A large yellow mining truck is the central focus of the image, parked on a dirt road in a quarry. The truck is massive, with a high cab and large, treaded tires. In the foreground, a group of about seven workers wearing hard hats and work clothes are standing in a line, looking towards the camera. The background shows a rugged, rocky landscape under a blue sky with scattered white clouds. The overall scene is industrial and outdoor.

Data Mining – quo vadis?

Sebastian Kuhn
Old Dominion University

Overview

- What data exist?
 - Which ones do we already have converted?
- What signals have we proposed to look for?
 - Which ones have already looked at?
 - Successes and failures
 - Ongoing and planned analyses
- What else can we do?
- What is needed most urgently?
- What should we do in the future?

What data exist?

| Run Period | Beam type | Beam energy (GeV) | Targets | Status |
|------------|-----------|-------------------|---|-------------|
| E2a | e | 1.16, 2.26, 4.46 | ^3He , ^4He , C, Fe | In progress |
| E2b | e | 0.98, 4.46, 4.71 | ^3He , Fe | Available |
| E6 | e | 5.77 | ^2H | Available |
| EG1a | \vec{e} | 2.5, 4.2 | $\vec{\text{NH}}_3$, $\vec{\text{ND}}_3$, C | |
| EG1b | \vec{e} | 1.6 to 5.7 | $\vec{\text{NH}}_3$, $\vec{\text{ND}}_3$, C | |
| E5 | e | 2.56, 4.23 | ^1H , ^2H | |
| EG2 | e | 4.0, 5.0 | ^2H + (C, Al, Ni, Fe, Pb) | Available |
| E1e | e | 2.04 | ^2H | |
| EG3 | γ | < 5.76 | ^2H | |
| E8 (BoNuS) | e | 1.1 to 5.4 | ^1H , ^2H | |
| EG6 | e | 1.1 to 5.4 | ^4He | |

- Photon runs? e.g. g11. Also, EG4 and EG1DVCS for additional nuclear+polarized nucleon/deuterium data
- “Available” = coatjava version on ODU cluster (M.W., ...?)
- New EVIO format conversion starting over (Gagik, Lamiaa, Mariana): e2a, e2b, eg2a,...
- Others use root, ntuple or BOS files directly from JLab farm

What projects exist?

https://clasweb.jlab.org/wiki/index.php/Nuclear_Data_Mining

Current Data Mining Projects

List of Data Mining projects:

- [DashyanPb](#) (eg2, Natasha Dashyan)
- [QuasifreeDelta0](#) [↗](#) (e2b, Dan Protopopescu) ... I might move this over here [QuasifreeDelta0](#) if the DM really takes off
- [E6 re-analysis Wiki](#) [↗](#) (Sebastian Kuhn, Michael Mayer,...)
- [EG1b analysis of d\(e,ep\)n](#) [↗](#) (Sebastian Kuhn, Michael Mayer,...)
- Design of new CLAS analysis framework - [GooLibrary](#) (Gagik Gavalian)
- d(e,e'Delta Delta) from eg2 - Chris Wooten (with L.B. Weinstein)
- [EenAnalysis](#) 3He(e,e'n)/(e,e'p) ratio and (e,e'pn) and (e,e'pp) - Mariana Khachatryan (with L. Weinstein)
- A(e,e' backward protons) - (L. Weinstein, not currently active)
- 'Tagged' EMC Effect - [TaggedEMCEG2](#), Barak Schmookler (with S. Gilad)
- d(e,e'p) and d(e,e'p gamma gamma) - M. Kamel (with W. Boeglin)
- e and gamma induced multiproton knockout from 208Pb - Lorenzo Zana and Dan Watts
- e and gamma induced pp pi+pi+ knockout from various nuclei (C -> Pb) - Lorenzo Zana and Dan Watts
- Hadronization of the Λ^0 baryon (EG2 datasets) - Krishna Adhikari and Lamiaa El Fassi
- Tel Aviv U projects:
 - Extraction of mean field transparency ratios.
 - Extraction of np/pp ratio in heavy nuclei.
 - Extraction of pp and np scaling in nuclei (i.e. np_A / np_C).
 - Extraction of pp-SRC pairs relative and c.m. motion in nuclei.
 - Extraction of the Pm dependence of the pp-SRC probability (i.e. (e,e'pp)/(e,e'p)).
 - Extraction of neutrons detection efficiency and momentum correction
 - Extraction of neutron mean-field (SRC?) transparency ratios.
- Theory: optimal strategies including building a MC for discovering Delta's in the deuteron - Sargsian and Strikman
- please add here your project and set up the wiki for it

