Hall A Software & Analysis

Ole Hansen

Jefferson Lab

JLab 12 GeV Software Review
Morning Session
June 7, 2012

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Outline

1. Overview
2. Data & Analysis Flow
3. Software Components
4. Status, Tasks, Manpower
5. Computing Requirements
6. Summary

Hall A Collaboration

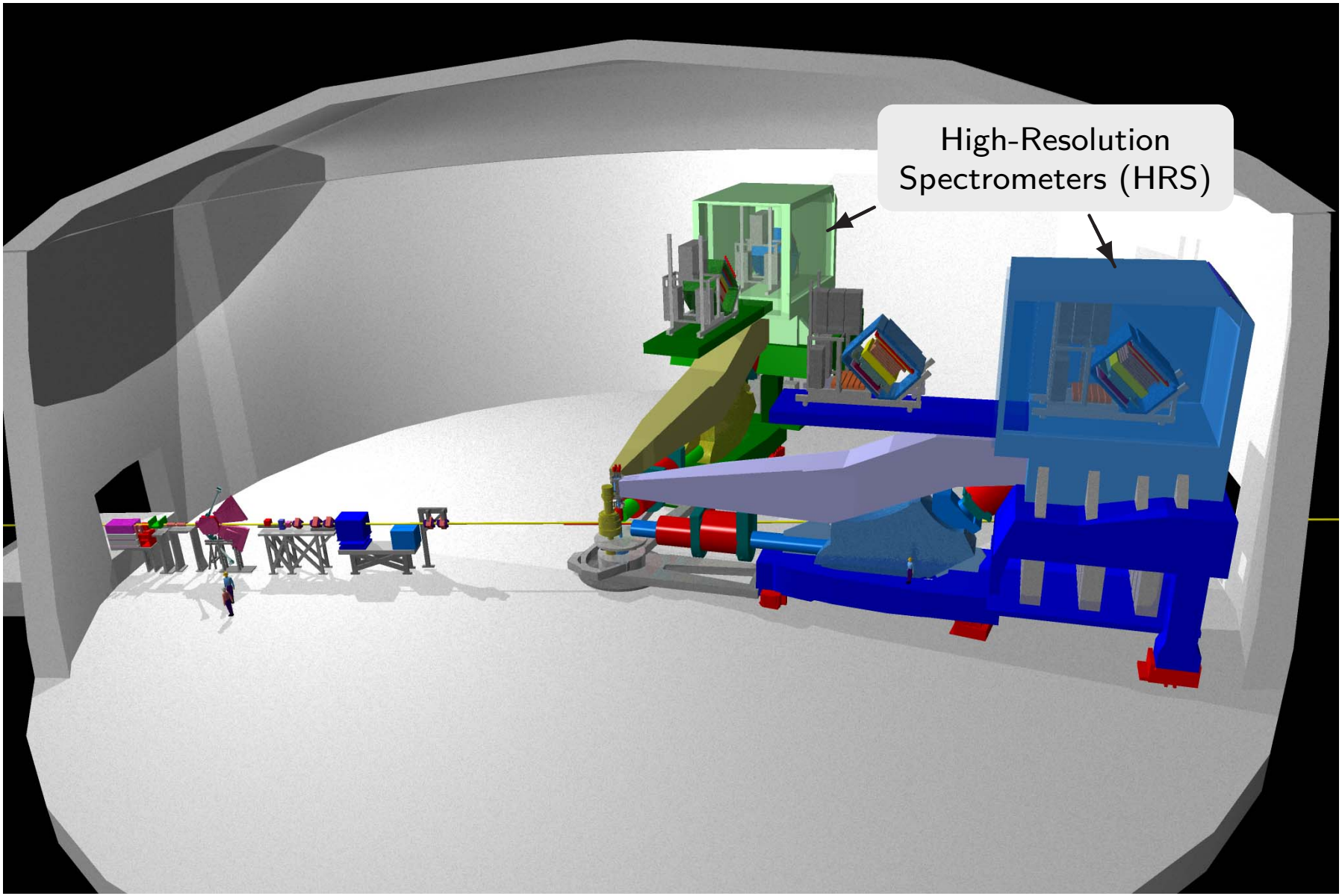
(insert list of Hall A Collaboration institutions here)

