HPC User Forum Update

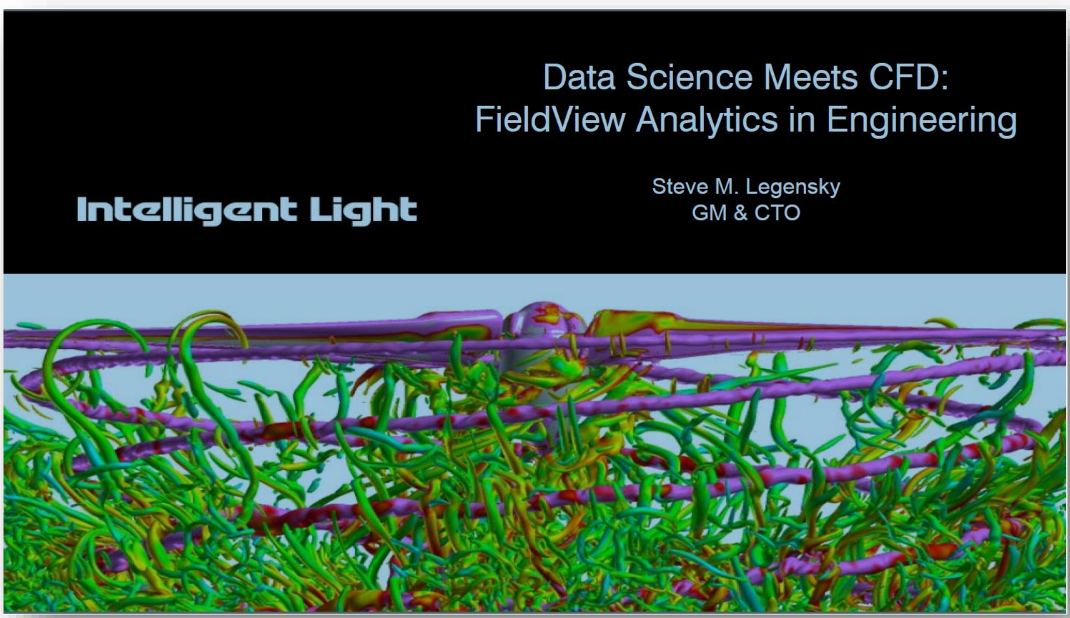
Data Science Meets CFD, HPC User Forum Dearborn Michigan

Bob Sorensen
March 2019

IN THIS UPDATE

The HPC User Forum was established in 1999 to promote the health of the global HPC industry and address issues of common concern to users. In September 2018, Hyperion Research hosted its 70th User Forum in Dearborn, Michigan. This update summarizes a presentation from that meeting in the session on Automated Driving Systems, Automotive Engineering, Sensor Networks, entitled *Data Science meets CFD,* given by Steve Legensky, General Manager and CTO, Intelligent Light.

