

METHODOLOGY

Open Access



# The Hiperwall tiled-display wall system for Big-Data research

Muhammad Saleem<sup>1</sup>, Hugo E. Valle<sup>2\*</sup>, Stephen Brown<sup>1</sup>, Veronica I. Winters<sup>3</sup> and Akhtar Mahmood<sup>1</sup>

\*Correspondence:

hugovalle1@weber.edu

<sup>2</sup> School of Computing,  
Weber State University, 3848  
Harrison Blvd, Ogden, USA  
Full list of author information  
is available at the end of the  
article

## Abstract

In the era of Big Data, with the increasing use of large-scale data-driven applications, visualization of very large high-resolution images and extracting useful information (searching for specific targets or rare signal events) from these images can pose challenges to the current display wall technologies. At Bellarmine University, we have set up an Advanced Visualization and Computational Lab using a state-of-the-art next generation display wall technology, called Hiperwall (Highly Interactive Parallelized Display Wall). The 16 ft × 4.5 ft Hiperwall visualization system has a total resolution of 16.5 Megapixels (MP) which consists of eight display-tiles that are arranged in a 4 × 2 tile configuration. Using Hiperwall, we can perform interactive visual data analytics of large images by conducting comparative views of multiple large images in Astronomy and multiple event displays in experimental High Energy Physics. Users can display a single large image across all the display-tiles, or view many different images simultaneously on multiple display-tiles. Hiperwall enables simultaneous visualization of multiple high resolution images and its contents on the entire display wall without loss of clarity and resolution. Hiperwall's middleware also allows researchers in geographically diverse locations to collaborate on large scientific experiments. In this paper we will provide a description of a new generation of display wall setup at Bellarmine University that is based on the Hiperwall technology, which is a robust visualization system for Big Data research.

**Keywords:** Hiperwall, High-resolution displays, Information visualization, Collaborative visualization, Big data visualization, Visual analytics

## Introduction

One of the greatest scientific challenges of the 21st century is to understand the plethora of information from the various Big Data sources. Therefore, finding new information and patterns in Big Data is the current imperative [1]. Large-scale data driven applications are on the rise, and so is the need to extract useful information from terabyte and petabyte scale data sets (Big Data). Additionally, the amount of data from large scientific instruments as well as the resolution of large images are also increasing which creates visualization challenges that require a larger viewing area to capture the details [2]. So, visualizing multidimensional, time-varying data sets is both a challenge to the computational infrastructure and to the current display wall technologies.

Large tiled-display walls are enabling a new era in large-scale visualization and scientific explorations, for the various research and scientific communities. These large

display walls are very useful in order to interpret and visualize the enormous amount of experimental data that is being collected by large science experiments. These tiled-display walls allow users and researchers to explore features that are either very hard or practically impossible to identify using a conventional PC monitor.

This paper focuses on a novel visualization and display wall technology for displaying large high resolution images from very large data sets. A single computer monitor can constrain the way in which we can analyze large volumes of data. We can view either a low-resolution image that represents the entire extent of the data and thus risk missing the smaller details, or we can view small portions of the data in high resolution while losing the full context. In order to extract new information from Big Data, we need advanced visualization and new data analytics tools for Big Data environments [3]. In general, scientific data can be meaningless if it cannot be represented in a meaningful way, not only to the scientific community, but also to the general public. With Hiperwall, we can see both the broad view of the data sets and the details concurrently which enables the shared view of complex results. With our Hiperwall visualization system, we can display very large high-resolution images, conduct interactive analysis of large data sets, view multiple images or parameters of large data sets, and conduct a comparative view of many data sets at once and stream HD videos from remote collaborators (e.g. from CERN and LSST) [4, 5].

## Methods

### Visualization methods

Data visualization refers to the techniques that help the user transform data in the form of visual objects contained in graphics. Visualization is defined as a mental image or visual representation of an object or scene or person or abstraction that is similar to visual perception [6]. Visualization of the data can be a challenge to the user's perception and capabilities [7]. Other definitions of visualization most commonly used in the literature are "the use of computer-supported, interactive, visual representation of data to amplify cognition", where cognition is the acquisition or use of knowledge [8].

Over the past decades, data visualization has evolved and it has been categorized differently by different authors. This section presents a summary of the information visualization techniques and some of the data visualization categories used in this analysis. More categories could be found in the periodic table of visualization, which is an interactive chart displaying various data visualization methods [9]. Information visualization is an interactive interface of data to increase cognition or perception ability. It utilizes graphics to assist people in comprehending and interpreting the data [7, 10]. Data visualization represents quantitative data with or without axes in schematic or diagrammatic form such as a table, line chart, pie chart, histogram, etc. These methods become more important when the data is massive, which needs to be analyzed and represented in a meaningful form, which is easy to understand and interpret [10]. A recent research study in the white paper by the SAS Institute showed that 98% of the most effective companies working with Big Data are presenting results of the analysis via visualization [11].

Visual Analytics combines both visualization and data analysis [12–14]. In data visualization, a variety of techniques are commonly used by data scientists. A table is the simplest and the easiest data representation technique. The information is structured in

rows and columns. Rows represent variables and columns represent records with a set of values [15]. The role of the table is the essential part of research and data analysis [10]. Next, pie charts, which is a circle divided into number of slices. Each of the slices describe a numerical proportion. Pie chart visualization could be difficult when comparing different sections of different pie charts. The next category is 2D/3D representations. This category includes various types of graphs and bar charts [16], histograms, and line charts. Bar charts represent a single data series, and related data points are grouped in one series [10]. Histograms, which are commonly used in statistics, represent an accurate distribution of numerical data. Line charts represent information in connected points or “markers” by straight line segments. Line charts are often used to visualize a trend in data over time intervals [10]. These techniques however, offer some limitation when complex data structures need to be visualized. Next, there is geometric transformation techniques which include the scatter plot and parallel coordinate plot. Scatter plot is a graphical display of a set of data using Cartesian coordinates to display the relationship between two variables for a set of data. Typically, one of the variables represents the horizontal axis, and the second variable represents the vertical axis of the data points. These plots are useful to determine trends in data and to identify outliers easily [10, 17]. The parallel coordinate plot helps the user to visualize multi-dimensional data [18]. Next, we have hierarchical image plots, which is a great method to visualize hierarchical structure data [19]. Finally, we have methods that focus on pixels, which is a visualization technique that focuses on the pixels [20]. The main purpose of this technique is display a single value per pixel. In our case, this is an excellent technique to visualize large data sets using the big display walls such as the Hiperwall (see “[Visual and data analytics with Hiperwall](#)”).

Information Visualization focuses on the use of visualization methods to help the user understand the data. Information visualization is the transmission of abstract data through the use of interactive visual interfaces [21]. Olshannikova et al. [7] presents a detailed list of visualization techniques such as tree map, tag cloud, clustering, motion charts and dashboard which is used in text analysis to depict keyword metadata, typically on websites [7].

The third area for data visualization is interactivity techniques. In these techniques, the user interacts directly with the data through mechanisms that make it possible to manipulate visualization effectively. When the user needs to reveal more details hidden in the data, the zooming technique is a great choice, or when the users need “zoom-in” capabilities, a big display wall such as the Hiperwall is an excellent visualization tool for this technique (see “[Visual and data analytics with Hiperwall](#)”). Another technique is the overview + detail which uses multiple views at the same time. This technique works for 2D and 3D views. Others methods such as zoom + pan, focus + context or fish eye, offer different interaction techniques to users [10].

### **Visual and data analytics with Hiperwall**

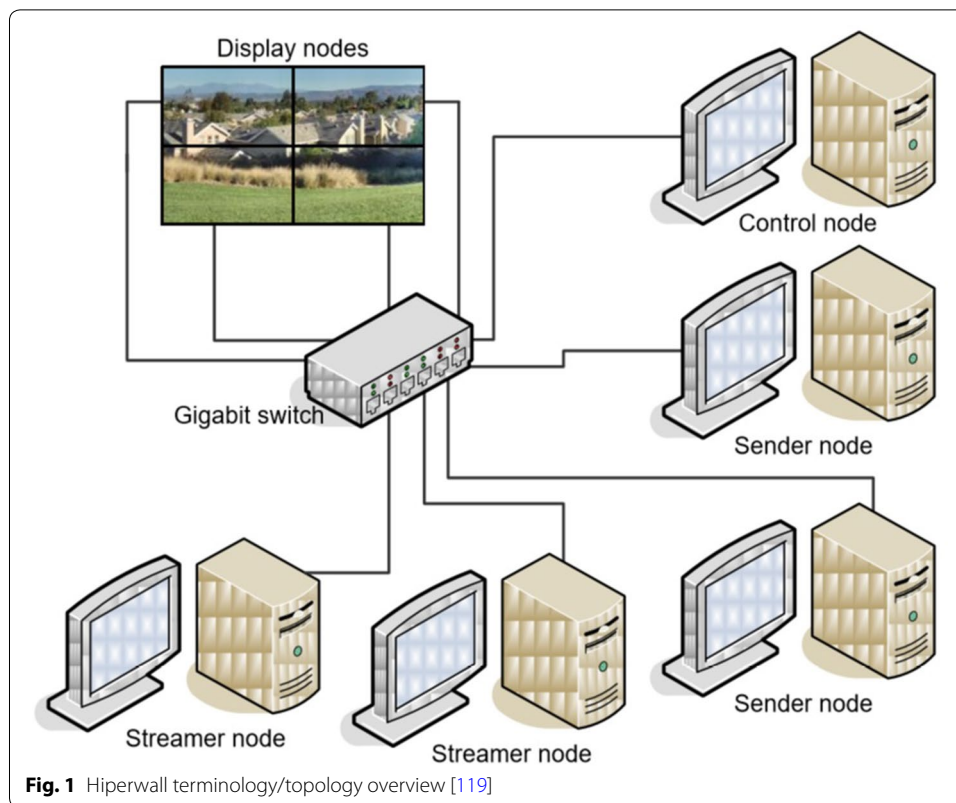
Large display walls have traditionally been the domain of air-traffic control centers, NASA command centers, and a few big budget corporations [22]. Nowadays, large high-resolution display walls that show large quantities of information at a single glance can be seen everywhere—in shopping malls, airports, athletic arenas,

museums, large corporations, and at TV/News stations and broadcasting studios. The uses of display walls are indeed very broad. In the commercial arena, digital signage is by far the most common usage of display walls. In recent years, a number of universities have also installed these display walls in sports arenas and VIP lounges to help attract students and sports fans to their institution to create a better fan experience and for advertising from sponsors. In academia, in the past few years, a number of universities have installed display walls in classrooms, libraries and lobbies. A few universities have installed display walls mostly for conducting interactive data visualization and analysis in a collaborative research environment.

Our tiled-display wall setup, known as Hiperwall (Highly Interactive Parallelized Display Wall) was inspired by an earlier tiled-display wall design architecture, called the OptlPortal that was developed almost 10 years ago [23]. The OptlPortal design used regular LCD PC monitors that were driven by PCs (which was attached to back of each LCD monitor), with optimized graphics processors and network interface cards. Our Hiperwall display was built using commercial-brand 55" Samsung LED (backlit) HDTVs (instead of LCD PC monitors used in the OptlPortal design) and is driven by a separate rack-mount PCs—one PC for each Hiperwall display tile (instead of each PC attached to back of each display-tile). Hiperwall Inc. is a spinoff company that was created from the NSF funded OptlPuter project at UC Irvine's California Institute for Telecommunications and Information Technology (Calit2) [24, 25]. It was designed to be a collaborative visualization platform capable of displaying information in real time. The Hiperwall software runs as an application operating on standard PCs connected via a standard 1 Gbps network switch (see Fig. 1 for terminology/topology overview of the Hiperwall visualization system).

Hiperwall is a high performance, high resolution visualization system. Hiperwall's visualization system software architecture applies parallel processing techniques to overcome the performance limitations typically associated with the display of large multiple images on multiple display devices (flat-panel tiles) simultaneously. Hiperwall is a collaborative visualization platform that is designed to visualize enormous data sets and allows viewers to see the details, while retaining the context of the surrounding data. This allows a group of scientists/researchers/students to collaborate and share detailed information. The Hiperwall system allows the user to display a wide variety of high-resolution 2D and 3D images, animations, movies, and time-varying data in real time all at once on multiple display-tiles that can be configured according to the needs of the user. Users can project a single image across the entire display area or many different images simultaneously. Hiperwall's middleware even allows researchers in geographically diverse locations to collaborate on Big Science Experiments that generate Big Data.

Members of this research team from Bellarmine University (BU) have recently joined the Large Synoptic Survey Telescope (LSST) project and are part of LSST's Dark Energy Science Collaboration (DESC). Equipped with a 3.2 Gigapixel camera (the world's largest digital camera), the goal of the LSST project is to conduct a 10-year survey of 37 billion stars and galaxies that will deliver large volumes of images and data sets (astronomical catalogs) that is thousands of times larger than previously compiled to address some of the most pressing questions about the structure and



evolution of the universe, such as understanding the mysterious Dark Energy that is driving the acceleration of the cosmic expansion.

Starting from 2022, LSST will produce 15 Terabytes of raw data images per night, that's about 200 Petabytes of imaging data over the 10 years of operation, which will be the largest data set in astronomy/astrophysics and cosmology [29, 30]. BU's Hiperwall visualization system will be used for some of the LSST data analysis tasks to study and visualize some of these large LSST images in order to extract useful and specific information from these large high resolution images. Table 1 shows the current single display technology resolutions which are much smaller than most current astronomical cameras' full resolution. Figure 2 shows a portion of a high resolution image of a supernova remnant displayed on the Hiperwall Display Wall.

In the field of experimental High Energy Physics (HEP), the four big HEP experiments (ATLAS, CMS, LHCb, and ALICE) at the Large Hadron Collider (LHC) at CERN in Switzerland are producing colossal volumes of data (Big Data). As of June 2018, these HEP experiments at the LHC have recorded over 250 Petabytes of raw data that are stored permanently at the CERN Data Center (DC) [31].

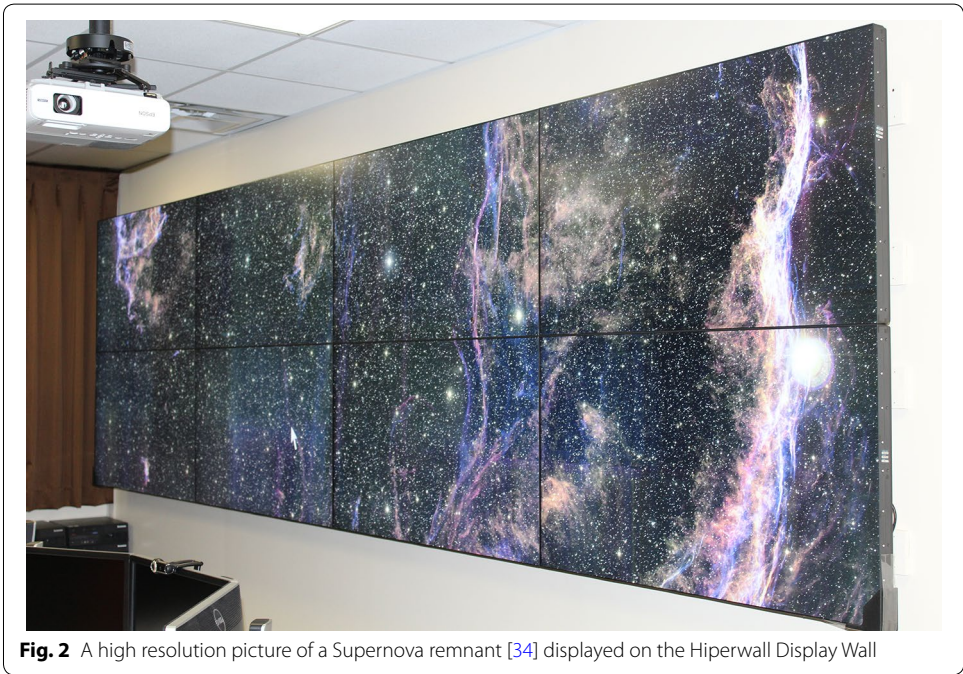
When proton beams collide head-on at high energies inside the LHC detectors, new subatomic particles are created that decay in complicated ways as they travel through layers of the detector. The collision energy is converted into mass that create these short-lived subatomic particles. Currently, protons collide head-on inside the LHC detectors at an unprecedented rate of 1 billion times per second [31]. Most (99%) of the raw data events are filtered out by the LHC experiments, and the "interesting" data events are



**Table 1** Table adapted from Meade et al. [26] shows the increasing resolution in Megapixels of instrument/imaging devices versus the resolution of common displays

	Resolution	Megapixels (MP)
<i>Instrument/Imaging Device</i>		
HST Advanced Camera for Surveys [26]	2 × 2048 × 4096	16
The QUEST Large Area CCD Camera [27]	112 × 600 × 2400	116
Skymapper [26]	32 × 2048 × 4096	268
DECam [26]	62 × 2048 × 4096	520
Subaru Hyper Suprime-Cam [26]	104 × 2048 × 4096	870
LSST Camera [28]	189 × 4096 × 4097	3200
<i>Display Device</i>		
Standard desktop display	1680 × 1050	1.7
Full high-definition (FHD) desktop display	1920 × 1200	2.3
iPad (with retina display)	2048 × 1536	3.1
Dell UltraSharp desktop display	2560 × 1600	4.1
Laptop (Macbook Pro)	2880 × 1800	5.2
4K ultra-high definition (UHD) display	3840 × 2160	8.3

Note that no current display format can display a complete image at the original resolution



**Fig. 2** A high resolution picture of a Supernova remnant [34] displayed on the Hiperwall Display Wall

stored for further analysis. The filtered LHC data are stored at the CERN Data Center, where the initial data reconstruction is performed [31]. Even after this data reduction, the CERN Data Center (Tier0 Grid site) processes on average about one petabyte of data per day the equivalent of around 210,000 DVDs [32]. As of June 2018, even after filtering out 99% of the raw data, around 50 Petabytes (50 PB) of data have been stored [33] for data analysis. That's 50,000 Terabytes of data, the equivalent to nearly 15 million HD movies [33]. This 50 PB of data that have been stored can be retrieved and accessed by thousands of LHC physicists worldwide for analysis via the Worldwide LHC Computing

Grid (WLCG) and the Open Science Grid (OSG) cyberinfrastructure that is based on tiered grid computing technology. On any given day, on average, more than two million grid jobs run on WLCG and OSG for the LHC experiments. These subatomic particles that decay within the detector, leaves tracks or signatures of their presence that are registered by converting the particles' paths and energies into electrical signals to create a digital snapshot of the "collision event".