Physics Topics in Hall A

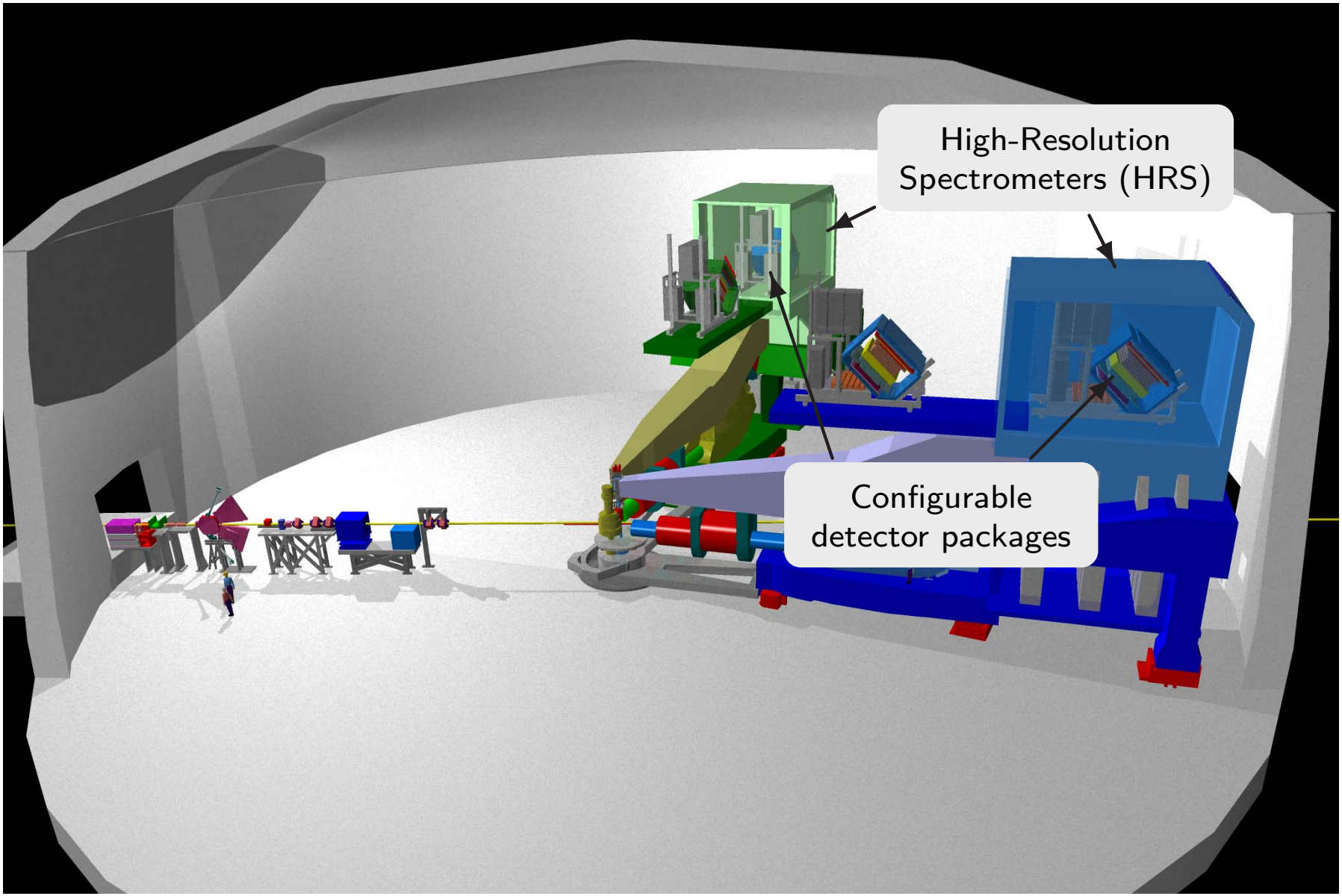
1. Transverse Hadron Structure (EM form factors)
2. Longitudinal Hadron Structure (PDFs, valence quark structure)
3. 3D Hadron Structure (GPDs, TMDs)
4. Hadrons and Cold Nuclear Matter (NN correlations, medium modifications, hypernuclear physics, few-body physics)
5. Low-energy tests of the Standard Model and Fundamental Symmetries (PREX, APEX)

