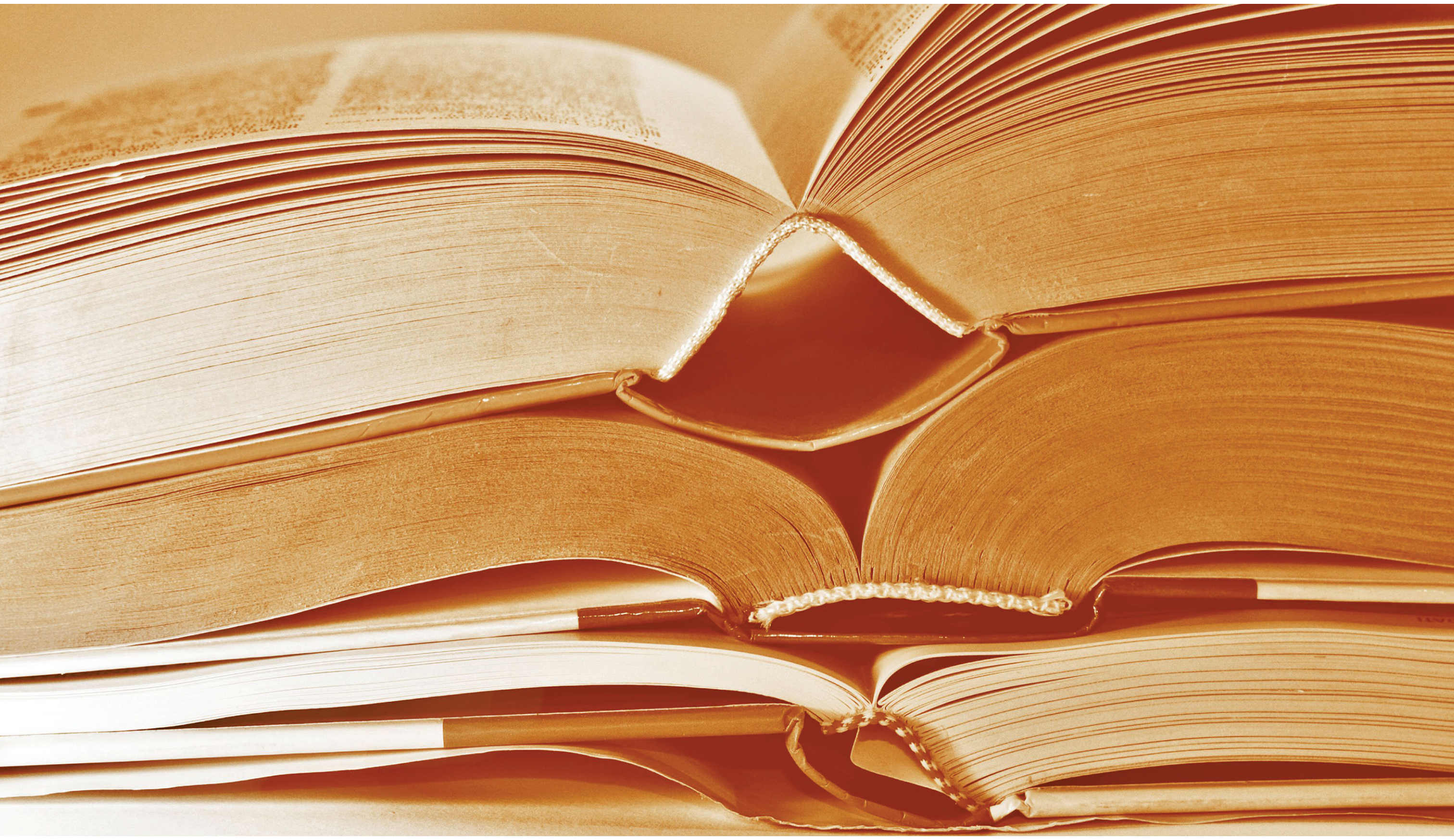
Editors: Joao Comba, UFRGS, comba@inf.ufrgs.br, and Daniel Weiskopf, weiskopf@visus.uni-stuttgart.de



Audience-Targeted Design Considerations for Effective Scientific Storytelling

Franz Sauer, Tyson Neuroth, Jacqueline Chu, and Kwan-Liu Ma
University of California, Davis

Scientific storytelling involves constructing a coherent narrative to effectively communicate a topic in science. Just like a story,[1] an effective visualization has a setting, characters, and a plot. The setting is the relevant background information the viewer needs to know to put the visualization into context. The characters are the visual representations that capture the viewer's attention, and the way they interact and evolve over time represents the visualization's plot. Whether creating a traditional story or a visualization using storytelling elements, creators must carefully consider its intended purpose and target audience.

Effective scientific storytelling can serve a variety of purposes. First, it can help uncover the subtle and complex patterns found in scientific datasets. By mapping values and parameters to visual representations, scientists can gain an intuitive understanding of their system of study. In other words, a visualization can tell a story to researchers about how their data is behaving. Presenting this data in different ways tells the story from different perspectives, and being able to study a system from multiple viewpoints is necessary to gain a complete understanding of the underlying phenomena.

