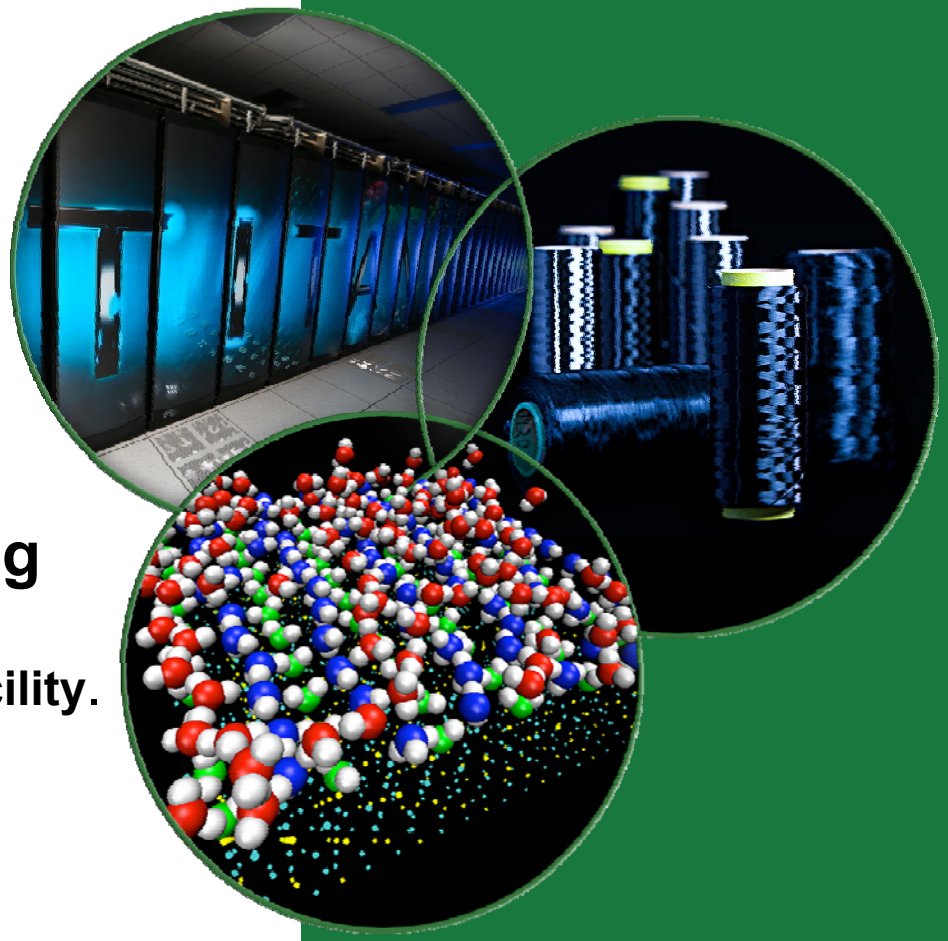
The Oak Ridge Leadership Computing Facility



Visualization Technologies
and the challenge of
High Performance Computing
at the
Oak Ridge Leadership Computing Facility.

2017 SMFoLD Workshop

Jamison Daniel, Ben Hernandez

ORNL is managed by UT-Battelle
for the US Department of Energy

Jamison Daniel
Presented October 3
, 2017

OAK RIDGE National Laboratory | OAK RIDGE LEADERSHIP COMPUTING FACILITY

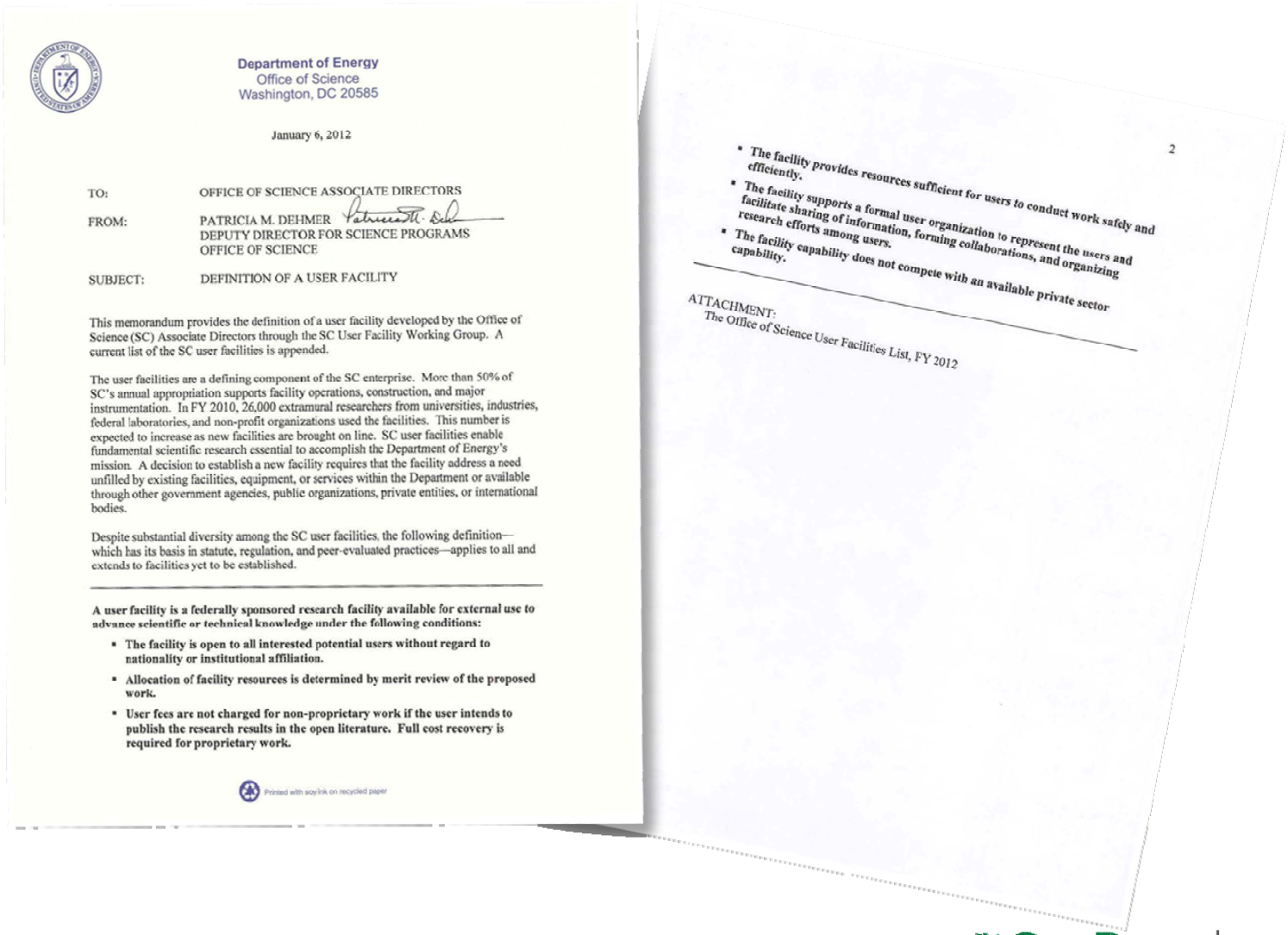
Oak Ridge National Laboratory (ORNL)

- Established in 1943 as part of the Manhattan Project.
- Employs over 4,400 staff, including scientists and engineers in more than 100 disciplines.
- Operating budget of $1.4 billion.
- 30,000-acre tract of land in East Tennessee.
- Largest DOE science and energy laboratory.

The Oak Ridge Leadership Computing Facility (OLCF)

- Established in 2004.
- Mission of scientific discovery by providing HPC resources to high-priority research projects.
- Funded through congressional appropriations for construction and operations.
- Operates Titan, the nation's largest supercomputer for open science research.

Definition of a DOE User Facility



Department of Energy
Office of Science
Washington, DC 20585

January 6, 2012

TO: OFFICE OF SCIENCE ASSOCIATE DIRECTORS

FROM: PATRICIA M. DEHMER
DEPUTY DIRECTOR FOR SCIENCE PROGRAMS
OFFICE OF SCIENCE

SUBJECT: DEFINITION OF A USER FACILITY

This memorandum provides the definition of a user facility developed by the Office of Science (SC) Associate Directors through the SC User Facility Working Group. A current list of the SC user facilities is appended.

The user facilities are a defining component of the SC enterprise. More than 50% of SC's annual appropriation supports facility operations, construction, and major instrumentation. In FY 2010, 26,000 extramural researchers from universities, industries, federal laboratories, and non-profit organizations used the facilities. This number is expected to increase as new facilities are brought on line. SC user facilities enable fundamental scientific research essential to accomplish the Department of Energy's mission. A decision to establish a new facility requires that the facility address a need unfilled by existing facilities, equipment, or services within the Department or available through other government agencies, public organizations, private entities, or international bodies.

Despite substantial diversity among the SC user facilities, the following definition—which has its basis in statute, regulation, and peer-evaluated practices—applies to all and extends to facilities yet to be established.

A user facility is a federally sponsored research facility available for external use to advance scientific or technical knowledge under the following conditions:

- **The facility is open to all interested potential users without regard to nationality or institutional affiliation.**
- **Allocation of facility resources is determined by merit review of the proposed work.**
- **User fees are not charged for non-proprietary work if the user intends to publish the research results in the open literature. Full cost recovery is required for proprietary work.**

Printed with soy ink on recycled paper

2

- **The facility provides resources sufficient for users to conduct work safely and efficiently.**
- **The facility supports a formal user organization to represent the users and facilitate sharing of information, forming collaborations, and organizing research efforts among users.**
- **The facility capability does not compete with an available private sector capability.**

ATTACHMENT:
The Office of Science User Facilities List, FY 2012

User Facility Requirements

- Federally sponsored research facility available for external use to advance scientific or technical knowledge.
  - The facility is open to all interested potential users without regard to nationality or institutional affiliation.
  - Allocation of facility resources is determined by merit review of the proposed work.
  - User fees are not charged for non-proprietary work **if** the user intends to publish the research results in the open literature, **otherwise** full cost recovery is required.
  - Cannot compete with existing private business.

Strategic Advantage of the OLCF

- Unique scientific instrument not available elsewhere in industry or government.
- Designed for *capability* high performance computing.
  - Capability is achieved through delivering:
    - 27 petaflops of compute cycles:
      - Unique Gemini system interconnect (very fast network).
      - 18,688 nodes (nodes can exchange data quickly using the Gemini network).
      - 32 petabyte filesystem that can deliver 1 terabyte per second performance (record simulation results for post-analysis).