## The control node

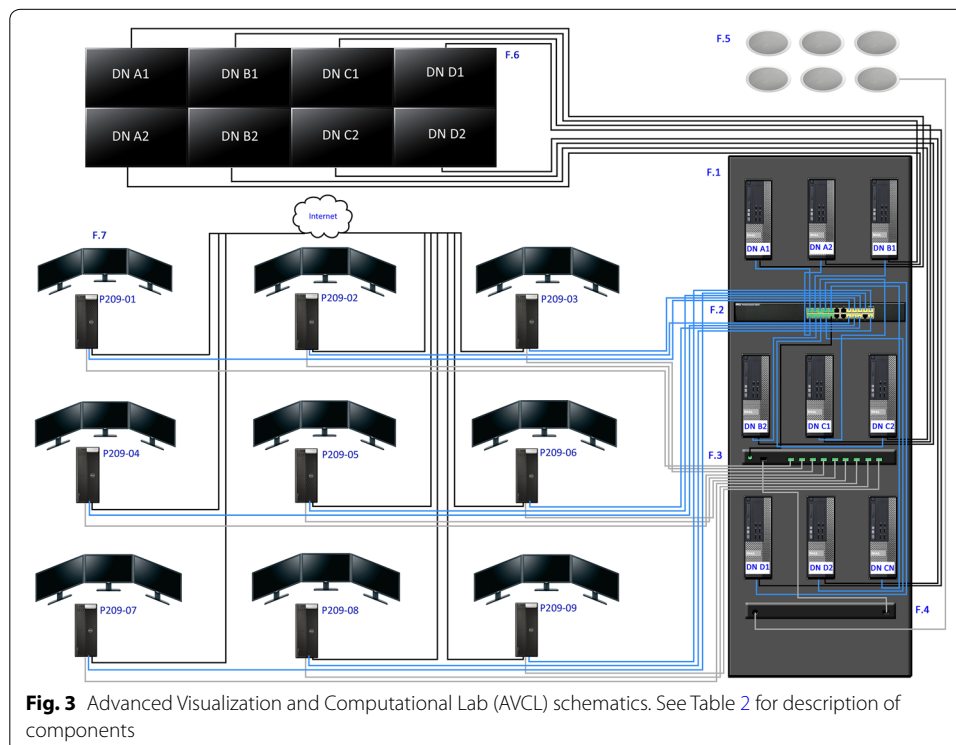
The control node (labeled as DN CN in Fig. 3) is the brain of the Hiperwall system. The control node gives the user the ability to control what, where, how, and when the contents are displayed. The user interface shows a miniature view of the eight Hiperwall display-tiles. The interactive Hiperwall drag-and-drop simplicity gives the user the ability to place contents on the display-tiles.

### The secondary control node

The secondary control node provides the user the ability to control the Hiperwall system (nodes) remotely from a distant location via a portable device, such as a tablet. The secondary control node is not shown in Fig. 3.

## The display node

The 8 display node PCs in the rack (labeled as DN A1 through DN D2) as shown in Fig. 3.F.1 are connected to the Hiperwall display-tiles (one display node for each display tile). Each display node consists of a display device that is driven by a PC (mini-tower or rack-mount) running the Hiperwall software. The Hiperwall software runs as



**Table 2 Hiperwall visualization system hardware components and specifications as shown in Fig. 3**

Figures	Hiperwall visualization system hardware components and specifications
Figure 3.F.1	<p>Hiperwall Visualization Cluster (1 Control and 8 Display Nodes)</p> <p>A. <i>Rack-Mounted Control and Display Nodes</i> 9 Dell Optiplex 7010 PCs on a 40U Rack (1 Control Node and 8 Display Nodes) Specs for each node: 3rd Gen Intel Core 3.4 GHz i7-3770 Processor; 4 GB 1600 MHz DDR3 RAM; 1 GB RAM AMD RADEON HD 7570 Graphics Card; 500 GB Hard Drive (Total RAM: 9 GB; Total Hard Drive Space of 4.5 TB). Operating System: Windows 7 (64-bit)</p> <p>B. <i>Secondary Control Node (Not Shown in Figure 3)</i> Dell Latitude 10 Tablet. Specs: 1.8 GHz Z2760 Intel Atom processor; 2 GB RAM; 533 MHz Intel Graphics Media Accelerator Graphics Card; 2 GB DDR2 SDRAM RAM; 64 GB Hard Drive. Operating System: Windows 8 Pro (32-bit)</p>
Figure 3.F.2	<i>1 Gbps 24-port Network Switch</i> 1U Rack-Mount Dell 1 Gbps 24-Port PowerConnect 2824 Switch
Figure 3.F.3	<i>Audio Digital Signal Processor</i> Biamp Nexia-CS-10 mic/line inputs and 6 mic/line outputs
Figure 3.F.4	<i>Audio Power Amplifier</i> 2U Rack-Mount TOA 900 Series-II P-912MK2
Figure 3.F.5	<i>Ceiling Speakers</i> 6 Ceiling-Mounted Speakers, Atlas Sound FAP62T (Strategy-II series) 6", 32W
Figure 3.F.6	<p><i>Display Wall</i> 8 commercial-brand 55" Samsung LED (backlit) HDTV (Model UD55C). Left/Top Bezel Size = 3.7 mm and Right/Bottom Bezel Size = 1.8 mm. Display Wall Layout: 4 × 2; Total resolution (4 × 1920) × (2 × 1080) = 7680 × 2160 (16.5 MP)</p> <p>Display Wall Dimensions: Width = 16 ft (4.9 m) and Height = 4.5 ft (1.4 m); Bezel-to-Bezel Size = 0.2" (5.5 mm)</p> <p>Connection from Control/Display nodes to display wall: HDMI</p>
Figure 3.F.7	<p><i>Sender/Streamer Nodes (9 Data Analysis Workstations)</i></p> <p>9 Dell T5600 Workstations each connected to three 24" monitors</p> <p>Specs for each Workstation: Each Node has a Dual-Four Core 3.3 GHz Xeon E5-2643 processor; 64 GB DDR3 RDIMM RAM; 2 GB × 2 GB RAM NVIDIA Quadro Graphics Cards; two CUDA-768 cores (1536 CUDA cores) per workstation; 4 TB Hard Drive (Total RAM: 576 GB; Total Hard Drive Space: 36 TB). Operating System: Dual-Boot Windows 7 (64 bit) and Scientific Linux 6.4</p>

an application operating on a standard PC connected via a standard 1 Gbps network switch. Hiperwall's tiled display system software architecture applies parallel processing techniques to overcome the performance limitations typically associated with the display of large multiple images on multiple display devices (flat-panel tiles) at once. A display node software receives content over a 1 Gbps network switch from a sender or a streamer node and projects it on the display wall tile via a HDMI cable. All display-node computers work in parallel, thus giving flexibility and scalability while also allowing the user to place contents on one or more multiple display-tiles all at once.

### The display wall

Each of the 8 display-tiles (labeled as DN A1 through DN A2) as shown in Fig. 3.F.6 are connected to the corresponding display node PC in the rack. The display wall projects the contents from the user's Workstation (sender and streamer nodes) onto the display wall (flat-panel tiles) via the control node. Each tile is a high-end commercial-brand 55" (UD55C) Samsung LED (backlit) HDTV with a very narrow bezel size. The top and the left bezel size is 3.7 mm (about 0.15"), whereas the bottom and the right bezel size is 1.8 mm (about 0.07"). Each of the display node PCs are responsible for displaying and rendering a portion of the overall image that is being shown/displayed on the Hiperwall. Thus, all the PCs work in parallel to render the total image.

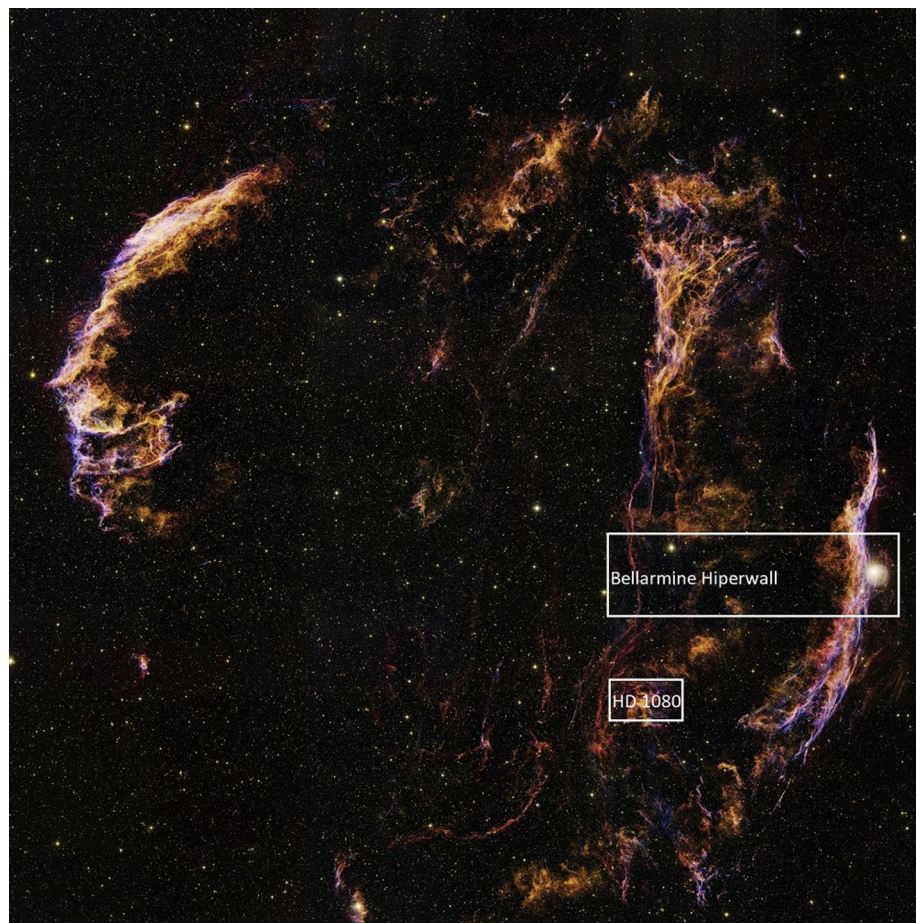
The Hiperwall software runs on each display node and renders images though the control node which transfers the data displayed on Hiperwall display-tiles via the display nodes. The Hiperwall system provides robust visualization of multiple forms of display and monitoring capabilities simultaneously with a lot of user flexibility. With the



Hiperwall visualization system we can display various types of images (1 GB or larger) and contents on the display wall. We can add, resize, scale, rotate, zoom and reposition large images, charts, plots and HD video feeds for visualization, simultaneously. Users can spread one large piece of image or spread a variety of images across the entire display wall. In our setup, the 8 display-tiles are arranged in a  $4 \times 2$  format. The  $16' \times 4.5'$  display wall has an effective bezel-to-bezel size of only 5.5 mm (0.2"). The total resolution of the display wall is  $7680 \times 2160$  pixels which is about 16.5 Megapixels (MP). Figure 4 shows an actual  $24457 \times 24576$  pixels (600 MP) large astronomical image of a supernova remnant. In comparison, the total viewing area of our Hiperwall Display Wall and a standard HD  $1920 \times 1080$  (2 MP) monitor are shown as white boxes. A single LSST image [34] will be about five times larger than this astronomical image.

#### The sender/streamer nodes

The sender and streamer nodes deliver content to the Hiperwall display-tiles via the 1 Gbps network switch. The sender/streamer nodes are high-end workstations. It is



**Fig. 4** The total viewing area of the  $7680 \times 2160$  pixels (16.5 Megapixels) Hiperwall Display and a standard HD  $1920 \times 1080$  (2 Megapixels) monitor superimposed (shown as white boxes) on top of a  $24457 \times 24576$  pixels (600 Megapixels) large astronomical image of a supernova remnant (SNR) [34]. Note that a single LSST image [120] will be about five times larger than this astronomical image

called a sender node if it is running the Hiperwall sender software license for displaying still images and animations. A streamer node runs the Hiperwall streamer software license for displaying HD videos. Sender nodes allow users to view the contents of the monitor(s) connected to the users' workstation. The sender/streamer nodes can run multiple applications of the Hiperwall sender/streamer software. A Java application allows the output of the user's application to be transported across the network for viewing on the Hiperwall display wall tiles. Users can send the entire screen to the display wall or divide it up into rectangular regions (tiles) and deliver each region to the Hiperwall as an independent object. Figure 5 shows a high resolution picture of the surface of Mars taken by NASA's Curiosity rover that was displayed/projected from one of the sender nodes (workstation) onto the Hiperwall Display Wall. Figure 6 shows students analyzing particle tracks from HEP proton-proton collision events from the ATLAS experiment using the Hiperwall Display Wall.

The screen sender software can also run remotely. At BU, we have 11 sender licenses and 1 streamer license, that allows us to have up to 12 display projections at a time. Streamer software can provide high frame rate feeds to the Hiperwall displays and are used for HD videos and animations. The streamer system can stream up to 60 frames per second (FPS) at a resolution of 1080P or higher. FPS is the frequency (rate) at which an imaging device produces consecutive images. 60 FPS is typically used because high definition televisions (HDTV) are designed around 60 Hz signals as a common base capability. 60 FPS is optimal to the eye and gives a much more realistic sense of motion for what is happening on the screen [35]. A high resolution image or HD video feed can be enlarged and stretched across the entire display wall or multiple arrangements of many images or HD video feeds can be arranged in any size or position across the display-tiles and zoomed without loss of clarity.

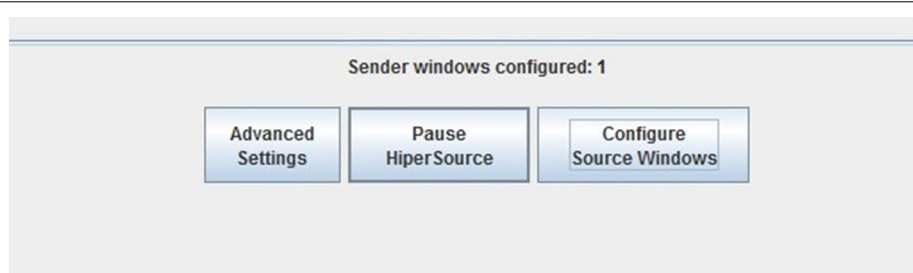


**Fig. 5** A high resolution picture of the surface of Mars taken by NASA's Curiosity rover displayed on the Hiperwall Display Wall [121]





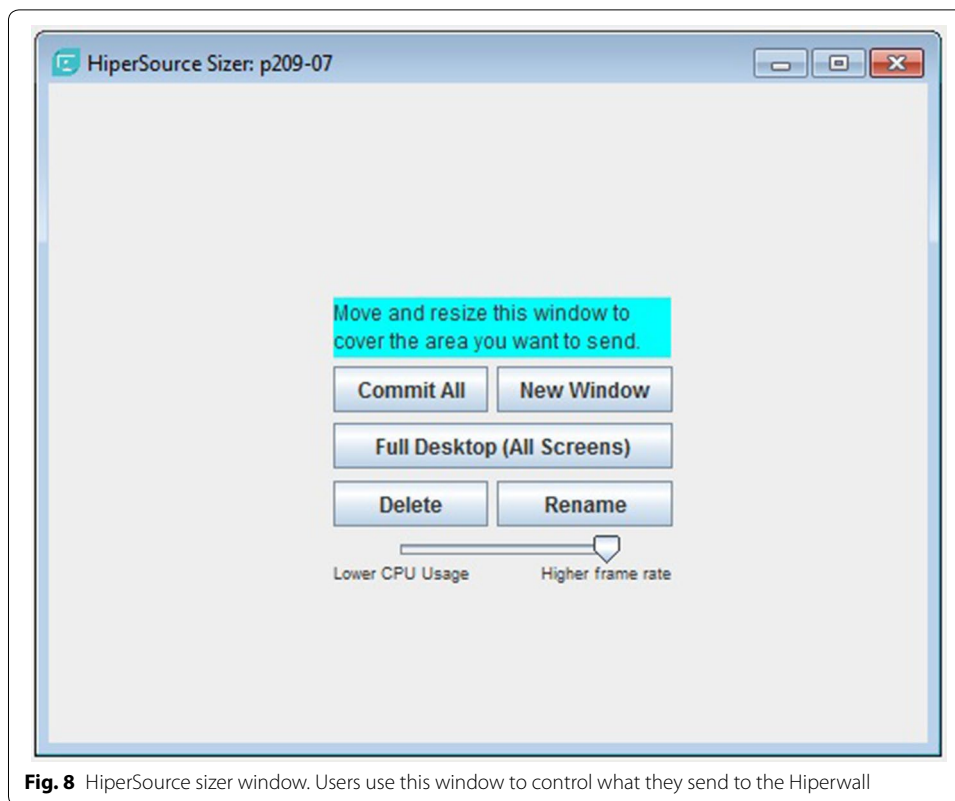
**Fig. 6** Students analyzing particle tracks from HEP proton-proton collision events



**Fig. 7** Hiperwall sender window

### Hiperwall user portal

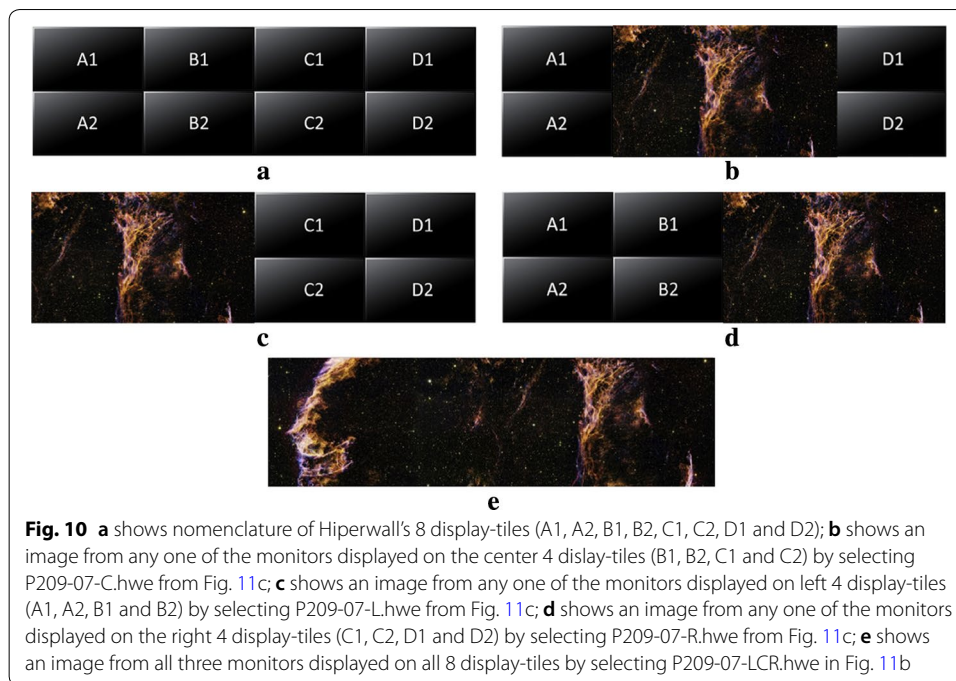
The original version of the Hiperwall software had numerous challenges as described further in section challenges while using the Hiperwall in educational and research pursuits. When any user needs to display an image from any one of their monitors to the display-tiles, the user first clicks on the Sender Icon on their workstation's taskbar to open a window as shown in Fig. 7. Then the user selects the "Configure Source Windows" button to open a window, labeled HiperSource Sizer (shown in Fig. 8), which the user can move and resize to select full space or "dragged area" of their workstation's monitors to send to the Hiperwall control node. The user has a choice to select the contents of one monitor, two monitors, or all three monitors. The layout of the workstation's three monitors are shown in Fig. 9. Regardless of how many monitors the user sends or the contents of each of the monitors, Hiperwall treats them as a single image. Users can use the HiperSource Sizer window in Fig. 8 to resize and maximize the HiperSource Sizer to send the contents of a single monitor (any one of the workstation's three monitors) to the control node. Users can also click and drag the HiperSource Sizer window border to select any two or all three adjacent monitors' display area. The HiperSource Sizer allows a click and drag operation to select any subset of the total viewing area, but for simplicity, we only present to our users the options of one, two or three



monitors. Besides resizing to customize the sent area, there is also a button for sending all screens, “Full Desktop (All Screens)” as shown in Fig. 8. This will skip the “Commit All” step and immediately broadcast all three monitors. For the other methods of resizing, the user needs to click “Commit All” after they have determined and selected what they want to send to the control node.