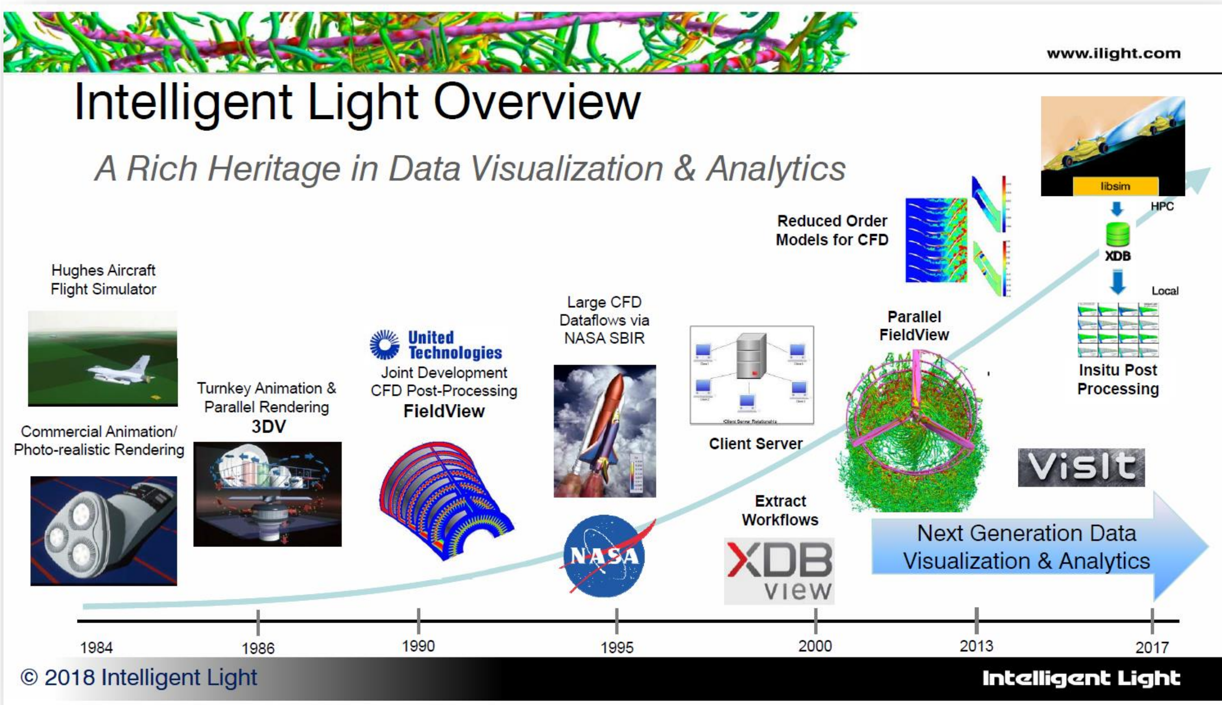
Legensky provided an overview of the current state of CFD and why it is as important now as ever. Legensky offered some solutions to the problems of increasing data storage requirements for more precise simulations using extracts, physics-based subsets of the entire calculation, to offer a more workable solution. The talk also covered the potential for bringing CFD techniques to the field of machine learning-based design.



PRESENTATION: DATA SCIENCE MEETS CFD, STEVE LEGENSKY, GENERAL MANAGER AND CTO, INTELLIGENT LIGHT

Legensky began his talk by giving a background on Intelligent Light, which has a strong 34-year history in developing software tools for computational fluid dynamics (CFD) post-processing, as well as capabilities for next-generation data visualization and analytics. Their main product is FieldView, which is a worldwide tool for CFD post-processing. Over the years Intelligent Light (IL) has built in additional FieldView data management and analysis tools. When IL started to work with the DOE they got involved with more open-source projects, and IL ultimately commercialized a version of the DOE's Visit software, which is a highly scalable visualization and data analysis tool. IL's most recent release was a combination of those two products, which they call HPCFieldView.