Covers 5 of the 6 JLab physics categories
Many A-rated experiments

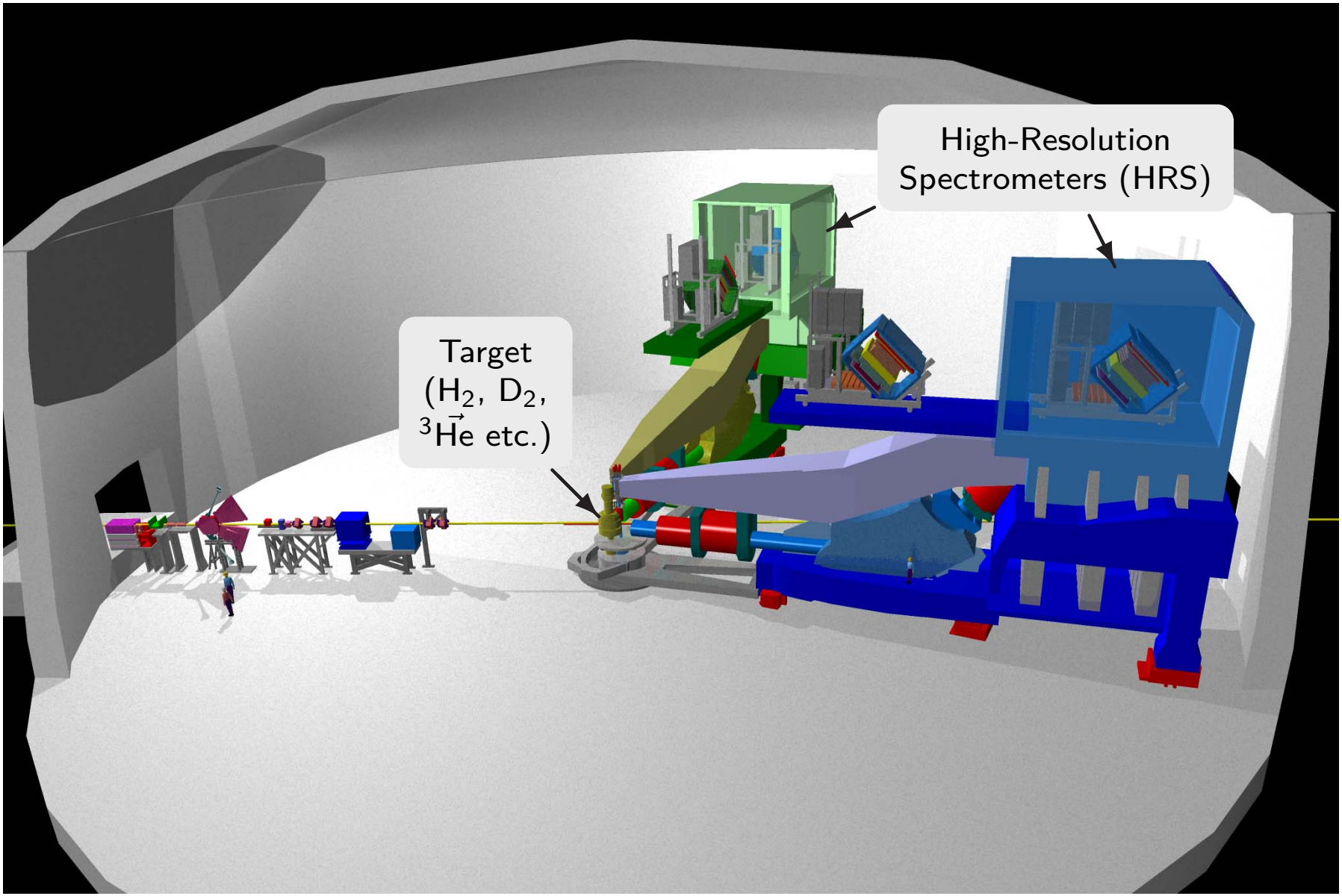
Hall A Base Equipment