Once the user has used the sender software (see “[The sender/streamer nodes](#)”) to choose what they will broadcast to the Hiperwall, they can then use the Hiperwall User Portal to decide how the broadcast will be displayed or placed on the display-tiles as shown in Fig. 10. We modified Hiperwall’s environment settings and developed our own User Portal, starting from the workstations and building from the Hiperwall software’s





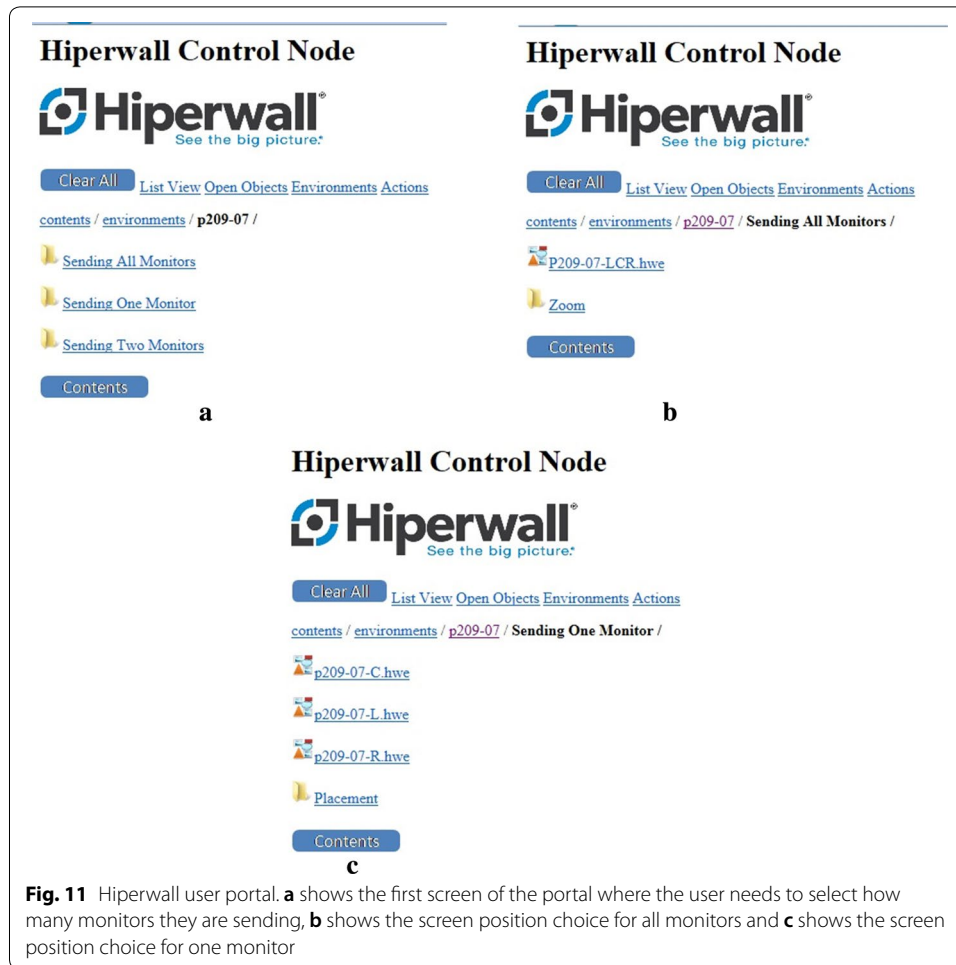
own environment. We were able to individually make display environments for each specific workstation that has access to the Hiperwall as shown in Fig. 11a–c. In Fig. 11a, the user is given the links to choose how to display on the Hiperwall their broadcast, depending on if the user is sending one, two, or all three of their workstation's monitors. We had to decrease some of the control features of the control node for the users in order to make the system more user friendly. For example, they are not able to rotate or resize or manually move the displayed areas on the Hiperwall. Figure 10 shows the nomenclature of Hiperwall's 8 display-tiles, Fig. 10b shows an image from any one of the monitors displayed on the center 4 display-tiles, Fig. 10c shows an image from any one of the monitors displayed on the left 4 display-tiles, Fig. 10d shows an image from any one of the monitors displayed on the right 4 display-tile, and Fig. 10e shows an image from all three monitors displayed on all 8 tiles. Fig. 12 shows the scenario where the user has selected to send all three monitors and then choose the LCR option in Fig. 11b to center the broadcast on the Hiperwall. This leaves black bands on the top and the bottom of the displayed image which can be removed by using the Zoom link option in Fig. 11b.

## Results and discussion

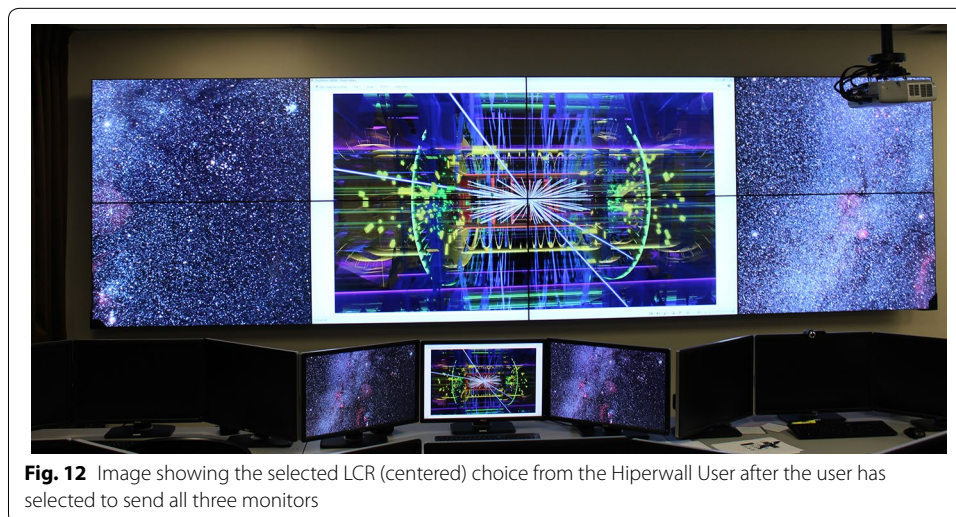
### Display wall cost benefit analysis

In the US, planar-tiled projected displays were first built in the 1990s at Princeton University and at Argonne National Laboratory [25]. These display walls were all prototypes. The first screen display, called PowerWall was built in 1994 at the University of Minnesota in collaboration with Silicon Graphics as a single  $6' \times 8'$  screen illuminated from the rear by a  $2 \times 2$  matrix of video projectors that were driven by 4 RealityEngine2 graphics engines [36]. Around the same time a number of academic research groups began building display walls using commodity DLP or LCD projectors [37].





**Fig. 11** Hiperwall user portal. **a** shows the first screen of the portal where the user needs to select how many monitors they are sending, **b** shows the screen position choice for all monitors and **c** shows the screen position choice for one monitor



One of the first-generation display walls was built by Princeton University in 1998, that used a  $18' \times 8'$  rear projection screen and eight Proxima LCD commodity projectors [38]. This system had a resolution of  $4096 \times 1536$  pixels (6.3 MP) that was driven

by a network of eight Pentium II PCs running Windows NT. In 2000, this display was later scaled up to a resolution of  $6144 \times 3072$  pixels (19 MP) with 24 Compaq MP1800 digital light-processing (DLP) projectors and a network of 24 Pentium III PCs running Windows 2000 [38].

A notable early LCD tiled-display wall was built in August 2002 at the NASA Ames Research Center as a  $7 \times 7$  array flat panel LCD screens powered by a Beowulf cluster [23]. In September 2002, the Electronic Visualization Laboratory (EVL) [39] at the University of Illinois at Chicago built a 100 MP LambdaVision display wall with NEC monitors using custom-built PCs [23]. In 2005, a group at UC-Irvine built a 200 MP 50-tiled display wall out of 3000 Apple Cinema displays and Apple Power Mac G5s that was later adopted by the OptIPuter project that used PCs driving the 24" LCD monitor displays. In 2008, a group at UC-Irvine's California Institute for Telecommunications and Information Technology (Calit2) received funding from NSF for their OptlPortal open-source project [25] and built a 27 MP 12-tile OptlPortal display wall. The OptlPortal project used desktop LCD tiles, PCs to power the display-tiles (one PC per tile), graphics card, and local networking to build a 200 MP OptlPortal tiled display wall at a cost of about \$1000/MP [23]. So the 200 MP OptlPortal display wall would have costed about \$200,000. Since the OptlPortal 50-tiled display used regular  $1920 \times 1080$  (2 MP) 24" LCD displays for desktop computers, the group at Calit2 was able to keep the cost down. The main disadvantage with the OptlPortal tiled-display was that the LCD displays had thick bezels size that created broad mullions (horizontal and vertical strips where there is no visible image) in between two adjacent tiles which the users found to be distracting [23]. Back then, the typical 24" desktop LCD monitors had a bezel size of 15 mm (about 0.6"); so bezel-to-bezel size was 30 mm about (1.2").

A notable display wall, later dubbed as the "IQ-Wall" was installed at Indiana University's (IU) Advanced Visualization Laboratory (AVL) in 2004, utilizing relatively expensive high-end components and a professional audio/video integrator for uses varying from research visualization to collaborative presentations to artistic creations [40]. The "IQ-Wall" struck a balance between cutting-edge technology, good performance, and cost [40]. IU has been a pioneer in building display walls in academia. From 2004 to 2015, various groups at IU have built nine separate display walls with  $2 \times 2$  ( $8' \times 4.5'$ ),  $3 \times 3$  ( $10' \times 6'$ ),  $3 \times 4$  ( $10' \times 8'$ ),  $4 \times 4$  ( $13.5' \times 8'$ ),  $6 \times 2$  ( $24' \times 4.5'$ ),  $6 \times 4$  ( $24' \times 9'$ ), and  $8 \times 2$  ( $27' \times 4'$ ) tile configurations and size, with display wall resolution ranging from 2.1 MP to 42 MP [40].

As mentioned earlier, our  $16' \times 4.5'$  ( $4 \times 2$  tile configuration) 16.5 MP custom-made tiled-display wall setup was built upon the OptlPortal display wall's PC-based architecture to power each display tile. The installation of our Hiperwall visualization system was completed by October 2013 in the Advanced Visualization and Computational Lab (AVCL). Since we received \$209,347.00 from our NSF grant, our total budget for everything (including the shipping and handling costs, and other miscellaneous expenses) had to be within that amount.

The cost for the hardware components of the BU's Hiperwall set-up that includes the 8-tiled display-tiles (each display tile is a 55" Samsung commercial-brand UD55C HDTV), Rack-Mounted Control and Display Node (9 Dell Optiplex PCs), Secondary

Control Node (1 Dell Latitude Tablet), 1 Gbps Dell 24-port Network Switch, Audio Digital Signal Processor, Audio Power Amplifier, Ceiling Speakers, plus the HDMI and Network cables and wires was \$70,830. We also paid \$54,194 (\$10,839/year) for the Hiperwall software licenses and full-technical and software support for 5 years, of which \$33,884 (\$6777/year) was for the Hiperwall software licenses and \$20,310 (\$4062/year) was for the full-technical/software support for 5 years. The Hiperwall software license cost is for the Hiperwall Master license, 1 Secondary Control license, and 11 Sender licenses for the 9 Sender nodes (Data Analysis workstations), the Hiperwall Control Node and the Secondary Control Node, and 8 Display licenses for the 8 Hiperwall Display Nodes. Additionally, the hardware installation cost was \$11,285. So all the Hiperwall hardware components, software licenses and technical/software support, including the installation expenses was \$136,309.

We also purchased 9 high-end Dell T5600 Data Analysis Workstations (Sender/Streamer Nodes) each with 64 GB RAM, Dual-CUDA-768 cores and 4TB HD with triple 24" monitors that is separate from the Hiperwall display wall at a cost of \$64,846. These Data Analysis Workstations were configured to run on both Windows and the Linux OS environments, simultaneously. Therefore, the total of cost of the 8-tiled 16'  $\times$  4.5' (4  $\times$  2 configuration) 16.5 MP (7680  $\times$  2160 pixel) Hiperwall visualization system that includes the 9 Dell T5600 Data Analysis Workstations and the Hiperwall software licenses and full-technical/software support for 5 years and the installation expenses was \$201,155. The most expensive hardware component of our Hiperwall setup was the eight 55" 1920  $\times$  1080 pixel resolution (2 MP or 2K) Samsung commercial-brand UD55C HDTV for the 8 display-tiles, which was \$5994 each (\$47,952 for the eight HDTVs). The reason we decided to go with these 55" Samsung commercial-brand UD55C HDTV was due to its narrow bezel size. Bezel size is perhaps the single most critical aspect of any display wall since thick bezels can be distracting when viewing images that create the appearance of a mullion between two adjacent display-tiles. Moreover, there is loss of information from displayed images in the bezel sections when images are displayed on multiple display-tiles. Even though these HDTVs were pricey, at that time these were the least expensive 55" commercial-brand HDTVs in the market which had the narrowest bezel size (Left/Top bezel size of 3.7 mm and Right/Bottom bezel size of 1.8 mm); so a bezel-to-bezel size would be 5.5 mm (about 0.2"). We also looked at the non-commercial-brand (regular) 55" (2 MP or 2K) Samsung UE55A HDTV which were cheaper (\$3750 each) but these UE55A models have twice the bezel size, a bezel-to-bezel size of 10 mm (about 0.4") compared to the UD55C model that has a bezel-to-bezel size of 5.5 mm (about 0.2").

Even though the hardware expenses for our Hiperwall set up was relatively low, the only added expense with Hiperwall is the cost for the annual software license and technical/software support of the Hiperwall Master license, 1 Secondary Control license, and 11 Sender licenses for the 9 Sender nodes (Data Analysis workstations), the Hiperwall Control node and the Secondary Control Node, and 8 Display licenses for the 8 Hiperwall Display Nodes which is \$10,839/year. For the future, when we renew our license for another 5 years from October 2018 to October 2024, we will not purchase the technical/software support since we are now very familiar with the Hiperwall software and the system.

Additionally, since we rarely used the Secondary Control Node (the Tablet) to control the Hiperwall, we will no longer renew the license for the secondary control. Moreover, we hardly used all 9 sender nodes (the 9 Data Analysis workstations) all at once to display images and contents to the display-tiles for visualization. So, we have decided to purchase 6 sender licenses for the next 5 years from 2018 to 2024. Therefore, we will be spending \$5237 annually (which is reasonable) from 2018 to 2024 for the Hiperwall Master license, 7 Sender licenses for the 6 Sender nodes (Data Analysis workstations) and the Hiperwall Control node, and 8 Display licenses for the 8 Hiperwall Display nodes.

Before we decided to go with Hiperwall, we explored RGB Spectrum visualization system's MediaWall 4200 Display Processor for the same 8-tiled  $16' \times 4.5'$  ( $4 \times 2$  tile configuration) 16.5 MP ( $7680 \times 2160$  pixel) display wall. At the time, the cost of this expensive high-end display wall processor was \$50,000. The total cost of RGB Spectrum's MediaWall would have been about \$20,000 more than our budget amount of about \$200,000 (excluding shipping/handling and any miscellaneous expenses) for a comparable 8-tiled  $16' \times 4.5'$  ( $4 \times 2$  tile configuration) 16.5 MP ( $7680 \times 2160$  pixel) Hiperwall visualization system that includes the nine Dell T5600 Data Analysis Workstations using the same eight Samsung commercial-brand UD55C display-tiles.

Recently this year, the School of Computing at Weber State University set up a comparable 8-tiled 16.5 MP ( $7680 \times 2160$  or  $8K \times 2K$  pixel resolution) Visualization system from Extron Electronics [41], along with four T5600 Dell Data Analysis Workstations (each with 16 GB RAM and 0.5TB HD) at a cost of \$95,913. Their  $16' \times 4.5'$  display wall uses the Extron Quantum Connect Display Wall Processor [42] to power the 8 display-tiles that cost \$17,815, which is relatively inexpensive for a display wall processor. For the display-tiles, Weber State University purchased the latest 55" Panasonic TH55LFV70U commercial display ( $1920 \times 1080$  pixel) HDTV at \$5190 apiece, that has a Left/Top bezel size of 2.25 mm and Right/Bottom of 1.25 mm; so bezel-to-bezel size would be 3.5 mm (about 0.14") which is considered among the very best for a tiled-display wall. The main advantage is that there are no annual license fees for the Extron display wall which is a significant cost saving. This Extron display wall is a very good option for a control/surveillance command center. The setup can be easily expanded up to 14 display-tiles. Unlike Hiperwall, Extron is not the best display wall to conduct research tasks as a group to carry out visualization studies of images to study rare events in high resolution since zooming an image on the display wall across all the 8 tiles results in a reduction of the resolution image since Extron's software does not run on each display tile to render the images, instead projects the source images to the display wall. Even though, the Extron display wall has some limitations in retaining the resolution of zoomed images, it gives the best bang for the buck as a general-purpose display wall due to its ultra-narrow bezel size of the display-tiles, the relatively cheap display wall processor and no annual license fees.

In the US, in academia, in the past few years, Hiperwall display walls have been installed at Clemson University, La Sierra University, Gonzaga University, UC-Irvine, and Arizona State University. Hiperwall is also being used in Command and Control Centers at Santa Ana Police Department, Kansas City Police Department, Los Angeles Metro-Rail and Bus System, Western States Information Network (WSIN), and at the

US Department of Veterans Affairs. Notably, in entertainment, the Blue Man Group uses Hiperwall to enhance the entertainment experience to transform their digital signage at their Las Vegas venue and by the T3 Enterprise in Media and Advertising.