FIGURE 1

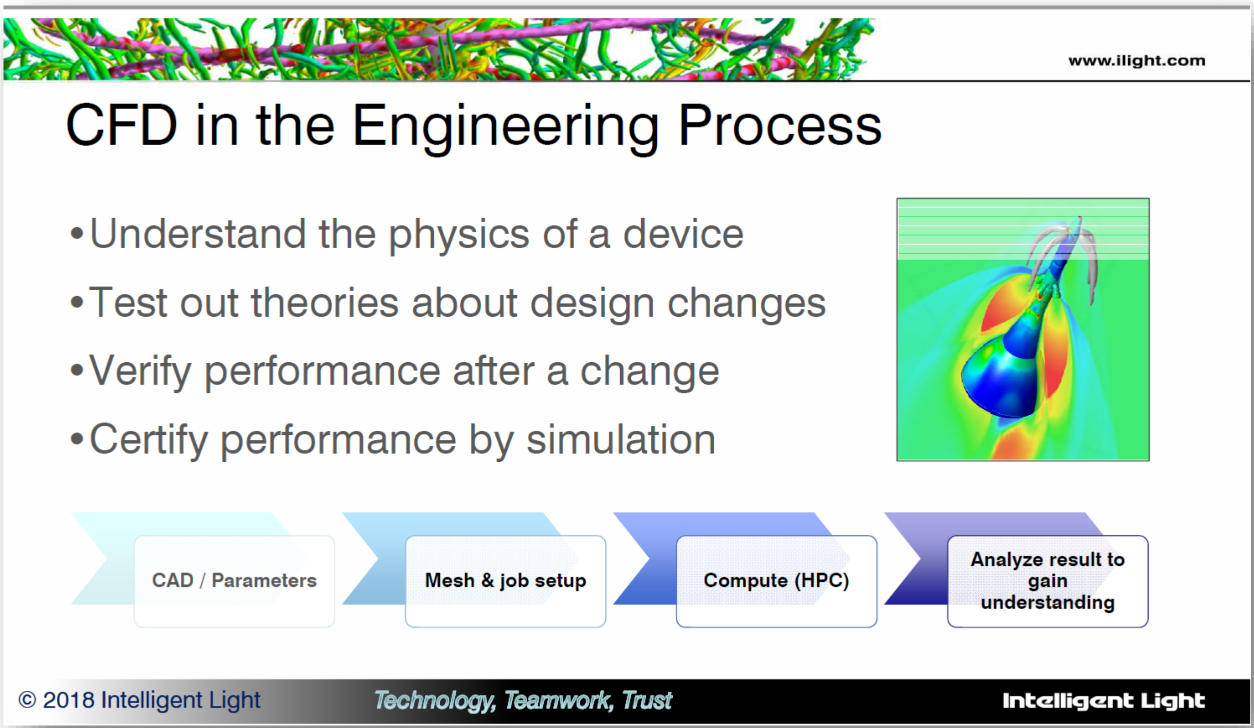


Source: www.light.com and Hyperion Research, 2019

Legensky provided an overview of the current state of CFD and why it is as important now as it ever has been. He noted that to improve the design of any complex device, the physics of it must be well understood, and that any theories concerning changes in design to increase performance must be tested against those physical governing equations. For example, if a Formula One car is being designed and the front wing is being changed, designers need to know how that change will affect performance going forward.

Legensky noted that nobody wants to build a test airplane if they don't have to. Indeed, the Holy Grail of design is being able to verify performance entirely by simulation. There are strong incentives to reduce the billions of dollars spent on testing airplanes to certify them and the testing of cars by simulation rather than doing so much physical testing. However, for all these pieces to come together the simulation must be believable, and the uncertainty must be well understood. The general steps to accomplish this start with some CAD and geometry to create a discretized version of the simulation that can then be computed on an HPC resource in an effective way to allow those results to be analyzed.

