



DATA ANALYSIS AND ITS NEEDS @ EUROPEAN XFEL

WG3:

Filipe Maia, Peter van der Meulen, Abbas Ourmazd, Ivan Vartanyes, Gabor Bortel, Krzysztof Wrona, Massimo Altarelli, Gosta Huldt, Daniel Larsson, Rafael Abela, Veit Elser, Tomas Ekeberg, Kewish Cameron, David van der Spoel, Hidetoshi Kono, Fengling Wang, Pierre Thibault, Adrian Mancuso + 3 more people

BIO-IMAGING DATA

Data Rate of Bio-Imaging Experiment		
	LCLS	EU-XFEL
Rep Rate	120	5000
Detector Size (Megapixel)	1-4	1-4
Intensity Depth (bit)	32	16
Data Rate (Gigabit/s)	4-16	80-320
Data Production (Pb/day)	0.04-0.16	0.9-3.5

SYSTEM INFRASTRUCTURE

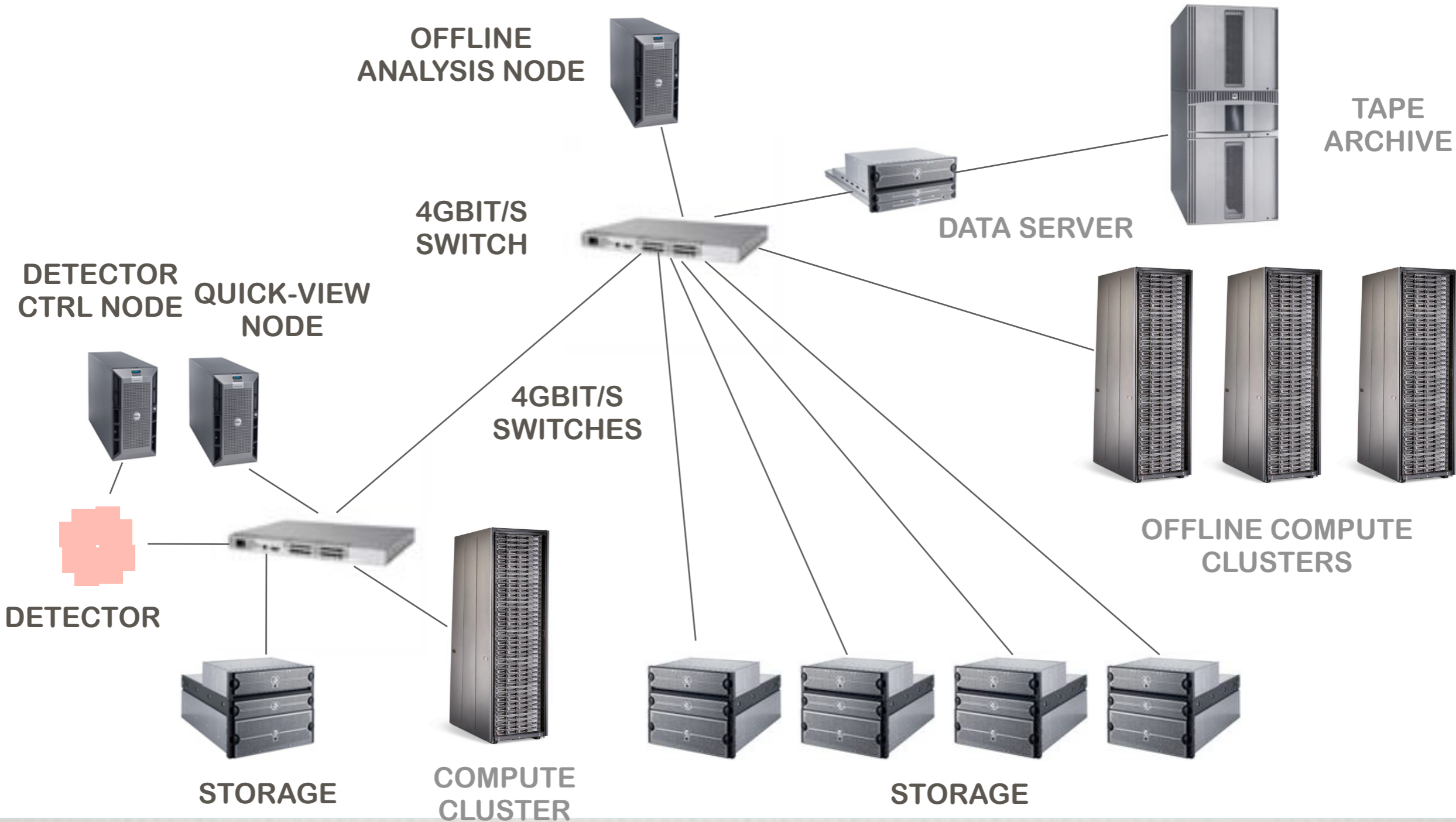


IMAGE FROM YIPING FENG, SLAC

CHALLENGES

- DATA RATE (SWITCHES)**
- ON-LINE PROCESSING (HIT OR MISS?)**
- LONG TIME STORAGE (1-4 PB/MONTH)**

DATA REDUCTION

Method	Effectiveness Estimate
Compression	50-90%
Incremental data processing	?
Intelligent data acceptance/ filtering	?
Automatic correlation with other data (TOF)?	?

DATA STORAGE ISSUES

- HOW LONG TO STORE DATA @ EU-XFEL**
- SHOULD THERE BE DATA PROCESSING FACILITIES @ EU-XFEL?**
- GRID COMPUTING**
- PARALLEL FILESYSTEMS**
- DATA FORMATS: COMMON FORMAT, E.G. BASED ON HDF5 – ONGOING EFFORT SPONSORED BY EU**