Second, scientific storytelling can be used as a means of presenting new findings to the scientific community. In this case, domain scientists have already uncovered new information from their data and wish to share it with their peers. Consequently, an effective presentation must be able to intuitively describe their results to a group of educated scientists who might not be experts in the specific area of study. Such a visualization must place the discovery in context

  Copublished by the IEEE CS and the AIP November/December 2016

with background information so that the viewer can understand and appreciate its importance.

Finally, scientific storytelling is an effective way of educating the general public, including children. Unlike the previous two cases, in which results are presented to scientists, many of the characteristics of this type of visualization focus on using simple visuals and constant interactivity to maintain viewer interest and excitement. In many cases, the general public (especially children) lose interest in a topic if it's too difficult to understand or presented in a mundane fashion. As a result, the medium, such as an interactive museum exhibit, must be able to educate its users through active engagement that nurtures their natural curiosity.

In this article, we present three different projects that each use scientific storytelling to achieve one of the above described goals: a visualization of particle accelerator data for domain scientists, a presentation of fusion science findings for general scientists, and an interactive exhibit of plankton populations for the public. Although each example uses techniques from scientific storytelling, the design considerations we use depend heavily on both the area of study and the target audience. Figure 1 shows an overview of the process involved in developing an effective visualization. As we discuss in the following examples, certain visualization types will rely more heavily on different steps in the workflow.

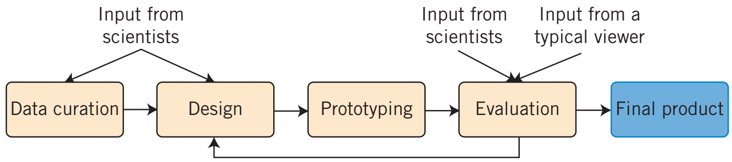


Figure 1. A flowchart describing the major steps involved in designing an effective visualization. In some cases, we have to go through numerous redesigns based on input from scientists or typical viewers.

Visualization of Particle Accelerator Data for Domain Scientists

Our first example is a visualization of particle accelerator data designed specifically for domain scientists. We focus on the steps involved in designing such a visualization and what characteristics can make it effective for scientific discovery. Finally, we describe the challenges we faced working with accelerator data and how we chose to overcome them. The visualization takes the form of a video that we designed and pregenerated. In the future, we plan to investigate how we can use other techniques[2] to create a system that lets scientists generate their own videos based on any desired viewing angles or parameter selections.

Background Information

We teamed up with researchers at the SLAC National Accelerator Laboratory who are studying the effects of dark current in their acceleration devices. They use a simulation called ACE3P to simulate electromagnetic dynamics in cryomodule devices.[3] These devices consist of a set of resonating cavities driven by a radio frequency (RF) antenna. Electromagnetic field fluctuations build on themselves within the cavities and accelerate charged particles to near light speed.

One issue with these devices is the presence of dark current, in which charged particles are emitted from the cavity surfaces and can enter the accelerating beam. The unstable motion of these particles can cause them to become deposited downstream, resulting in a collision with device walls that could cause damage. A complete understanding of multiple aspects of the simulation data is necessary to study this phenomenon. Consequently, we developed a comprehensive visualization of the data, which, through the form of a visual story, can intuitively show scientists how their simulation behaves.

Designing a Visualization for Domain Scientists

One advantage of developing a visualization directly for domain scientists is that they're already experts in their particular field. Not only do we not have to worry about presenting relevant background information, but their interest in the topic matter is guaranteed. We can instead focus on representing the details within the data and tailor the visualization directly to their needs. This is done through multiple discussions about the specific phenomena and types of patterns they're looking for in their data. Our goal is to highlight as many of these patterns as possible throughout the visualization. Maintaining maximum scientific accuracy is of utmost importance because such a visualization could lead to new scientific discoveries. When considering the workflow in Figure 1, creating such a visualization relies most heavily on the data curation and initial design steps with heavy input from the scientists. It then goes through a few redesigns based on evaluations with the scientists.

Figure 2 gives an overview image of our particle accelerator visualization. Because it's important

