Amplitude Analysis Tools: Development of a Collaborative Analysis Toolkit

Ryan Mitchell
Indiana University

INT Workshop
November 13, 2009

Introduction

- **GOAL:** Develop a framework for amplitude analysis that...
  - ... is modular and independent of experiment.
  - ... scales to very large data sets.
  - ... accommodates increased computational demands.
  - ... allows amplitudes to be written by the physicist.
  - ... separates the computer science from the physics.
  - ... encourages a closer theory-experiment collaboration.

- **COLLABORATION:**
  - *Funded by the National Science Foundation (NSF) Physics at the Information Frontier (PIF) program.*
    - *Indiana University:* R. Mitchell, M. Shepherd, A. Szczepaniak, P. Guo, H. Matevosyan
    - *Carnegie Mellon University:* C. Meyer, (M. Williams)
    - *University of Connecticut:* R. Jones, I. Senderovich, (J. Yang)
    - plus one more postdoc opening

Design Philosophy

- The **"User" (a.k.a. physicist)**...
  - ... supplies 4-vectors (data and MC).

    experiment-independence
  - ... writes amplitudes.

    physics is not a "black box"
  - ... builds new applications using modules from the framework.

    flexibility
  - ... spends time doing and thinking about physics.

    the fun part

- The **Framework**...
  - ... calculates the likelihood.
  - ... performs the minimization.
  - ... provides a set of modules that can be used to develop new applications.

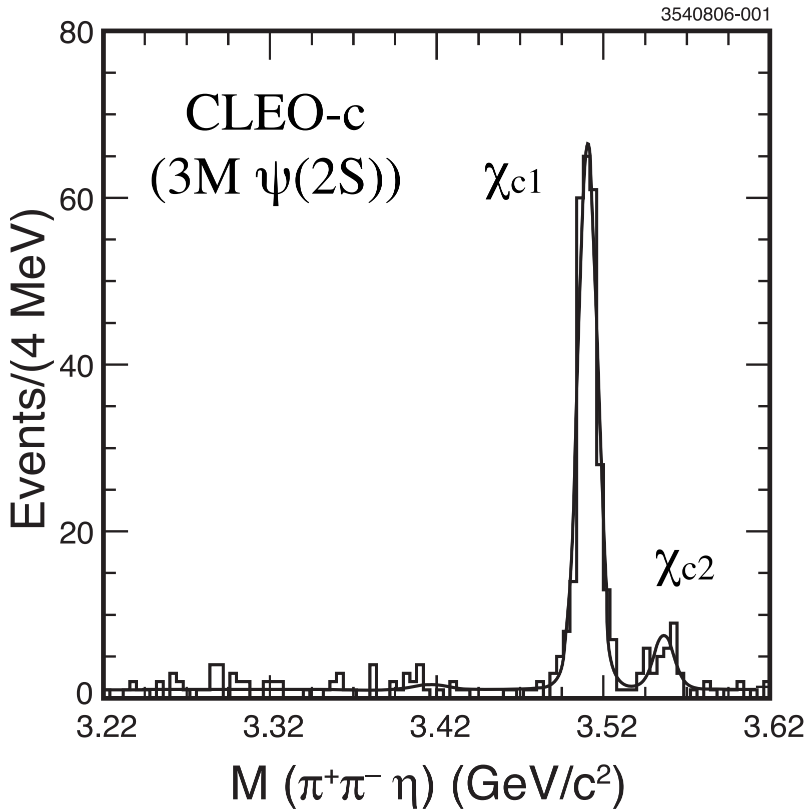
    visualization

    amplitude generation
  - ... thinks about how to cache data and optimize calculations.
  - ... handles the distribution of data and processing over multiple processors, the Grid, or GPU's.