Top 500 Supercomputers

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

Scientific Discovery - An Abstract Currency

- In 2015, OLCF users published 229 breakthrough publications in high-impact journals:
  - *Science, Nature*, *Nature Physics*, *Nature Communications*, *Nature Scientific Reports*, *Physical Review X*, *Physical Review Letters*, etc.
- These publication metrics are reported to the program owners at the Office of Science annually and are considered a significant measurement of mission success.

Who is the customer of the OLCF?

- USA taxpayer?
  - Initial direct cost of federal tax obligations.
  - Votes for congressional representatives.

- Program owner at the Office of Science in Washington D.C.?
  - Awards annual funding and designation of DOE User Facility to OLCF. Funds are received from Congress through annual appropriations bill.

- Users of the facility
  - Researchers from universities, government agencies, and private companies.

Algorithm Motifs

Table 6. Algorithm motifs for selected applications

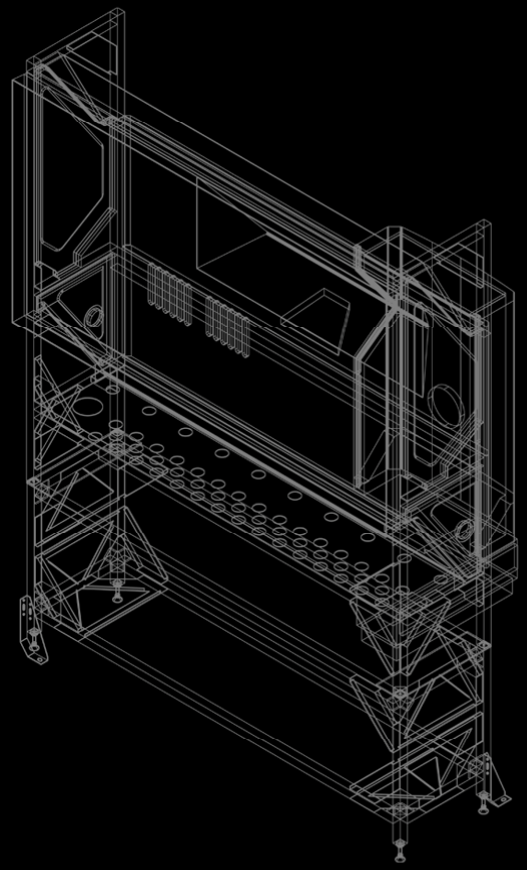
Application	Structured grids	Un-structured grids	FFT	Dense linear algebra	Sparse linear algebra	Particles	Monte Carlo
NWCHEM			●	●			
S3D	●			●	●	●	
XGC		●				●	
CCSM	●		●		●		
CASINO							●
VPIC	●					●	●
VASP			●	●			
MFDn					●		
LSMS				●			●
GenASiS		●			●		
MADNESS		●	●	●			
GTC	●				●	●	●
OMEN	●				●		
Denovo	●			●	●	●	●
CP2K	●				●	●	
CHIMERA	●			●	●	●	
DCA++				●			●
LAMMPS	●		●			●	
DNS	●			●	●	●	
PFLOTRAN	●	●		●	●		●
CAM	●		●	●	●	●	
QMCPACK						●	●
TOTAL	**12**	**4**	**6**	**11**	**12**	**11**	**8**

Scientific Visualization – tools and techniques

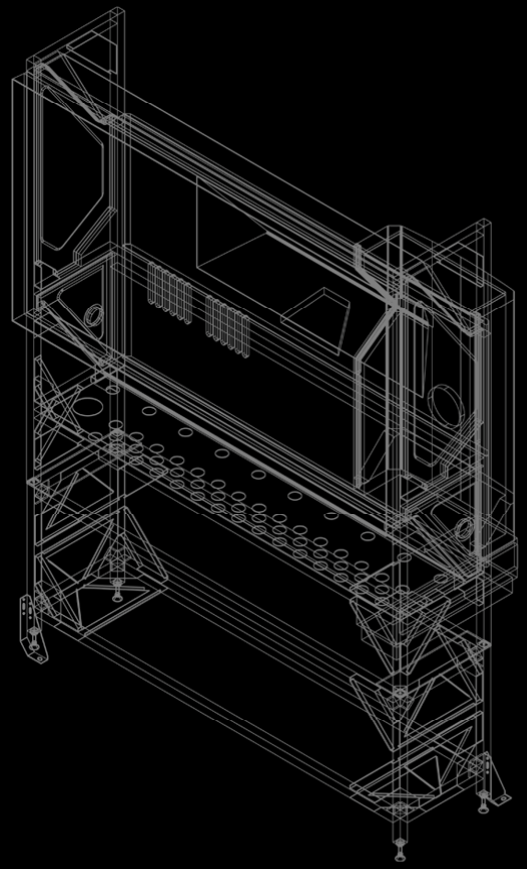
- **Analysis** – Do not know the answer. May or may not know the question. Visualization is leveraged to stream information to the human brain for processing. Answers are sought.
- **Communication** - Know the answer. Visualization is leveraged to stream information to the human brain for understanding and comprehension. Answers are disseminated.

BARCO OL-721

* On-screen contrast: 560,000:1
* Brightness: 235 Cd/m
* Display technology: Rear projection DLP
* Brightness uniformity: 95% ANSI 9
* Screen gap: Hor: 1mm, Vert. 0.7mm
* Color stability: Self calibration with spectrometer
* Light source: 6x redundancy for each of 3 LED's
* Light source lifetime: 60,000 hours
* Power: 230W
* Heat dissipation: 785 BTU/h

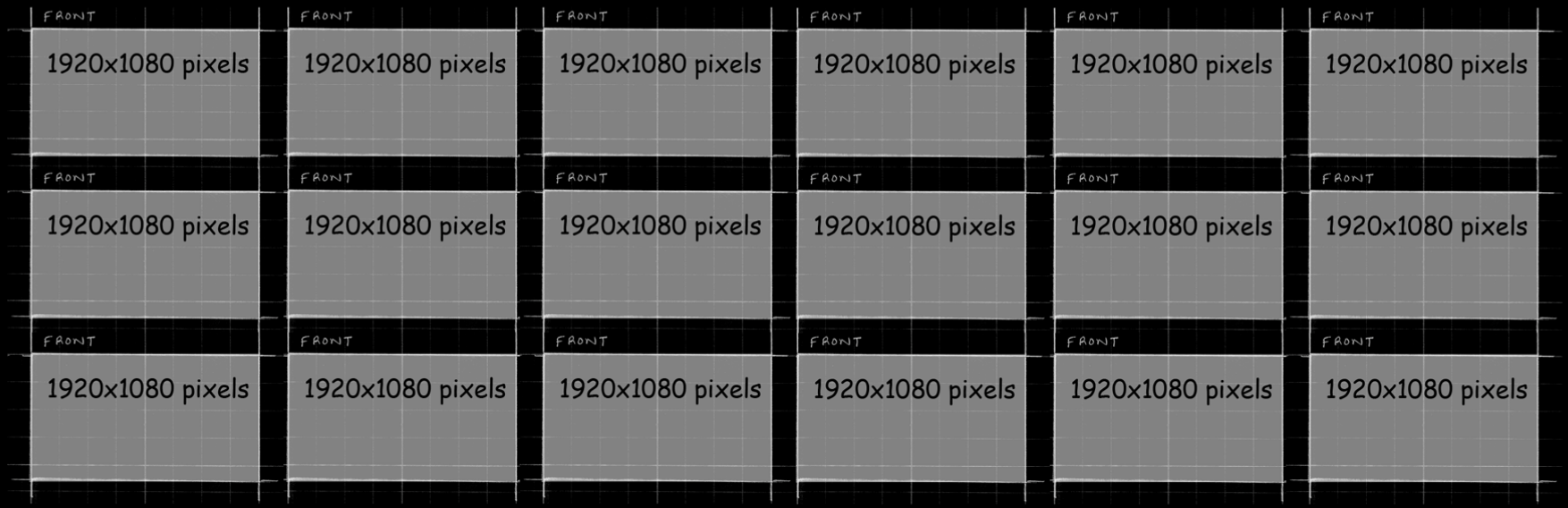


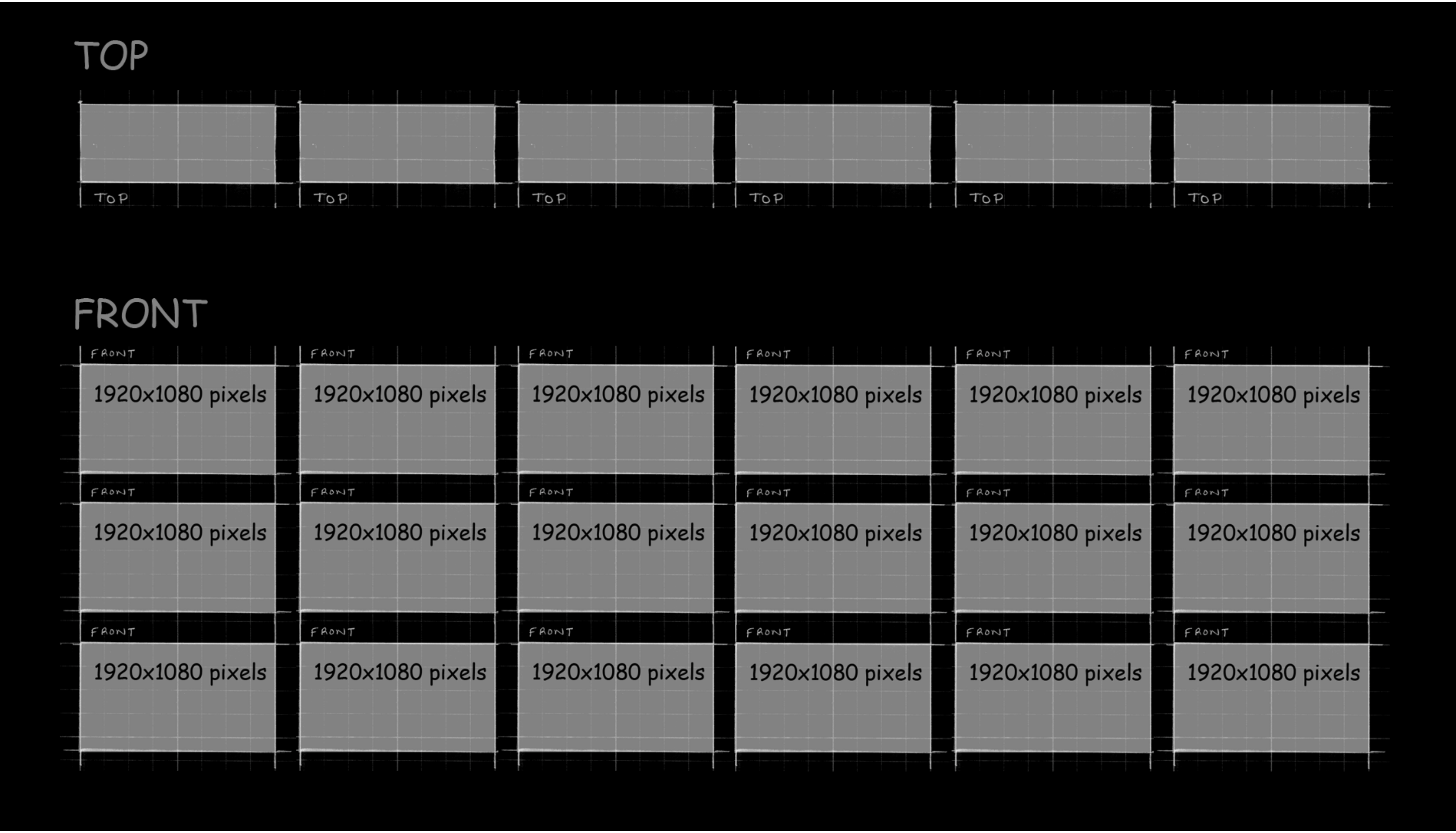

BARCO
OL-721
TOP
SIDE
FRONT
1920x1080 pixels
2.044ft.
2.868ft.
5.083 ft.