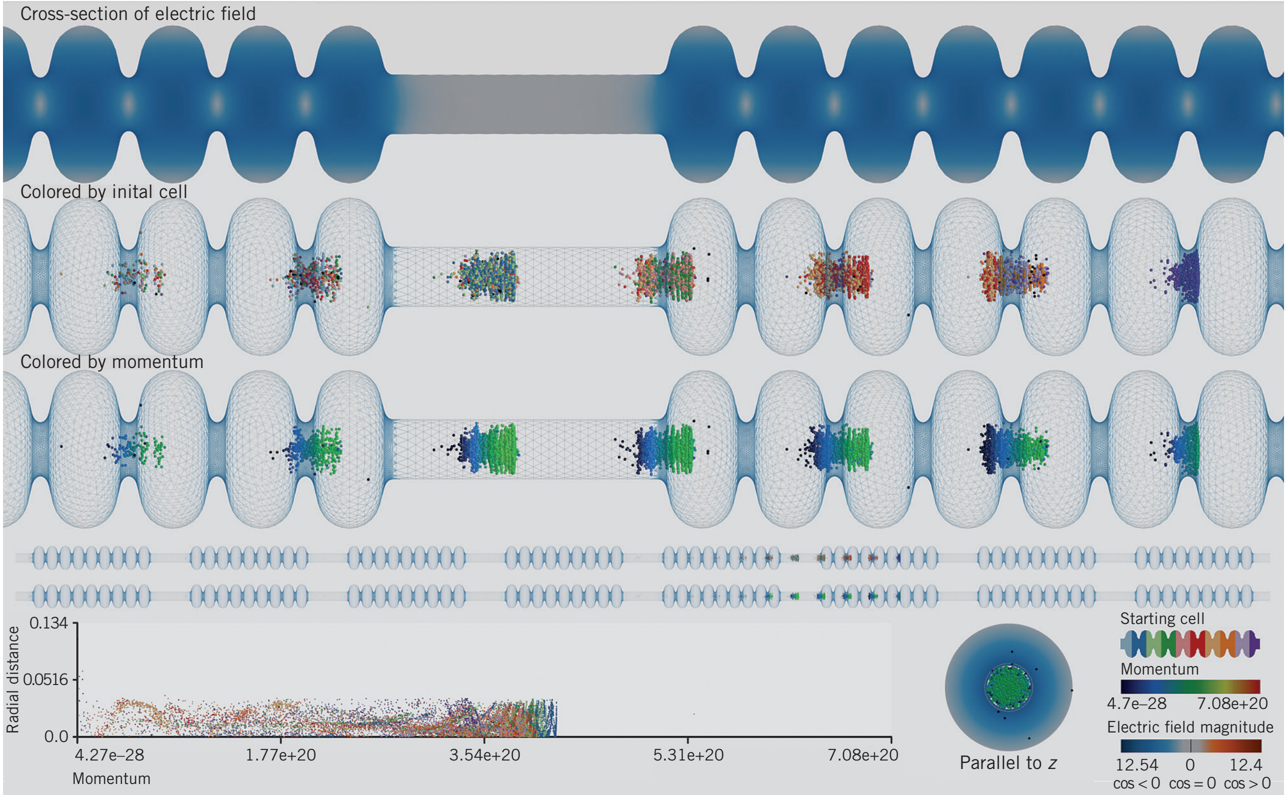


Figure 2. An overview of the particle accelerator visualization. From top to bottom: three zoomed-in views of the cryomodule geometry showing the background electric field strength and simulation particles colored according to their starting location and momentum, two zoomed-out views showing the full device with the same simulation particles, a phase space plot of the particles comparing momentum and radial distance, a view looking down the cryomodule device, and a color legend.

to view multiple aspects of the data simultaneously, we divide up the screen space into various sections, each presenting the data from a different perspective. Each view is carefully designed based on the scientists' needs and helps explain the system's underlying dynamics. Furthermore, animating the visualization allows us to intuitively represent data evolution over time.

Each of the parameters represented in the top views of Figure 2 serves a unique purpose. The electric field is what drives the particle motion, and it's important to see how the particles adjust their motion as the field changes over time. Because the particles are already moving at near light speed, adding energy increases their momentum rather than their velocity. Particles with more momentum can result in more damage to the device. The particle's starting location is also important because the end result of a particle's motion (a collision with the device wall) is based heavily on its initial conditions. Coloring particles this way helps identify such trends.

The bottom views present alternate ways of looking at simulation behavior. We use a phase space plot to show trends between nonspatial particle variables (radial position versus momentum). Figure 3 shows a larger version of this plot as well as the complex wake-like structures that can appear in such a view. Finally, we provide a view looking down the cryomodule device to indicate any potential variations in its rotational symmetry.

Overall, this visualization is characterized by a diverse set of detailed views into the simulation data. Because it's targeted at an audience consisting of expert scientists, the visualization is designed to present as much information as possible. Through repeated viewings of this information-dense "story," scientists can gain multiple insights into their data from just a single visualization.

A video of the visualization is at https://www.youtube.com/watch?v=3_5Vj_R9de8.