Completed Analyses

- pn/pp in heavy (asymmetric) nuclei
- Double spin asymmetry in $D(e,e'p)n$
(M. Mayer; PWG approved, paper about to be submitted)
- Transparency in SR pp pair emission from Nuclei

Ongoing Analyses

- Validation of neutrino cross section MCs (K.M., S.M,...; MichSU, URoch,)
- $\omega/\rho/\phi$ meson production in medium (M.W., Canisius)
- Production length (UTFSM)
- EMC ratio tagged on backward p in nuclei (B.S., S.G., MIT)
- p/n on ^3He as function of momenta (M.K., ODU)
- Λ^0 production in EG2 data (K.A., MissSU)
- $A(e,e'p)/^{12}\text{C}(e,e'p)$ vs. $A(e,e'n)/^{12}\text{C}(e,e'n)$ as function of momenta (M.D., Tel Aviv)
- pp pair c.m. momenta (E.C., Tel Aviv)
- 3N SRC (E.C., Tel Aviv)
- 5th SF for $D(e,e'p)$; G_{Mn} from e5 ? (G.G., Richmond)
- Δ^0 production on ^3He ? (Glasgow)
- $(e,e'p)$ transparency ratio $A/^{12}\text{C}$? (Tel Aviv)

Explored but abandoned channels

- $\Delta\Delta$ from $d(e,e')$ (eg2, e6)