In recent years, some of the universities have already installed display walls (from other vendors) for teaching and research, notably, the University of Michigan (75 MP— $12' \times 7'$  in a  $3 \times 3$  tile configuration), Georgia State University's Collaborative University Research & Visualization Environment Lab (25 MP— $14' \times 4.5'$  in a  $6 \times 2$  tile configuration), Dartmouth College (16.5 MP— $16' \times 4.5'$  in a  $4 \times 2$  tile configuration), Brown University's Digital Lab (24 MP— $16' \times 7'$  in a  $4 \times 3$  tile configuration), University of Texas at Austin's Texas Advanced Computing Center (328 MP— $36.5' \times 8'$  in a  $16 \times 5$  tile configuration with 30" Dell monitors), Stanford University's Rumsey Center (33 MP— $16' \times 7'$  in a  $3 \times 3$  tile configuration), University of Oregon's Visualization Lab (50 MP— $24' \times 9'$  in a  $6 \times 4$  tile configuration), and Weber State University's School of Computing (16.5 MP— $16' \times 4.5'$  in a  $4 \times 2$  tile configuration).

In Table 3 we have provided a detailed comparison chart of Hiperwall and eight other leading display-wall products [43]. The cost comparisons are for the display wall hardware only. Based on the hardware and the features of these eight display wall products, only one company (Useful Corporation) best matches the Hiperwall's features and capabilities. Based on our assessment, the PC cluster-based distributed display software architecture to power the display-tiles is an improvement over the display wall Processor-based systems. As evident by the comparison chart in Table 3, Hiperwall's features and capabilities are broader than others.

In summary, the final cost will depend on the physical size of the display wall, the number of pixels that can be packed into the wall (pixel density, i.e. pixel count per screen area) based on the HDTV model, the bezel size of the displays, and any custom-made features added to the display wall.

### Big Data analytics and techniques

Today large data sources are ubiquitous throughout the world due to the dramatic increase in human and machine generated data [7]. Big Data is generated from the large Big Science experiments/instruments such as LHC in HEP and LIGO in Gravitational Wave Astrophysics, other measuring devices, radio frequency identifiers, social media and social network portals such as Facebook, Twitter, and Instagram, search engines like Google, meteorological and remote sensing instruments, and data from various mobile devices, including a plethora of audio and video recordings from various devices. Currently, there are a number of different Big Data analysis techniques in use in a wide variety of areas that generate Big Data that are primarily based on tools used in statistics and computer science [7].

The Big data analysis tasks involving visual analytics is an emerging area in visualization. A typical visual analytics problem is often complex and could require multiple data analysis and visualization task. Visual analytics of Big Data requires both automated analysis and interactive visualizations for effective understanding, reasoning and decision making on the basis of a very large and complex dataset [53, 54]. A number of Big Data Visual Analytics techniques have been developed over the years. The rapid technical developments have greatly enhanced the use of visual analytics techniques to solve



**Table 3** Tiled-display wall comparison chart [43]

	Hiperwall [44]	Christie [45]	RGB Spectrum [46]	Planar [47]	Barco [48]	Userful [49]	Datapath [50]	Jupiter [51]	Cinemassive [52]
Cost (\$ low) (\$\$\$ high)	\$	\$\$\$	\$\$\$	\$\$\$	\$\$\$	\$	\$\$\$	\$\$\$	\$\$\$
Max.source resolution	No limit	4K	4K	4K	4K	8K	4K	4K	4K
Multiple users	Yes	Yes	No	No	No	Yes	No	No	Yes
User authentication	Yes	Yes	No	No	Yes	Yes	No	No	No
KVM control	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Multiple video-wall control	Yes	No	Yes	No	No	Yes	No	Yes	No
Advanced scheduling	Yes	No	No	No	No	Yes	No	No	No
API	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
Local/remote access	Yes	No	No	No	No	No	No	No	No
Source processing	Yes	No	Yes	Yes	No	Yes	Yes	No	No
Unrestricted content manipulation	Yes	No	No	No	No	No	No	No	No
Multiple simultaneous sources	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Setup complexity	Low	High	High	Medium	High	Low	High	High	High
Server location and relay anywhere within LAN *(within 330')	Yes	Yes	Yes	Yes*	Yes	Yes	Yes*	Yes	Yes
Maximum number of displays	400+	128	56	16	64	100	32	N/A	N/A
Seamless 24/7 fault tolerance	Yes	Yes	No	No	Yes	No	No	No	No
Non-proprietary hardware (display, PCs)	Yes	No	No	No	No	Yes	No	No	No
Mix of display sizes and aspect ratios	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes

**Table 3 (continued)**

	Hiperwall [44]	Christie [45]	RGB Spectrum [46]	Planar [47]	Barco [48]	Useful [49]	Datapath [50]	Jupiter [51]	Cinemassive [52]
Mixed display orientation support	Yes	No	Yes	YesYes	No	Yes	Yes	No	Yes
Full desktop/browser capture	Yes	No	No	No	No	Yes	No	No	No
Native support for real-time content	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drag and drop content manager	Yes	No	No	Yes	No	No	No	No	Yes
Mobile device access	Yes	No	No	No	No	Yes	No	No	No
Bezel correction	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes

real-world problems. Due to the growing need for visual analytics, there is an increasing need for a comprehensive survey covering the recent advances of the field.

As the data acquisition and processing rates are growing, traditional data analytics methods are being replaced by advanced Big Data analytics and visualization techniques using Big Data technologies such as Statistical Analysis System (SAS) [55, 56] for data visualization from a business and corporate perspective. In the scientific arena, there are number of data-driven advanced data analysis techniques used today to analyze large amounts of data (Big Data) such as Neural Networks (NN) [57–61], Data Mining [62–65], Signal Processing [66–68], Visualization Methods [69–71], and Machine Learning [72–74] that includes various Multivariate Analysis (MVA) techniques such as Boosted Decision Trees (BDT) [75–84], Likelihood [81, 84], Artificial Neural Networks (ANN) [62, 85, 86], and Bayesian Neural Network (BNN) [81, 87–89].

In HEP, to search for new and rare subatomic particles in ever larger data sets at the LHC experiments at CERN indeed challenges our ability to analyze and visualize more scientific information in a timely manner than we had ever encountered in science where we need to extract the maximum available information from the datasets. Extracting rare “discovery” signals by separating (filtering) background events from a large number of parameters in the petabyte-scale datasets is indeed a very challenging task both in terms of data analysis and visualization at the TeV scale LHC energies.

The subatomic particles that decay within the detector, leaves tracks or signatures of their presence that are registered by converting the particle paths and energies into electrical signals to create a digital snapshot of the “collision event”. At the LHC, data recorded by the ATLAS detector corresponding to proton-proton annihilation events can typically consist of about 2000 measurements of position, deposited energy and timing for a large number of tracks, etc. During each collision, many new particles are created. After passing through the reconstruction algorithm, and after being combined

with calibration constants, these measurements are typically reduced to about 200 reconstructed objects, such as tracks from all decay products, electromagnetic clusters, hadronic clusters, etc.

With the emergence of such large and high dimensional data sets, the task of data visualization has become increasingly important in both machine learning and data mining [89]. So, visualization is crucial for analyzing and exploring such large-scale complex multivariate data. The footprints or electronic signals that are recorded by the detectors during a collision that can be visualized are called “Event Displays”. These Event Displays are very useful for pattern recognition of particle tracks. A visual representation of the data as an Event Display shows the correspondence between the data and the detector geometry that provides a snapshot of what has occurred inside the detector. These Event Displays are dynamic images that enable us to see the various parameters/features about the subatomic particle, and shows how these subatomic particles traveled through the detector. The recorded Event Displays of the collisions are stored in the XML file format. The visualization of these collision data is essential for conveying the underlying physics of these data events. The Event Display contains objects such as Tracks and Hits. They are usually represented by simple (Hits) or unique (Tracks) geometrical objects, which can be displayed in special (2D) projections. Detector Display contains Detector elements (Chambers, and the various Detector layers). They are usually represented by replicated complex geometrical objects.

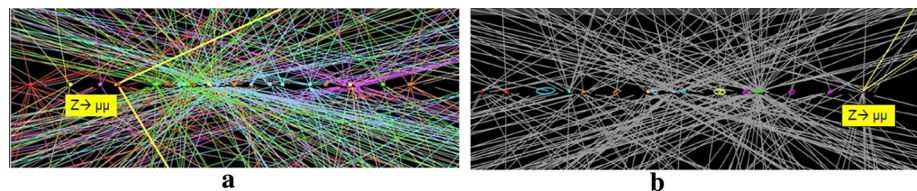
Visualization is especially useful when the information being portrayed is positional in nature and the spatial correlations are important. For example, color and other visual cues are employed to convey information about the amount of energy deposited at the spatial points. For the ATLAS experiment, visualization of the Event Displays can be done by the data analysis software packages/toolkits- ATLANTIS and HYPATIA (a student version of ATLANTIS) in 2-dimension and in 3-dimension using CAMELIA using a built-in XML event loader. All three visualization toolkits are a collection of software modules capable of displaying all types of data from the ATLAS experiment and have sophisticated graphics, GUIs, debugging tools and analysis capabilities for the visual investigation and physical understanding of all types of events inside the ATLAS detector.

All of these three data analysis toolkits are graphics packages for visual analytics that provides a graphical representation of what is happening in a collision event to understand the underlying physics processes inside the ATLAS detector which displays the trajectories of the particles produced by the proton-proton collisions. These toolkits allow us to interact with the data visually and study the traces of the particles in different parts of the ATLAS detector in order to discover new particles by reconstructing them from their decay products. These toolkits also help to develop reconstruction and analysis algorithm and are also used for debugging. These toolkits display the tracks together with their signature in the different parts of the ATLAS detector and provide information about the particles, and their energy etc. These toolkits also provide real-time interactive control of the display, processing and I/O of the data, and has all the basic modules of a complete visualization system that represent both the main data flow and the interactive links controlling program flow. The I/O streams are being accessed interactively.

One can select specific tracks for analyses, display event properties for detailed study, filter events, and display them in perspective, orthogonal and cut views. The reconstruction code combines data from the experiment with calibration constants and geometric information about the detector from the database into reconstructed objects such as clusters in the calorimeters. The user can interact “locally” to manipulate the rendering of the contents of the display list, rotating, shading, zooming the image, etc. Two-dimensional representations of the data are usefully employed to display a high resolution “fish-eye” view of the event so that a range of sub-detectors of vastly different sizes can be included in the same image panel [90, 91].

The Hiperwall visualization display wall allows us to carry out the visualization studies of the decay products of particle collisions using these data analysis software packages/toolkits installed in the workstations. In Hiperwall, these Event Display images can be displayed anywhere, within or across the display-tiles, in correct aspect ratio or stretched to fit, in whole or can be zoomed on specific features of an event to emphasize details and display its properties. A full array of features includes dynamic window sizing and positioning, smooth zooming within the images, custom borders, titling, programmable presets, and backgrounds. CAMELIA offers browsing and manipulating 3D data with full functionality for 3D display, such as 3D visual operations, like Rotation, Translation, Zooming, Scaling, Sheering and Skewing. With CAMELIA, we can pick a track in a current display and zoom in to show the position of the track with respect to the recorded hits and detector elements. We can also move the view in 3D to inspect the track from all angles and display the reconstructed properties and parameters of the fitted track.

At the LHC, pile-up (multiple interaction points) presents both a serious data analysis and visualization challenge when searching for new and rare types of subatomic particles from the complex and massive petabyte scale experimental datasets. At the LHC, decay particles containing electrons or muons provide clear trigger signals and rich event signatures for new particle searches but due to pile-ups, the identification of the correct interaction point to determine the origin of these electron and muon tracks is a challenge. Figure 13 shows the pile-up situation from the proton-proton collisions in the ATLAS detector at the LHC. The image in Fig. 13 that is part of the Event Display is a good example of visualization challenges in HEP. We are looking for only two particle tracks- a pair of muon and an anti-muon track from the decay of the Z Boson ( $Z^0 \rightarrow \mu^- \mu^+$ ) from just one interaction point where the rest of the tracks are simply background events. To complicate the scenario that creates a visualization challenge is



**Fig. 13** ATLAS Pile-up. **a** and **b** shows the pile-up (multiple interaction points) situation from proton-proton collision at the LHC inside the ATLAS detector in a region that is only about 5 cm wide. The pile-up in **a** has 25 interaction points and in **b** has 17 interactions points, where only one interaction point produced the decay tracks from the Z Boson. Image courtesy ATLAS Collaboration at CERN

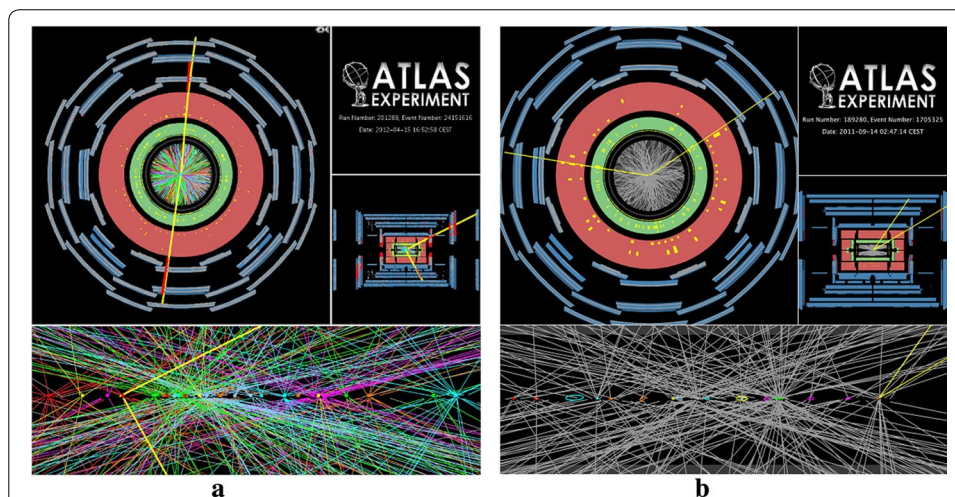
that very small (tiny) fraction of these background tracks are what we call “fake” muon-antimuon pairs, since these have decayed from hadrons (which are called hadronic jets) and are not from the Z Boson. So these “fake” muon-antimuon tracks which are part of the background events have to be filtered out. Such a scenario requires a high-resolution display wall such as Hiperwall which can assist with the challenging visualization task to filter out the signal events from the myriad background events.

Using the Hiperwall, we are able to zoom in to this 5 cm region of the detector to a size of  $16' \times 4.5'$  using all 8 display-tiles to create a 16.5 MP image without losing any resolution to determine and identify the correct particle track pairs (a muon and an anti-muon) from the decay of the Z Boson. Figure 14a and b show the full 2D Event Display of the corresponding to the Fig. 13a and b.

#### Signal extraction from large heterogeneous background events using Multivariate Analysis (MVA) techniques in HEP

Large-scale data driven applications are on the rise, and so is the need to extract useful information from terabyte to petabyte scale data sets (Big Data). The main goal in data analysis is to obtain the maximal information and extract the best possible results from the data. Multivariate analysis (MVA) methods are discriminating techniques for extracting small signals based on machine learning to separate the signal from the background data in very large datasets that are very useful in fields that generate Big Data.

Each MVA technique is well characterized mathematically and use different algorithms to approximate the same mathematical object. One of the reasons for applying MVA techniques is simply the lack of knowledge about the mathematical dependence of the quantity of interest on the relevant measured variables. So either there is no mathematical model at all and a detailed analysis is the only possibility of finding the correct dependence, or the known models are simply insufficient and MVA techniques provides a better description of the data.



**Fig. 14** **a** and **b** shows the visualization of a full 2D Event Displays of the corresponding Fig. 13a and b using the ATLANTIS toolkit. Image courtesy ATLAS Collaboration at CERN.



In HEP, for the large-scale data analysis and visual analytics tasks to search for new subatomic particles, Machine Learning and MVA techniques are becoming ever crucial in extracting rare signal events (new subatomic particles) from huge number of background events in the complex experimental HEP environments. As HEP data analysis is becoming more and more challenging, MVA methods are routinely being used to suppress background and for parameter estimation (regression), where a physical quantity is extracted from a set of directly measured observables. Typically one in a trillion data events will lead to a new particle discovery, such as the recent discovery of the Higgs Boson at the LHC. The main challenge is to identify the rare signal events involving multiple processes in the presence of pile-ups and huge background events from the experimental data sets that contains over 100 trillion events with a plethora of multiple and complex heterogeneous variables. The data analysis process, becomes even more challenging since some of the background events can look exactly like the signal events, which are “fake” signal events (signal-like events coming from various processes such as from other decay channels). The process of data analysis and visualization in HEP is indeed like a detective work. It's like *finding needle in a hay-stack*. The key task is to intelligently search for the specific particle of interest by distinguishing it from the background particles and processes by filtering them using numerous detector filter parameters, without throwing out any of the particle of interest (i.e. signal events).

Moreover, pile-up (multiple interaction points) as shown in Fig. 13 presents a challenging visualization environment. Due to the high resolution of the Hiperwall visualization display wall, we are able to zoom into this pile-up region of the HEP Event Displays and identify the right interaction point of these particle tracks. To obtain best possible results it is important to make use of maximum possible information in the data and hence employ optimal multivariate MVA methods for data analysis.

So, for the large-scale HEP data analysis tasks, MVA techniques are extensively being used for signal optimization studies in order to extract useful information about these rare and new subatomic particles from the myriad background events, thus maximizing the chances of a discovery. We have used MVA techniques to guide our signal optimization and visualization studies. The main idea of the MVA is the choice of input variables, where we can simultaneously analyze multiple measurements (variables) on the object being studied. These variables can be correlated in various complex ways.

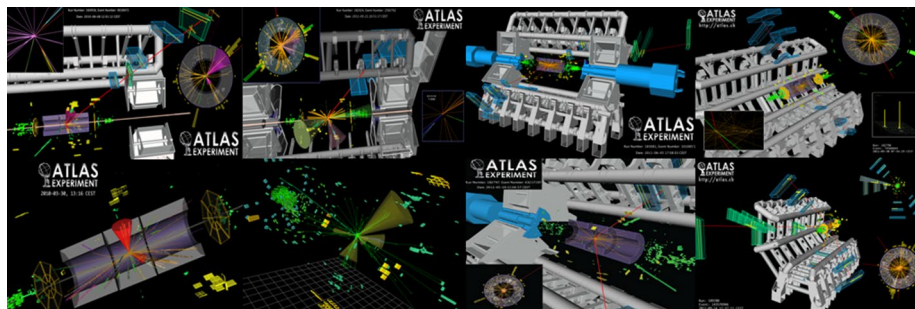
MVA techniques are particularly useful when dealing with large experimental data sets where multi-dimensional features of the data could be missed in a conventional data analysis. In conventional statistical techniques, one starts with a mathematical model and finds parameters of the model either analytically or numerically by using some optimization criteria. MVA addresses the problem of signal/background classification through supervised machine learning. Starting with samples of data of known type (signal or background), the data are divided randomly into training and testing sets, each of which consists of a mix of signal and background events. MVA techniques are also “trainable”, i.e. it can be designed to self-organize and optimize the classification ability by analyzing events of a known class. This trait makes MVA techniques very efficient for optimization of various parameters in order to separate signal from the myriad complex and heterogeneous background events. For the data analysis tasks, MVA techniques help to suppress the background events without further reducing the signal events.