Challenges Specific to Particle Accelerator Data

One of the main challenges of working with the accelerator dataset involved handling the elongated simulation domain associated with a cryomodule

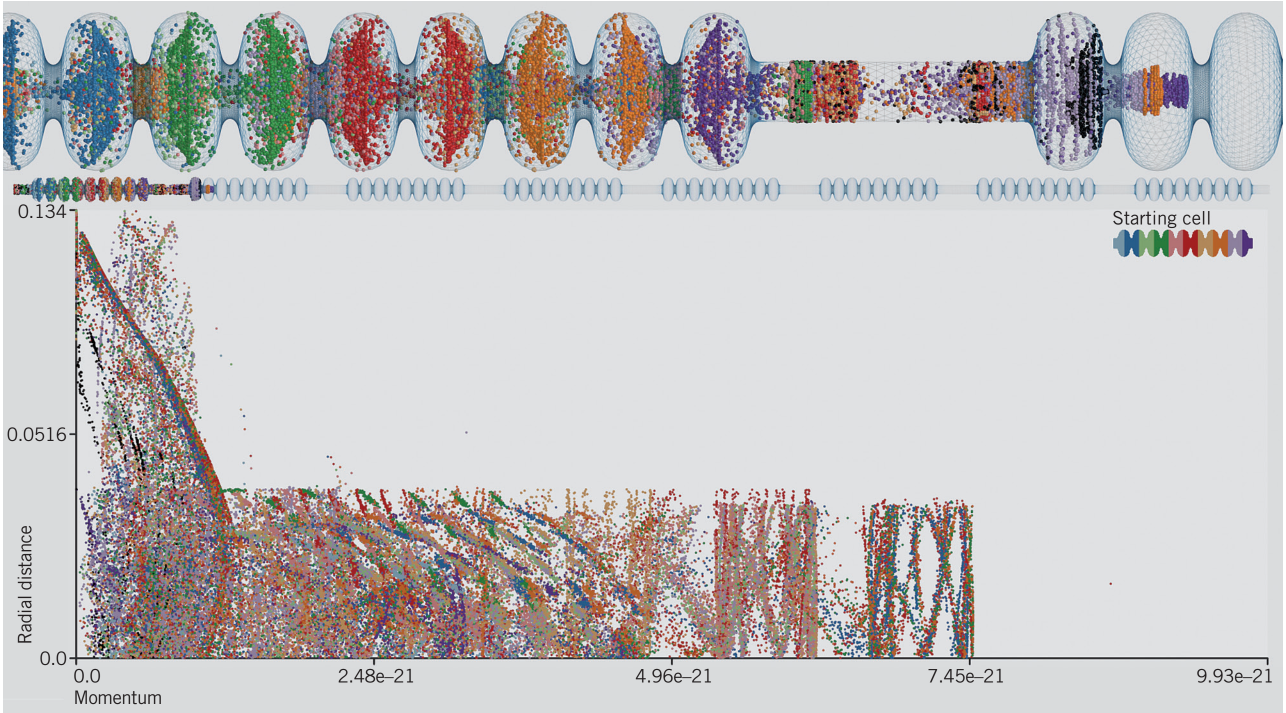


Figure 3. An alternate view showing a larger version of the phase space plot of the particles (radial position versus momentum). Many wave-like fluctuations in the plot can give insights into the underlying dynamics of the particle motions. Particles are colored according to their starting location with spatial context given in the two views above.

device. Because the domain is very long in one dimension, it's difficult to show detailed views of the data while also presenting an overview of the simulation as a whole. Consequently, we chose to use our split screen space to show both zoomed-out contextual views and zoomed-in detailed views of the most important simulation parameters. In this way, a viewer can gain the benefit of both perspectives simultaneously. Furthermore, automatically panning the zoomed views along the cryomodule device to follow the bulk of the simulation particles lets us capture as many details as possible throughout their evolution.

In addition, certain simulation parameters, such as the particle momentum, change by many orders of magnitude throughout the simulation, making it difficult to choose appropriate axis ranges in the phase space plot while still highlighting subtle trends at both the start and end of the simulation. We therefore dynamically scale the axis range at key points throughout the simulation according to values of the majority of the particles. This lets viewers see the minor variations occurring near the start of the run as well as the larger trends that present themselves near the end of it.

Presentation of Fusion Science Findings for General Scientists

Our next example involves a visualization we created to aid fusion researchers in presenting their latest findings to the scientific community. Through the design process, we were also able to help the scientists verify their results and gain additional insights into their discovery through our chosen representations of the data. Although the visualization in this case also takes the form of a video, its design is significantly different from the previous example. We focus here on what characteristics make this type of visualization different and what design choices we made to effectively demonstrate fusion results to nonexpert viewers. We also describe the types of challenges involved in visualizing fusion data.

Background Information

We worked closely with scientists at the Princeton Plasma Physics Laboratory (PPPL) who asked us to help present their new discovery at the 2015 Supercomputing conference. Their findings were generated by XGC1, a massively parallel fusion simulation for tokamak devices.[4] These devices are a promising way of electromagnetically confining the high temperature plasmas necessary for harnessing