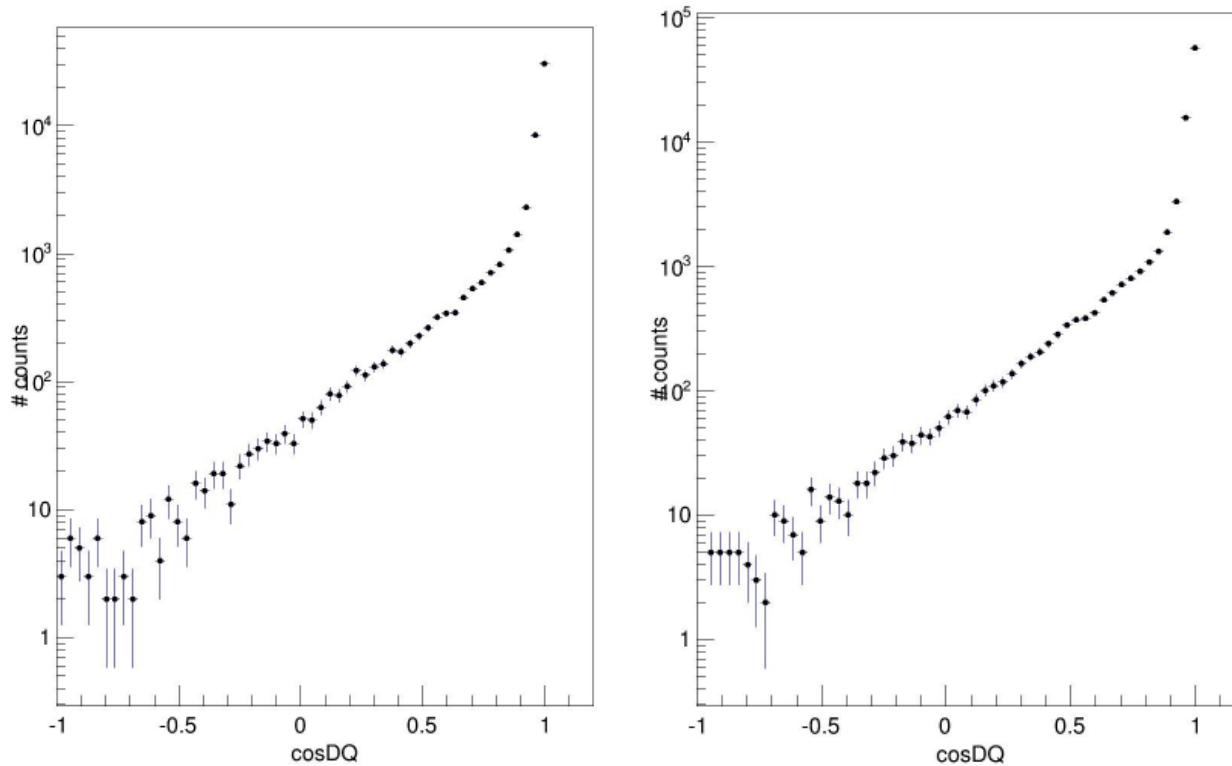


Figure 9: Preliminary angular distributions of final state Δ^{++} (left) and Δ^0 (right) relative to the direction of the virtual photon, from CLAS E6 data.

What have we promised for the future?

- $(e, e'pN)$ over a wide range of A , Z/N , Q^2 (incl. 0 GeV^2) and p_{cm} , p_{rel}
- 3N SRC
- pol. $D(e, e'p)n$ with EG4, EG1-DVCS; 5th SF with E5, E6, EG2, E1e
- $D(e, e'p_S)X$ for other data sets than E6
- Backward (and forward!?) Δ from nuclei
- CT studies: $D(e'ep/\Delta)$ in transverse kinematics, ρ production, S_{11} production on nuclei
- Hadronization (partially underway - $\omega/\rho/\phi, \Lambda$)

What else can we do?

- ???

What is needed most urgently?

- Continue conversion to newest (“final”?) EVIO format and CLARA/Java
- Software improvements?
- Most urgent new data sets?
- Most urgent cuts, corrections, ...?
- SIMULATIONS? (GEMC???)
- What else is needed to support ongoing analyses?

What should we do in the future?

- OPTION A:
Complete ongoing analyses, publish and declare victory
- OPTION B: Continue work of converting data, simulations etc. as part of overall CLAS/CLAS12 data preservation strategy (slow)
 - Gagik working with other run groups like eg1-dvcs
- OPTION C: Write another proposal to hire a postdoc to address the most urgent needs and future plans

2014 Proposal - Reviews

- 1 Without this initiative these data, that required serious financial and human efforts, would be probably lost. **Summary Score:** Strongly Encourage Funding (5-6) **6.0**
- 2 **[Panelist]** There is, in my opinion, a dramatic mismatch between a beautifully written summary of the physics background and the real task of the personnel for which funding is requested. My feeling is that the request is rather on the higher side and that the work could also be done by a talented PhD student. **Summary Score:** Encourage Funding (3-4) **3.0**
- 3 **[Panelist]** I admit that I found this proposal to be rather strange, and the panel discussions indicated that I was not alone with this impression. **Summary Score:** Encourage Funding (3-4) **3.0**
- 4 There is clearly an abundance of topics that can be addressed. These researchers have been involved in these studies for many years. The technique proposed has produced valuable results in the past. This is effort is likely to do so as well. The proposal address inclusion of a user-friendly simulation component to the already implemented data analysis capability. This is a necessary step to produce physics results; the data analysis is only half the job. **Summary Score:** Strongly Encourage Funding (5-6) **5.0**
- 5 Nevertheless, it is important to archieve the wealth of data obtained at CLAS6, and this proposal will certainly allow more researchers to take a quick look of existing data to check out their ideas or speculations. This proposal is definitely cost-effective in maximizing the physics output from JLab 6 GeV programs. **Summary Score:** Encourage Funding (3-4) **4.0**
- 6 It is impossible for the reviewer to gauge how much benefit, if any, the data-mining effort has had on these topics or how much benefit is to be expected. ...Given this reduction in scope of the project, however, it is surprising that this, now much less ambitious, program has not progressed much more rapidly. Only three of the eleven data sets have been processed, with a fourth being in progress. **Summary Score:** Encourage Funding (3-4) **4.0**
- 7 **[Panelist]** It is not sufficiently clear how much of the exiting data have been looked at and what has really come out of the existing effort? There are very little details about what the Postdoc will actually do, and thus there is very little basis to refute the conjecture that the personal costs seem inflated for this task. **Summary Score:** Discourage Funding (1-2) **2.0**

What should we do in the future?

These smart people (and the offsite participants) will have to decide:

