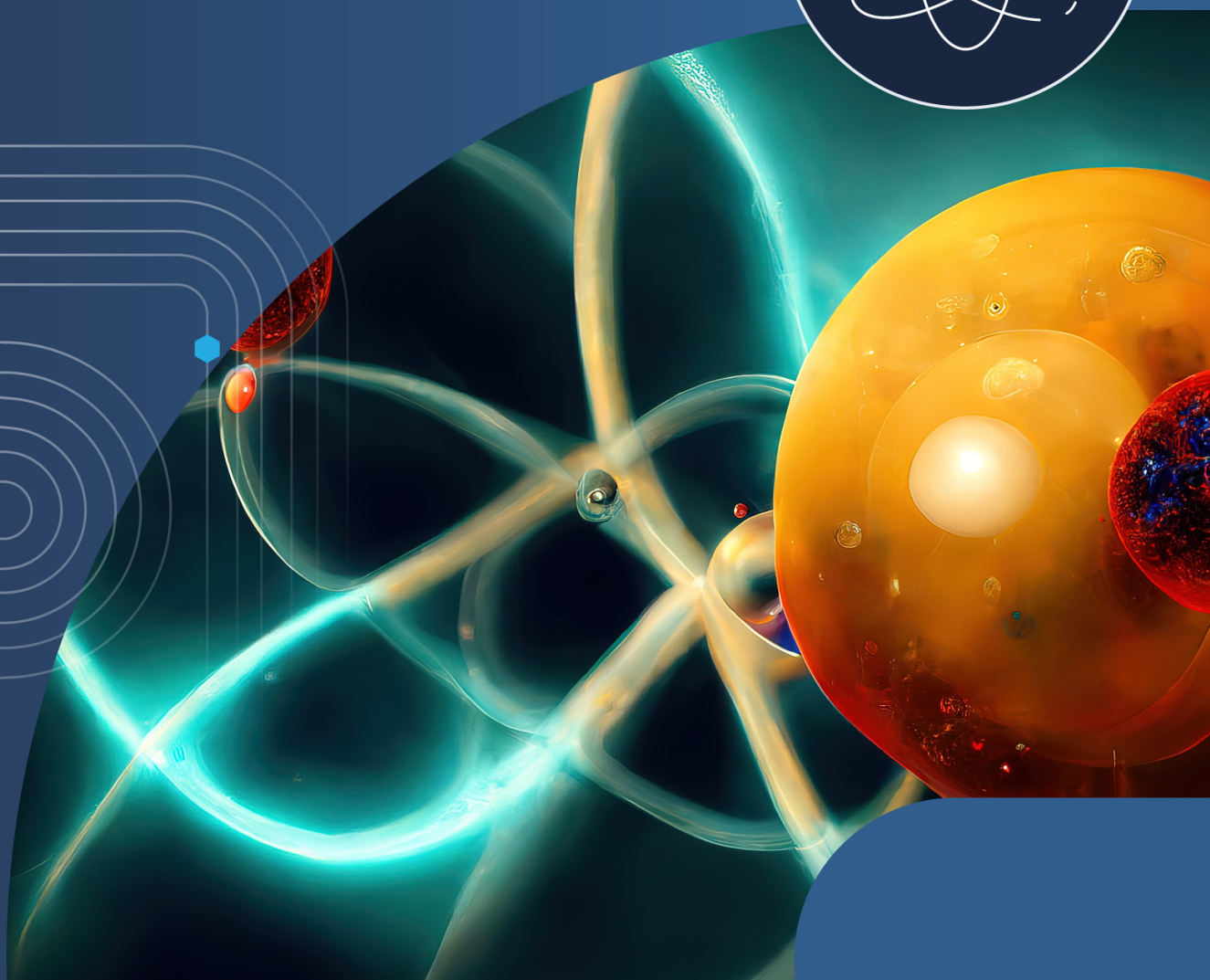
SambaNova SYSTEMS

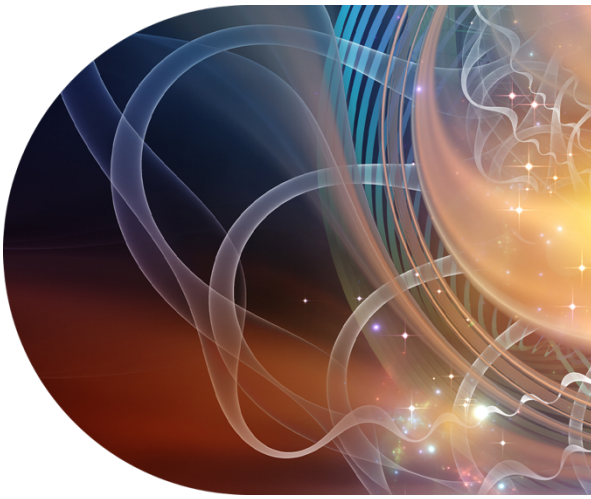
SOLUTION BRIEF

SambaNova DataScale® for Science

Accelerate AI for Science workloads to make more discoveries faster



World record performance to power world changing research



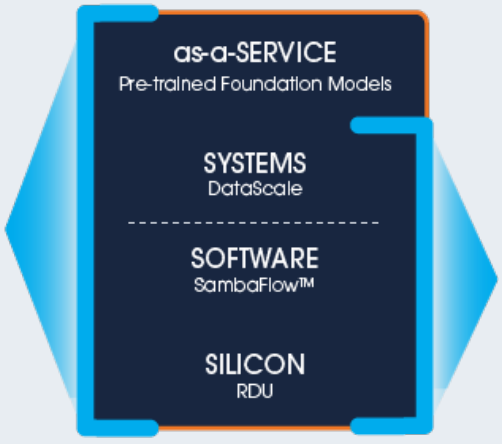
The world has entered a new era of AI, powered by foundation models and state-of-the-art deep learning. Foundation models make up the next generation of large scale models which are transforming what is possible with AI and they are enabling organizations to achieve new and exciting discoveries across numerous areas of research, including true resolution medical image analysis, large language models for science, and multi-physics simulation workloads. While the potential for using foundation models and deep learning to accelerate these discoveries is immense, these state-of-the-art models come with their own training and management challenges.

SambaNova Suite delivers value and innovation across the full AI stack - hardware, software, systems, and even pre-trained models – to enable research organizations to achieve a performance advantage over GPU based systems on the most challenging foundation models and deep learning workloads. For research organizations, this means more experiments and more discoveries with the potential to change the world.

The SambaNova Platform

Innovation at every layer of the stack

SambaNova Suite



DataScale

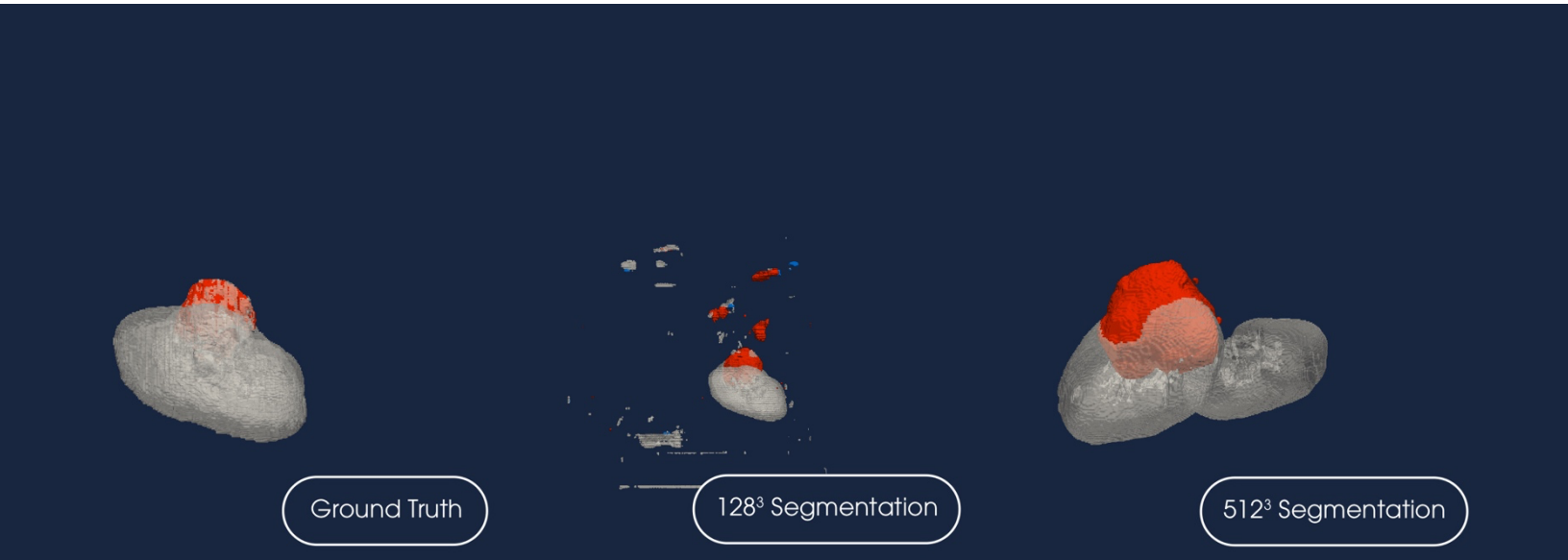
512^3 resolution medical image analysis

High resolution 3D computer vision has the potential to revolutionize the diagnosis of serious diseases while simultaneously accelerating and improving patient treatment and care. 512^3 resolution images both deliver a more accurate diagnosis prediction and enable medical professionals to view and assess a more complete picture of a patient's health than was previously possible.

The memory constraints inherent to GPU-based AI infrastructure limits the resolution of 3D medical images that can be analyzed with those systems, requiring tiling or down sampling, which reduces resolution, increases complexity and significantly lowers accuracy.

The SambaNova platform delivers industry leading performance and, with 12.8x more memory than GPUs, can analyze full 512^3 3D images and beyond without the need for complex tiling or decreasing image resolution. This enables medical professionals to make more accurate diagnosis in less time to have a dramatic impact on the quality and patient care and outcomes.

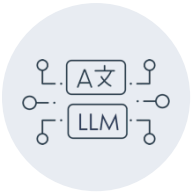
The improved accuracy from higher resolution computer vision models reduces both missed diagnosis, as well as false positive diagnosis, allowing medical experts to focus their time on patients who need treatment the most.



Large Language Models (LLMs) for Science

Large language models (LLMs) can unlock insights in unstructured data with human level accuracy and solve dozens of language tasks with a single model. Beyond traditional language tasks, these models have demonstrated potential in scientific domains by becoming 'experts' in specific topics, such as genomic data for Covid-19 research.

Deep learning is revolutionizing research by unlocking insights trapped in unstructured data, such as research reports, studies, scientific journals, and more. However, training custom LLMs on domain specific scientific topics is a complex and technically challenging process on GPU-based systems.



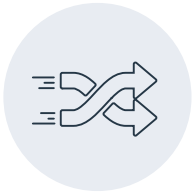
SambaNova's Dataflow architecture and large memory enable research organizations to train custom LLMs with a scientific corpus that can understand domain specific content, without the massive parallelization needed with GPU-based systems.

Delivering faster performance over legacy GPU-based systems, researchers can take advantage of the power of the SambaNova platform to make ground breaking discoveries faster than before, while leveraging the power of LLMs to further validate their findings before publication.