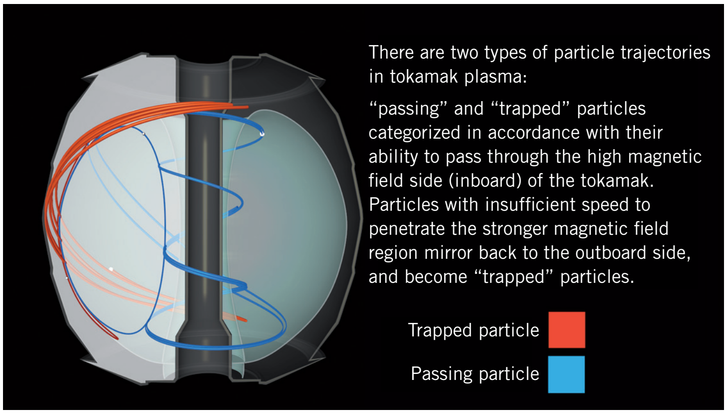


Figure 4. An image showing the difference between a trapped trajectory (orange) and a passing trajectory (blue) within a tokamak device. A projection of each of the two trajectories is also placed onto a white plane on the left side of the device. While the trapped trajectory reverses directions several times and is confined to the outer portion of the device, the passing trajectory is free to continuously circle throughout all parts of the tokamak.

fusion energy on Earth. However, more research is necessary before these devices can become practical enough as a reliable source of energy.

The research team at PPPL focused on studying the bootstrap current, which is caused by a strong flow of charged particles throughout the tokamak device. These charged particles are often classified as either trapped or passing based on the shape of their orbits through the domain. In the past, it was thought that only passing particles can carry the bootstrap current. However, experimental results have always shown a strong bootstrap current in the outer/edge region of tokamaks, which contains very few passing particles. Recent simulation results have shown that, in the edge region, most of the current is actually carried by particles exhibiting a trapped trajectory.[5] This new understanding of the bootstrap current in tokamak devices can be used by experimentalists to analyze and predict edge confinement and stability.

New Findings for General Scientists

Unlike the previous example, the scientific discovery has already been made, allowing us to focus the visualization on a relevant subset of data and choose the most appropriate visual representations. However, a general scientific audience might not have the same level of expertise and will require some background information to put the discovery into context. Consequently, a large portion of our visualization is dedicated to explaining various aspects of the system of study before presenting the actual discovery. Referring back to Figure 1, this type of visualization relies first on the initial design step based on what aspects of the discovery the scientists wish to portray. This then involves a handful of redesigns based on evaluations from the scientists as well as some typical viewers who are less familiar with fusion science.

The visualization is split up into two main components. First, background information is given to familiarize viewers with details about the area of study. Instead of overwhelming them with an in-depth introduction to the field, we instead choose to quickly explain only the most relevant pieces of information necessary to understand the discovery. This is done through visuals and explanatory text. Second, the discovery itself is explained along with a simple visualization of the relevant simulation data subset that revealed it. Once again, explanatory text is used to guide the viewer and explain the importance of the discovery. In this way, the nonexpert viewers themselves can identify the same trend in the data that was discovered by the scientists. This helps reinforce their understanding of the new findings.

One of the fundamental pieces of information necessary for understanding this discovery is the difference between the trapped and passing trajectories. Because the difference depends heavily on spatial motion, we visually explain this distinction by animating the motion of a trapped and passing particle throughout the tokamak geometry, as shown in Figure 4. Accompanying text helps to guide the viewer's attention through the visual animation and further explains this distinction. Furthermore, this representation helps to show viewers why the outer regions of tokamak devices are made up primarily of trapped particles.

To explain the discovery, we present the data subset that the scientists used to make the discovery in a simplified form so that it can be interpreted by the viewer as shown in Figure 5. In this case, we use a heatmap of the bootstrap current on two cutting planes in the tokamak geometry. One plane represents the current produced by the trapped particles, while the other represents the current produced by the passing particles. The viewer can clearly see that the current in the outer/edge region is much stronger for the trapped case as they make up most of the bootstrap current. We also supplement this with a line plot showing the current as a function of radius for each of the trajectory types as well as their combined total current. Once again, explanatory text helps guide the viewer and explains the significance of this discovery.

Unlike the previous example, this visualization is characterized by its use of simple but effective visuals

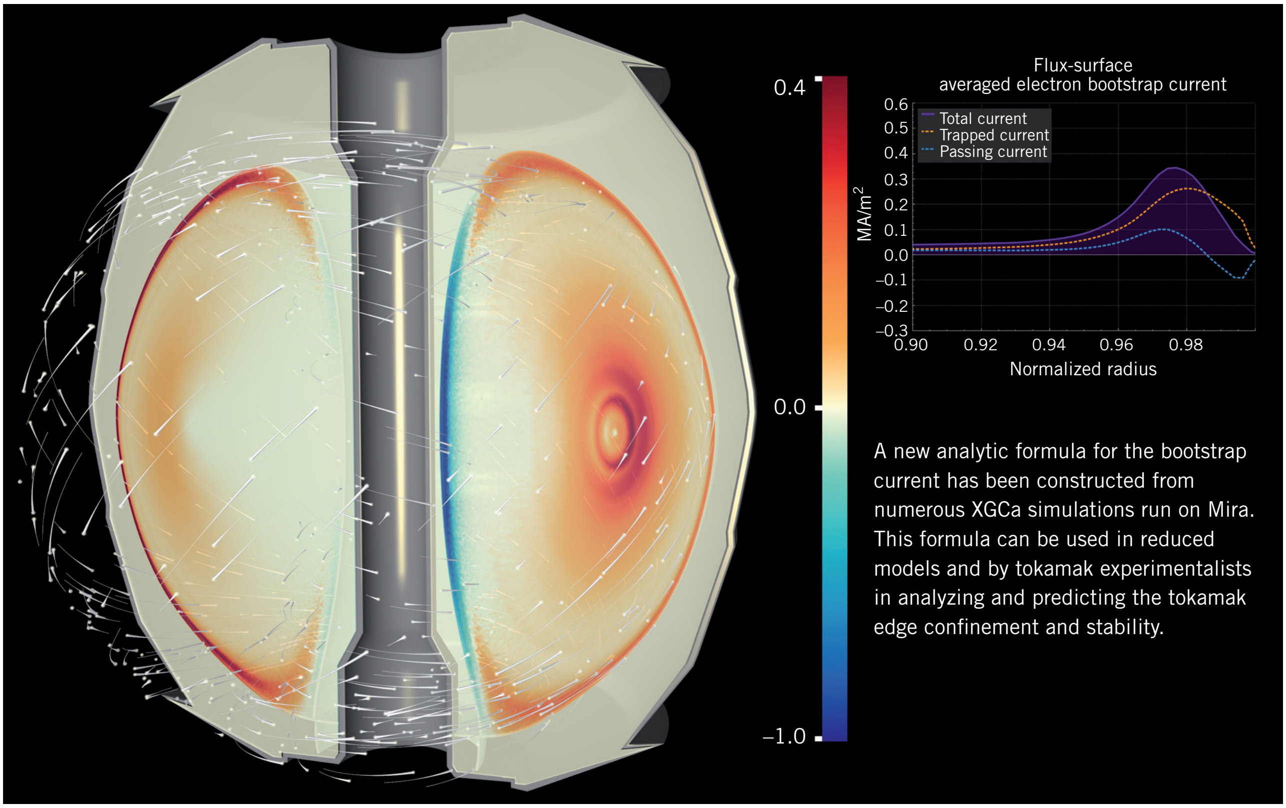


Figure 5. A view of the bootstrap current within the tokamak device projected onto two cutting planes. The plane on the left shows the current caused by the trapped particles, while the plane on the right shows the current caused by the passing particles. From the slices, it's clear that the bootstrap current in the outer/edge regions of the tokamak is much higher for the trapped particles, indicating that they make up most of this current. A line plot on the top right section shows the current as a function of radius for each of the trajectory types as well as their combined total current.

to explain a scientific discovery and relevant background information. This is because it's targeted at explaining a discovery to nonexpert scientists. Instead of being information dense, it focuses on telling a "story" in a more traditional sense by guiding the viewer through a series of visual examples.

A video of the visualization is at https://www.youtube.com/watch?v=a1ibTeOoyeE.

Challenges Specific to Fusion Data

Just like the cryomodule device, the tokamak device's simulation geometry presents a major challenge. In this case, it's because the device's donut-like shape is inherently prone to occlusion. Unless it's viewed directly from above or below, one side of the device will block the view of the other side. However, we can exploit the device's circular symmetry by cutting away certain portions and projecting patterns onto cutting planes. Although parts of the domain are no longer shown, the major trends seen on the projections represent the entire device volume.

Another challenge involves the data's low temporal resolution. Because fusion simulations tend to incorporate massive amounts of particles, data I/O becomes a heavy strain. Consequently, scientists are able to save information for only a limited number of time steps. This, coupled with the fact that the particles tend to move at high speeds in a loop-like fashion, makes it difficult to construct accurate and visually understandable trajectories. Simply using a linear interpolation scheme that connects the temporally distant points with lines results in a self-intersecting mess with sharp angles. We therefore implemented an alternative interpolation method that followed the curvature of the electromagnetic fields in the device. This smoothed out the trajectories and provided a more accurate representation of particle motion.

Interactive Exhibit of Plankton Populations for the Public

Our last example describes an interactive museum exhibit that uses visualization as a tool for visitors

Figure 6. Museum visitors interacting with the plankton visualization. The custom-built table houses a Dell OptiPlex 990 and a 55-inch MultiTaction Cell. It's placed in an open area so that visitors can access and interact with the visualization from all sides. Labels placed around the table help guide and inform the viewer about exhibit content.

to explore large datasets.[6] This visualization is designed for the general public, especially for children. We describe the process of developing such an exhibit from start to finish and the key characteristics that separate it from the previous two examples. Finally, we describe the unique challenges involved with using scientific storytelling to present information to this type of audience.

Background Information

This exhibit was developed through a close collaboration with the Exploratorium, a "hands-on" museum located in San Francisco, and the Center for Microbial Oceanography Research and Education (C-MORE). It represents one of three portions of the Living Liquid project, which aims to visualize ocean life, with this particular exhibit focusing on plankton. The data was provided by MIT's Darwin Project, which utilizes sophisticated supercomputer models to simulate the distribution of various phytoplankton species over time.[7]

The Exploratorium primarily focuses on using interactive exhibits to explain phenomena from the fields of art, science, and human perception for visitors who are at least 11 years of age. Unlike a traditional classroom, the museum has an informal learning environment in which visitors' experiences are described to be unmediated, short, and episodic. In particular, in the life sciences portion of the museum, visitors only spend an average of 30 seconds at each exhibit. These characteristics of the learning environment present several design challenges when creating an exhibit based on guests who often have limited time and little to no domain knowledge of the data.