The training data set is used to fix the parameters of the MVA classifier, a function that assigns to each input event a measure of its consistency with the signal or background hypotheses. The parameters are chosen to achieve the best separation between the signal and background training events. The trained classifier is then applied to the testing data set to obtain an unbiased evaluation of the classifier's detection performance from the number of correctly classified signal and background events. MVA techniques can mine the full parameter space of the events to better discriminate between signal and background.

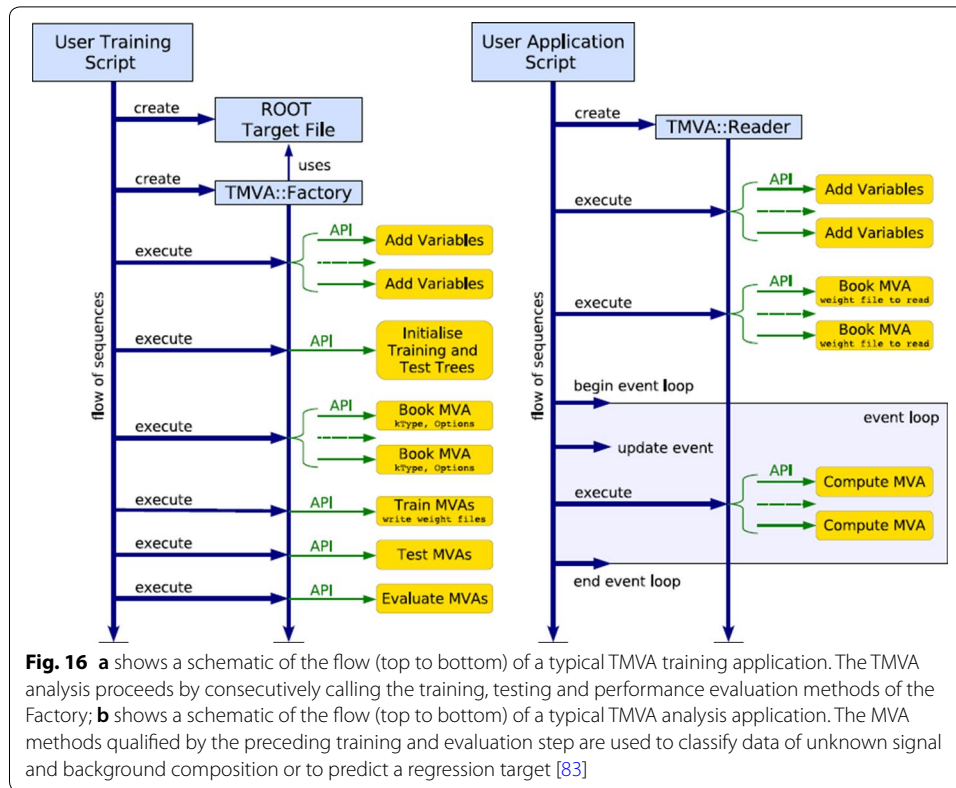
Subatomic particles that decay to produce multiple hadronic jets (e.g. quarks from Mesons and Baryons) such as in the decay of the Top quark and in the search for the charged Higgs ( $H^+$ ) Boson, a high resolution visualization display wall that can zoom into the various parts of the detector in high resolution to identify the particle tracks is crucial during data analysis. Using the Hiperwall, we can simultaneously analyze up to eight Event Displays at a time (each with a resolution of 2 MP) one on each of the eight display-tiles (as shown in Fig. 15) or two Event Displays one at a time (each with a resolution of 8 MP) on all eight of display-tiles (i.e. one event display per four display-tiles) or 1 event displays on at a time (with a resolution of 16 MP) on all eight of display-tiles.

Integrated into our HEP data analysis framework ROOT [92] is the Toolkit for Multivariate Analysis (TMVA) [83] which is designed for machine learning applications in HEP. The TMVA toolkit unifies highly customizable MVA classification and regression algorithms in a single framework. The latest version, TMVA 4.0.1 also features more flexible data handling allowing one to arbitrarily form combined MVA methods [83]. TMVA is specially designed for the complex HEP data analysis environments.

In this paper we will describe the TMVA framework for our data analysis tasks in HEP. TMVA provides an environment for the processing, parallel evaluation and application of multivariate classification and multivariate regression techniques, where all multivariate methods respond to supervised learning, i.e. the input information is mapped in feature space to the desired outputs [83]. The mapping function contains various degrees of approximations and may be a single global function, or a set of local models. TMVA provides an unbiased performance comparison between the various MVA methods using the same data. Training, testing, performance evaluation algorithms and visualization scripts of all available classifiers can be carried out simultaneously via the TMVA user-friendly interfaces. TMVA analysis consists of two steps [83] as shown in Fig. 16.



**Fig. 15** Shows eight event displays projected on Hiperwall's eight display-tiles



1. *Training phase* The ensemble of optimally customized MVA methods are trained and tested on independent signal and background data samples; the methods are evaluated and the most appropriate (performing and concise) ones are selected. In the training phase, the communication of the user with the data sets and the MVA methods is performed via a Factory object that is created at the beginning of the program. The user script can be a ROOT macro [92], C++ executable, or a python script. The TMVA Factory provides member functions to specify the training and the targeted data sets, to register the discriminating input variables, and to book the multivariate methods selected by the user. Subsequently the Factory calls for training, testing and the evaluation of the booked MVA methods. Specific result (“weight”) files are created in XML format after the training phase by each booked MVA method and the evaluation histograms are stored in the output file [83].

2. *Application phase* Selected trained MVA methods are used for the classification of data samples with unknown signal and background composition, or for the estimation of unknown target values (regression). The application of training results to a data set with unknown sample composition (classification)/target value (regression) is governed by the Reader object [83]. During initialization, the user registers the input variables together with their local memory addresses, and books the MVA methods that were found to be the most appropriate after evaluating the training results. As booking argument, the name of the weight file is given [83]. The weight file provides for each of the methods full and consistent configuration according to the training

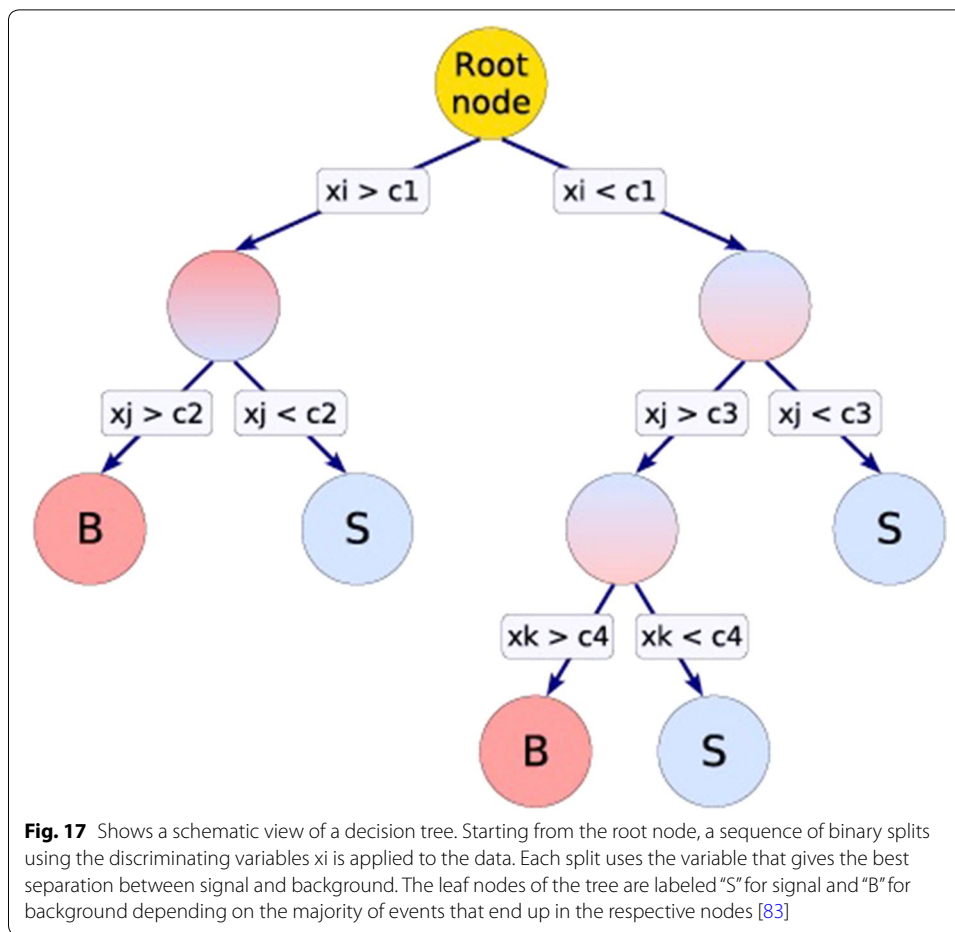
setup and results. Within the event loop, the input variables are updated for each event, and the MVA response values are computed [83].

During the Factory running, standardized outputs and dedicated ROOT [92] macros allow a refined assessment of each method's behavior and performance for classification and regression. Once the appropriate MVA method(s) have been chosen by the user, it can be applied to data samples with unknown classification or target values. The interaction with the methods occurs through a Reader class object created by the user. A method is booked by giving the path to its weight file resulting from the training stage [83]. Then, inside the user's event loop, the MVA response is returned by the Reader for each of the booked MVA method, as a function of the event values of the discriminating variables used as input for the classifiers. Alternatively, for classification, the user may request from the Reader the probability that a given event belongs to the signal hypothesis [83].

In HEP, data analysis tasks, typically include (i) classification (the process of assigning objects or events to one of the possible discrete classes), (ii) parameter estimation (extraction and measurements of track parameters, vertices, physical parameters such as production cross sections, branching ratios, and masses) and (iii) function fitting. Classification of objects or events is, by far, the most important analysis task in HEP, such as for example the identification of Electrons, Photons, Muons, Tau-leptons, etc. and the discrimination of signal events from those arising from background processes [81]. Optimal discrimination between classes is crucial to obtain signal-enhanced samples for precision physics measurements. In HEP data analysis, MVA techniques are mainly used for signal-to-background discrimination, variable selection (finding variables which give the maximum signal/background discrimination), finding regions of interest in the data and measuring parameters (regression). MVA algorithms can be also classified according to two main classes: Supervised Training, where a set of training events with correct outputs is given, and Unsupervised Training, where no outputs are given and the MVA algorithm has to find them.

Boosted Decision Tree (BDT) analysis is similar to the machine learning technique. Let's first start with a Decision Tree (DT) which is a binary tree structured classifier with two terminal nodes called leafs- the Signal "S" leaf and the Background "B" leaf. DT classifies events by following a sequence of decisions depending on the events variable content- S or B leaf node. Repeated "Pass/Fail" decisions are taken on one single variable at a time until a stop criterion is fulfilled [83]. The "tree" essentially provides a flowchart describing how to classify an observation. We start at the root of the tree, and the leaf ends up determining the classification as shown in Fig. 17.

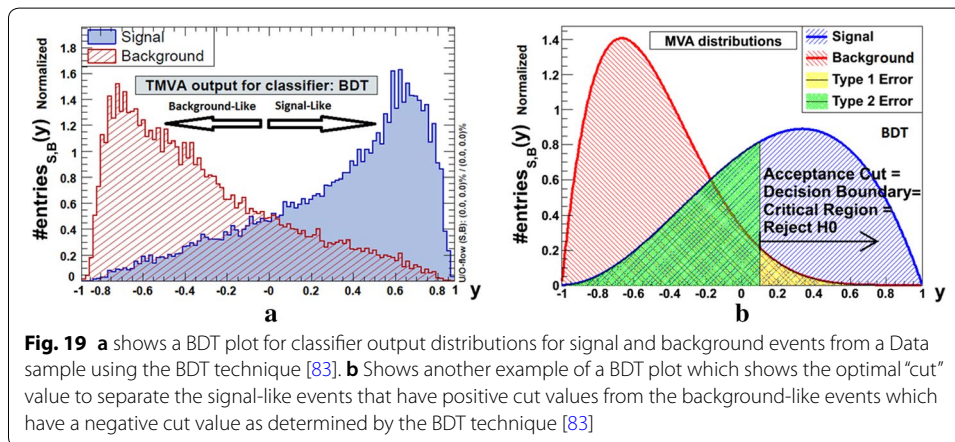
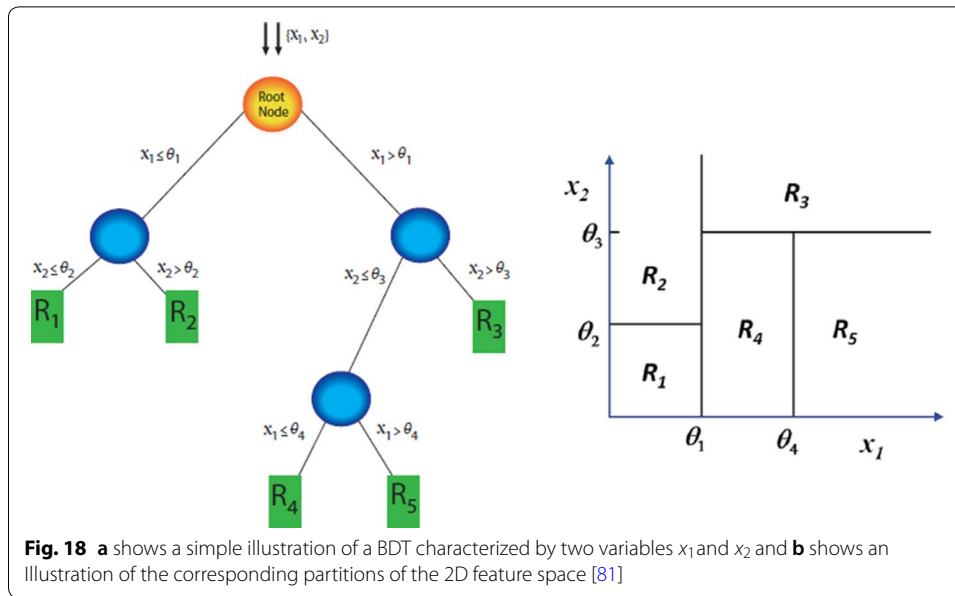
The decision trees are robust in many dimensions but needs a boost and hence referred as a Boosted Decision Tree (BDT). The boosting of a decision (regression) tree extends from one tree to several trees which form a forest of decision trees and an event is classified on the majority outcome of the classifications done by each tree in the forest [83]. In BDT, the training and building of a decision tree that starts with the root node is the process that defines the splitting criteria for each node, where an initial splitting criterion for the full training sample is determined. The split results in two subsets of training events that each go through the same algorithm of determining the next splitting iteration which is repeated until the whole tree is built [83]. At



each node, the split is determined by finding the variable and corresponding cut value that provides the best separation between signal and background. The node splitting stops once it has reached the minimum number of events which is specified by the person performing the data analysis. The leaf nodes are classified as signal or background according to the class the majority of events belongs to.

All trees in the forest are derived from the same training sample, with the events being subsequently subjected to so-called boosting, a procedure which modifies their weights in the sample. In BDT, the tree is split into many regions that are either classified as signal or background, depending on the majority of training events that end up in the final leaf node [83]. Boosting is done by re-weighting events in such a way (a weighted average of the individual decision trees) that higher weight is assigned for misclassified events and a lower weight for correctly classified events. Boosting thus increases the statistical stability of the classifier and also improves the separation performance compared to a single decision tree [83]. Once boosting is done it re-trains and repeats the given set of classifiers thus classifies data by a weighted vote of the classifiers. Boosting stabilizes the response of the decision trees with respect to fluctuations in the training sample and is able to considerably enhance the performance with respect to a single tree [83]. With the BDT technique, we are able to





detect signals that have different morphologies to those used in the classifier training, and improve on the false signal probabilities.

A simple illustration of a BDT is shown in Fig. 18 [81]: (a) characterized with two feature variables  $x_1$  and  $x_2$  with the resulting partition of the 2D feature space are shown in the schematics in (b).

BDT is especially useful to recover events that fail criteria in a traditional data analysis. In the BDT technique, for each variable, we first have to find splitting value (called a “cut”) with best separation between two “entities”—mostly signal (signal-like) in one, and mostly background (background-like) in the other as shown in Fig. 19a and b. The terminal node leaf is assigned a *Purity Value* =  $S/(S + B)$ . So the BDT output for each event is determined by the leaf Purity value which is closer to  $-1$  for background and closer to  $1$  for signal. The total of all trees in BDT is combined into a “score”. Once a variable (data parameter) is selected, it is then split into two values with the best separation to produce two “branches”. This process is repeated recursively on each node, which ends

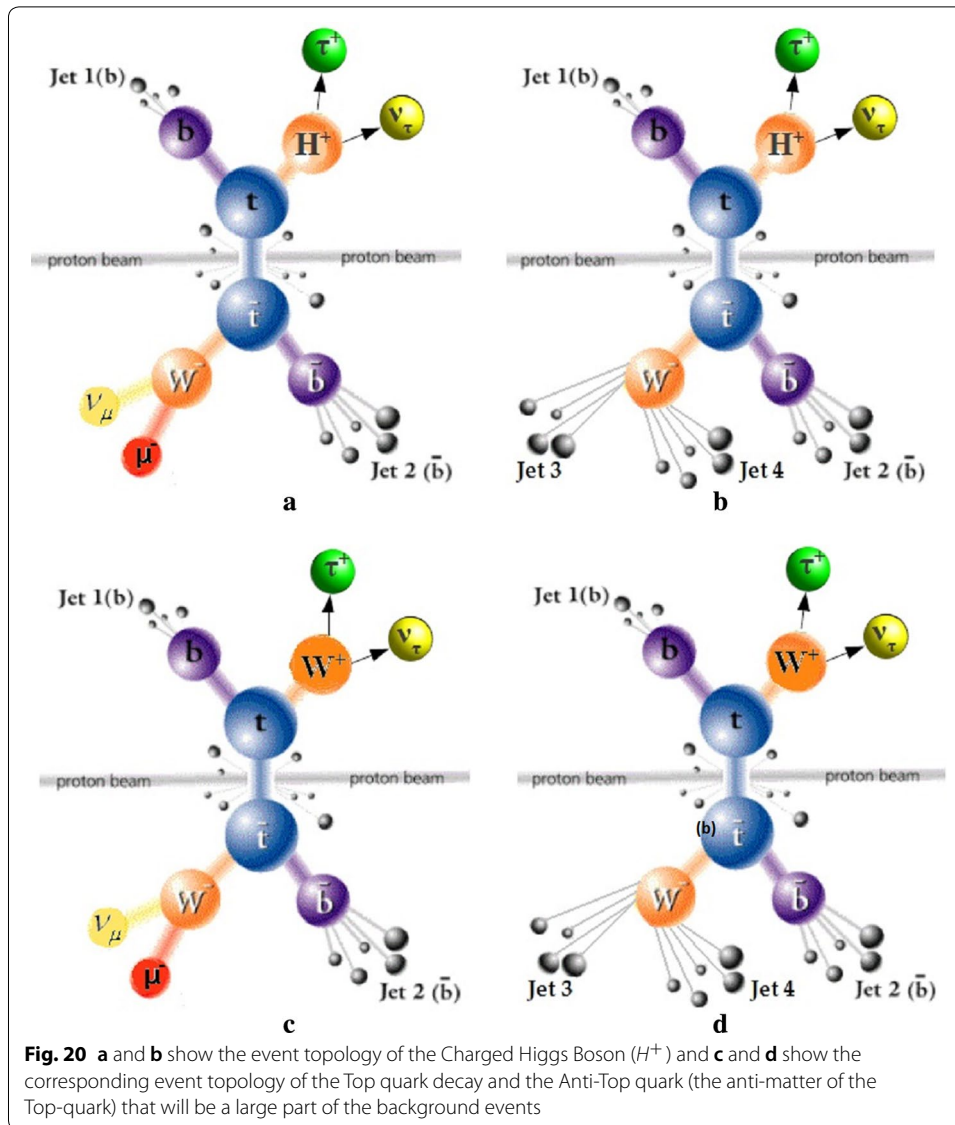
when there is no longer an improvement in the signal efficiency or when too few events are left. During the data analysis phase, we need to display this BDT plot of the datasets in high resolution using the full display-tiles so that we can accurately determine this splitting (“cut”) value that best separates the signal events from the background events. In that regard, high resolution visualization of the BDT plot is a critical task during data analysis.