FIGURE 2



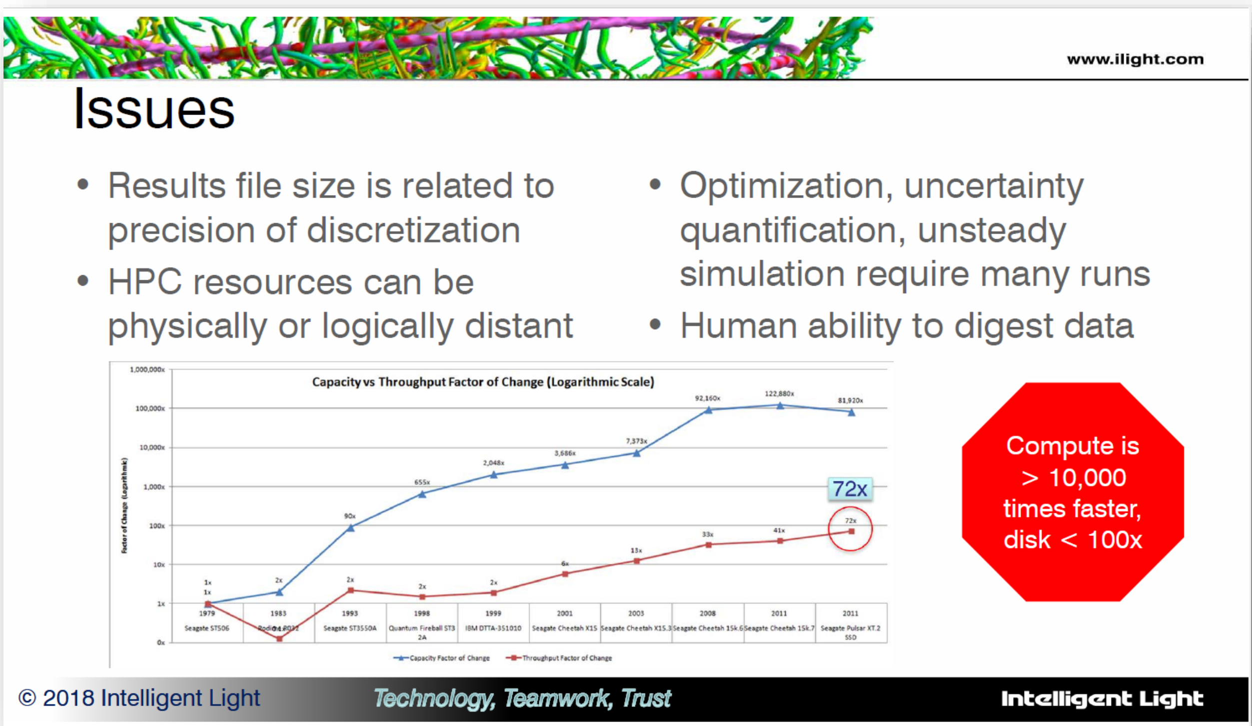
Source: www.light.com and Hyperion Research, 2019

Legensky noted that in looking at the big picture there is a cycle of decision regarding changes in design. After the geometry is created and evaluated, changes are down-selected, evaluated again (maybe it's a wind tunnel or a test), then something is built and tested. All this time there is a conceptual system model: "If the wings are fatter this will happen… If the tires are wider this will happen…" He noted that one of the things that they are really interested in is how machine learning could impact this conceptual system model so that one can use that to guide which geometry choices are going to be made ahead of time. Legensky offered that this could be an interesting opportunity to leverage machine learning for engineering.

Legensky outlined the various issues facing those involved in leading-edge CFD, noting that it's not a perfect world. For example, the results' file size is related to precision of the discretization: the more precision in the simulation, the larger the required files sizes -- and five to ten Gbytes is an average, single, steady-state run file. In addition, HPC resources can be physically or logically distant, which means one can't just reach out and touch the data, it must be copied or moved or something else altogether. It is all exacerbated by the fact that optimization, uncertainty quantification, and unsteady calculation require many files or many runs. He noted, for example, that it's not unusual for Boeing to run a thousand wing calculations as part of an optimization, or ten thousand calculations to position a pylon on a wing, raising issues of what can be done with ten thousand calculations and how does one digest that data.

Complicating that issue is the fact that storage, even though it can be bought, is not getting any faster. This is an issue at the level of raw disk performance, not even considering the file system on top of the hardware. Compute, over the period from 1980 to now, is ten thousand times faster on average in a company while disk has gotten faster by less than 100x. There is a convergence of problems, and the balance is going backwards. The faster that HPC can generate data, the worse it is to deal with that data.