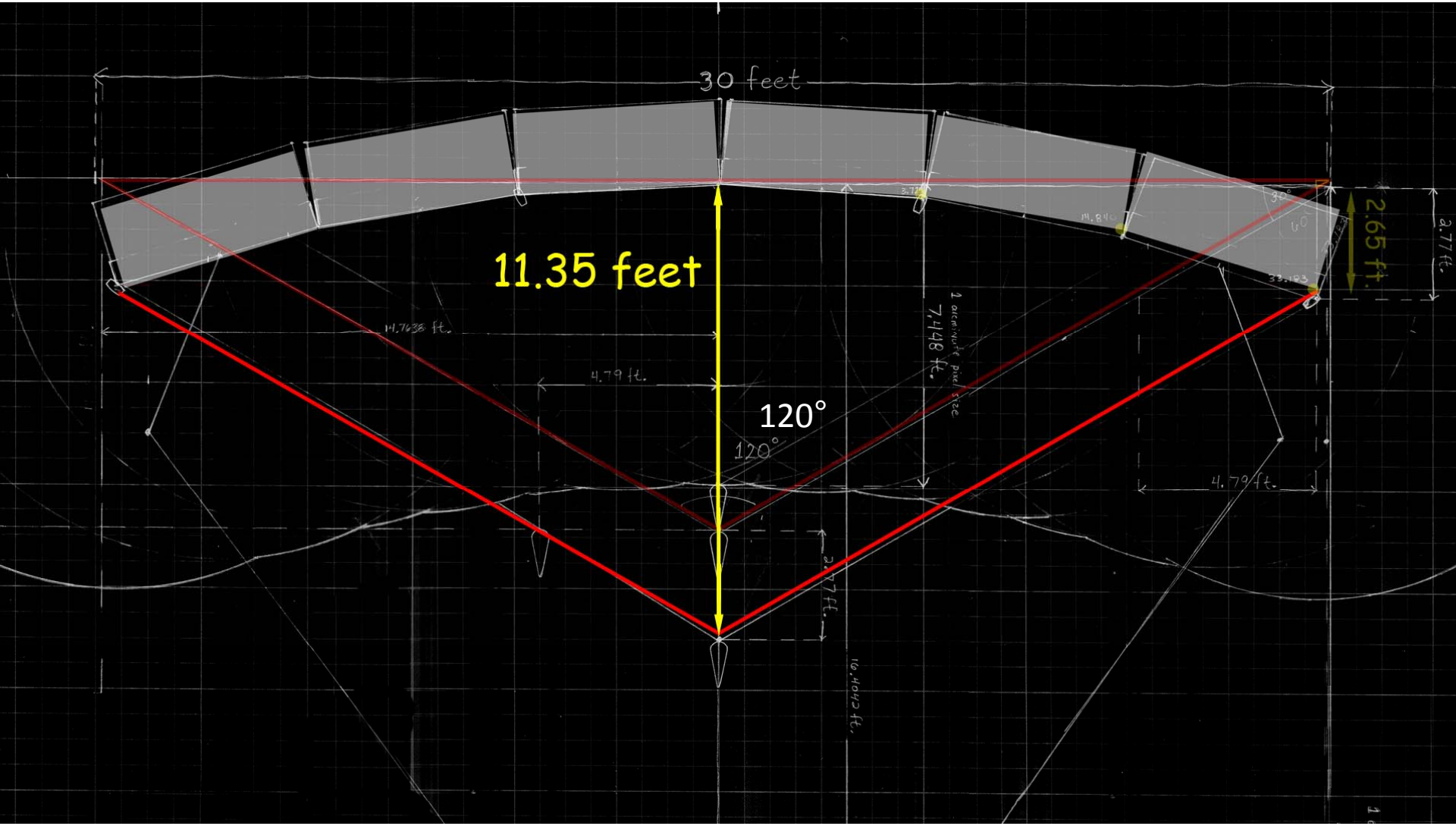


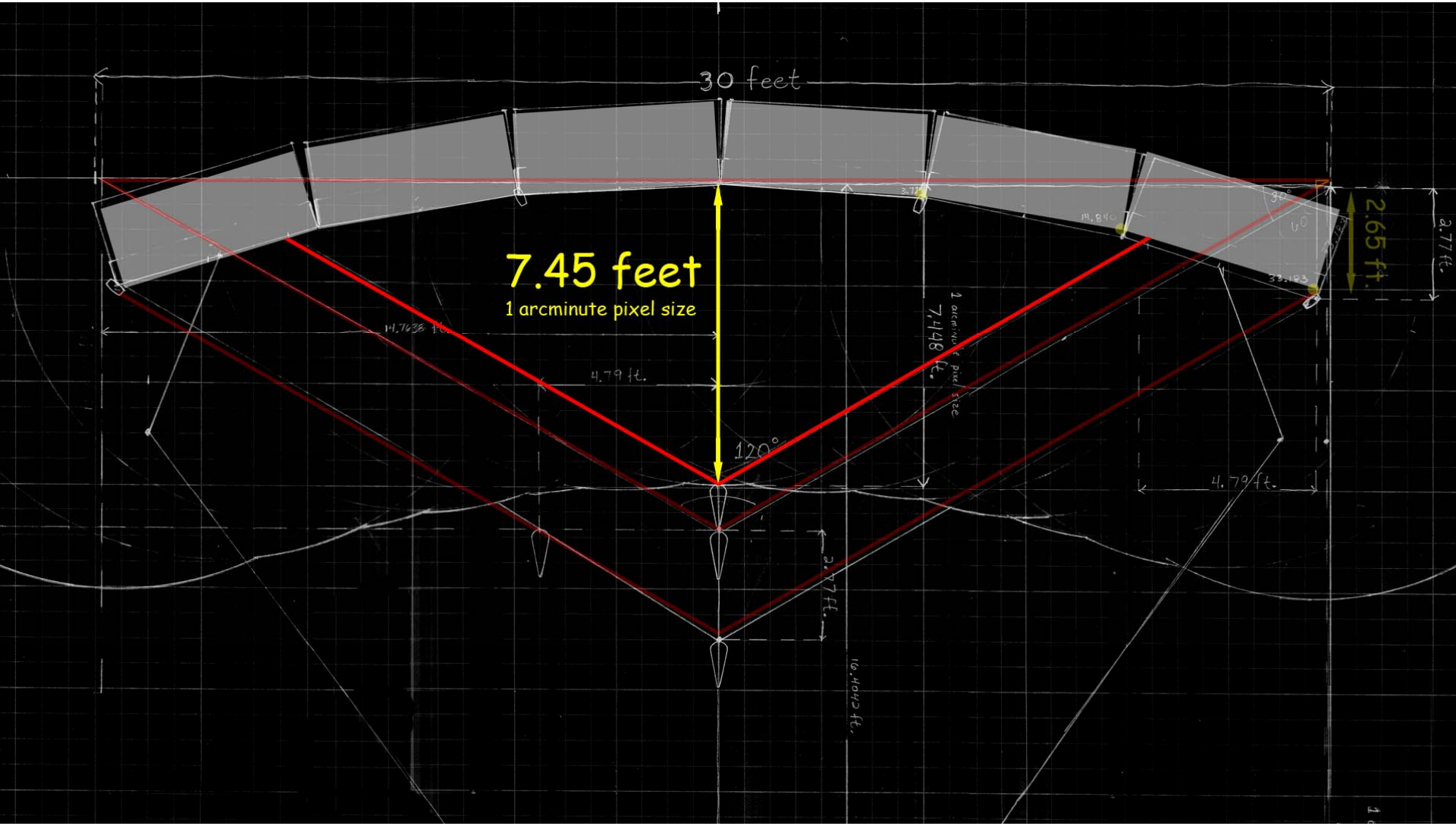
37,324,800 pixels

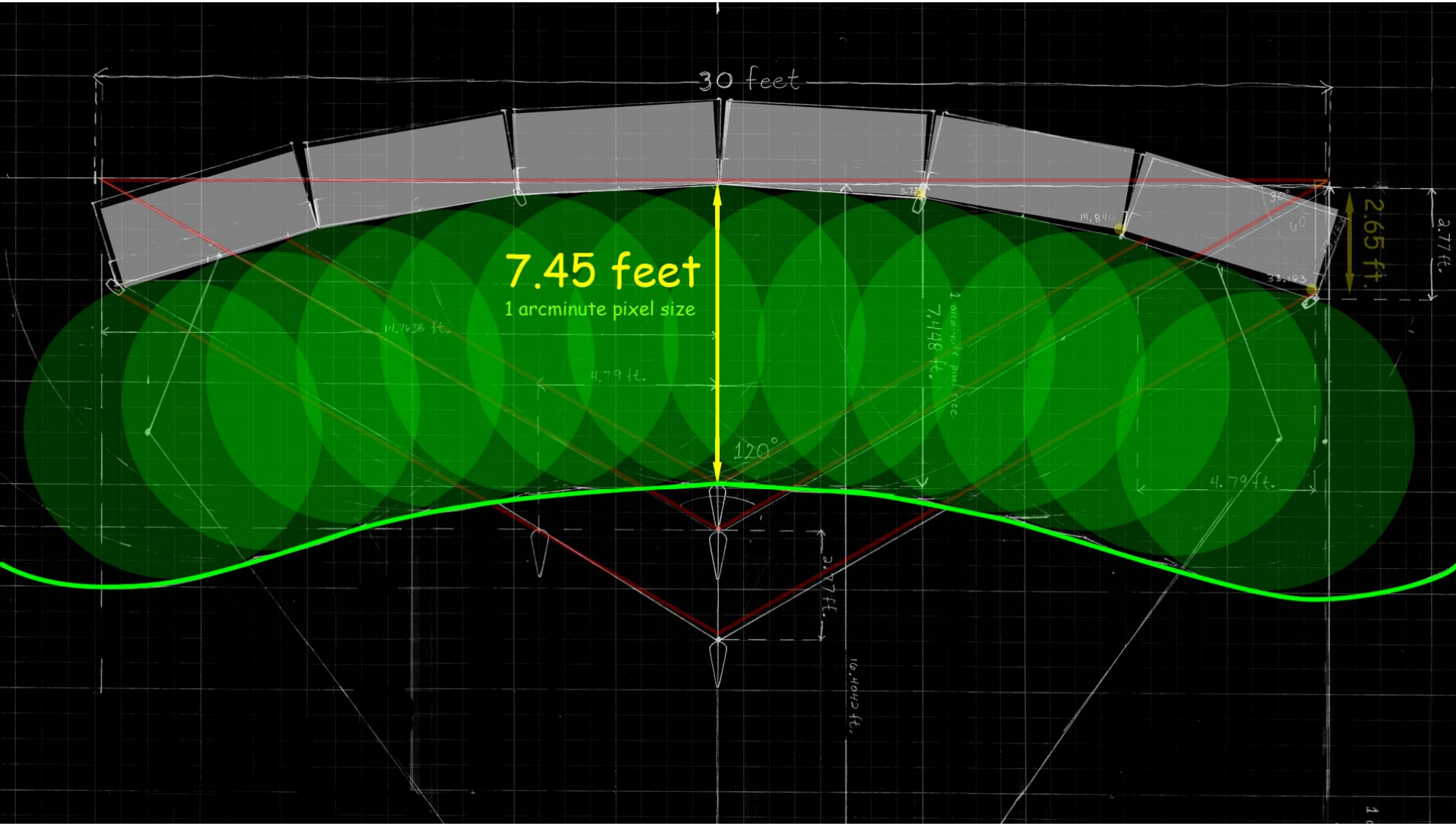
11,520 x 3240 pixels

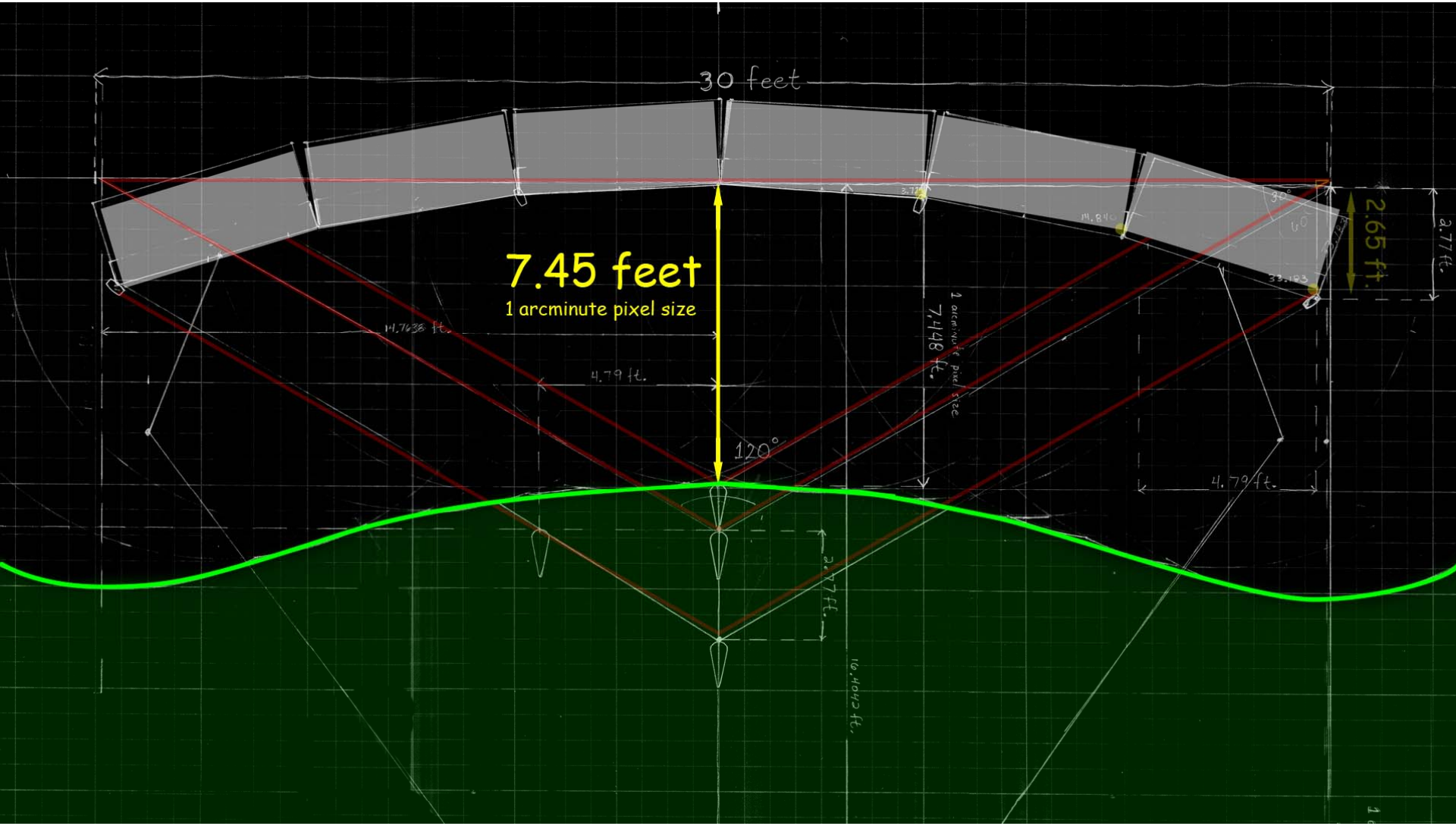


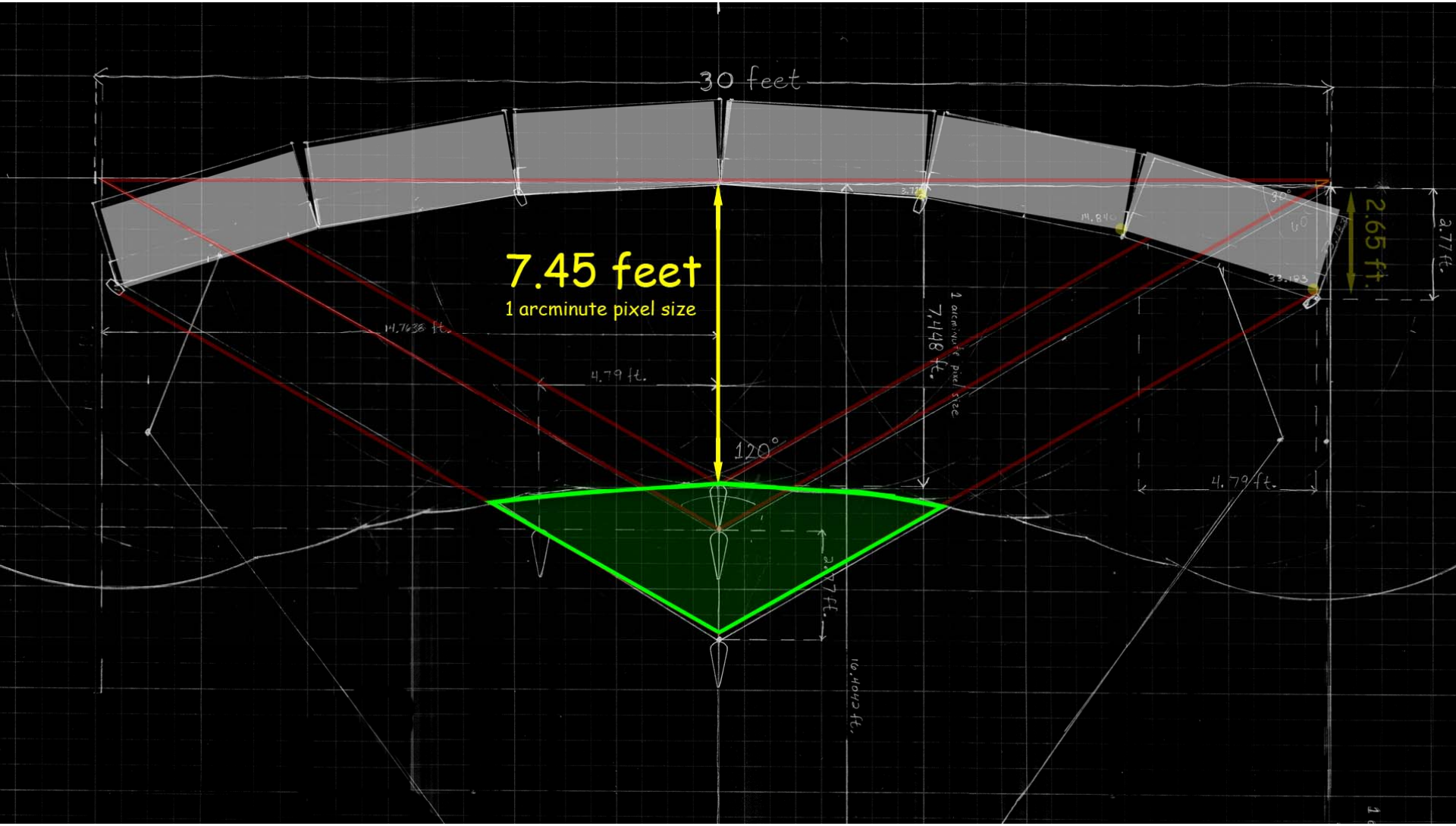

TOP
TOP
TOP
TOP
TOP
TOP
TOP
FRONT
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels
FRONT
1920x1080 pixels

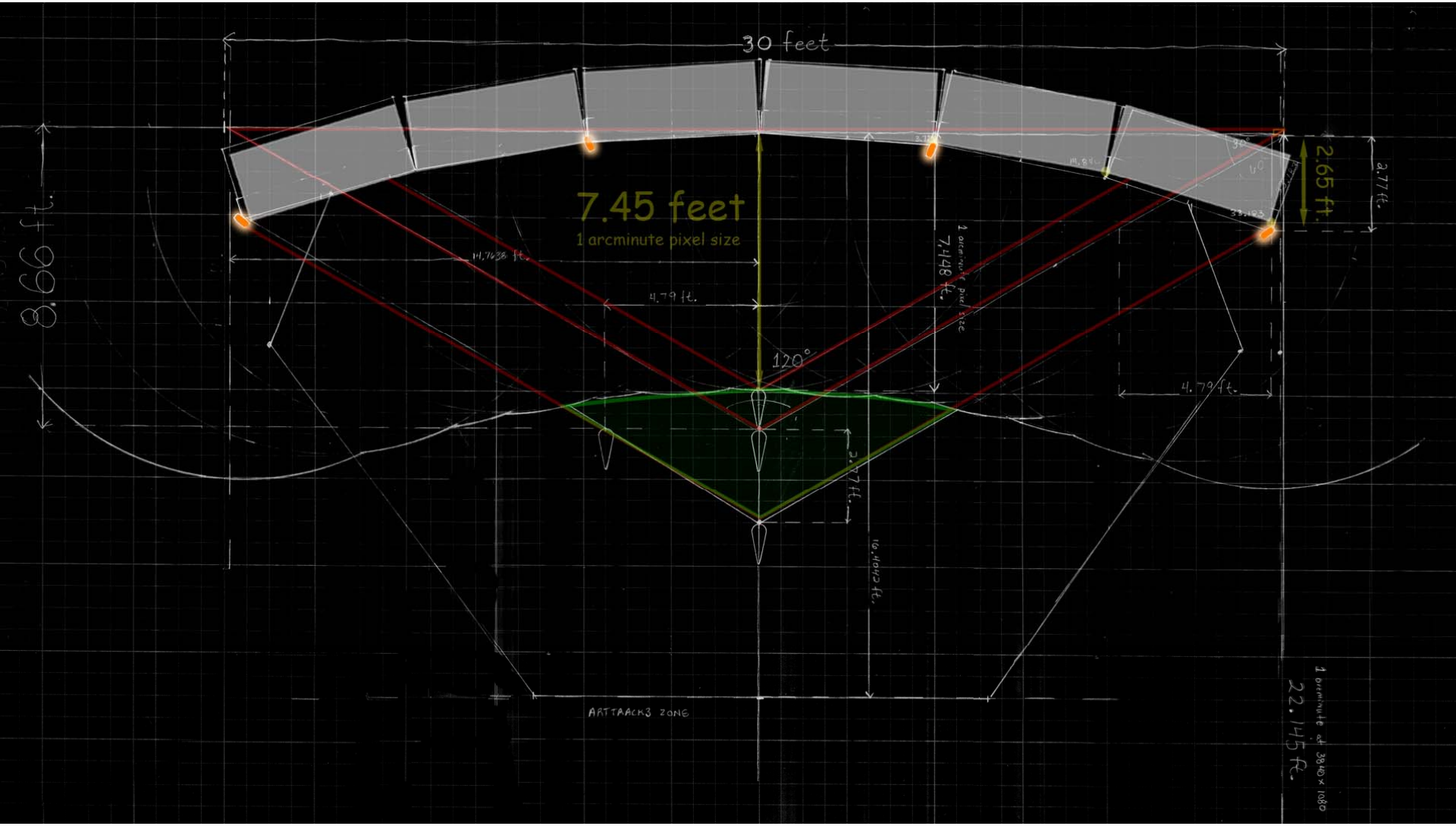

30 feet
11.35 feet
2.65 ft.
120°
120°
4.79 ft.
4.79 ft.
14.7638 ft.
7.4148 ft.
2.77 ft.
16.4042 ft.

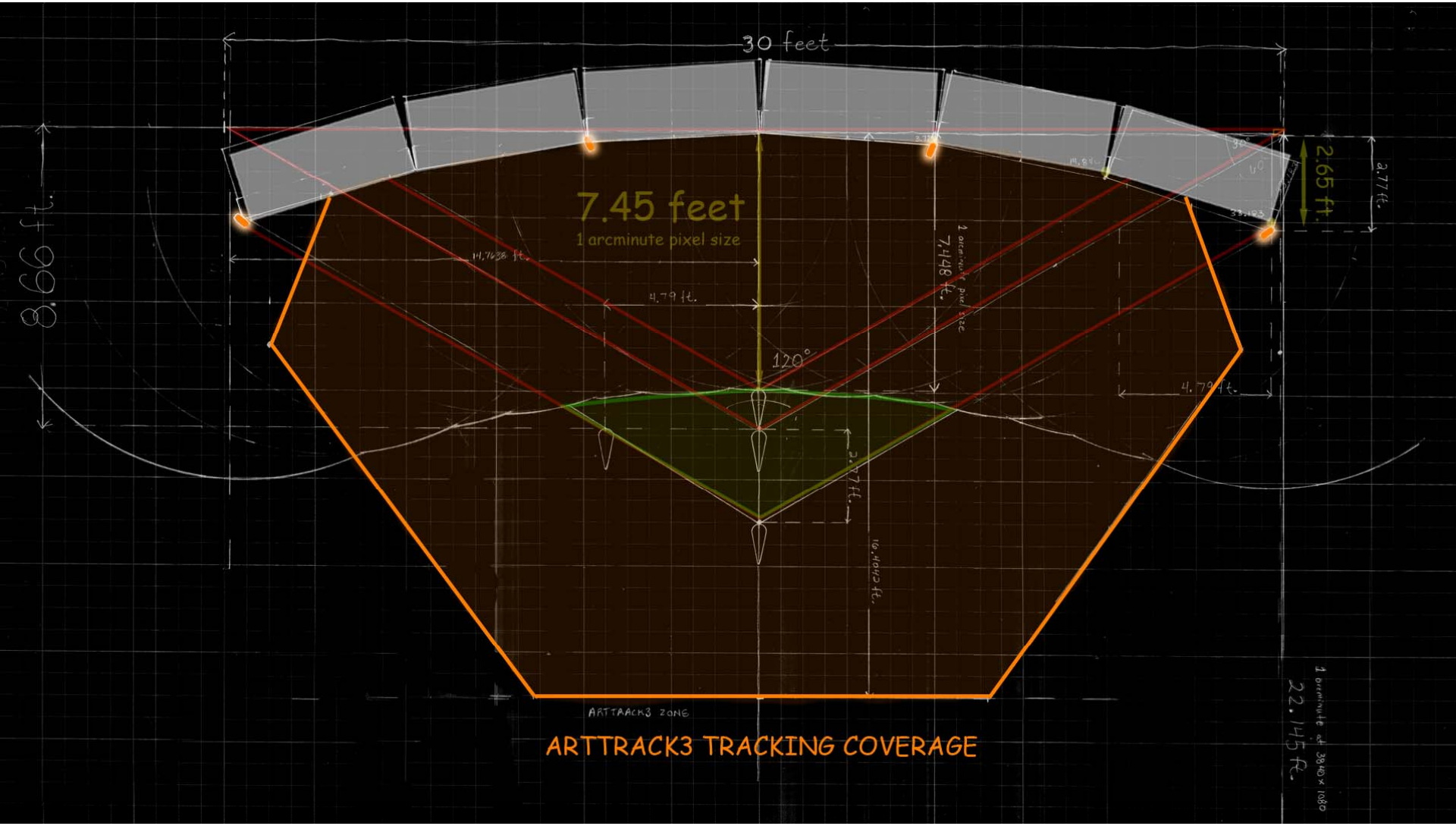

30 feet
7.45 feet
1 arcminute pixel size
2.65 ft.
4.79 ft.
120°
4.79 ft.


30 feet
7.45 feet
1 arcminute pixel size
2.65 ft.
2.77 ft.
11.1638 ft.
4.79 ft.
7.4748 ft.
1 arcminute pixel size
120°
4.79 ft.
2.77 ft.
10.4042 ft.

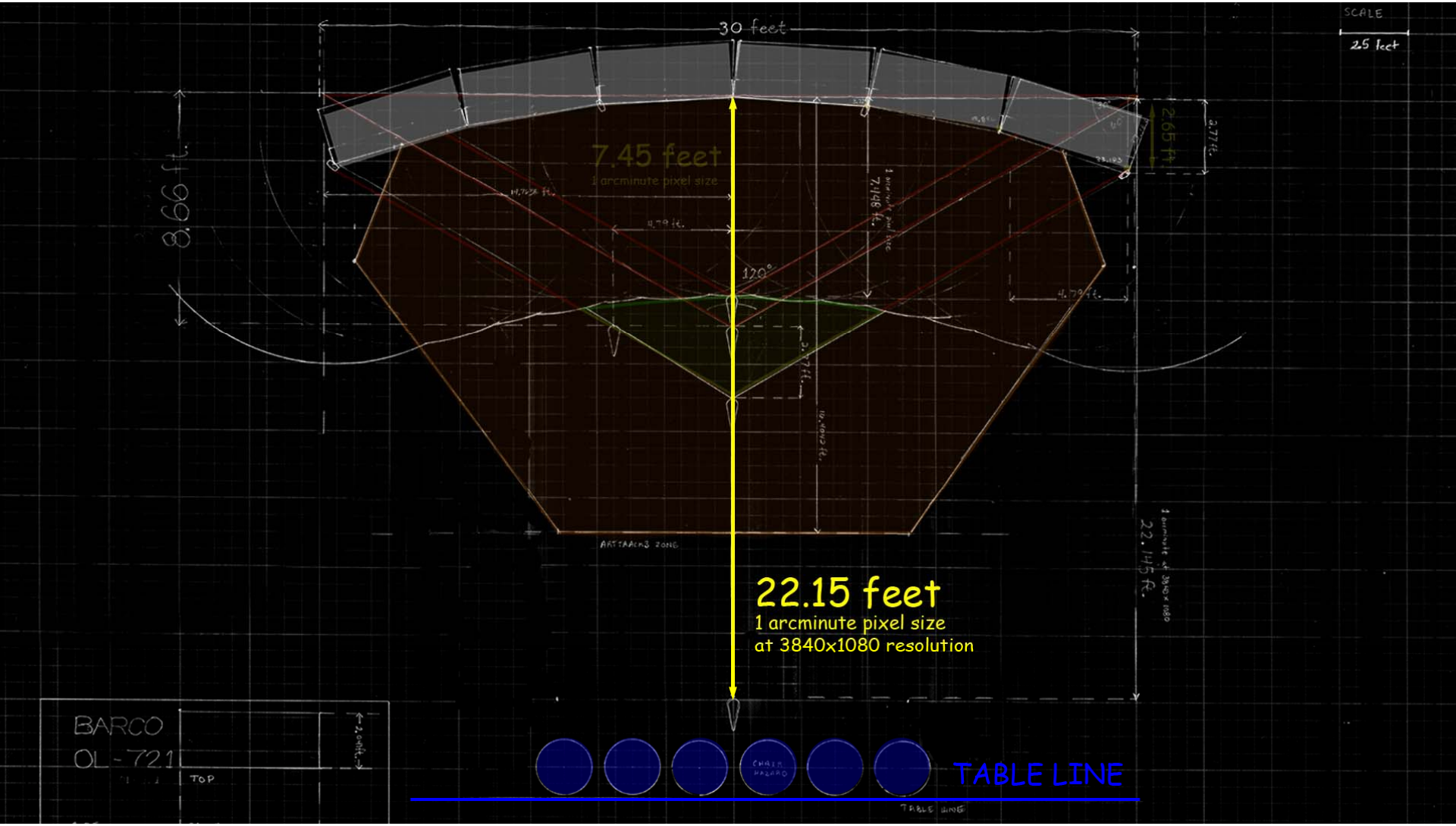

30 feet
7.45 feet
1 arcminute pixel size
2.65 ft.
14.7638 ft.
4.79 ft.
120°
7.4148 ft.
4.79 ft.


30 feet
7.45 feet
1 arcminute pixel size
2.65 ft.
14.7638 ft.
4.79 ft.
7.4148 ft.
120°
4.79 ft.
2.77 ft.


30 feet
7.45 feet
1 arcminute pixel size
2.65 ft.
2.77 ft.
14.7638 ft.
4.79 ft.
7.448 ft.
120°
4.79 ft.
2.77 ft.
16.1042 ft.
ATTACK3 ZONE
22.145 ft.



ARTTRACK3 TRACKING COVERAGE


30 feet
SCALE
2.5 feet
7.45 feet
1 arcminute pixel size
120°
22.15 feet
1 arcminute pixel size
at 3840x1080 resolution
BARCO
OL-721
TOP
CHAIR HAZARD
TABLE LINE

I/O Considerations…

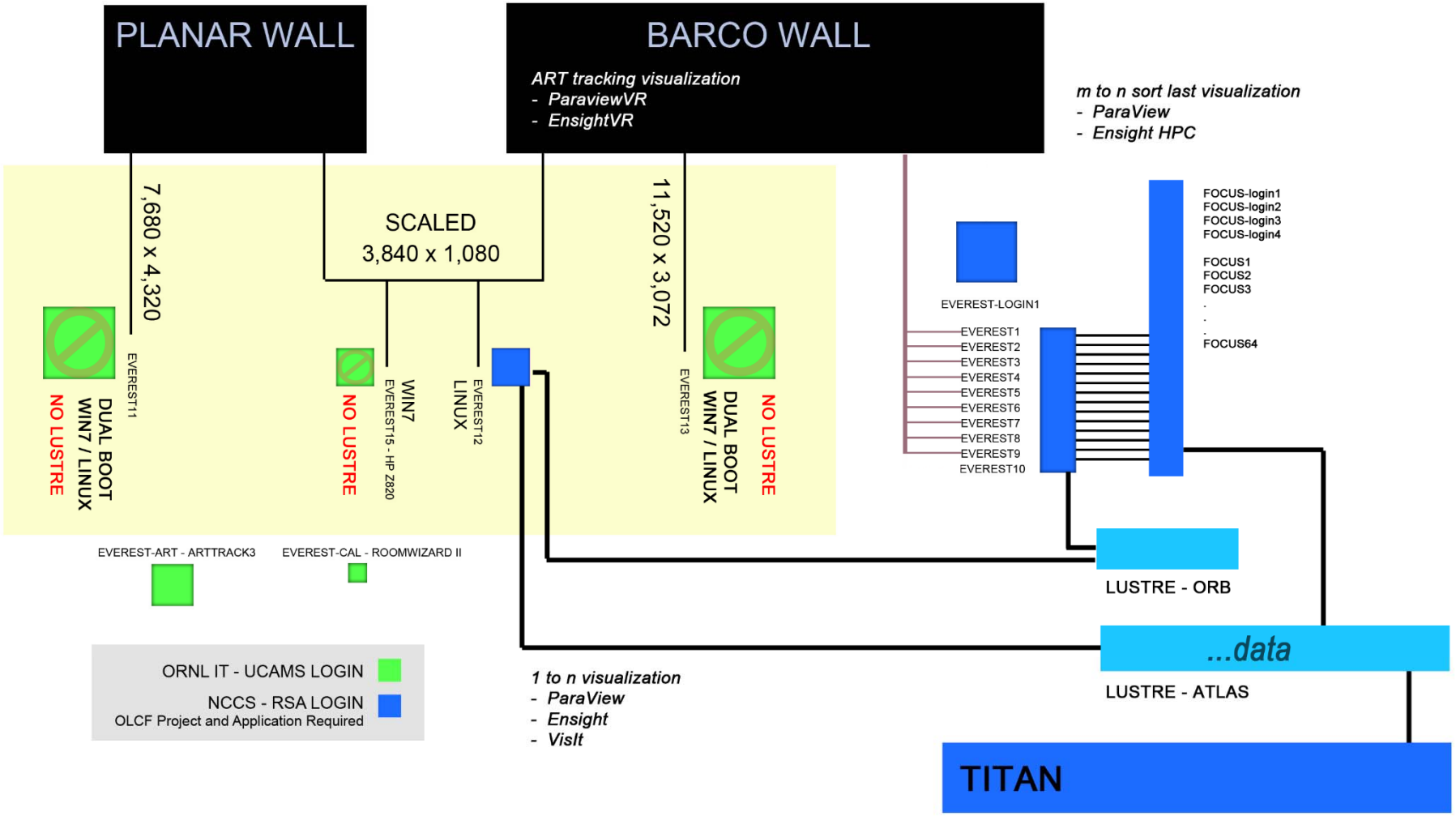
Uncompressed RGB Frames and Required Bandwidth

11,520 x 3,240:
24 frames per second (fps): 2,687 MB/s
48 fps (for stereo): 5,375 MB/s

3,840x1,080:
24 fps: 299 MB/s
48 fps: 597 MB/s

1920x1,080:
24 fps: 149 MB/s
48 fps: 299 MB/s

Solutions: Parallel file systems (lustre), Deep Neural Networks for resolution scaling, Hardware compression (NVENC).


PLANAR WALL
BARCO WALL
ART tracking visualization
- ParaviewVR
- EnsightVR
m to n sort last visualization
- ParaView
- Ensight HPC
7,680 x 4,320
SCALED
3,840 x 1,080
11,520 x 3,072
EVEREST11
DUAL BOOT
WIN7 / LINUX
NO LUSTRE
NO LUSTRE
WIN7
EVEREST15 - HP Z820
EVEREST12
LINUX
EVEREST13
DUAL BOOT
WIN7 / LINUX
NO LUSTRE
EVEREST-LOGIN1
EVEREST1
EVEREST2
EVEREST3
EVEREST4
EVEREST5
EVEREST6
EVEREST7
EVEREST8
EVEREST9
EVEREST10
FOCUS-login1
FOCUS-login2
FOCUS-login3
FOCUS-login4
FOCUS1
FOCUS2
FOCUS3
FOCUS64
EVEREST-ART - ARTTRACK3
EVEREST-CAL - ROOMWIZARD II
LUSTRE - ORB
...data
LUSTRE - ATLAS
ORNL IT - UCAMS LOGIN
NCCS - RSA LOGIN
OLCF Project and Application Required
1 to n visualization
- ParaView
- Ensight
- VisIt
TITAN

Focus – Visualization Analysis Engine (18 nodes, 11TB, 576 cores)

Quantity: 4 (4 TB, 128 cores)
CPU: 4 x Intel Xeon E5-4620v2, 2.6GHz (8-Core, HT, TB, 20MB Cache, 95W) 22nm
RAM: 1TB (32 x 32GB DDR3-1333Controller: Dual-Port Intel I350 GigE plus 2 Ports 6Gb/s
GPU: NVIDIA GTX 780Ti 3GB GDDR5 PCIe 3.0
FDR: Mellanox ConnectX-3 FDR InfiniBandAdapter

Quantity: 14 (7 TB, 448 cores)
CPU: 4 x Intel Xeon E5-4620v2, 2.6GHz (8-Core, HT, TB, 20MB Cache, 95W) 22nm
RAM: 512GB (32 x 32GB DDR3-1333Controller: Dual-Port Intel I350 GigE plus 2 Ports 6Gb/s
GPU: NVIDIA GTX 780Ti 3GB GDDR5 PCIe 3.0
FDR: Mellanox ConnectX-3 FDR InfiniBandAdapter

Remote Visualization

- Imposer Libraries
  - NICE DCV (commercial)
  - Chromium (GPL, no longer supported)
  - VirtualGL
- Virtualization
  - ThinLinc (commercial)
- Client/Server
  - ParaView, Ensight, VisIt
  - Sight (Benjamin Hernandez, OLCF)
- Image Based
  - NVENC (hardware H.265)
  - VNC (software)

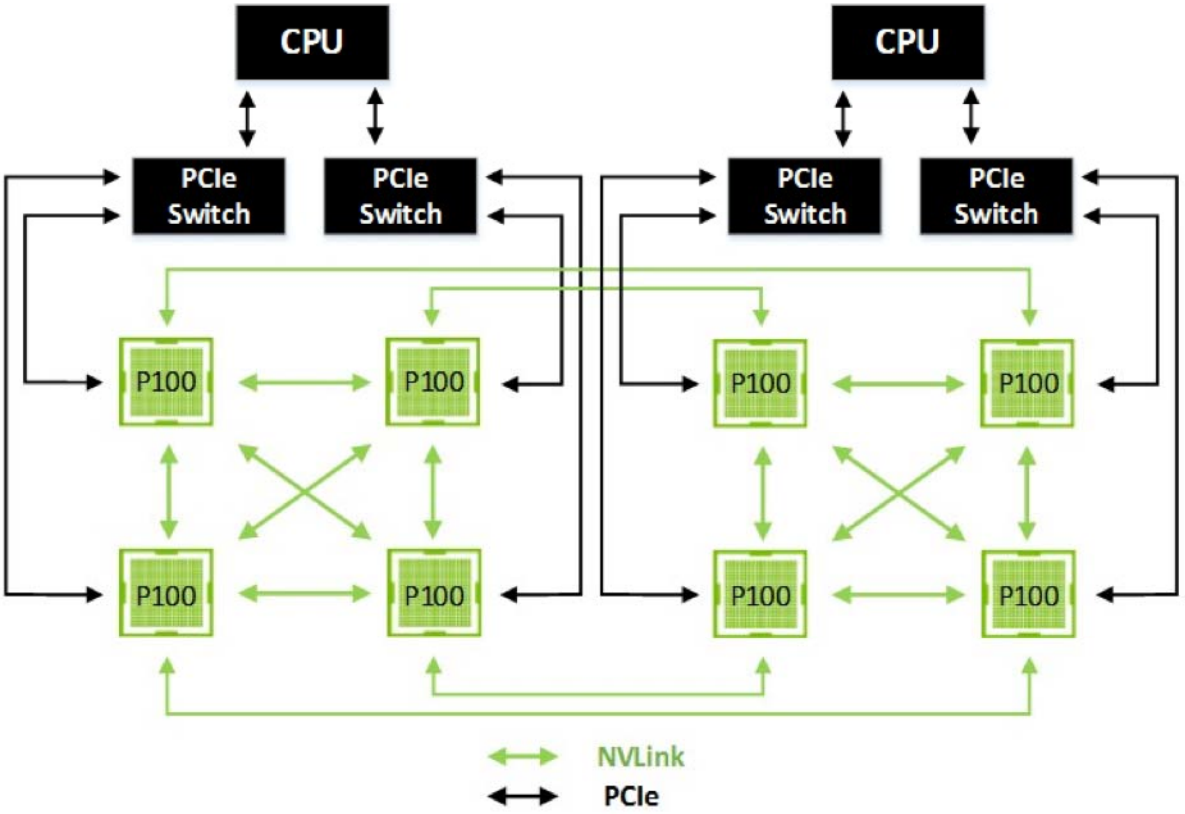
Future Technologies

- Nvidia V100 Volta GPU, NVLink.
- Deep Learning and Neural Networks for analysis and resolution scaling.
- In-Situ Visualization – VTK-m.
- Augmented Reality – video and optical passthrough technologies.
- Field-of-Light display technologies.