By TMVA convention, signal events have a large classifier output value, while background events have small classifier output values. Hence, retaining the events with splitting value larger than the cut requirement selects signal samples with efficiencies that decrease with the cut value, while the significance of the signal increases. Since the splitting criterion is always a cut on a single variable, the training procedure selects the variable and cut value that optimizes the increase in the separation index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events [83]. In principle, the splitting could continue until each leaf node contains only signal or only background events.

BDT is used for signal optimization in our data analysis to determine the optimal filter parameters in order to efficiently extract signal events from the plethora of complex and heterogeneous background events in LHC datasets by improving the Signal to Background ratio. BDT allows us to study the complicated patterns and parameters in the datasets and make intelligent decisions about the signal extraction by finding new ways to tune and to combine classifiers for signal optimization and gain in performance.

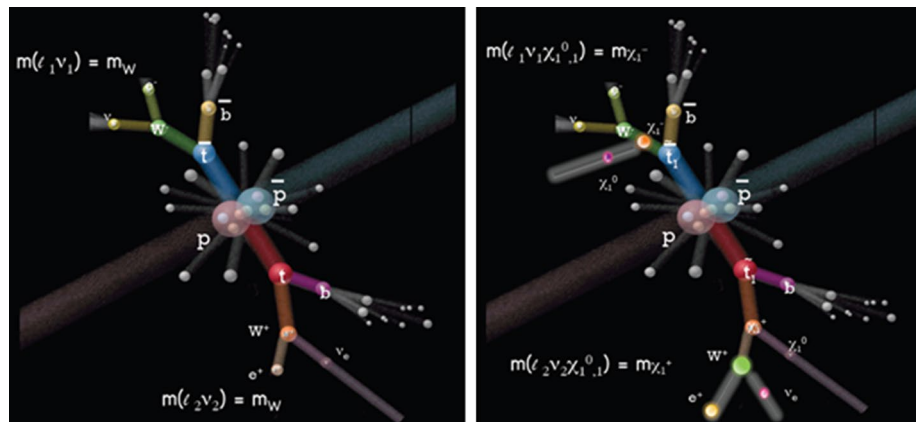
In addition to the recently discovered Higgs Boson, according to Minimally Supersymmetric Standard Model (MSSM) [93–96], there ought to also be a charged Higgs Boson. Finding the charged Higgs Boson would be an enormous scientific breakthrough. Undoubtedly, the observation of the charged Higgs Boson, would indicate physics beyond the Standard Model and provide pathways to new frontiers in HEP. The MSSM models predict that the charged Higgs Boson will decay into a Tau Lepton and a Tau-Neutrino as long its mass is less than the mass of the Top quark. The data analysis of the charged Higgs Boson is immensely challenging since the LHC produces a lot of Top quark events which will be a large background for the charged Higgs Boson search. So sophisticated MVA techniques such as Boosted Decision Trees (BDT) are needed to separate signal from background, by maximizing the signal acceptance. Figure 20a and b show the event topology of the charged Higgs Boson and Fig. 20c and d show the corresponding event topology of the Top quark decay, which mimics the charged Higgs Boson signal, and are therefore background events to the charged Higgs Boson search. The main difference between the data sample in the BDT plot in Fig. 19 and a data sample in the BDT plot of a charged Higgs Boson is that the background events will be overwhelmingly large while the signal events will be very small.

Another example of the usefulness of the Hiperwall display system is for the search of the lightest Supersymmetric (SUSY) particles (LSPs) [97–111], such as the Neutralinos and the Charginos which remedies many of the shortcomings of the Standard Model in HEP involving the instability of the mass of the Higgs Boson. Fig. 21a shows a Top quark decay at the LHC which has strikingly very similar visualization characteristics of the lightest SUSY particles (LSPs), the Neutralinos and the Charginos, as depicted by the subtle purple and orange dots (for visualization purposes), respectively, as shown

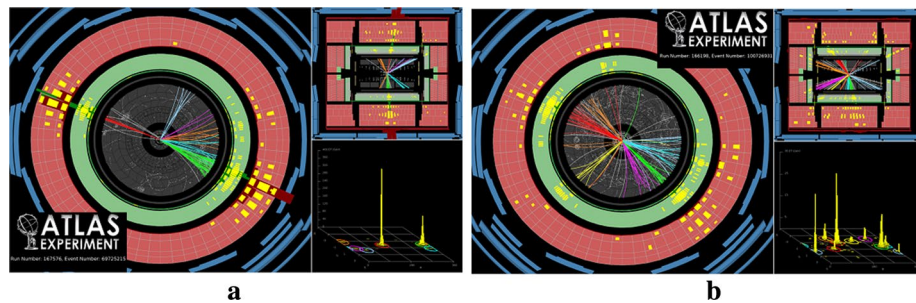


in Fig. 21b. To detect these rare subatomic particles, the Neutralinos and the Charginos, the Top quark decay as depicted in Fig. 21a would be an overwhelming background when searching for these very rare SUSY particles-Neutralinos and the Charginos in the LHC data. Such complex data visualization studies for this search requires a high resolution visualization system like Hiperwall.

Another example of the usefulness of the BDT technique is for the separation of the different hadronic jets that are produced from the Top quark (t-quark) and other subatomic particles, where we need to separate the hadronic jets that are coming from the Up, Down and the Strange quarks (u, d, and s quarks) called the “light-jets” from the hadronic jets that are coming from the heavy quarks- the Beauty and the Charm quark (b and c quarks), which is a very challenging task. We use a BDT based technique called *b-tagging* [112–114] to identify a jet associated with a b-quark by reconstructing the decay vertex of a B-hadron within the jet. It turns out that the b-quarks



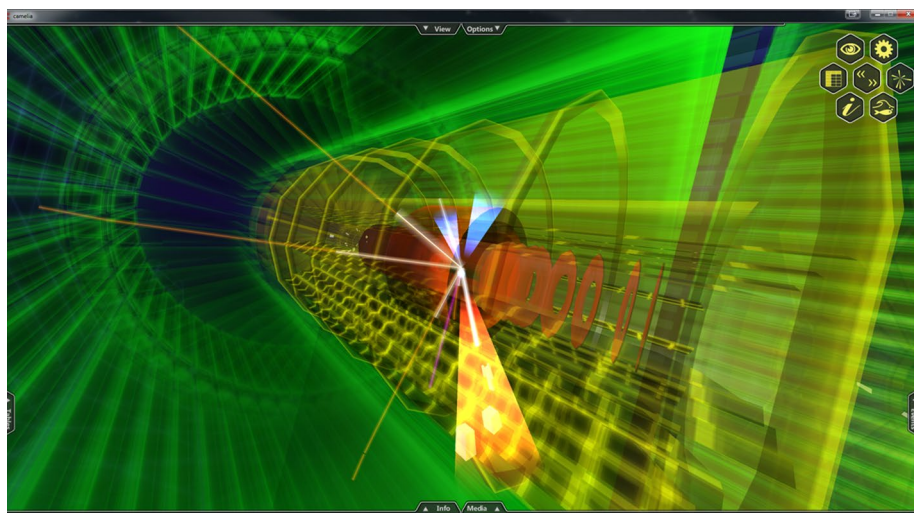
**Fig. 21** **a** shows the visualization characteristics of the Top quark decay at the LHC and **b** shows the visualization characteristics of the lightest SUSY particles (LSPs), the Neutralinos and the Charginos, as depicted by the subtle purple and orange dots, respectively. The Neutralinos and the Charginos signal events as depicted in **b** needs to be extracted from the background Top quark decay events using the BDT Multivariate Analysis (MVA) technique. Image courtesy ATLAS Collaboration at CERN



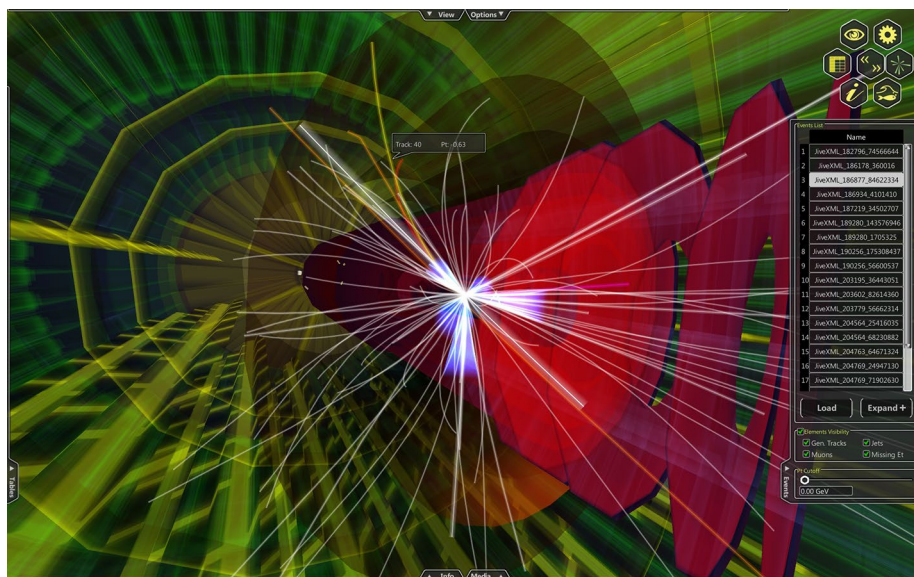
**Fig. 22** **a** and **b** show ATLAS Event Displays of a subatomic particles that has produced different jets (shown in different colors) which has been tagged using the BDT-based *b*-tagging technique. Image courtesy ATLAS Collaboration at CERN

form B-hadrons that result in b-jets, which then decay into lighter hadrons that result in the light-jets. This complex task of the identification of b-jets in presence of pile-ups is a major visualization challenge in ATLAS. Since decays of the Top quark, Higgs Boson, and SUSY particles involves decay to b-quark, we need to tag the jets that come from a b-quark and from all other background tracks and light-jets. The identification of b-jets originating from b-quarks is an effective way to reject backgrounds events. Figure 21a and b show a sketch of two Top quark decay channels that shows the tagged b-jets from the other jets using the BDT-based *b*-tagging technique. Figure 22a and b show ATLAS Event Displays of a subatomic particles that has produced different jets (shown in different colors) which has been tagged using the BDT-based *b*-tagging technique. Figure 23 shows a 3D visualization using the CAMELIA toolkit displayed on the Hiperwall that shows a subatomic particle decaying into several hadronic jets and leptons inside the ATLAS detector and Fig. 24 shows a zoomed 3D visualization using the CAMELIA toolkit displayed on the Hiperwall that shows a





**Fig. 23** A 3D visualization using the CAMELIA toolkit displayed on the Hiperwall that shows a subatomic particle decaying into several hadronic jets and leptons inside the ATLAS detector



**Fig. 24** A zoomed 3D visualization using the CAMELIA toolkit displayed on the Hiperwall that shows a subatomic particle decaying into several hadronic jets and lepton in presence of large background events inside the ATLAS detector

subatomic particle decaying into several hadronic jets and leptons in presence of large background events inside the ATLAS detector.

#### Collaborative research and learning environment using Hiperwall

As an extension of the desktop display environment, tiled displays are still an emerging technology for constructing custom-made visualization environments capable of displaying high-resolution images and visualizing complex data. Tiled display technologies



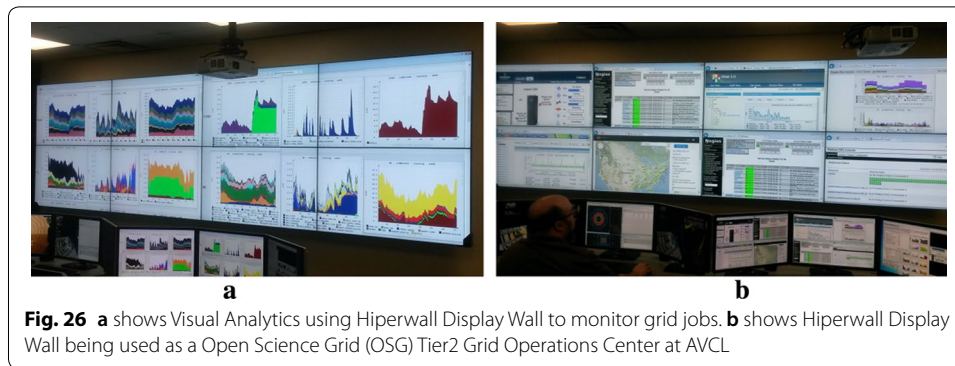


**Fig. 25** Students interacting with the high-resolution image displayed on the Hiperwall to determine and identify the specific particle tracks from the decay of the Higgs Boson from an ATLAS Event Display

offer a wide range of opportunities for exploring scalable large displays for high-resolution scientific visualization. A high-resolution display wall can show the details much more clearly than a low-resolution display of the same physical size that could be critical in decision making and scientific discovery. These next-generation tiled-display walls are increasingly becoming affordable for building visualization display environments.

The implementation of the Hiperwall visualization system has facilitated an advanced visualization and data analysis environment beyond what is typically possible on desktop systems. In a single desktop monitor, we are limited by the resolution of a single display. The total viewing space (area) of the Hiperwall display wall is over 30 times larger than a 24" desktop monitor and the total resolution of the Hiperwall display wall is eight times that of a typical 24" desktop monitor. At BU, the Hiperwall visualization system has made a significant impact and has indeed created an inspiring vision of the future of interactive visualization studies.

Using Hiperwall, our users, mainly undergraduate students can resize images and juxtapose multiple windows to run applications and conduct interactive visualization. Students can display word documents, excel spreadsheets, simulations, programming code, webpages, images, and videos, including video conferencing, on different display-tiles of the Hiperwall simultaneously. Nowadays, visual representation of Big Data is crucial for its interpretation. In that regard, Hiperwall has enabled students to display a lot of related information from various sources or applications at the same time that is not possible with a typical desktop monitor. Using Hiperwall students can visualize the ATLAS Event Displays in 3D, at a level of detail that is impossible to obtain on a single desktop monitor. Figure 25 shows students interacting with the high-resolution image displayed on the Hiperwall to determine and identify the specific particle tracks from the decay of the Higgs Boson from an ATLAS Event Display. In HEP, these Event Displays often provide the only means of communication between the complex and abstract world of particle physics and the general public. So, visualization is a valuable educational and public relations tool. As shown in Fig. 26, we have also used the Hiperwall



visualization System as an OSG Tier2 Operations and Command center at BU, like the NASA mission Operations/Command center.

Visualization is increasingly becoming a requirement for computational-enabled discovery and is an essential component of data-intensive research fields, where collaborative visualization of very-large datasets across distributed sites is becoming ever increasingly important for making new discoveries. Visualization can enable us to analyze and understand unprecedented amounts of experimental and simulated data. Indeed, human-computer interaction is one of the most basic features of scientific visualization. The insights provided through the interrogation of the data and the visualization of the observed objects will help in scientific discoveries, create new theories, techniques, and methods. Also, visualization facilitates learning, and increases communication among research collaborators.

In fields such as HEP and Astronomy, complex scientific data needs to be presented on a large screen with high enough resolution so that the finest of details are preserved. Current HEP datasets are too large to understand in any way other than statistical analysis and visualization. So, HEP experiments must rely on sophisticated simulation and visualization systems for exploring large-scale, complex, multidimensional systems of subatomic particles. This is where our Hiperwall is very useful and also impressive, since in a tiled display, each tile displays a section of the total image which is displayed in high resolution. Interactive visualization requires high-resolution display walls. Using Hiperwall users can place and arrange images/objects in a desired location using anywhere from 1 to 8 display-tiles within the embedding space/frame of the display wall and can adjust the images to their preference for the interactive visualization. Tiled displays can increase both the quality and the granularity of displayed information and allow more information, more applications, and high-resolution images than that can be viewed on a typical single desktop monitor.

From the perspective of the students, the Hiperwall visualization system has made a significant impact to conduct their data analysis tasks. Before we implemented the Hiperwall visualization system, we only had just one Dell Precision T7500 workstation with dual 24" monitors to conduct the data analysis tasks. We had issues when multiple full-Events Displays did not fit on the two adjacent 24" desktop monitors. Moreover, with just two desktop monitors, the complex physics processes in the Event Displays were too inadequate to be understood due to the limited viewing space of the 24" desktop monitors.