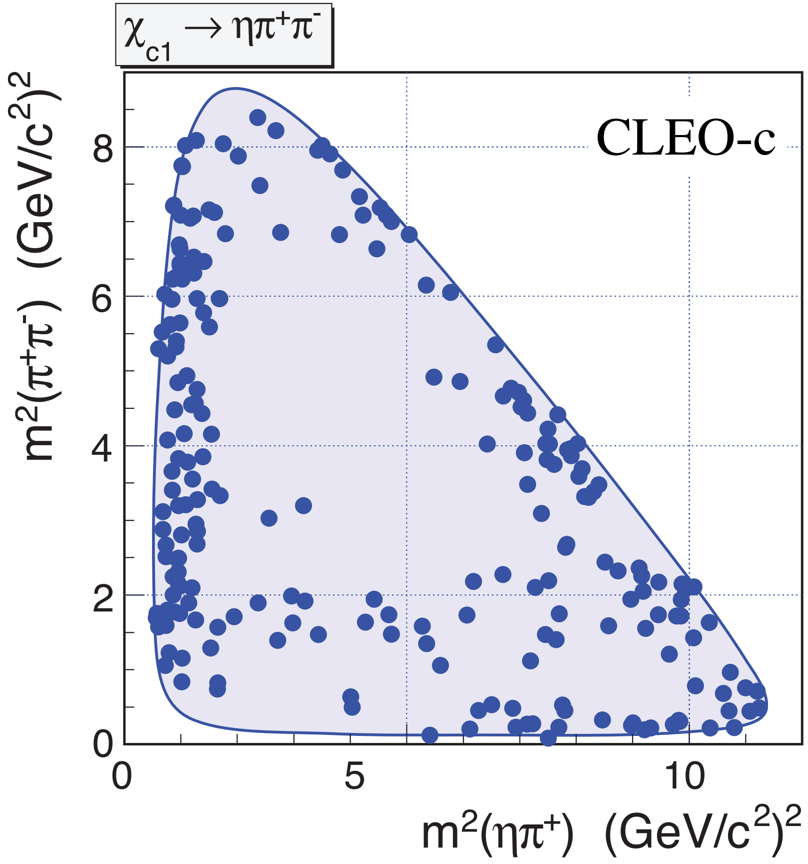
Physics Example 1: CLEO-c / BES III

$e^+e^- \to \psi(2S)$; $\psi(2S) \to \gamma\chi_{c1}$; $\chi_{c1} \to \eta\pi^+\pi^-$

$\psi(2S)$ decays are a clean source of χ_{cJ}:



χ_{cJ} decays are rich in substructure:

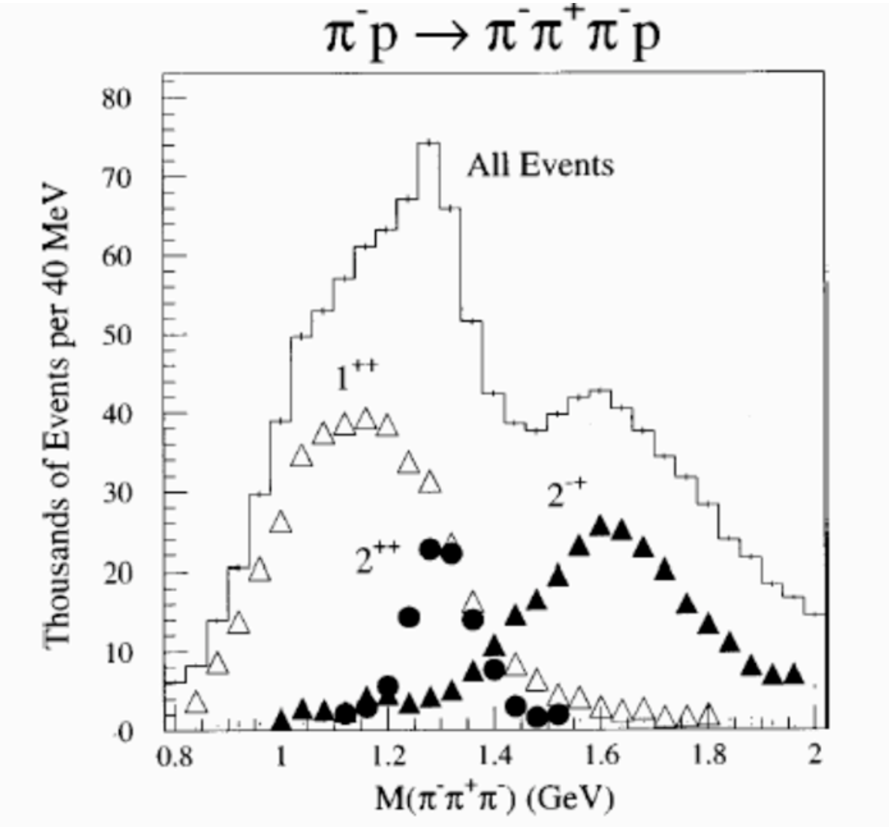


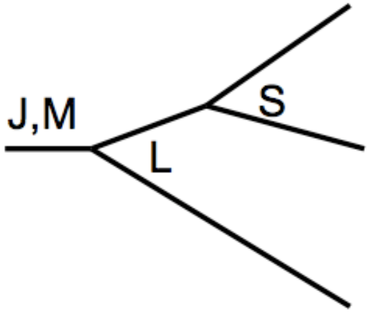
Pick from a large number of final states to isolate different features…

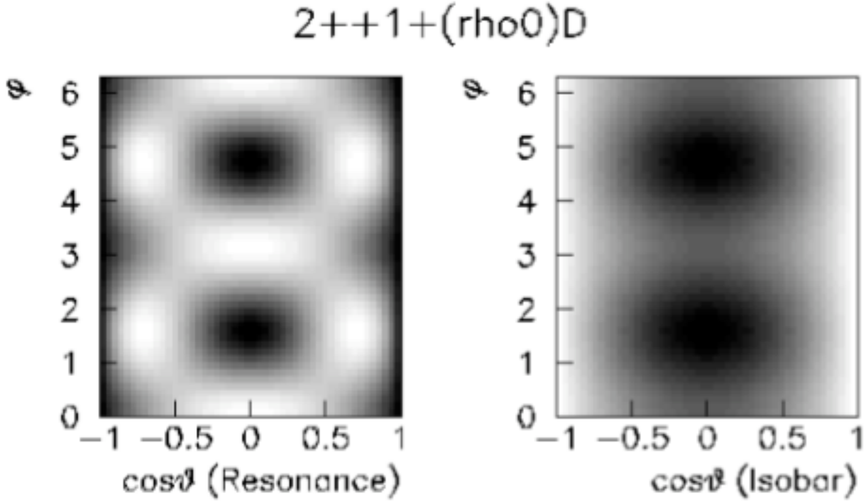
Physics Example 2: E852 at BNL

$\pi^- p \to \pi^-\pi^-\pi^+ p$ at 18 GeV/c

Perform amplitude analyses in bins of M($\pi^-\pi^-\pi^+$):