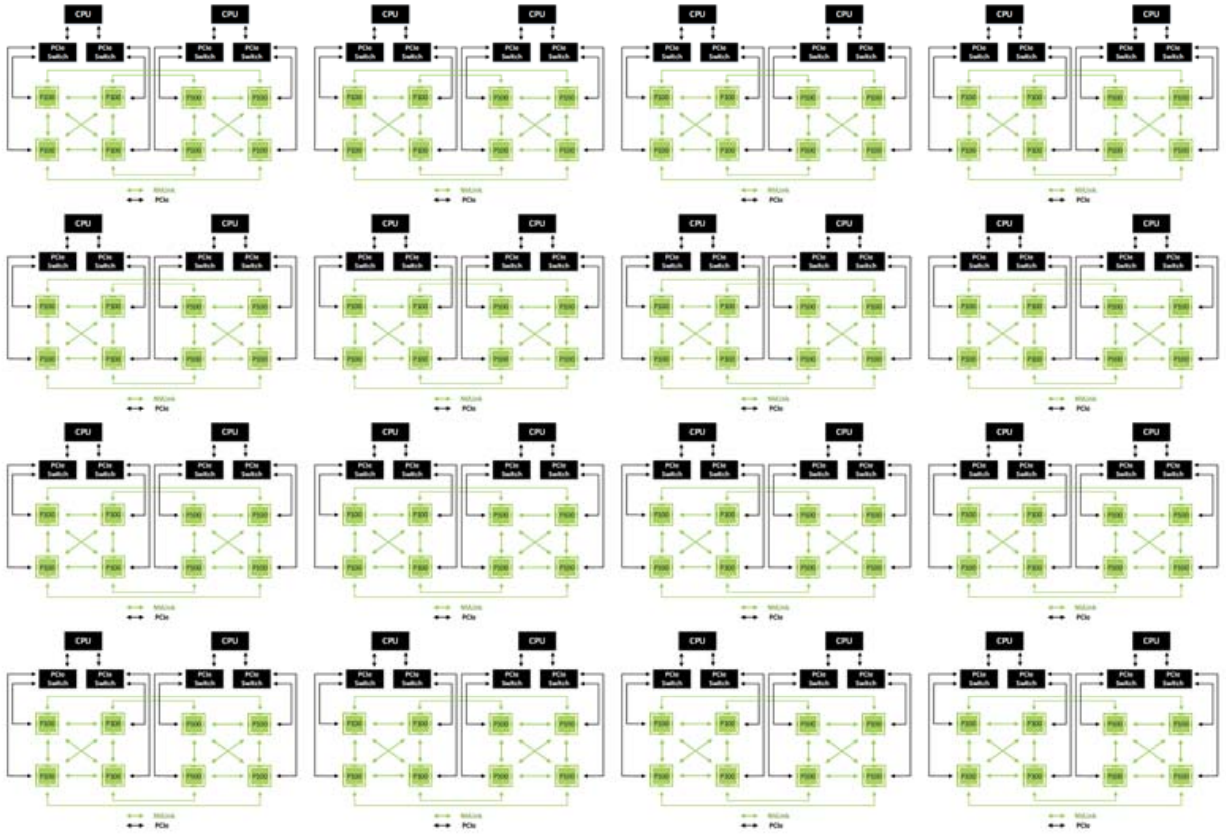
Nvidia's DGX-1 Deep Learning Supercomputer

- 8 Tesla P100 GPUs
- GPU Mem: 16 GB per GPU
- System Mem: 512 GB
- Storage: 4x 2 TB SSDs
- Out of the box containers (tensorflow, theano, caffe, cuDNN, cuBLAS, etc)

Nvidia's DGX-1 Deep Learning Supercomputer

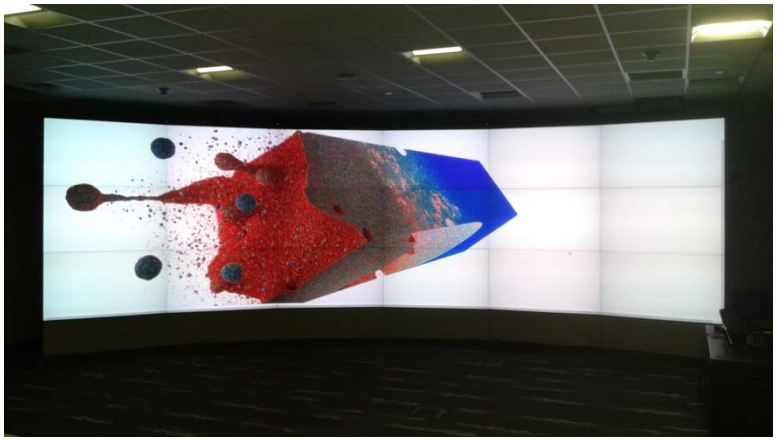


Risks – Scalability and Complexity

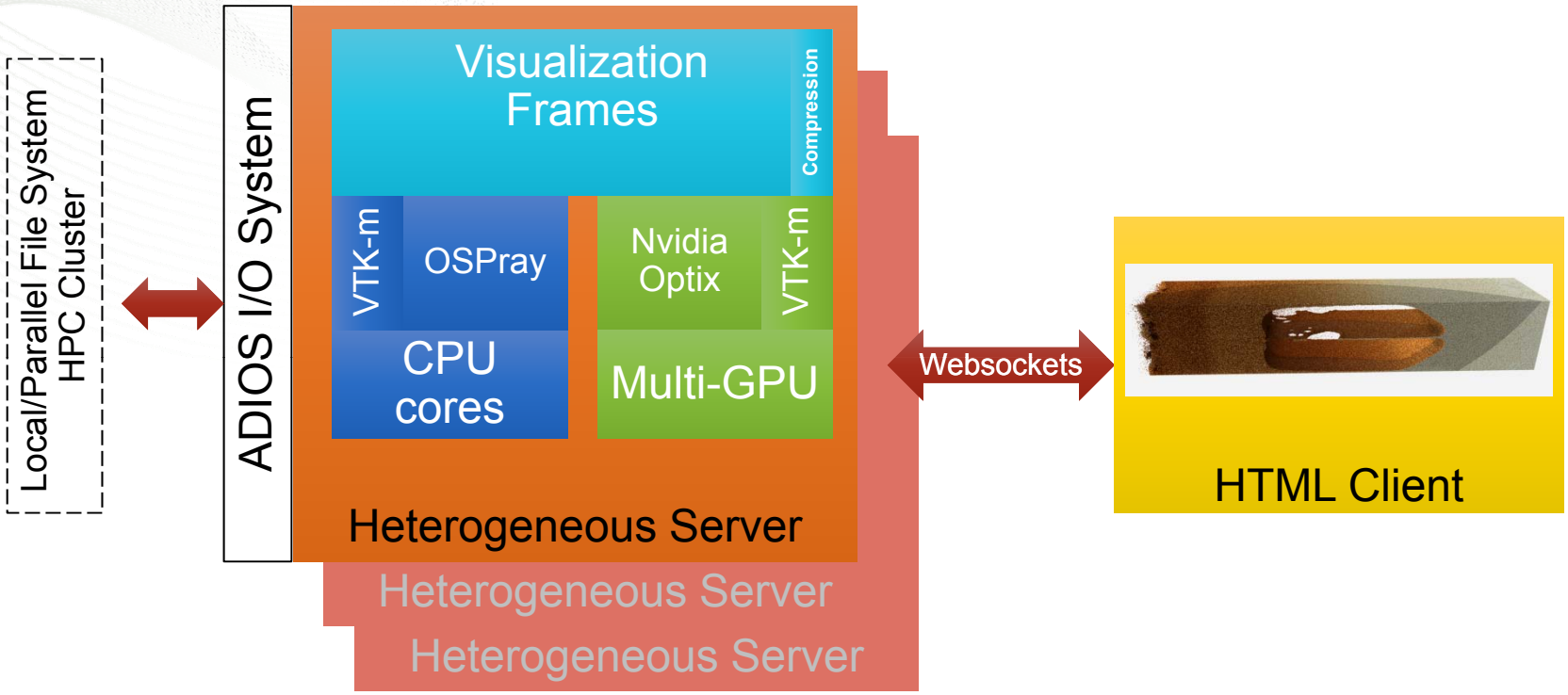


Sight: Exploratory Visualization of Scientific Data
Features

- Server/Client architecture to provide high end visualization in laptops, desktops, and powerwalls.
- Heterogeneous scientific visualization
  - Take advantage of both CPU & GPU resources within a node: DGX-1 use case.
  - Advanced shading to enable new insights into data exploration.
- Parallel I/O & Data Staging
  - "Pluggable" for in-situ visualization
- Lightweight tool
  - Load your data
  - Perform exploratory analysis
  - Visualize/Save results