Deep learning accelerated surrogate models

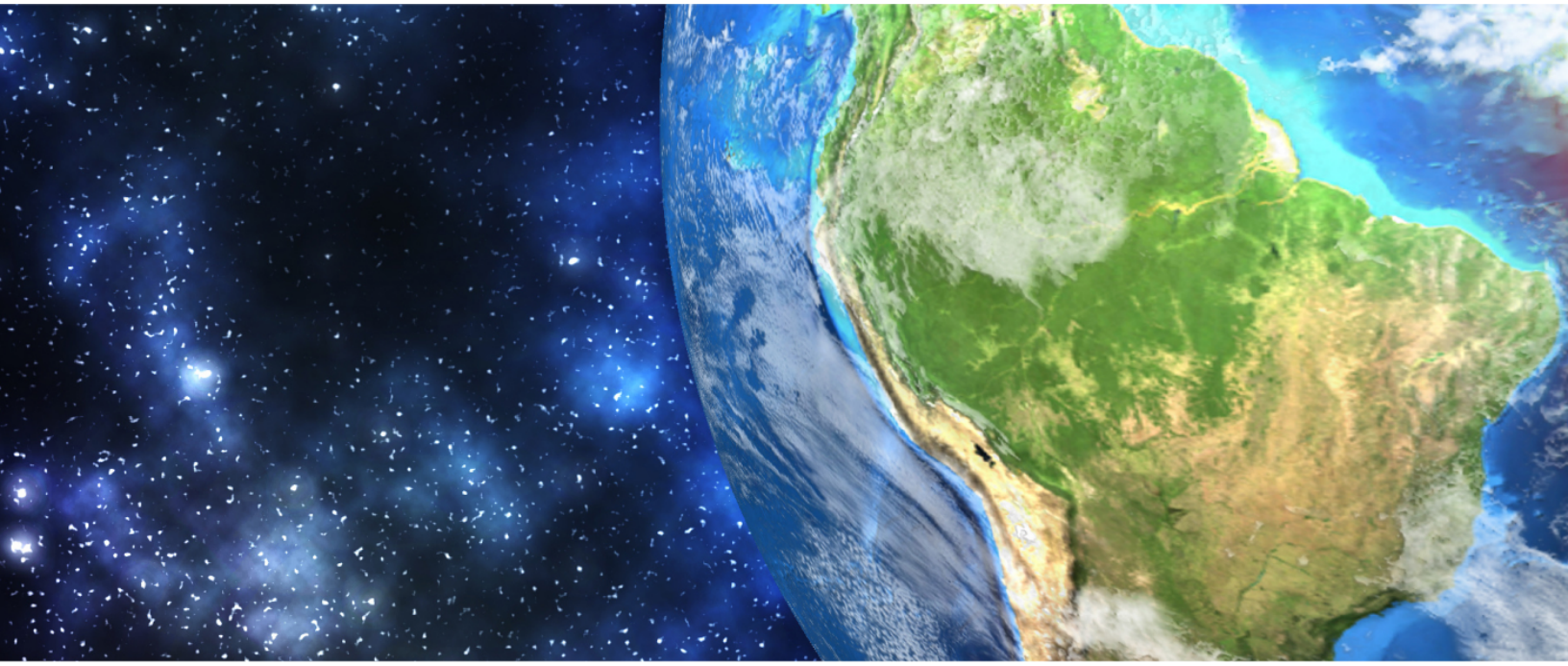
Surrogate models are deep learning models that replace one or more components of larger multi-physics simulation workloads, such as computational fluid dynamics or weather forecasting. However, the GPU/CPU-based architectures used to compute these simulation workloads struggle to deliver performance on sparse, detailed deep learning models.

Research organizations must either operate both types of workloads on the same architecture, or tradeoff gained speed with added latency from shifting data between different systems optimized for each of the different workloads.



The SambaNova platform significantly improves the performance of sparse, detailed deep learning models compared to GPU-based architectures, even taking into consideration the additional latency requirements of two different systems.

Utilizing the SambaNova platform, researchers can efficiently run deep learning workloads for advanced scientific workloads, such as computational fluid dynamics simulations, weather forecasting, and physics informed neural networks



SambaNova SYSTEMS

Next steps

If you would like to learn more about how organizations are preparing for the future, visit SambaNova.ai

Learn more at SambaNova.ai

linkedin.com/company/sambanova

@SambaNovaAI

info@sambanova.ai

Customers turn to SambaNova to quickly deploy state-of-the-art generative AI capabilities within the enterprise. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing.

Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others.