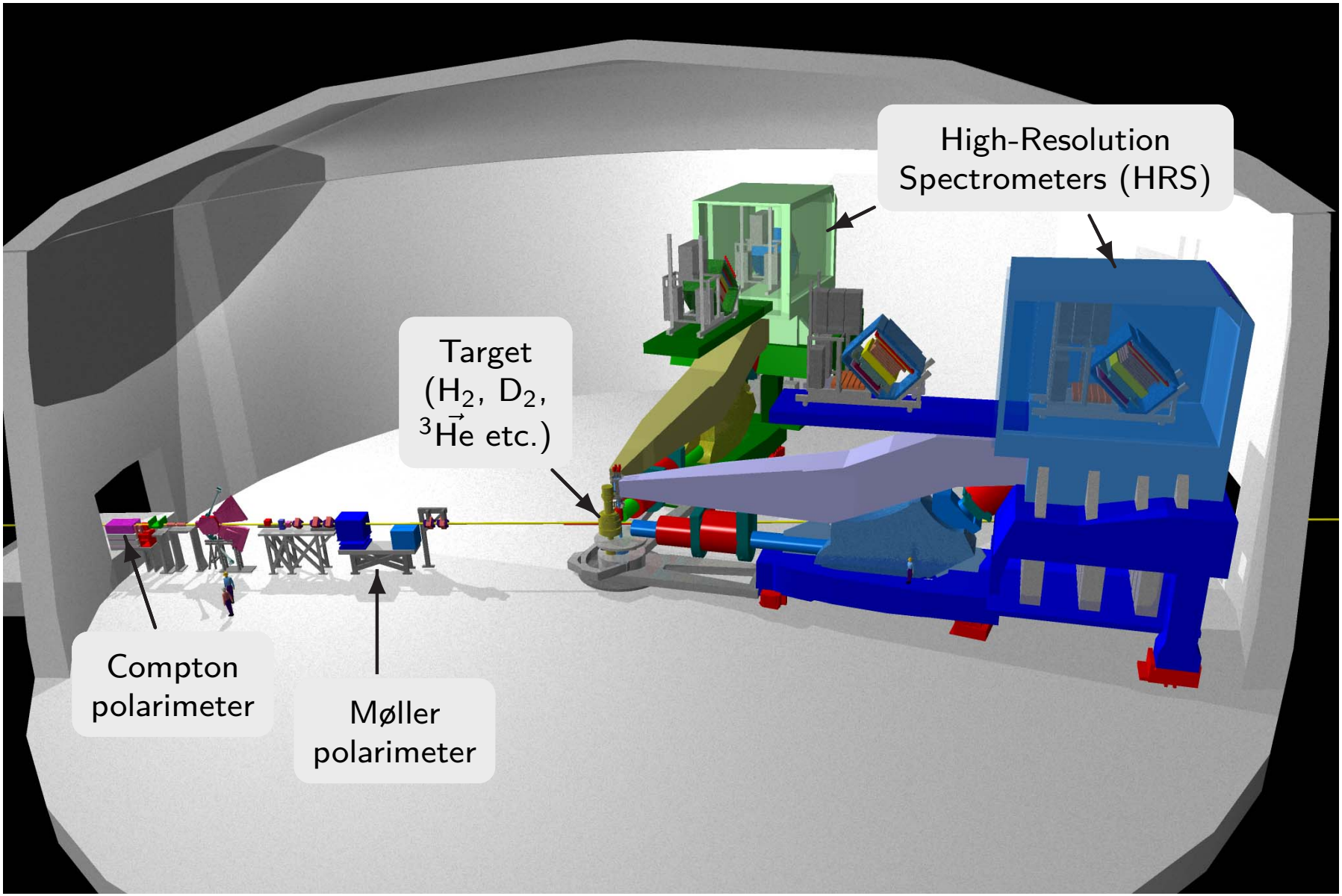
Hall A Base Equipment



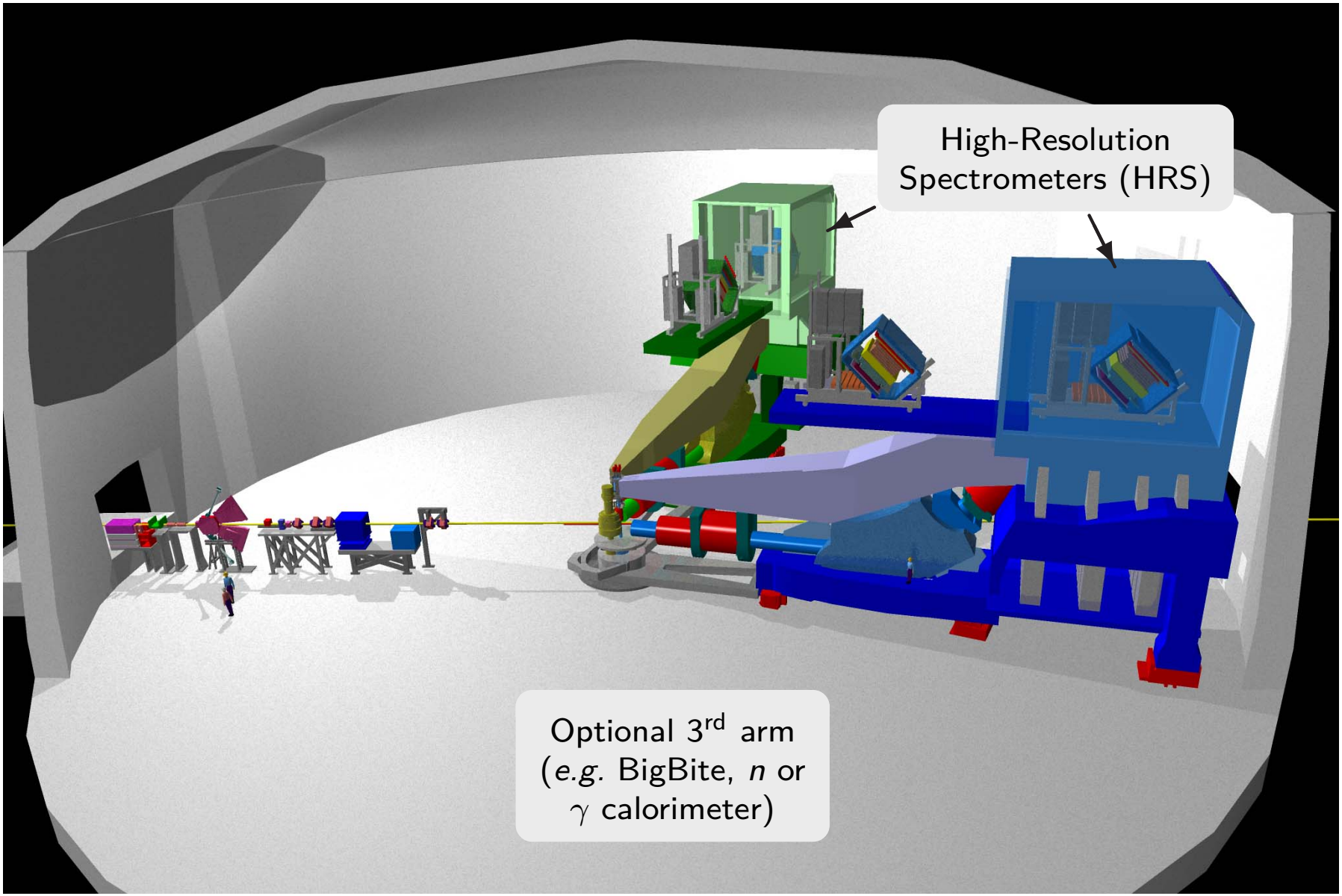
Hall A Base Equipment



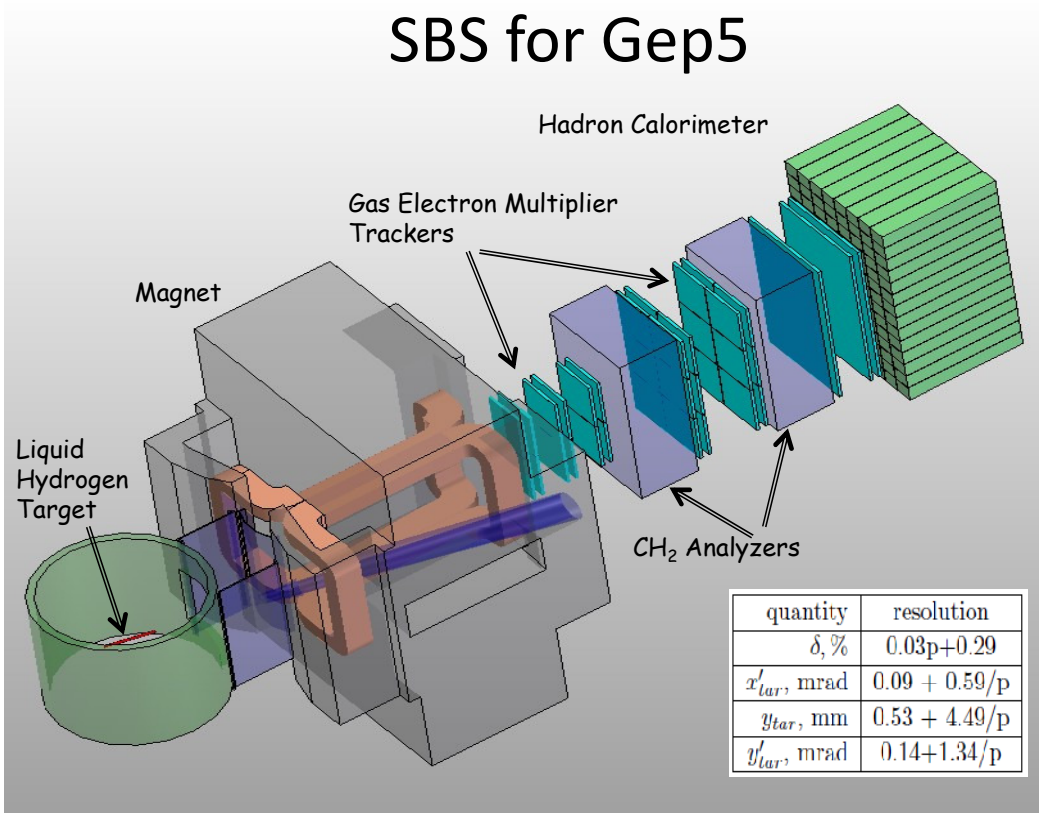