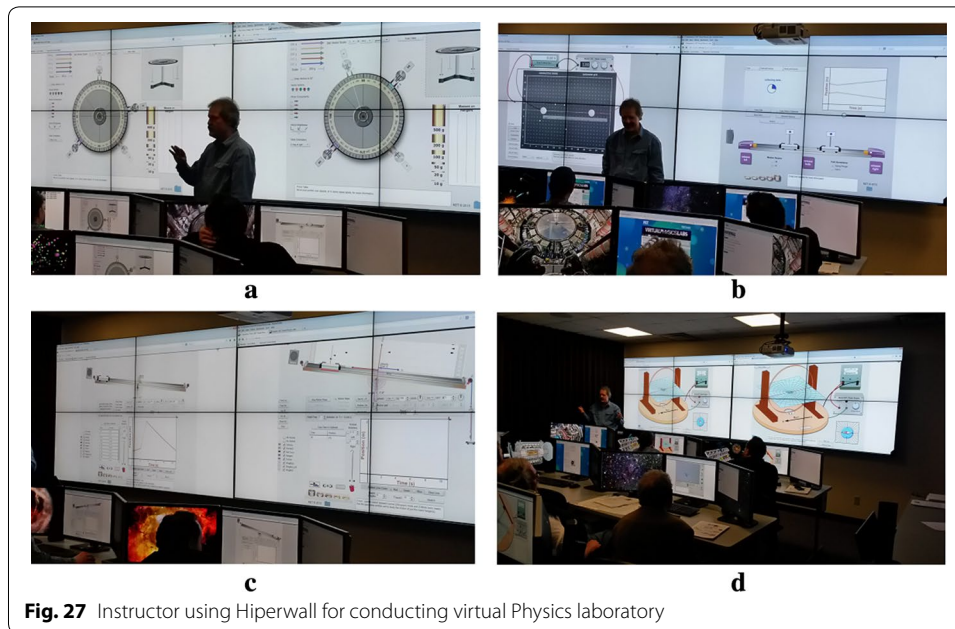
While the size of various datasets is still increasing, how to filter the large HEP datasets while visualizing them is a visualizing challenge. As shown in Fig. 20a and b, the discovery of the rare charged Higgs Boson will rely on correctly observing the decay patterns of the charged Higgs Boson from that of the background processes/events that look similar to the charged Higgs Boson signal, which is in itself a visualization challenge. This is where our 16.5 MP 16'  $\times$  4.5' Hiperwall visualization system is very useful, since without the use of high resolution display wall, the identification of charged Higgs Boson signal events by filtering and separating these background events from the large terabyte-scale datasets can be a daunting task. Therefore, the acquisition of this Hiperwall display has become an invaluable visualization platform for enabling scientific discovery. Using the Hiperwall visualization system, we can interact with the data visually and interpret the data efficiently, which can accelerate the discovery process. Using Hiperwall, our students, are developing expertise in large-scale data analysis and visualization skills and techniques that will be highly valuable for their future careers as they become part of the 21st century STEM workforce.

This research team and our students have greatly benefited from having a large high-resolution visualization system that also enabled us to see in real-time Event Displays that are generated at CERN on BU's Hiperwall. The Hiperwall visualization system has indeed created an exciting atmosphere for undergraduate students to participate in data-intensive research and has created a multi-user collaboration environment with a sense of visual immersion.

At BU, using the Hiperwall Visualization System we have implemented the first classroom-based display wall system in a Physics Department at an undergraduate institution. Hiperwall has provided a rich and immersive learning environment for our students which has enabled sharing of information between the students' data analysis workstations and the Hiperwall. Contents from student workstations can be displayed on the Hiperwall while they collaborate in small groups. The instructor also has the option to display the content of any of the student workstations on the Hiperwall at any time thus allowing the entire class to watch and discuss each others' work. So due to Hiperwall's flexibility, it enables the class to work collectively on challenging problems, as well as data analysis.

The Hiperwall visualization system has been instrumental in terms of both innovative and collaborative approach to interactive classroom and laboratory instruction. Using the Hiperwall, we have tested the setup to successfully conduct virtual Physics labs and virtual Physics simulations, animations and demonstrations as shown in Fig. 27. Using the Hiperwall visualization system, we can thus develop on-line virtual Physics labs for lower-division Physics courses.

The Hiperwall visualization system at BU has made a significant impact at this institution. The acquisition of this visualization system has provided an exceptional opportunity to engage undergraduate STEM students in research involving Big Data that is at the forefront of data-intensive computational science, visualization and HEP. The Hiperwall system has created a state-of-the-art and a dynamic/interactive visualization environment by engaging our students in inquiry-based visual learning, and exploratory data analyses activities. The Hiperwall technology has allowed students and researchers to study and visualize large-scale data events with high resolution



**Fig. 27** Instructor using Hiperwall for conducting virtual Physics laboratory

and conduct interactive analysis of very large datasets. One of our goals is to make visualization tools more transparent to the undergraduate students and ease usability and accessibility, and make it possible for a wider range of users to explore the large volumes of data generated by experiments in HEP.

We are continuing to train more undergraduate students on how to use the Hiperwall system for large-scale data analysis and visualization. Our plan is to convert the AVCL that houses the Hiperwall visualization system into a university-wide facility so that faculty and students from other departments that have visualization needs can use the Hiperwall display system. In fact, the School of Environmental Science is interested in using the Hiperwall visualization system to teach their Geographic Information System (GIS) course.

We have written a User's Manual for the Hiperwall Visualization system with input from undergraduate students who are our primary users. We have also designed and developed a web-portal interface that allow users to select specific display-tile arrangements for visualization at the click of a button, based on what is displayed on the user's Data Analysis Workstation's (sender/streamer nodes) triple monitors.

### Challenges while using the Hiperwall in educational and research pursuits

While we have had success in using Hiperwall for both education and research at BU, there have been some challenges for the users. In this section, we will discuss some of these challenges and our solutions to overcome these challenges.

First, and most important, the control scheme for the Hiperwall software is not user-friendly for the average user from the workstations (the sender and streamer nodes as shown in Fig. 1). Typically, an experienced Hiperwall user needs to actually sign in remotely to the Hiperwall control node in order to control the display contents on the Hiperwall (as shown in Fig. 1). While in the last couple of years Hiperwall has added a



control panel-like app that is accessible from the workstations, this is still a non-intuitive control scheme. Every semester we have trained new incoming students and the non-intuitiveness of the Hiperwall control environment was evident. The user of the sender/steamer nodes must decide what they want to display and then to actually implement the broadcast, they then need someone—usually another experienced user—to log into the control node. Then the experienced user needs to control the Hiperwall software to implement how it is displayed on the main screen.

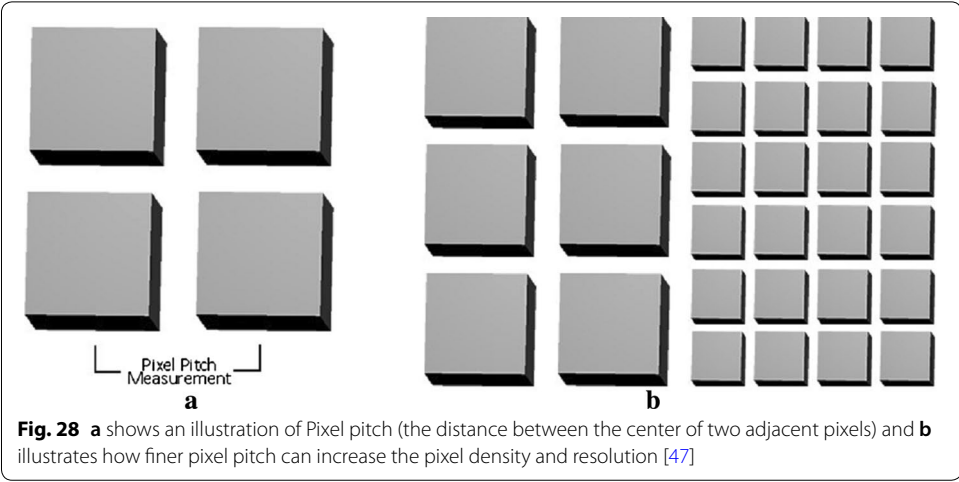
Second, the ability to control multiple workstations with a more intuitive control scheme (as described in section Hiperwall user portal) was written by our team that utilizes a web-based approach to allow the workstation's user to decide how the content of the workstation monitors are displayed on the Hiperwall as shown in Fig. 10. Based on the content environment already present in the Hiperwall software, our team added the ability for any user at a specific workstation to click on a link as shown in Fig. 11 to display the contents of the monitors on the Hiperwall. These links are subdivided by number of screens the user is broadcasting, then by the location on the Hiperwall display-tiles where they want their screens to be shown. And finally, users can choose to zoom in so that one of their workstation's monitors completely covers all the Hiperwall display-tiles. A significant drawback to this methodology is that some of the power of the software is reduced in order to make the system more user friendly.

It must also be noted that the purpose of the Hiperwall is not about increasing the resolution. Once the original image has been created the resolution can not be changed. The Hiperwall can increase the image size by adding pixels by interpolation methods, but the original resolution is not increased. Therefore, the purpose of the Hiperwall is to be able to show more details while also showing more display area. The user gains more display area of the image without losing the details. However, the system does add or subtract the number of pixels displayed in the frame of view of the user on the main screen of the Hiperwall. While we do not have access to the propriety code of the Hiperwall software, one can intelligently assume [23] it uses a closest neighbors approximation for pixel values as pixels are added to the image as the image is zoomed larger than its inherent resolution. As an example, stretching and zooming into a small thumbnail image ( $240 \times 240$ ) out to a full HD ( $1920 \times 1080$ ) sized image requires the addition of about 2 million pixels that did not exist in the original image.

The pixel count or resolution of LCD-based displays are 2K ( $1020 \times 1080$  pixels) for full HD or 4K ( $3840 \times 2160$  pixels) for full UHD. However, since LED-based display walls are built with modular components, the resolution (total number of pixels) of any particular installation can vary, which is why the most common specification for pixel density is referenced in the industry as “pixel pitch”. Pixel pitch is the distance between the center of two adjacent pixels as shown in Fig. 28a. So the smaller the pitch, the more pixels one can fit into a fixed dimension of the display wall. The smaller the pixels, the more one can fit into the same physical dimensions and thus increase the pixel count which delivers more details and smoother edges to the image s. So finer pixel pitch means that more pixels can be packed closely together, thus increasing the pixel density and resolution as shown in Fig. 28b.

Also, as the pixel density increases, one can stand closer to the display wall without seeing the pixels, which is especially important as displays get bigger or people move





**Table 4** Table shows the ideal viewing distance based on the pixel pitch of a display wall. Table adapted from [47]

Ideal viewing distance (ft)	Pixel pitch
25	2.5
18	1.8
15	1.5
12	1.2
9	0.9

closer to the display wall with the addition of interactive technologies [47]. Table 4 shows the ideal viewing distance based on the pixel pitch of a display wall.

Finally, we have consistently run up against the situation with a lecturer not being able to control the Hiperwall while standing in front of the workstations (see Fig. 27) or off to the side. Currently, a guest speaker/lecturer needs to have a “helper” logged in onto one of the workstations or use a tablet to be able to control the Hiperwall. As a solution to this issue our team has attempted to implement two technologies coming from the recent Virtual Reality (VR) hardware revolution, HTC Vive’s hand controller [115] and an infrared hand tracking detector Leap Motion [116]. While we are still working to get the Vive controllers to control the Hiperwall system, we were able to get the Leap Motion hand-tracking system working with basic controls like “grab” this image stream and move it to another position on Hiperwall. It is still possible to add other features and options for even more advanced controls.

Overall, the success and benefits of the Hiperwall greatly outweigh any of the drawbacks and challenges we have expressed with using the Hiperwall display system either as a research or an educational tool. The authors would advise other institutions as they develop their own visualization system to keep the above issues of control, usage, and manipulation in consideration.

## Conclusion

Data visualization is growing in a rapid pace with new and evolving visualization methods. Visualization of useful information in Big Data can indeed significantly improve our understanding of the vast amount of information in large datasets. The transformation of data into graphical images that can be visualized on a high-resolution display wall can be of significant benefit during data analysis of complex scientific data. It is often easier to extract specific information from images by interacting with it. Scientific visualization and visual representations enhances the clarity of the displayed information and thus allow users to extract useful information from the vast amount of data from these large science experiments that are often complex and heterogeneous. Scientists often face the challenging task of uniting the abstract world of data and the physical world through visual representations. So, visualization techniques are a useful tool for transforming information into a understandable form for human perception. In science, 3D visual representation and 3D simulations that are commonly used for scientific visualization applications can identify hidden patterns or anomalies in the datasets from Big Data. As the number of petascale datasets is increasing, a combination of data analytics and data visualization will enable scientific data exploration and allow improvements in outcomes and decision-making.

High resolution Tiled-Display Walls create an engaging and a pervasive collaborative environment for research and active learning in academia. These visualization systems are especially helpful for research groups analyzing Big Data from large science experiments. We broke new ground by being the first undergraduate institution in the US (and the only institution in Kentucky) to implement the Hiperwall technology in the field of HEP. The Hiperwall technology has an architectural advantage since Hiperwall runs on standard PCs connected to standard Ethernet network, thus eliminating the need for specialized switches and servers. Each display node renders its own portion of the display, thus eliminating the bottlenecks associated with central servers that make Hiperwall a robust visualization system. This paper suggests a cost-effective solution for visualizing high resolution large-scale data-driven images and applications with a broad view using the Hiperwall visualization system based on the setup at BU's AVCL.

We have described a novel Hiperwall visualization system setup at BU that increases the display size of large images in Astronomy and for interactive data analysis of Event Displays in HEP from the ATLAS experiment at the LHC at CERN. The key breakthrough of the Hiperwall system is the distributed software architecture that overcomes performance limitations typically associated with the display of large contents on multiple display-tiles. The Hiperwall visualization system is an effective way to display a large image or multiple images on display-tiles. Hiperwall facilitates detailed data analysis tasks and the aggregation and juxta-positioning of a variety of information and allows the examination of details within the context of the entire data set. Hiperwall is ideal for scientific visualization, visual analytics, and for viewing high-resolution 2D and 3D images, models and simulations, virtual reality and streaming videos, including Google Earth, Google Maps, and Web browsers.

We have used the Hiperwall visualization system successfully for conducting workshops on virtual Physics Labs, as well as interactive workshops on Python in a classroom setting. So, we have demonstrated that the Hiperwall visualization system can

also be used as a collaborative and interactive learning environment in classrooms. Hiperwall visualization system is also useful for collaborative research work with other institutions, where researchers, teachers, students can visualize and analyze the data with high resolution and to conduct interactive analysis for large data sets in real-time.

In summary, Hiperwall is a cost-effective option for implementing a tiled display wall that has the flexibility to run on both the Windows and the Linux platform. Hiperwall is a novel visualization technology that can be used in any discipline. The Hiperwall is an effective visualization system for Big Data applications and for conducting visualization studies of large-scale images in Astrophysics and Astronomy, and Event Displays in experimental HEP. At BU, Hiperwall has indeed improved the data analysis tasks and has provided an ideal collaborative visualization environment that can accelerate the discovery process. Our future plans are to continue to grow the number of Hiperwall users in classrooms and educational settings, including workshops/seminars and to enhance collaborative research environments.

Additional information about BU's Hiperwall visualization system can be found on the AVCL webpage [117]. Research work of our students at BU using the Hiperwall system titled "Students tackle Big Data in Advanced Visualization and Computational Lab" that was written for the general audience was published in the Summer 2015 issue of BU Magazine [118].

#### Authors' contributions

SM contributed to study design, methodology and manuscript preparation. HV contributed to study design, methodology and manuscript preparation. SB contributed to study design, provided data for Table 1, Figure 4, methodology and manuscript preparation. VW contributed to study design, provided data for Figure 3. AM contributed to study design, methodology and manuscript preparation. All authors read and approved the final manuscript.

#### Authors' information

SM is an Assistant Professor of Physics at Bellarmine University. He received his Ph.D. in Physics in 2005 and MSc. in Physics in 2001, both from the State University of New York at Albany (SUNY-Albany). His research interests are in High Energy Physics, and High Performance Computing. He is a member of LSST's (Large Synoptic Survey Telescope) Dark Energy Science Collaboration (DESC). He is also a member of the ATLAS collaboration at the LHC that discovered the Higgs Boson.

HV is an Assistant Professor of Computer Science at Weber State University. He received his Ph.D. in Physics in 2008 and MSc. in Physics in 2006, both from Vanderbilt University (Nashville, TN). His research interests are in High Energy Physics, Data Visualization, Software Development for particle detectors, sensors, microelectronics, and embedded systems. Previously, he was a member of the PHENIX collaboration at RHIC.

SB is a Systems Manager for the Center for Supercomputing and Visualization (CSV) at Bellarmine University. He received his MS in Physics in 2009 from Indiana University-Purdue University Indianapolis (IUPUI) and his BS in Physics in 2005 from University of Missouri—St. Louis. His research interests are in Astrophysics, and High Performance Computing. He is a member of LSST's (Large Synoptic Survey Telescope) Dark Energy Science Collaboration (DESC).

VM completed her BSc. in Physics in 2017 from Bellarmine University. She is currently pursuing a Ph.D. in Physics at Vanderbilt University. She worked as an Undergraduate Research Assistant on this project.

AM is a Professor of Physics at Bellarmine University and established the Center for Supercomputing and Visualization (CSV). He received his Ph.D. in Physics in 1998 and MSc. in Physics in 1994, both from the State University of New York at Albany (SUNY-Albany). He received his B.Sc. in Computer Science (Honors) in 1992 from Edinboro University of Pennsylvania. His research interests are in High Energy Physics, Astrophysics, Data-Intensive Science and Cluster/Grid Computing. Currently, he is a member of LSST's (Large Synoptic Survey Telescope) Dark Energy Science Collaboration (DESC). He is also a member of the ATLAS collaboration at the LHC that discovered the Higgs Boson.

#### Author details

<sup>1</sup> Department of Physics, Bellarmine University, 2001 Newburg Rd, Louisville, USA. <sup>2</sup> School of Computing, Weber State University, 3848 Harrison Blvd, Ogden, USA. <sup>3</sup> Department of Physics and Astronomy, Vanderbilt University, 2201 West End Ave, Nashville, USA.

#### Acknowledgements

The authors would like to thank the National Science Foundation (Grant #1229306 and #1154454) and the Henry Luce Foundation (Clare Boothe Luce Program) for supporting this research project.

#### Competing interests

The authors declare that they have no competing interests.

**Availability of data and materials**

Not applicable.

**Consent for publication**

We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal.

**Ethics approval and consent to participate**

Not applicable

**Funding**

National Science Foundation (Grant #1229306 and #1154454) and the Henry Luce Foundation (Clare Boothe Luce Program).