Decay amplitudes have been written using the isobar model and helicity formalism…

Physics Example 3: GlueX at JLab

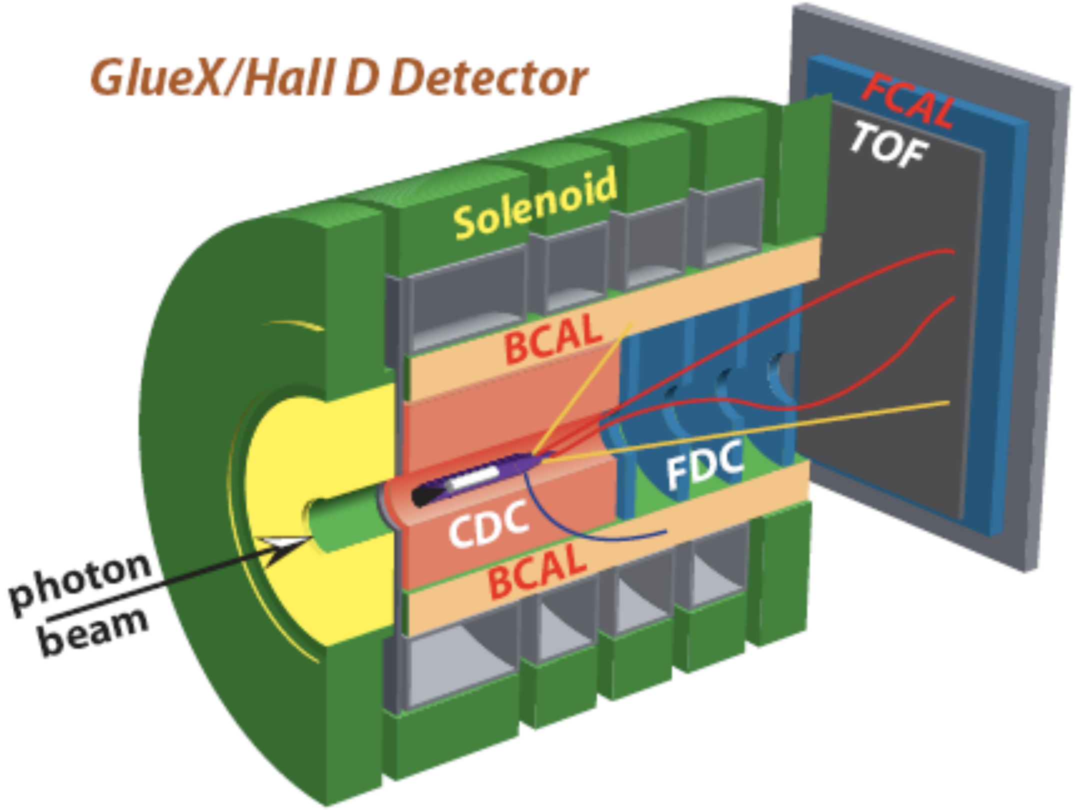
$\gamma p \rightarrow Xp$ at 9 GeV

A search for gluonic excitations requiring:

(1) innovations in phenomenology...

(2) the computational ability to deal with sophisticated amplitudes…

(3) the ability to handle larger data sets…



Common Features of the Problem

In each case, the experimental data is described by a series of amplitudes:

Fit Parameters

$$I(\Omega) = \sum_{\alpha} \left| \sum_{\beta} V_{\alpha,\beta} A_{\alpha,\beta}(\Omega) \right|^2$$

Decay Model
From Theory

and we want to minimize a function that looks like this:

$$-2\ln(L) = -2\sum_{data} \ln(I(\Omega_i)) + 2\sum_{MC} I(\Omega_i)$$

Three Challenges

Computational Challenge I:

We want to do fits with sophisticated (computationally intensive, sometimes) amplitudes on large data sets.

⇒ **GPU**

Computational Challenge II:

We need large MC data sets, and we need to be able to move around large amounts of data.

⇒ **Open Science Grid**

The Phenomenological Challenge:

For reliable results, it will be crucial to move beyond the isobar model, to explore more models, and to be open to phenomenological flexibility.

⇒ **Theoretical Input**

CHALLENGE I: Fit Speeds

In each case, the experimental data is described by a series of amplitudes:

$$I(\Omega) = \sum_{\alpha} \left| \sum_{\beta} V_{\alpha,\beta} \boxed{A_{\alpha,\beta}(\Omega)} \right|^2$$

Potentially
Expensive
Calculations

and we want to minimize a function that looks like this:

$$-2\ln(L) = -2\sum_{\boxed{data}} \ln(I(\Omega_i)) + 2\sum_{\boxed{MC}} I(\Omega_i)$$

Big Numbers

Multi-Node Amplitude Analysis

The form of the likelihood naturally lends itself to distributed computing:

$$-2\ln(L) = -2\sum_{data}\ln(I(\Omega_i)) + 2\sum_{MC} I(\Omega_i)$$

(1) The "master" distributes data (and possibly MC) over multiple "nodes" (or looks for "nodes" that currently store data).

(2) The "master" distributes copies of an Amplitude Calculator.

(3) The "master" sends the "nodes" the fit parameters at each fit iteration.

(4) The "nodes" send the "master" partial contributions to the likelihood.

(5) The "master" sums the partial contributions.

(Complication: If one node is lost, the fit fails.)

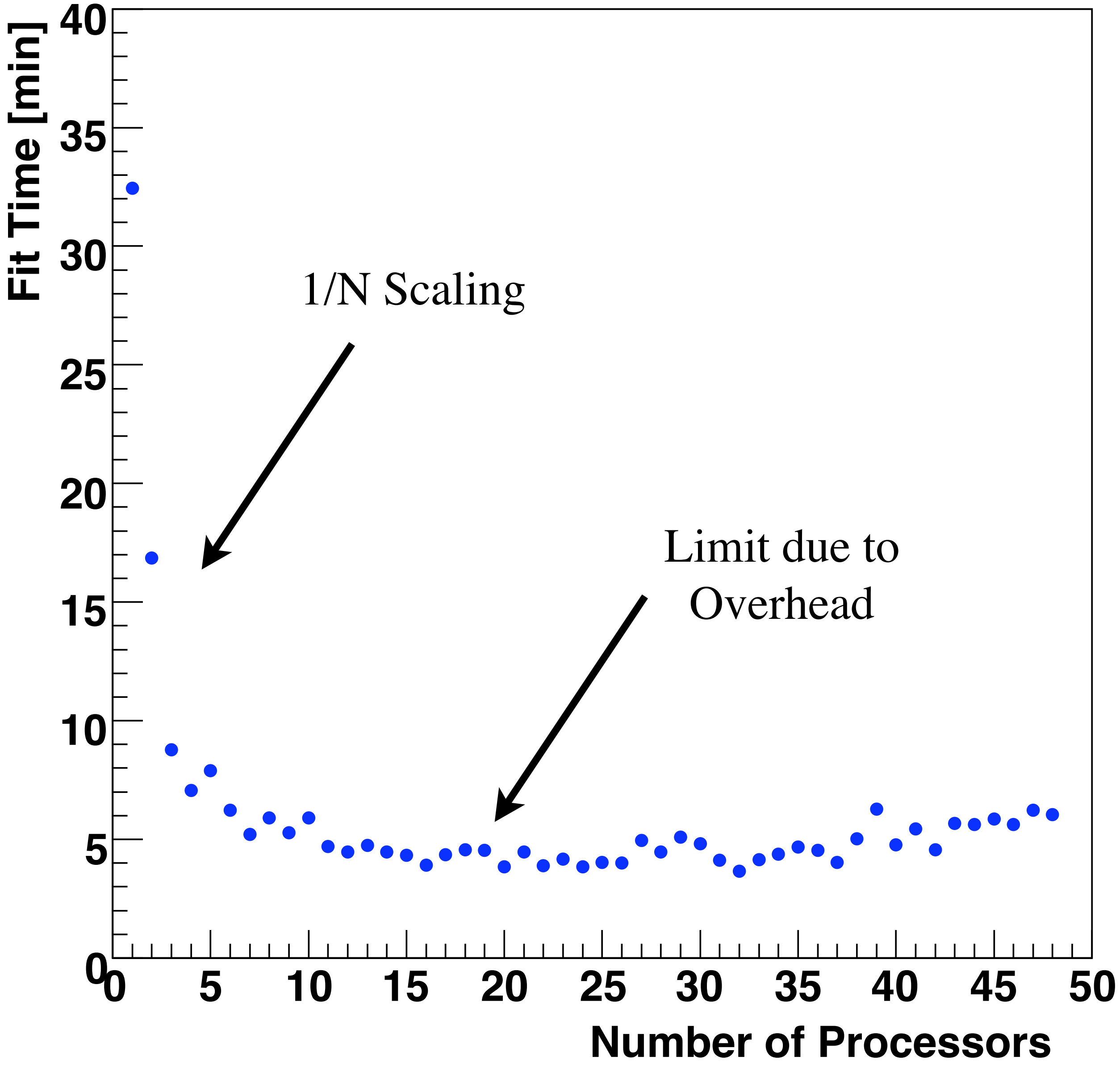
Using Multiple CPU with MPI (OLD)

Perform a simple fit using CLEO-c data (1000’s of events).

Vary the number of processors and record the fit time.

We find 1/N scaling initially, before overhead becomes important.