FIGURE 3

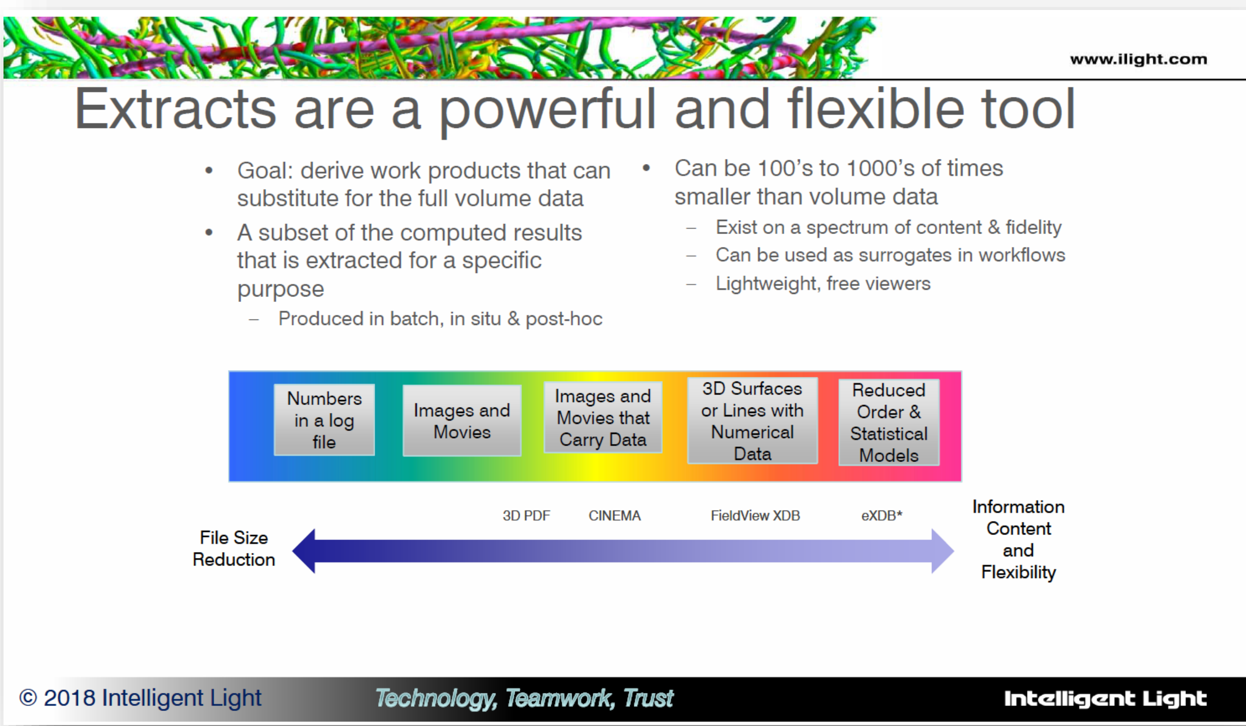


Source: www.light.com and Hyperion Research, 2019

Legensky offered some solutions to these problems noting that extracts, physics-based subsets of the entire calculation, offer a workable solution to the compute/storage problems. On the HPC side, when a batch job is run, users can extract only the critical parts of the problem that are meaningful. For example, in a gas turbine simulation there are mixing planes and the blade surfaces and the hub and the shroud, but that represents a very small amount of the data. Using an extract process, the smaller extracted database has full physical fidelity of the CFD mesh and the results, and one can store all the variables there along with the mesh geometry. Then a user can bring that over and make movies or perform additional analysis.

Legensky noted that this process, where critical data in the memory space is extracted by the solver code during execution, has been around for quite a while. They've done this for FLUENT, and for people who have in-house codes at gas turbine manufacturers, and aerospace entities such as NASA and JAXA. The idea is that the number of volume files that must be saved is reduced. The disk problem can be circumvented by simply not creating a mess in the first place. These extracts can be a thousand times smaller than a volume file.