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 April 2018 Accepted: 4 October 2018

Published online: 28 October 2018

**References**

- Oracle. The big data imperative-compressing the analysis-to-action life cycle; 2017. [https://www.oracle.com/webfolder/s/delivery\\_production/docs/FY16h1/doc35/big-data-imp-Jan17.pdf](https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc35/big-data-imp-Jan17.pdf). Accessed 8 July 2018.
- Hu H, Wen Y, Chua T, Li X. Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access*. 2014;2:652–87.
- Ball R, North C. Effects of tiled high-resolution display on basic visualization and navigation tasks. In: CHI'05 extended abstracts on human factors in computing systems. New York: ACM; 2005. p. 1196–9.
- Andrews C, Endert A, Yost B, North C. Information visualization on large, high-resolution displays: issues, challenges, and opportunities. *Inf Vis*. 2011;10(4):341–55.
- Endert A, Bradel L, Zeitz J, Andrews C, North C. Designing large high-resolution display workspaces. In: Proceedings of the international working conference on advanced visual interfaces. New York: ACM; 2012. p. 58–65.
- Webster Dictionary; 2018. <https://www.webster-dictionary.org/definition/visualization>. Accessed 8 July 2018.
- Olshannikova E, Ometov A, Koucheryav Y, Olsson T. Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *J Big Data*. 2015;2(1):22.
- Teyseyre AR, Campo MR. An overview of 3D software visualization. *IEEE Trans Vis Comput Graph*. 2009;15(1):87–105.
- Lengler R, Eppler MJ. Periodic Table of Visualization Methods; 2018. [https://www.visual-literacy.org/periodic\\_table/periodic\\_table.html](https://www.visual-literacy.org/periodic_table/periodic_table.html). Accessed 8 July 2018.
- Khan M, Khan SS. Article: data and information visualization methods, and interactive mechanisms: a survey. *Int J Comput Appl*. 2011;34(1):1–14.
- SAS. Data visualization: making big data approachable and valuable. Market Pulse: White Paper; 2013. [https://www.sas.com/en\\_us/home.html](https://www.sas.com/en_us/home.html). Accessed 8 July 2018.
- Dill J, Earnshaw R, Kasik D, Vince J, Wong PC, editors. Expanding the frontiers of visual analytics and visualization. London: Springer; 2012.
- Simoff SJ, Böhlen MH, Mazeika A, editors. Visual data mining. Berlin: Springer; 2008.
- Zhang Q, Segall RS, Cao M, editors. Visual analytics and interactive technologies: data, text and web mining applications. Hershey: IGI Global; 2010.
- de Oliveira MCF, Levkowitz H. From visual data exploration to visual data mining: a survey. *IEEE Trans Vis Comput Graph*. 2003;9(3):378–94.
- Tory M, Kirkpatrick AE, Atkins MS, Moller T. Visualization task performance with 2D, 3D, and combination displays. *IEEE Trans Vis*. 2006;12:2–13.
- Utts JM. Seeing through statistics. 3rd ed. Boston: Brooks/Cole; 2004.
- Oliveria SRM. Geometric data transformation for privacy preserving clustering. Department of Computer Science; 2003.
- Kamel MCA. Hierarchic image classification visualization. In: Proceedings of the 10th international conference on image analysis and recognition. ICIAR; 2013.
- Keim DA. Designing pixel-oriented visualization techniques: theory and applications. *IEEE Trans Visual Comp Graph*. 2000;6:59–78. <https://doi.org/10.1109/2945.841121>.
- Bederson BB, Shneiderman B, editors. The craft of information visualization: readings and reflections (interactive technologies). Burlington: Morgan Kaufmann; 2003.
- New ATLAS-New Technology and Science News. Hiperwall software video wall uses normal PCs and monitors; 2010. <http://newatlas.com/hiperwall-software-video-wall-uses-normal-pcs-and-monitors/14251/>. Accessed 18 Sept 2017.
- Defanti TA, Leigh J, Renambot L, Jeong B, Verlo A, Long L, et al. The OptiPortal, a scalable visualization, storage, and computing interface device for the OptiPuter. *Future Gener Comput Syst*. 2009;25(2):114–23.
- Coalition S. The science coalition: sparking economic growth 2.0—companies created from Federally Funded University research fueling American innovation and economic growth; 2018. [http://www.sciencecoalition.org/downloads/AMI\\_v2\\_10-21-13.pdf](http://www.sciencecoalition.org/downloads/AMI_v2_10-21-13.pdf). Accessed 18 Sept 2018.
- NSF Award. ITR: The OptiPuter; 2017. [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=0225642&HistoricalAwards=false](https://www.nsf.gov/awardsearch/showAward?AWD_ID=0225642&HistoricalAwards=false). Accessed 8 July 2018.

26. Meade BF, Fluke CJ, Manos S, Sinnott RO. Are tiled display walls needed for astronomy? *Publ Astron Soc Aust*. 2014;31:e033.
27. Baltay C, Rabinowitz D, Andrews P, Bauer A, Ellman N, Emmet W, et al. The QUEST Large Area CCD Camera. *Publ Astron Soc Pac*. 2007;119(861):1278.
28. Kahn SM, Kurita N, Gilmore K, Nordby M, O'Connor P, Schindler R, et al. Design and development of the 3.2 gigapixel camera for the large synoptic survey telescope. In: McLean IS, Ramsay SK, Takami H, editors. *Ground-based and airborne instrumentation for astronomy III*. vol. 7735 of society of photo-optical instrumentation engineers (SPIE) conference series; 2010.
29. LSST Dark Energy Science Collaboration, et al. Large synoptic survey telescope: dark energy science collaboration. *arXiv preprint*. [arXiv:1211.0310](https://arxiv.org/abs/1211.0310); 2012.
30. Ivezić Z, Tyson J, Abel B, Acosta E, Allsman R, Alsayyad Y, et al. LSST: from science drivers to reference design and anticipated data products. *arXiv preprint*. [arXiv:0805.2366](https://arxiv.org/abs/0805.2366); 2008.
31. Melissa Gaillard. CERN Data Centre Passes the 200-Petabyte Milestone; 2017. <https://home.cern/about/updates/2017/07/cern-data-centre-passes-200-petabyte-milestone>. Accessed 28 Sept 2017.
32. CERN Computing. CERN computing; 2017. <https://home.cern/about/computing>. Accessed 28 Sept 2017.
33. WLCG. World wide computing grid; 2017. <http://wlcg-public.web.cern.ch/about>. Accessed 28 Sept 2017.
34. Rector TA, Cool R, WYN/NOAO/AURA/NSF. Modified from image "Giant supernova remnant, Cygnus Loop"; 2017. [https://www.noao.edu/image\\_gallery/html/im1138.html](https://www.noao.edu/image_gallery/html/im1138.html). Accessed 18 Sept 2017.
35. Lifewire. How are video frame rate and screen refresh rate different; 2017. <https://www.lifewire.com/video-frame-vs-screen-refresh-rate-1847855>. Accessed 18 Sept 2017.
36. University of Minnesota. PowerWall; 1998. <https://www.lcse.umn.edu/research/powerwall/powerwall.html>. Accessed 8 July 2018.
37. Li K, Chen H, Chen Y, Clark DW, Cook P, Damianakis S, et al. Building and using a scalable display wall system. *IEEE Comput Graph Appl*. 2000;20(4):29–37.
38. Wallace G, Anshus OJ, Bi P, Chen H, Chen Y, Clark D, et al. Tools and applications for large-scale display walls. *IEEE Comput Graph Appl*. 2005;25(4):24–33.
39. University of Illinois at Chicago. Electronic Visualization Lab; 2018. <https://www.evl.uic.edu>. Accessed 1 Jan 2018.
40. Boyles M, Gniady T, Wernert E, Eller C, Reagan D, Rogers J. IQ-Wall: an open standard for tiled video walls that balances flexibility, usability, performance, and cost. In: *EdMedia: World conference on educational media and technology 2016*. Bloomington: Association for the Advancement of Computing in Education; 2016. p. 604–11.
41. Extron Electronics. Quantum Connect; 2018. <https://www.extron.com>. Accessed 8 July 2018.
42. Extron Electronics. Quantum connect; 2018. <https://www.extron.com/product/quantumconnect>. Accessed 1 Jan 2018.
43. Hiperwall. Hiperwall comparison chart; 2018. <https://www.hiperwall.com/video-wall-solution-competitive-comparison-chart>. Accessed 1 Jan 2018.
44. Hiperwall; 2018. <https://www.hiperwall.com>. Accessed 1 Jan 2018.
45. Christie Digital; 2018. <https://www.christiedigital.com/>. Accessed 1 Jan 2018.
46. RGB Spectrum; 2018. <https://www.rgb.com/>. Accessed 1 Jan 2018.
47. Planar. How much does a led video wall cost; 2018. <http://www.planar.com/>.
48. Barco; 2018. <https://www.barco.com/>. Accessed 8 July 2018.
49. Useful; 2018. <https://www.useful.com/>. Accessed 8 July 2018.
50. Datapath; 2018. <https://www.datapath.co.uk/>. Accessed 8 July 2018.
51. Jupiter by Infocus; 2018. <https://www.infocus.com/jupiter-by-infocus>. Accessed 1 Jan 2018.
52. Cinemassive; 2018. <https://www.cinemassive.com>. Accessed 1 Jan 2018.
53. Keim D, editor. Mastering the information age. Goslar: Eurographics Association; 2010.
54. Sun M, North C, Ramakrishnan N. A five-level design framework for bicluster visualizations. *IEEE Trans Vis Comput Graph*. 2014;20:1713–22.
55. Agrawal D, Das S, Elabbadi A. Data management in the cloud. No. 32 in *synthesis lectures on data management, Literaturverz*. S. San Rafael: Morgan & Claypool; 2013. p. 107–17.
56. SAS. Statistical Analysis System (SAS); 2017. [https://www.sas.com/en\\_us/home.html](https://www.sas.com/en_us/home.html). Accessed 8 July 2018.
57. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117.
58. Omidvar O, Dayho JE, editors. *Neural networks and pattern recognition*. San Diego, Calif: Academic Press; 1998.
59. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, editors. *Advances in neural information processing systems 27*. Curran Associates, Inc.; 2014. p. 3104–12.
60. Rojas R. *Neural networks: A systematic introduction*. Feldman J, editor. Berlin, Heidelberg: Springer; 1996.
61. Gurney K. *An introduction to neural networks*. Boca Rotan: CRC Press; 1997.
62. Larose DT. *Discovering knowledge in data: an introduction to data mining*. Hoboken: John Wiley & Sons; 2014.
63. Chu WW. Erratum: Data mining and knowledge discovery for Big Data. Berlin: Springer; 2014.
64. Berry MJA, Linoff GS. *Data mining techniques: for marketing, sales, and customer support*. Hoboken: Wiley; 1997.
65. Zaki MJ, Meira W Jr, Meira W. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge: Cambridge University Press; 2014.
66. Vetterli MKJ. *Foundations of signal processing*. Cambridge: Cambridge University Press; 2014.
67. Giannakis GB. Signal processing for Big Data. In: *2014 signal processing: algorithms, architectures, arrangements, and applications (SPA)*; 2014. p. 9–9.
68. Xhafa F, Barolli L, Barolli A, Papajorgji P. *Modeling and processing for next-generation Big-Data technologies*. Cambridge: Springer; 2014.
69. Shneiderman B. The big picture for big data: visualization. *Science*. 2014;343(6172):730.
70. Marr B. *Big Data: Analytics and metrics to make better decisions and improve performance*. Hoboken: Wiley; 2015.



71. Minelli M, Chambers M, Dhiraj A. Big Data, big analytics: emerging business intelligence and analytic trends for today's businesses. Hoboken: Wiley; 2012.
72. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of machine learning. Adaptive computation and machine learning series. Cambridge: MIT Press; 2012.
73. Murphy K. Machine learning: a probabilistic perspective. Adaptive computation and machine learning series. Cambridge: MIT Press; 2012.
74. Alpaydin E. Introduction to machine learning. Adaptive computation and machine learning series. Cambridge: MIT Press; 2014.
75. Roe BP, Yang H, Zhu J, Liu Y, Stancu I, McGregor G. Boosted decision trees as an alternative to artificial neural networks for particle identification. Nuclear Instrum Methods Phys Res Sect A Accel Spectrom Detect Assoc Equip. 2005;543:577–84.
76. Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: Saitta L, editor. Proceedings of the 13th international conference of machine learning (ICML 96). Morgan Kaufmann, San Francisco; 1996. p. 148.
77. Cowan G. Topics in statistical data analysis for HEP. CERN Yellow Report CERN-2010-002; 2009. p. 341–69.
78. Voss H. Multivariate Data Analysis and Machine Learning in High Energy Physics. Multivariate Data Analysis and Machine Learning in High Energy Physics (II). Heidelberg: MPI-K; 2009. p. 11.5–5.5.
79. Voss H. Decision trees and boosting. Heidelberg: MPIK; 2011.
80. Therhaag J. Defeating the black box neural networks in HEP data analysis, University of Bonn. In: TMVA Workshop, CERN; 2011.
81. Bhat P. Advanced analysis methods in particle physics, Fermilab-PUB-10-054-E. Advanced computing and analysis techniques in physics research. In: VII international workshop, ACAT 2000; 2000.
82. Zemla A. Bayesian neural network and its application in ATLAS—searching a needle in the Haystack: multivariate analysis of HEP data, L Cracow School of theoretical physics. Krakow: Institute of Nuclear Physics PAS; 2010. p. 09–19.
83. Hoecker A, Speckmayer P, Stelzer J, Therhaag J, von Toerne E, Voss H. TMVA—toolkit for multivariate data analysis with ROOT, TMVA version 4.1. arXiv preprint; 2018. [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
84. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer; 2001.
85. Sethi IK, Jain A. Artificial neural networks and statistical pattern recognition: old and new connections (machine intelligence and pattern recognition). Amsterdam, NL: North Holland; 2014.
86. Araghinejad S. Data-driven modeling: using MATLAB in water resources and environmental engineering: 67 (water science and technology library). New York: Springer; 2013.
87. Neal RM. Bayesian learning for neural networks (lecture notes in statistics). New York, NY, USA: Springer; 1996.
88. Lampinen J, Vehtari A. Bayesian approach for neural networks—review and case studies. Neural Networks. 2001;14(3):257–74.
89. Iwata T, Houlsby N, Ghahramani Z. Active learning for interactive visualization. In: Carvalho CM, Ravikumar P, editors. Proceedings of the sixteenth international conference on artificial intelligence and statistics. Proceedings of machine learning research, vol. 31. Scottsdale, Arizona, USA: PMLR; 2013. p. 342–50.
90. Drevermann H, Grab C. Graphical concepts for the representation of events in high energy physics. Int J Mod Phys C. 1990;1:147–63.
91. Breuker H, Drevermann H, Grab C, Rademakers AA, Stone H. Tracking and imaging elementary particles. Sci Am. 1991;265(2):58–63. <http://www.jstor.org/stable/24938680>. Accessed 1st July 2018.
92. Brun R, Rademakers F. ROOT: An object oriented data analysis framework. Nucl Instr Meth Phys Res Sect A. 1997;389(1):81–6. <http://www.sciencedirect.com/science/article/pii/S016890029700048X>. Accessed 1 July 2018.
93. Inoue K, Kakuto A, Komatsu H, Takeshita S. Aspects of grand unified models with softly broken supersymmetry. Prog Theor Phys. 1982;68(3):927–46.
94. Kane GL, Kolda CF, Roszkowski L, Wells JD. Study of constrained minimal supersymmetry. Phys Rev. 1994;D49:6173–210.
95. Gunion JF, Haber H, Kane G, Dawson S. The Higgs Hunter's guide. Boulder: Westview Press; 2000.
96. Gunion JF, Haber HE. Higgs bosons in supersymmetric models. 1. Nucl Phys. 1986;B272:1. [Erratum: Nucl Phys B402,567(1993)].
97. Brein O, Djouadi A, Harlander R. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. Phys Lett. 2004;B579:149–56.
98. Ciccolini ML, Dittmaier S, Kramer M. Electroweak radiative corrections to associated WH and ZH production at hadron colliders. Phys Rev. 2003;D68:073003.
99. Han T, Valencia G, Willenbrock S. Structure function approach to vector boson scattering in p p collisions. Phys Rev Lett. 1992;69:3274–7.
100. Berger EL, Campbell JM. Higgs boson production in weak boson fusion at next-to-leading order. Phys Rev. 2004;D70:073011.
101. Figy T, Oleari C, Zeppenfeld D. Next-to-leading order jet distributions for Higgs boson production via weak boson fusion. Phys Rev. 2003;D68:073005.
102. Beenakker W, Dittmaier S, Kramer M, Plumper B, Spira M, Zerwas PM. Higgs radiation off top quarks at the Tevatron and the LHC. Phys Rev Lett. 2001;87:201805.
103. Reina L, Dawson S. Next-to-leading order results for t anti-t h production at the Tevatron. Phys Rev Lett. 2001;87:201804.
104. Dawson S, Jackson CB, Reina L, Wackerroth D. Exclusive Higgs boson production with bottom quarks at hadron colliders. Phys Rev. 2004;D69:074027.
105. Harlander RV, Kilgore WB. Higgs boson production in bottom quark fusion at next-to-next-to leading order. Phys Rev. 2003;D68:013001.
106. Campbell JM, Ellis RK, Maltoni F, Willenbrock S. Higgs-Boson production in association with a single bottom quark. Phys Rev. 2003;D67:095002.

107. Dawson S, Jackson CB, Reina L, Wackerroth D. Higgs boson production with one bottom quark jet at hadron colliders. *Phys Rev Lett*. 2005;94:031802.
108. Dittmaier S, Kramer M, Spira M. Higgs radiation off bottom quarks at the Tevatron and the CERN LHC. *Phys Rev*. 2004;D70:074010.
109. Dawson S, Orr LH, Reina L, Wackerroth D. Associated top quark Higgs boson production at the LHC. *Phys Rev*. 2003;D67:071503.
110. Stirling WJ, Summers DJ. Production of an intermediate mass Higgs boson in association with a single top quark at LHC and SSC. *Phys Lett*. 1992;B283:411–5.
111. Maltoni F, Paul K, Stelzer T, Willenbrock S. Associated production of Higgs and single top at hadron colliders. *Phys Rev*. 2001;D64:094023.
112. The ATLAS Collaboration. Calibrating the b-tag efficiency and mistag rate in 35 pb<sup>-1</sup> of data with the ATLAS detector. In: ATLAS-CONF-2011-089. Geneva: CERN; 2011.
113. The ATLAS Collaboration. Performance of the ATLAS secondary vertex b-tagging algorithm in 7 TeV collision data. In: ATLAS-CONF-2010-042. Geneva: CERN; 2010.
114. The ATLAS Collaboration. Calibrating the b-tag and mistag efficiencies of the SV0 b-tagging algorithm in 3 pb<sup>-1</sup> of data with the ATLAS detector. In: ATLAS-CONF-2010-099. Geneva: CERN; 2010.
115. HTC Corporation. Controller; 2018. <https://www.vive.com/us/accessory/controller/>. Accessed 8 July 2018.
116. LEAP MOTION, INC. Leap motion hand tracking; 2018. <https://www.leapmotion.com/>. Accessed 8 July 2018.
117. Mahmood A. AVCL; 2016. <http://www.bellarmino.edu/faculty/amahmood/avcl/avcl.html>. Accessed 11 Sept 2018.
118. Bellarmine Magazine. Students tackle Big Data in Advanced Visualization and Computational Lab; 2014. <http://www.bellarminemagazine.com/students-tackle-big-data-in-advanced-visualization-and-computational-lab/>. Accessed 11 Sept 2018.
119. Hiperwall. Hiperwall User's Guide for software version 4.0; 2015. <http://www.bellarmino.edu/faculty/amahmood/HiperwallUser'sGuidev3.1.pdf>. Accessed 8 July 2018.
120. Hansen C, Johnson CR, Pascucci V, Silvia CT. Visualization for data-intensive science; 2011. <http://image.sciencenet.cn/olddata/kexue.com.cn/upload/blog/file/2011/1/201111122173654135.pdf>. Accessed 3 Mar 2018.
121. NASA/JPL-Caltech/MSSS. Strata at base of mount sharp; 2015. <https://www.nasa.gov/image-feature/jpl/pia19839/strata-at-base-of-mount-sharp>. Accessed 8 Oct 2017.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---