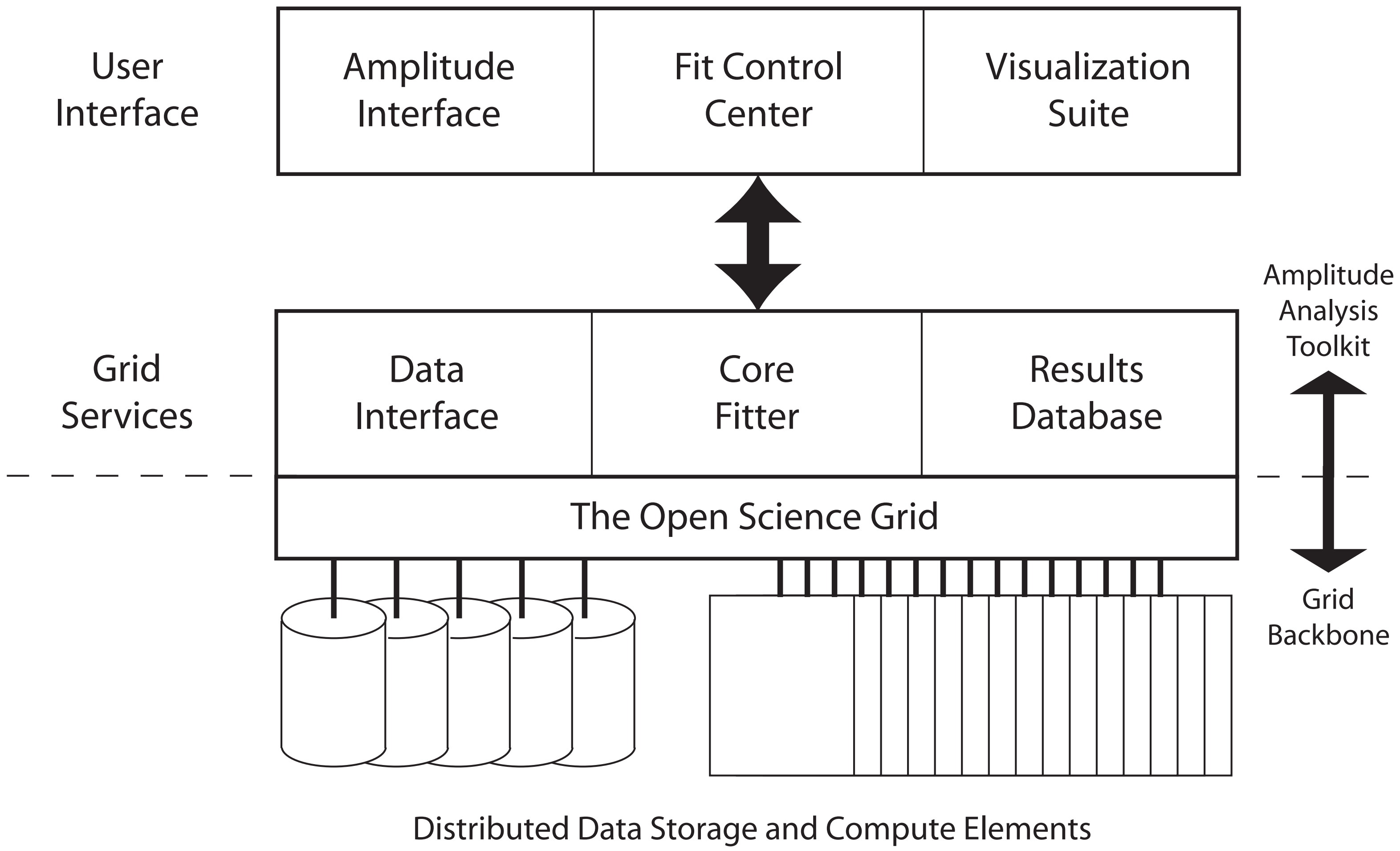
Larger data sets will allow more processors before reaching the overhead limit.



1/N Scaling

Limit due to Overhead

Using the Grid (OLD)



VERY difficult problems: *(i) Synchronization, (ii) losing compute elements, (iii) transferring information between compute elements, etc.*

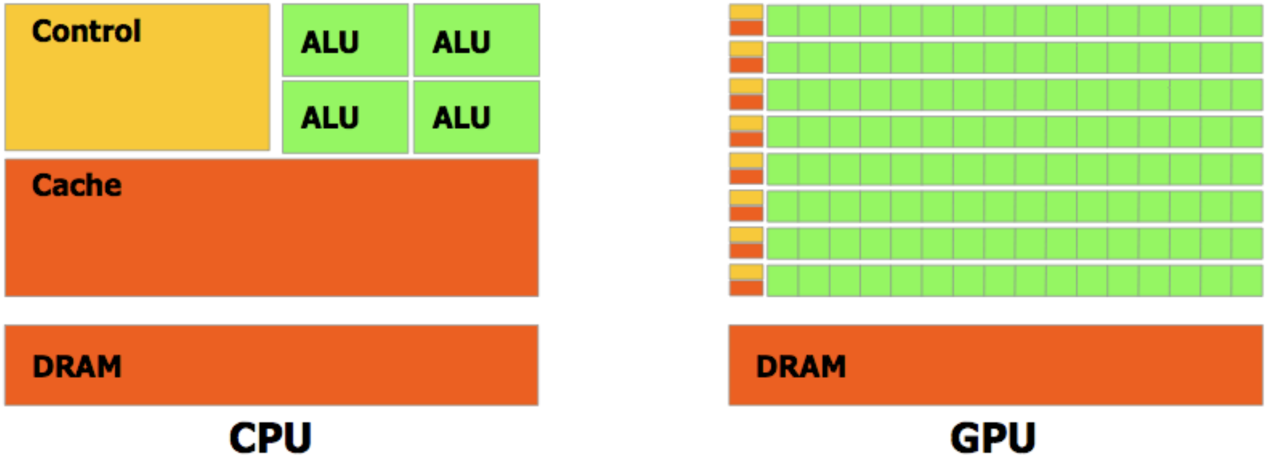
Moving to GPU's (NEW!)