This project's main goal was to have visitors use visualization as a tool to analyze large, authentic data through sustained inquiry. The ideal outcome is for a viewer to explore the data by asking questions, making comparisons, and evaluating their hypotheses. Because the exhibit is designed for the general public, it's important to incorporate background information into the visualization while avoiding distractions from the data to be explored. Consequently, design considerations must focus more on maximizing data understandability rather than accuracy details. This is one of the few exhibits that leverages visualization as a tool to allow visitors to behave like scientists. As a result, it has sparked the development of other exhibits that use a similar format, such as an exhibit the lets users explore the behavior of marine predators.[8]

Designing an Exhibit for the Public

Because the museum experience is unmediated, the exhibit design must be comprehensive and self-contained, from the visualization to the interaction design to physical aspects such as the printed labels that help guide visitors. Situated in a dynamic environment, the exhibit must attract visitors, especially in a setting with variable crowd densities and other interesting exhibits present. The visualization design must also consider that the visitor will likely spend a short amount of interaction time and have varying levels of interest and prior experience with the scientific content. Referring back to the workflow in Figure 1, this type of visualization relies most heavily on the evaluation step because it's difficult to predict how a viewer will respond. It's typical to go through a large number of redesigns based on how a museum visitor interacts with the prototype.

To minimize the decoding load on the visitor, we employ a visualization design guided by principles from education and cognitive thinking. We also use tangible user interfaces (TUIs) that bring aspects of the visualization to the physical domain. We conducted a study to determine the most effective type of TUI for interacting with the exhibit,[9] but the design process was iterative and guided by the results of formative evaluations.

The visualization runs on a Dell OptiPlex 990 with a MultiTaction Cell that can detect multiple finger touches as well as fiducial markers using infrared sensing. A custom-built table houses the 55-inch display shown in Figure 6. Multiple visitors can approach the table from any side and begin interacting with the exhibit simultaneously. The visualization displays a world map with swirling patterns of color, each representing a different plankton type. This plays and loops dynamically as a timeline displays the current time of year. Labels on the outer edges of the table help guide and inform the viewer about the nature of the data.

Visitors can interact with the exhibit by using the physical viewing lens shown in Figure 7. This lens can be placed anywhere on the ocean and shows icons representing the morphology and relative size of the different types of plankton in that area over time. If the lens is placed on land, the visualization reminds the visitor that plankton can't be displayed. An infrared-reflective fiducial marker engraved on the lens indicates its position and orientation; a virtual interface follows under the lens and can display more information on each plankton species. Three are available for viewing by multiple visitors at any time.

Figure 7. A close-up of one of the physical viewing lenses used as a tangible user interface to view the types of plankton species in various parts of the ocean. An infrared reflective fiducial marker engraved on the lens lets the table know its position and orientation. This is coupled with a virtual interface that moves under the physical lens and can provide more information about each species.

The information presented by this visualization example is much more distilled compared to the previous two examples, which were tailored to professional scientific audiences. Instead, this design process involves creating the proper abstractions for visitors to effectively explore and analyze the data. This exhibit makes use of interactive visualization and TUIs to capture the attention of the general public and encourages involvement in an informal learning setting. Although the content is presented with the focus of understandability over scientific robustness, this type of visualization still grants visitors the flexibility of choosing to learn more about the information, regardless of their interests and background.

A video describing the exhibit is at www.exploratorium.edu/visit/east-gallery/plankton-population.

Challenges Specific to the Museum Setting

Unlike the previous examples, in which many of the challenges involved a domain-specific area, most challenges in this case come from developing a visualization for visitors who likely have no prior knowledge of the domain. One major challenge is finding visual encodings for the data that require very little deciphering and cognitive load on the viewer. Consequently, it's more important to present the information understandably, even at the expense of simplifying details in the data. For example, the swirling colors used on the world map only represent the most dominant plankton type in that location. This is a simplification against using too many colors to represent areas that have many combinations of species in the same location. It isn't until users choose to explore the data with the viewing lens that they notice more complex mixtures of species. Furthermore, very simple glyphs represent the relative size and shape of plankton species in the viewing lens, whereas the actual plankton species can have more detailed differences.

Another challenge comes from the long iterative process necessary of designing an effective exhibit that uses visualization for scientific storytelling. Because interest and usability are the key factors, numerous evaluations were conducted before deciding on the topic matter. In earlier evaluations, noninteractive animations of different parameters in the data set were presented to visitors in various forms. The results have guided our decision to display dominant plankton populations on a global 2D map. Another challenge involves determining the right abstractions that will guide viewer interpretation, without placing too many constraints on a viewer's ability to freely explore the data. This is achieved by layering the accessibility to different types of data, allowing viewers to focus on simple relationships before moving on to more complex details.