FIGURE 4

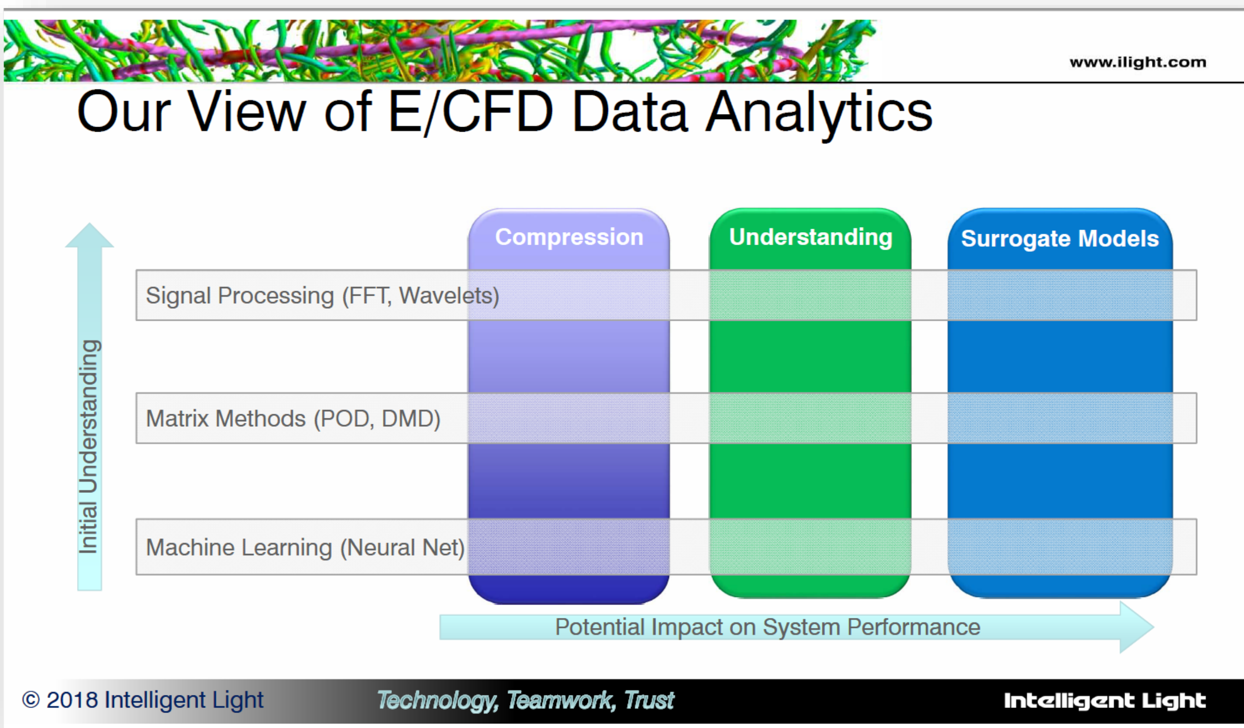


Source: www.light.com Hyperion Research, 2019

Finally, Legensky covered how related capabilities in data science could be helped by some of the FCD techniques. Referring to Figure 5, he indicated that machine learning (neural net), is on the bottom, matrix methods (such as POD, DMD) are in the middle, signal processing (FFT, wavelets) is at the top in terms of the initial level of understanding. The more that is known about a given problem space, the further up on the matrix one can go. As a neural network is being trained, it's being fed a bunch of images with the desired result of ten possible outcomes. It calculates gradients, builds up matrices, and there is linear algebra in there, too. It does the evaluation and ends up with a matrix of weights and a vector of biases and the machine is trained.

What Legensky is interested in is using matrix methods both for surrogate models and for understanding. IL is looking at problems in a completely different way: they're not thinking of a cutting plane and an isosurface of pressure. Instead, they are thinking of persistent, dominant, high-energy features and how they impact each other in the flow field. What they bring to the table in terms of data science is understanding the math part of CFD and fluid mechanics, and getting that into matrix systems. He noted, however, that this process is neither easy nor trivial.

FIGURE 5



Source: www.light.com and Hyperion Research, 2019

For more information or to view this and other presentations given at HPC User Forums dating back to 2008, visit www.hpcuserforum.com.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