- GPU's are massively parallel.
- The CUDA toolkit is maturing as an extension of C for programming on GPU's.
- Very preliminary PWA work has been done and it looks very promising.
- IU just received ~$45K (jointly with lattice theory) to build a GPU farm.
- Still exploratory. Expect much more in the future...

Raw Amplitude Calculation

(preliminary benchmarks)

Sample	CPU Intel Core i7 (1 core)	GPU nVidia GTX 285 (240 cores)
8M evts.; 3 Amps	48 s	0.29 s
8M evts.; 1 Amp	16 s	0.10 s



The Experts: Hrayr Matevosyan, Nik Berger

Parameters Inside A(Ω)

Recall the form of the intensity:

$$I(\Omega) = \sum_{\alpha} \left| \sum_{\beta} V_{\alpha,\beta} A_{\alpha,\beta}(\Omega) \right|^2$$

and the likelihood:

$$-2\ln(L) = -2\sum_{data} \ln(I(\Omega_i)) + \boxed{2\sum_{MC} I(\Omega_i)}$$

Case 1.

When there are *no* free parameters *inside* A(Ω),
the *V* can be factored out of the sum over MC,
and the sum over MC can be calculated *once* and stored.

Case 2.

When there *are* free parameters inside A(Ω),
we can solve the problem with *brute force*:
distribute MC and sum over it at every iteration.

This will allow for a new sophistication in phenomenology.

CHALLENGE II: MC and Data Set Size

Very large data sets ⇒

... Large MC samples

... Large amounts of compute time

... Both data and MC also present storage and logistical problems.

⇒ The Open Science Grid

“OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S...”

GlueX has recently become a “Virtual Organization”

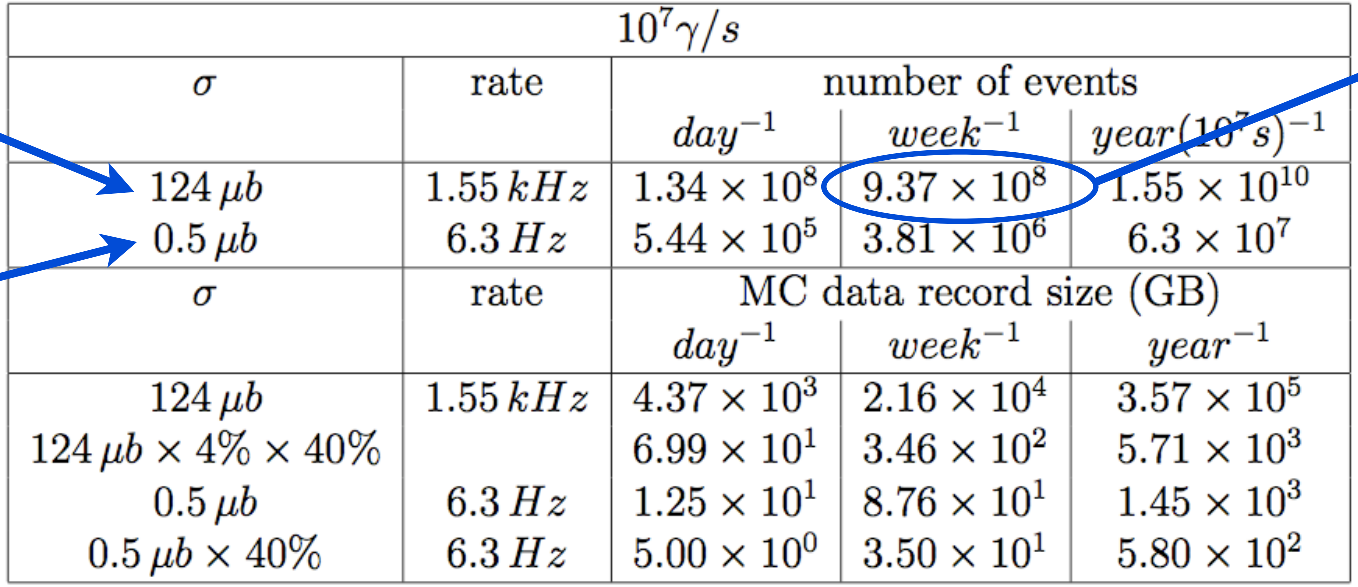
An Example From GlueX

MC Studies of $\gamma p \rightarrow \eta\pi^0 p$
(Blake Leverington)

With O(1) second/event for MC,
29 years!