We find that to create an effective visualization, design decisions must primarily focus on the setting and target audience. When producing a visualization for domain scientists, there's no need to provide background information or to try to hold viewer attention. When presenting results to general scientists, a visualization must follow a more traditional story narrative by presenting key concepts in a particular order, whereas designing a museum exhibit for the general public requires a new approach altogether. Such a visualization must focus primarily on capturing viewer interest through simple visuals and dynamic interactions. In general, an effective visualization using scientific storytelling must find an appropriate balance among background information, interest-inducing dynamics, and in-depth details into the data. ■

Acknowledgements

We thank all the collaborators who worked closely with us in developing each of our visualizations. For the accelerator visualization, we thank the research team at the SLAC National Accelerator Laboratory, especially Lixin Ge, Cho-Kuen Ng, and Kwok Ko. For the fusion visualization, we thank the research team at Princeton Plasma Physics Laboratory, especially Randy Michael Churchill, Seung-Hoe Ku, Robert Hager, and Choong-Seock Chang. Finally, for the plankton exhibit, we thank the entire staff at the Exploratorium, especially Joyce Ma, Jennifer Frazier, and Amy Snyder. Our research has been sponsored in part by the UC Davis RISE program; Gordon and Betty Moore Foundation; US National Science Foundation via grants DRL-1323214, IIS-1528203, and IIS-1320229; and US Department of Energy via grants DE-FC02-12ER26072 and DE-SC0012610.

References

1. K.-L. Ma et al., "Scientific Storytelling Using Visualization," *IEEE Computer Graphics and Applications*, vol. 32, no. 1, 2012, pp. 12–19.
2. I. Liao, W.-H. Hsu, and K.-L. Ma, "Storytelling via Navigation: A Novel Approach to Animation for Scientific Visualization," *Smart Graphics*, 2014, pp. 1–14.
3. O. Kononenko et al., "Progress on the Multiphysics Capabilities of the Parallel Electromagnetic ACE3P Simulation Suite," *Proc. 31st Int'l Rev. Progress Applied Computational Electromagnetics*, 2015, pp. 1–2.
4. M. Adams et al., "Scaling to 150k Cores: Recent Algorithm and Performance Engineering Developments Enabling XGC1 to Run at Scale," *J. Physics: Conf. Series*, vol. 180, no. 1, 2009, article no. 012036.
5. R. Hager and C.-S. Chang, "Gyrokinetic Neoclassical Study of the Bootstrap Current in the Tokamak Edge Pedestal with Fully Non-linear Coulomb Collisions," *Physics of Plasmas*, vol. 23, 2016, article no. 042503.
6. J. Ma et al., "Living Liquid: Design and Evaluation of an Exploratory Visualization Tool for Museum Visitors," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no. 12, 2012, pp. 2799–2808.
7. M. Follows and S. Dutkiewicz, "Modeling Diverse Communities of Marine Microbes," *Ann. Rev. Marine Science*, vol. 3, 2011, pp. 427–451.
8. C.H. Hsueh et al., "Fostering Comparisons: Designing an Interactive Exhibit That Visualizes Marine Animal Behaviors," *Proc. IEEE Pacific Visualization Symp.*, 2016, pp. 259–263.
9. J. Ma et al., "Using a Tangible Versus a Multi-touch Graphical User Interface to Support Data Exploration at a Museum Exhibit," *Proc. 9th Int'l Conf. Tangible, Embedded, and Embodied Interaction*, 2015, pp. 33–40.

Franz Sauer is a graduate student in the VIDI lab at the University of California, Davis, studying computer science and scientific visualization under Kwan-Liu Ma. His research interests include scientific data visualization, large-scale simulations, computer graphics, and physics. Sauer received a BS in physics from the California Institute of Technology. Contact him at fasauer@ucdavis.edu.

Tyson Neuroth is a graduate student in the VIDI lab at the University of California, Davis, studying computer science and scientific visualization under Kwan-Liu Ma. His research interests include scientific data visualization and analysis, high-performance visualization, and human-computer interaction. Neuroth received a BS in computer science from the University of California, Davis. Contact him at taneuroth@ucdavis.edu.

Jacqueline Chu is a graduate student in the VIDI lab at the University of California, Davis, studying computer science and scientific visualization under Kwan-Liu Ma. Her interests include scientific data visualization, computer graphics, and human-computer interaction. Chu received a BS in Computer Science from University of California, Irvine. Contact her at sjchu@ucdavis.edu.

Kwan-Liu Ma is a professor of computer science and the chair of the Graduate Group in Computer Science (GGCS) at the University of California, Davis, where he leads the VIDI research group and directs the UC Davis Center for Visualization. His research interests include visualization, high-performance computing, and user interface design. Ma received a PhD in computer science from the University of Utah. He's an IEEE Fellow. Contact him at ma@cs.ucdavis.edu.