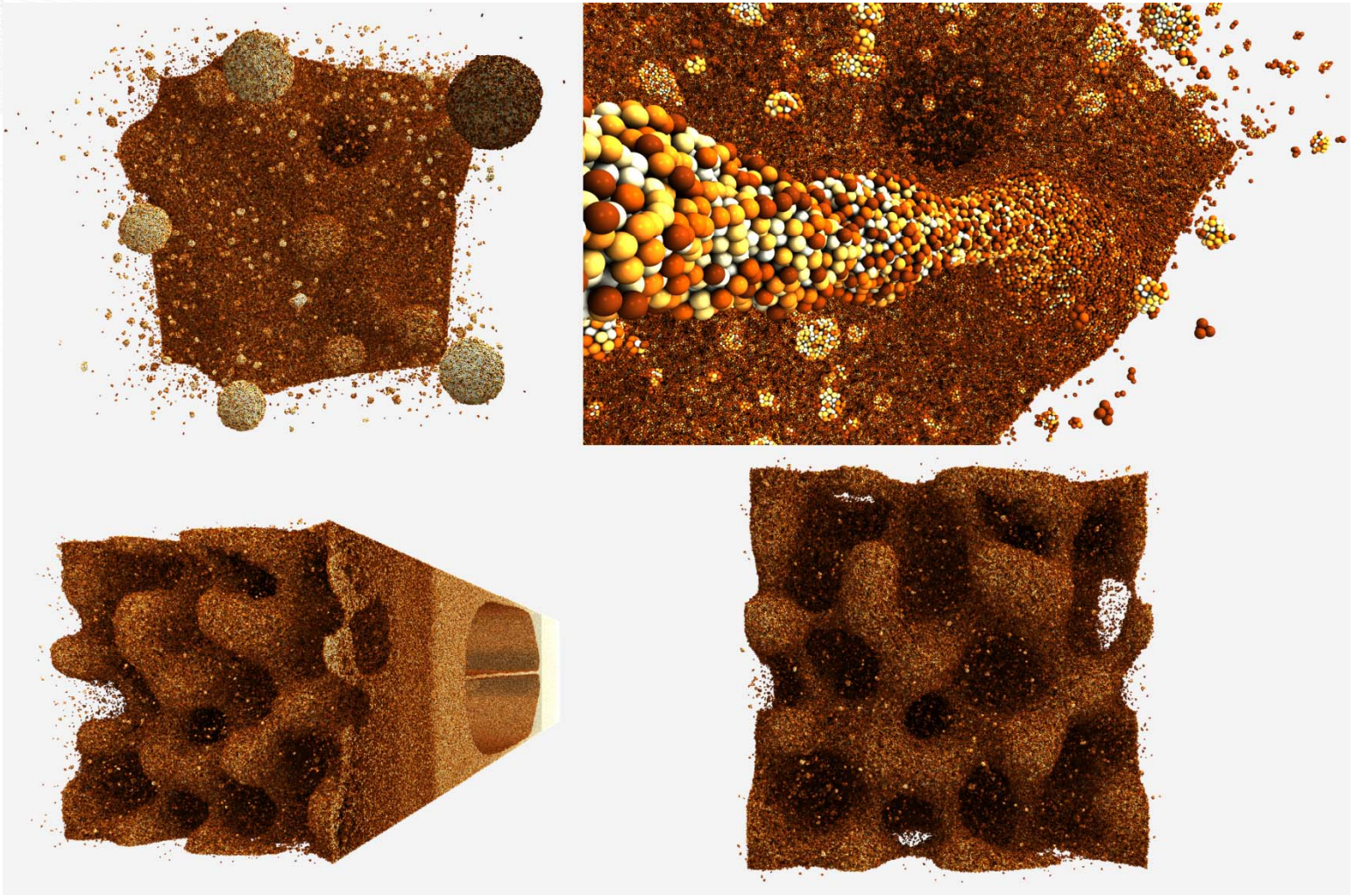
Sight System Architecture (in progress)



*OSPray and Nvidia Optix are finely tuned libraries for Ray Tracing in multicore and manycore architectures

Reference:
Benjamin Hernandez, "Exploratory Visualization of Petascale Particle Data in Nvidia in DGX-1". Nvidia GPU Technology Conference 2017, Sillicon Valley, CA.

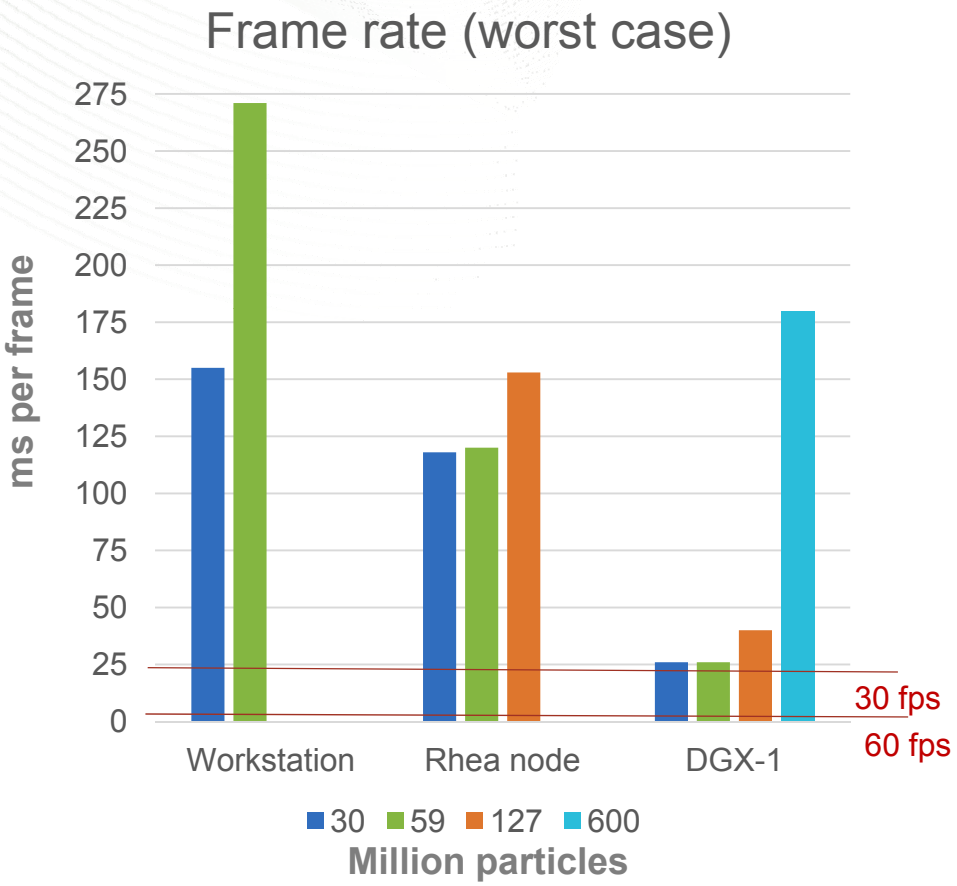
Preliminary Results (particle data)

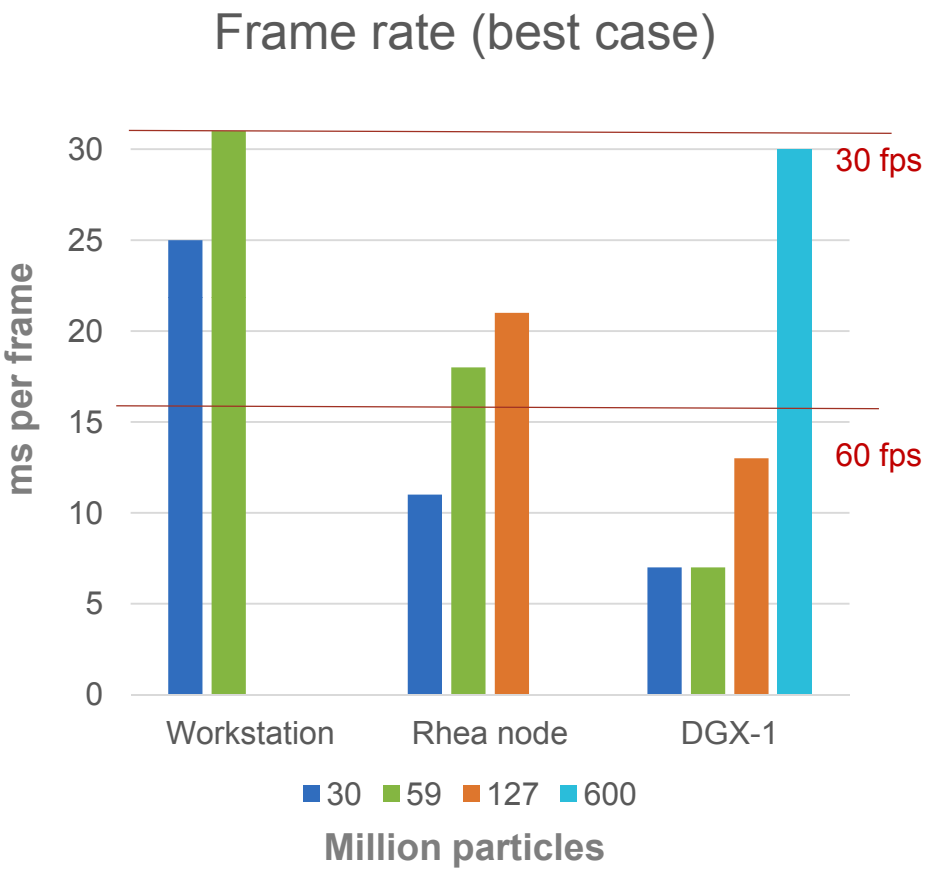


Simulation: OLCF INCITE 2017 "Petascale Simulations of Short Pulse Laser Interaction with Metals" PI Leonid Zhigilei, University of Virginia

Vis: Benjamin Hernandez, OLCF-ORNL

Preliminary Results using Optix API Performance

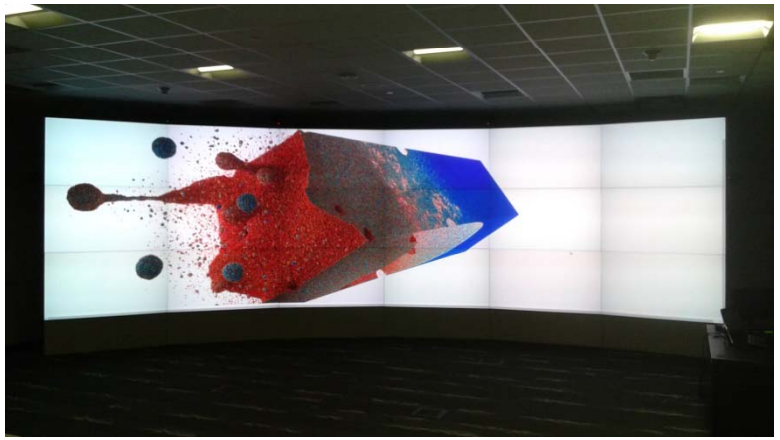




Reference:
Benjamin Hernandez, "Exploratory Visualization of Petascale Particle Data in Nvidia in DGX-1". Nvidia GPU Technology Conference 2017, Sillicon Valley, CA.

Discussion

- DGX-1 can handle **particle systems** up to **10x larger** in our test environment.
- For particle systems of the same size **DGX-1** is **10x faster** than the workstation system and 4.6x faster than Rhea node
  - We expect for larger image resolutions DGX-1 speed up will increase.
- Our preliminary tests showed DGX-1 has enough compute power to drive a powerwall
  - 3840 x 1080 @ 60 – 120 fps
    - Test larger resolution
  - Researchers usually are happy when they can explore datasets even at 1 fps



Questions?