Hall A Base Equipment



Hall A Base Equipment



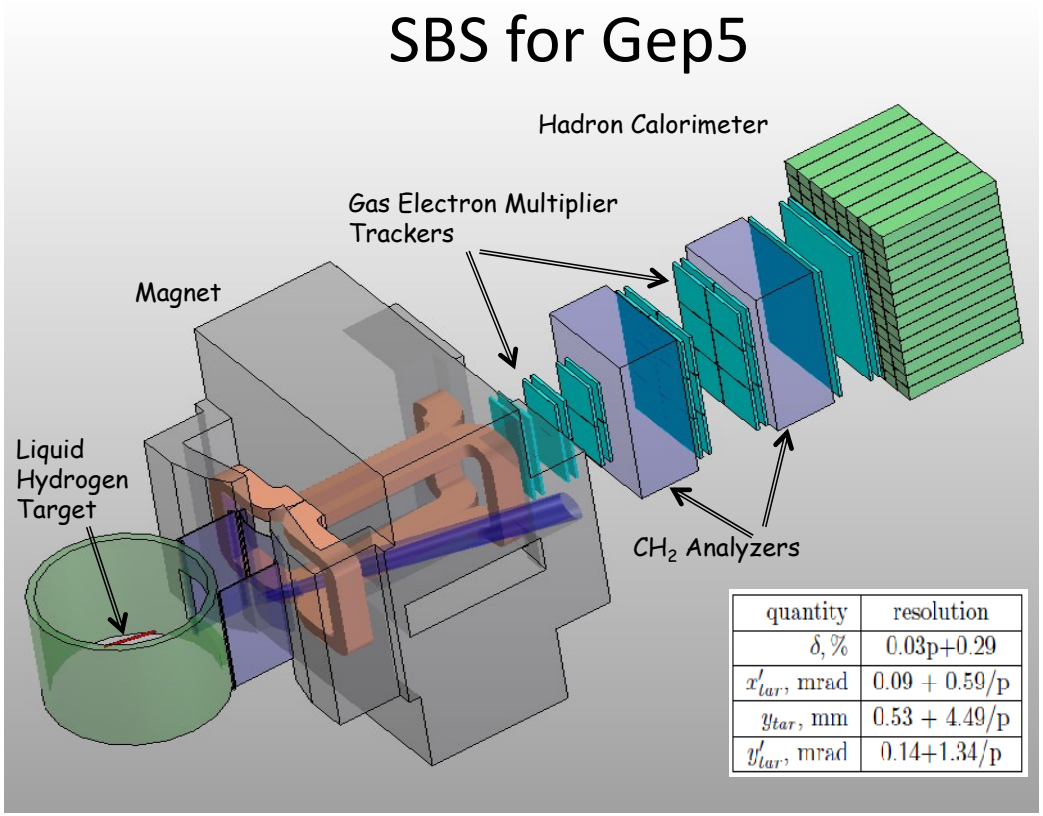
Hall A SuperBigbite Spectrometer (SBS) Project