Hadronic
BG Rate

Signal
Rate



$10^7\gamma/s$				
σ	rate	number of events		
		day^{-1}	$week^{-1}$	$year(10^7 s)^{-1}$
$124\,\mu b$	$1.55\,kHz$	1.34×10^8	9.37×10^8	1.55×10^{10}
$0.5\,\mu b$	$6.3\,Hz$	5.44×10^5	3.81×10^6	6.3×10^7
σ	rate	MC data record size (GB)		
		day^{-1}	$week^{-1}$	$year^{-1}$
$124\,\mu b$	$1.55\,kHz$	4.37×10^3	2.16×10^4	3.57×10^5
$124\,\mu b \times 4\% \times 40\%$		6.99×10^1	3.46×10^2	5.71×10^3
$0.5\,\mu b$	$6.3\,Hz$	1.25×10^1	8.76×10^1	1.45×10^3
$0.5\,\mu b \times 40\%$	$6.3\,Hz$	5.00×10^0	3.50×10^1	5.80×10^2

Table 2: Expected event rates for various production cross sections in photo-production. The σ column is the cross section. The rate is the equivalent experimental data rate given a beam rate of $10^7\gamma/s$ and the $30cm$ target. A beam rate of $10^8\gamma/s$ will increase the magnitudes of the events and data sets by one order.

The first GlueX OSG jobs will be submitted very soon.

CHALLENGE III: Phenomenology

- The framework encourages “open access” to the data by easily incorporating externally written amplitudes.

- Indiana is currently exploring the problem with CLEO-c analyses of charmonium decays.

Experimentalists:	**Theorists:**
Ryan Mitchell	Adam Szczepaniak
Matthew Shepherd	Peng Guo
Mihajlo Kornicer	Hrayr Matevosyan
Claire Tarbert	

- Active analyses: (i) $J/\psi \to \rho\pi$, (ii) $\chi_{c1} \to \eta^{(\prime)}\pi\pi$, (iii) $\chi_{c0} \to 4K$

Physicist-Defined Amplitudes

- **Physicist-defined amplitudes...**
  - ... ensure the physics is transparent (presumably to the author).
  - ... allow the amplitudes to be as complicated as required.
  - ... encourage a new level of collaboration between theory and experiment.
  - ... provide a new level of access to experimental data.
- **The "user" (physicist)...**
  - ... writes an amplitude function that inputs kinematics and parameters and outputs a complex number.
  - ... specifies constraints among free parameters.
  - ... writes in any of a number of different languages, using hooks that are provided to interface with the framework.
- **A stand-alone MC generator will allow a physicist to...**
  - ... develop amplitudes independent of experimental data.
  - ... view how these amplitudes project onto a given set of kinematics.
  - ... study how experimental acceptance may modify these distributions.

Rescattering Effects in J/ψ → ϱπ

Paper draft in progress (Peng Guo)...

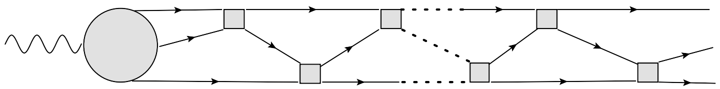


FIG. 2: Summation of rescattering contributions is demanded by unitarity and analyticity.

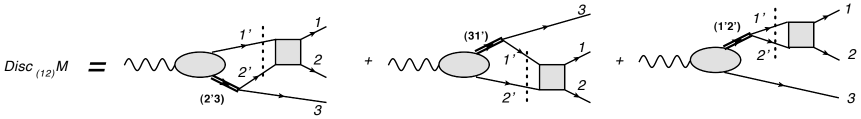
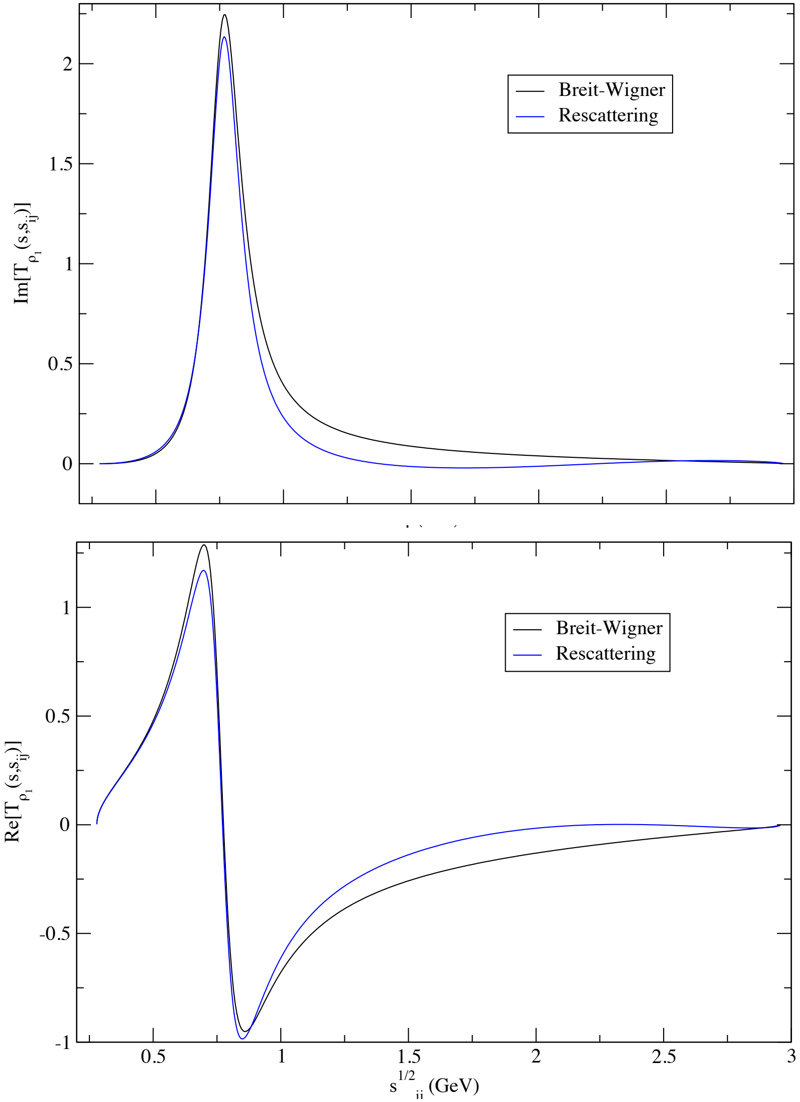