- EUROPEAN DATA NETWORK ON NEUTRONS AND PHOTONS (PROPOSED TO EU)**

DATA PROCESSING

- ALGORITHMS FOR DATA PROCESSING**
- CLASSIFICATION / AVERAGING**
- ORIENTATION**
- RECONSTRUCTION**
- DYNAMICS / DAMAGE**
- OURMAZD / ELSER / BORTEL / SHRINKWRAP / OTHER?**

BOTTLENECKS?

- OURMAZD, ELSE: COMPUTER POWER, SCALING OF ALGORITHMS**
- ALL: SIGNAL TO NOISE, DEPENDING ON MODEL?**
- DYNAMICS / ALIGNMENT / RADIATION DAMAGE NOT TAKEN INTO ACCOUNT**
- SOFTWARE DEVELOPMENT: MODULAR APPROACH IS NEEDED (E.G. FILE FORMATS!)**

IMAGE RECONSTRUCTION BENCHMARK

- DATA SET OF N NOISY IMAGES
- RECONSTRUCT TO AS GOOD RESOLUTION AS POSSIBLE (AUTOMATICALLY)
- TAKE, ALIGNMENT CONFORMATIONAL HETEROGENEITY AND RADIATION DAMAGE INTO ACCOUNT!
- START WITH EACH OTHERS DATA SETS, THEN MORE COMPLICATED SYNTHETIC DATA, AND FINALLY REAL DATA
- [HTTP://XRAY.BMC.UU.SE/SPBWIKI](http://xray.bmc.uu.se/spbwiki)

CONCLUSIONS

- DATA STORAGE AND ANALYSIS IS FEASIBLE AT DIFFERENT LEVELS DEPENDING ON POLICY AND \$\$\$**
- WE CAN LEARN A LOT FROM LCLS: BUT WE ARE PART OF LCLS**
- IMAGE PROCESSING AND RECONSTRUCTION: THERE ARE SOME VERY PROMISING ALGORITHMS – TEST THEM ON REAL DATA!**