quantity	resolution
δ, %	0.03p+0.29
x'_{tar}, mrad	0.09 + 0.59/p
y_{tar}, mm	0.53 + 4.49/p
y'_{tar}, mrad	0.14+1.34/p

- Set of components for flexible medium-acceptance spectrometer configuration
- GEM trackers for high-rate tracking
- Propsals approved for
  - EM form factor measurements to very high Q^2
  - SIDIS/Transversity
- Project Management Plan (PMP) approved by DOE in 2011
- Data taking tentatively in FY16

Hall A SuperBigbite Spectrometer (SBS) Project

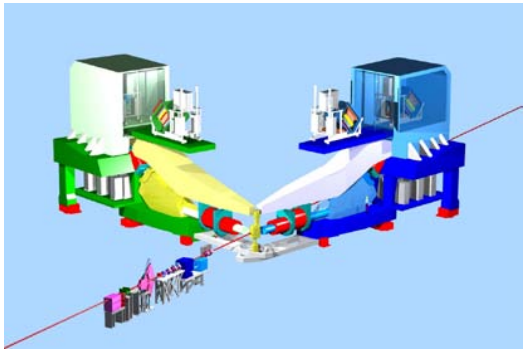


quantity	resolution
δ, %	0.03p+0.29
x'_{tar}, mrad	0.09 + 0.59/p
y_{tar}, mm	0.53 + 4.49/p
y'_{tar}, mrad	0.14+1.34/p

- Set of components for flexible medium-acceptance spectrometer configuration
- GEM trackers for high-rate tracking
- Propsals approved for
  - EM form factor measurements to very high Q^2
  - SIDIS/Transversity
- Project Management Plan (PMP) approved by DOE in 2011
- Data taking tentatively in FY16

Great variety of different experimental configurations in Hall A

Timeline of Early 12 GeV Hall A Experiments (from Bob Michaels)



Tentative!

16 mo. Shutdown

12 GeV Commissioning

Early Experiments

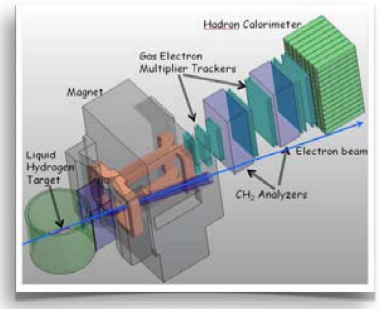
APEX and/or PREX
Dark Matter Neutron Skin

A1n / DVCS
Neutron Spin Structure Deeply Virtual Compton Scatt.

SBS Experiments
EM Form Factors at high Q^2

SBS Project

SuperBigbite



Beam 1st to Hall A

11 GeV

FY 2013 | FY 2014 | FY 2015 | FY 2016 | FY 2017

Hall A Data Acquisition

- Early experiments (through FY15):
  - Fastbus (existing, tested)
  - Non-pipelined VME (existing, tested)
  - Some pipelined JLab 12GeV ("GlueX") electronics (TBD)
- SBS (FY16–FY18):
  - Pipelined JLab 12GeV electronics, trigger system
  - Custom pipelined VME electronics for trackers
  - Some legacy Fastbus/VME (TBD)

Raw Trigger & Data Rates

Preliminary schedule & numbers

Experiment	APEX[1]	A_1^n	DVCS	SBS			
				G_E^n	G_M^n	$G_E^p(5)$	Transv
PAC number 12-	10-009	06-122	06-114	09-016	09-019	07-109	09-018
Config	L+R(CI)	L+R(SA)	L+γCal	BBG+ND	BBG+ND	SBS+BC	SBS+BB
PAC days	34	23	88	58	48	60	64
Schedule		—— 2014/2015 ——			—— 2016/2018 ——		
Evt size (kB)	4	2	30	30	20	120	5
Trig rate (kHz)	5	2×10	0.5	2	2	1	5
Data rate (MB/s)	20	40	15	60	40	120	25