FIG. 3: Subenergy discontinuity. Unitarity relates interactions in various sub-energy channels. The naive isobar model depicted in Fig.1 contains contributions from the last diagram only.



We're trying out these new phenomenological methods (and others) on CLEO data.

Analysis of $\chi_{c1} \to \eta^{(\prime)}\pi^+\pi^-$

Structure of the amplitude:

$\psi(2S) \to \gamma\chi_{c1}$ piece

$$
\begin{aligned}
A_a^R = & \sum_{\lambda_\chi=-1}^{1} C_a(M_\psi, \lambda_\gamma, \lambda_\chi) BW_\chi \\
\times & \sum_{M_\chi M_L, M_S} D^*_{M_\chi,-\lambda_\chi}(\phi_\gamma, \theta_\gamma, 0)\langle 1M_\chi | LM_L, SM_S\rangle p_R^L Y^*_{LM_L}(\theta_R, \phi_R) q_h^S Y^*_{SM_S}(\theta_h, \phi_h) \\
\times & \langle I0|11, 1-1\rangle T^{S,I}_{R(h_1' h_2')\to h_1 h_2}(s_{h_1 h_2})
\end{aligned}
$$

angular piece

theory input

Mihajlo (experiment) and Adam (theory) are in close collaboration…

Tools for Writing Amplitudes

A few amplitude writing tools are being developed to accommodate formalisms that would otherwise be difficult to code.

CMU (M. Williams and C. Meyer) has developed code for tensor algebra...

EXAMPLE

$\chi_{c0} \to KK\pi\pi$ via

$\chi_{c0} \to K_1^*K;\ K_1^* \to K^*\pi;\ K^* \to K\pi$

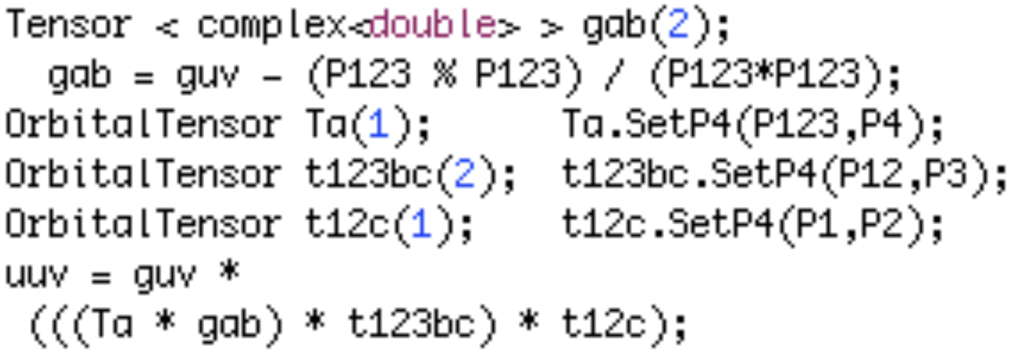
complicated angular structure!

Write Lorentz-covariant decay kinematics:

Dulat and Zou
EPJ A26, 125

$$g_{\mu\nu}\tilde{g}_{\alpha\beta}(p_{K_1^*})\tilde{T}^{(1)\alpha}_{[K_1^*\bar{K}]}\tilde{t}^{(2)\beta\sigma}_{(K^*\pi)}\tilde{t}^{(1)}_{(14)\sigma}$$

easy to code



```
Tensor < complex<double> > gab(2);
  gab = guv - (P123 % P123) / (P123*P123);
OrbitalTensor Ta(1);        Ta.SetP4(P123,P4);
OrbitalTensor t123bc(2);  t123bc.SetP4(P12,P3);
OrbitalTensor t12c(1);     t12c.SetP4(P1,P2);
uuv = guv *
 (((Ta * gab) * t123bc) * t12c);
```

Status and Outlook

- We have another full year left on our three year NSF grant.

- A lot of progress has been made and we continue to prototype software on both real and MC analyses:
  - CLEO-c χ_{cJ} decays.
  - $J/\psi \rightarrow \varrho\pi$
  - GlueX MC

- We are fostering experiment-theory collaboration.

- The framework is maturing... Use the Grid for MC generation and data storage... Use GPU’s for the fitting...

- We welcome new ideas regarding usability (especially once the code starts becoming more widely available... soon).

- We hope to have a well-documented package within a year.