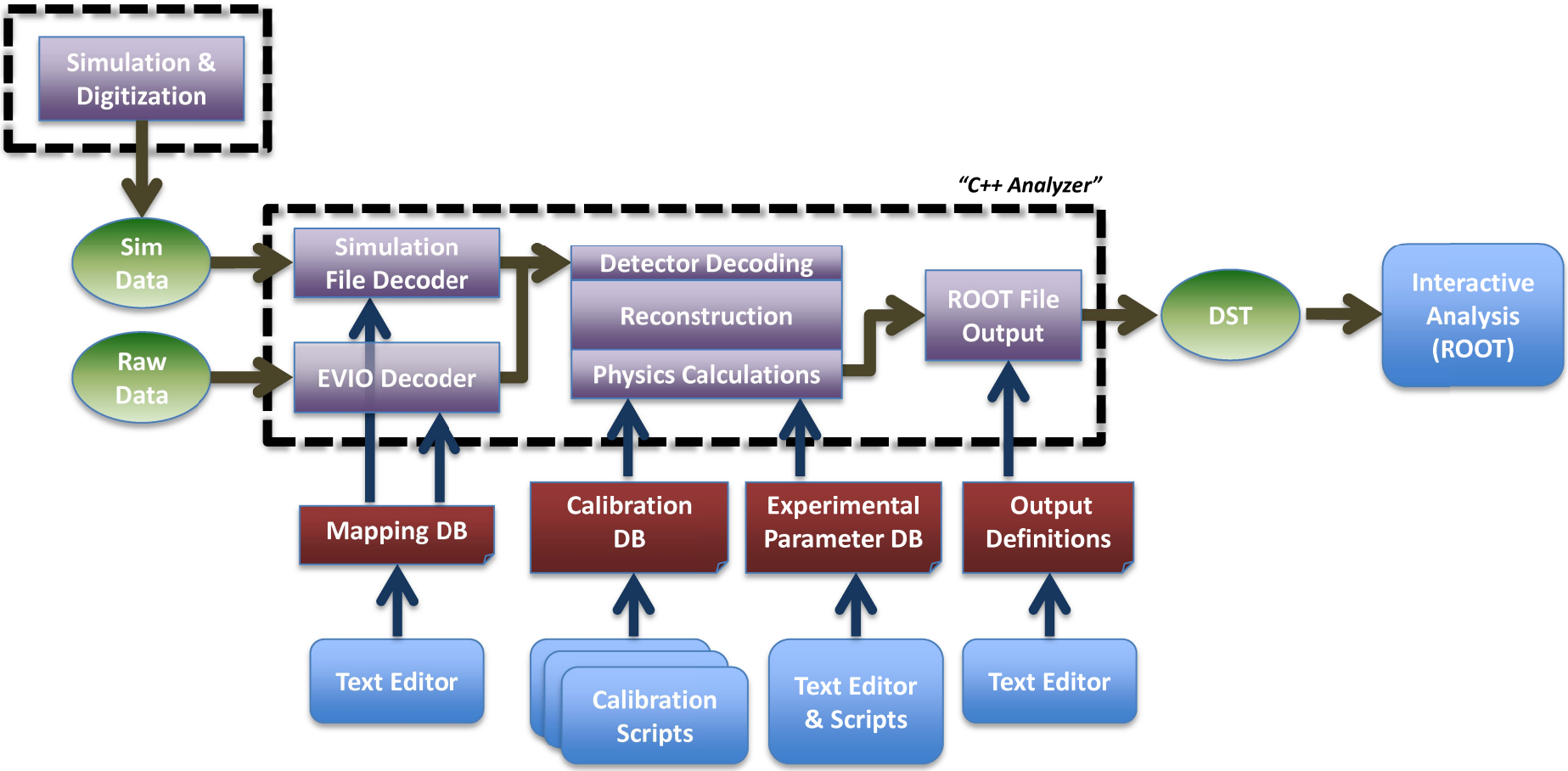
L: L-HRS, R: R-RHS, BB: BigBite, BBG: BB(GEM), ND: neutron det, SBS: SuperBigBite, BC: BigCal, CI: coinc., SA: sing. arm

[1] PREX requirements are negligible compared to APEX

Data Flow

(insert insanely complex illustration of data flow from DAQ $\rightarrow$ farm)

Analysis Software Components



Hall A Analysis (Reconstruction) Software

- Hall A analyzer ("Podd"), in production use since 2003
  - C++/ROOT-based
  - Highly modular. Many experiments write custom modules for their special requirements. SDK available.
  - Supported on Linux, Mac OS X, Solaris (deprecated)
  - Fully 64-bit compatible
  - Limitations
    - ★ single-threaded → Plan: multi-threaded by FY13
    - ★ not distributed (like *e.g.* Hall B) → no need anticipated
  - Development shared with Hall C
  - Performs "Reconstruction"
  - Physics "Analysis" typically done interactively on desktops using ROOT
- Custom software (user-supported)
  - Parity experiments: Parity Analyzer (PAN), an independent C++/ROOT development
  - DVCS: extensive custom code in addition to standard analyzer

Calibrations, Data Quality Checks, Prompt Analysis

- Approach
  - Instant replay of raw data from local disks on counting house cluster
  - Replay usually in real time $\rightarrow$ expected to be possible up until high-rate SBS experiments
  - Calibrations
    - ★ standard scripts available
    - ★ required custom scripts written for specific experiments by users
  - Data Quality Checks via customizable interactive viewer
- Counting House Cluster Resources
  - 12 dedicated CPU cores $\rightarrow$ 64 cores by FY15
  - 1.5 + 6 TB (raw + DST) local disk $\rightarrow$ 15 + 60 TB by FY15, 30 + 120 TB by FY18 to hold $\approx$ 10% of raw/analyzed data
  - RHEL5 $\rightarrow$ RHEL6 by FY13
  - Managed by Hall A staff
  - Funding out of Hall A operations
  - Expect IT support for routine backups & networking

Simulations

- Typically low-volume
- Typically run off-site or on user desktops
- Existing frameworks
  - SIMA (Hall C's SIMC adapted to Hall A): matrix optics for transport through spectrometers
  - MCEEP: different development, similar to SIMA/SIMC, unmaintained
  - GEMC: Geant4-based Hall B development, adopted by some future large-installation experiments in Hall A (SoLID)
  - SBSsim: custom Geant4 development, maintained by/run at INFN
- Spectrometer-based experiments (*i.e.* early exp'ts through FY15) well simulated with matrix-based codes
- GEMC computing requirements being collected, modest so far but expected to increase

Software Status, Tasks

- Basic reconstruction software ready for early HRS-based 12 GeV experiments
- Main open items
  - High-rate VDC track reconstruction for APEX
    - ★ successfully addressed with test run data
    - ★ still need to integrate into mainline analyzer
  - SBS analysis software (GEM tracking, calorimeter)
    - ★ part of SBS management plan
    - ★ external institutions responsible (CMU, INFN, UVa, etc.)
  - Analyzer parallelization/multi-threading
  - Support for pipelined JLab 12GeV electronics

Milestones

Month/Year	Goal
7/2012	Move to `git` code management system (for Hall C)
10/2012	Object-oriented decoder implementation (for Hall C)
12/2012	Parallel architecture of analyzer implemented
3/2013	APEX VDC code integrated
6/2013	Decoding of pipelined electronics ready
9/2013	Pipelined electronics decoding tested

Software Management & Manpower

- One Hall A staff (OH) dedicated to software management, coordination & development ($\approx$ 0.5 FTE)
- New developments typically carried out by external users, with guidance from Hall A staff. Specifically, this will be the case for SBS
- Setup and configuration of online and offline replay routinely done by students/postdocs from experiment user groups
- No formal structure or assigned responsibilities for early 12GeV experiments (through FY15) since software considered ready
- SBS project software responsibilities assigned in separate SBS management plan

Collaboration Resources

- Annual "Analysis Workshop" in conjunction with collaboration meeting
- Extensive web resources at `http://hallaweb.jlab.org/podd/`
  - Web-based user guide
  - Example scripts
  - ROOT `THtml` reference documentation
  - Software development kit (SDK)
- Bi-weekly meetings with Hall C

Estimating Hall A Computing Requirements

(insert brilliant explanation of the assumptions for next slide's table)

Summary of Hall A Computing Requirements

Anticipated SciComp resources (not yet updated for current schedule)

	2013 g2p replay	2014 COMISS	2015 HRS/BB	2016 SBS	2017 SBS
Time per event/core (ms)	5	5	20	40	60
Passes through data	1	2	3	3	3
Output size/input size	1	2	1	1	1
Years to analyze	1	1	3	3	3
Replay duty factor	50%	50%	50%	75%	75%
Output held on work disk	10%	20%	20%	20%	10%
CPU time per year (s)	1.9e8	1.8e8	6.6e8	1.4e9	2.0e9
Dedicated farm cores	12	12	42	60	84
Cooked data to tape (TB)	245	174	132	510	1641
Work disk storage (TB)	13	25	25	23	26
Avg bandwidth (MB/s)	16	31	31	20	17
Totals					
Farm cores (2011 vintage)	3	12	12	12	42
New cores each year	0	9	0	0	30
Raw+cooked to tape (PB)	.26	.36	.25	.19	.26
Disk storage (TB)	13	25	25	23	26
Storage bandwidth (MB/s)	25	41	31	23	34

Summary & Conclusions

- Hall A has a mature software framework, in production use for over 8 years
- Hall A reconstruction software essentially ready for early 12GeV experiments through FY15
- FY16-17 SBS project requires extensive software development, planned to be carried out by several external institutions
- Hall A computing resource requirements through FY17 relatively modest
- Unlike other halls, the exact Hall A requirements and specifications are harder to forecast due scheduling uncertainties and multiple possible experimental configurations