Farming Data for the HyperCP Experiment
A small experiment handling a large dataset

Timothy Holmstrom
University of Virginia

International Conference on
Computing in High Energy and Nuclear Physics
Beijing, China, September 05, 2001
The HyperCP Collaboration

A. Chan, Y.C. Chen, C. Ho, and P.K. Teng
Academia Sinica, Nankang, Taipei 11529, Taiwan, Republic of China

W.S. Choong, Y. Fu, G. Gidal, P. Gu, T. Jones, K.B. Luk, B. Turko, and P. Zyla
Lawrence Berkeley Laboratory and University of California, Berkeley, CA 94720, USA

C. James and J. Volk
Fermilab, Batavia, IL 60510, USA

J. Felix
Universidad de Guanajuato, León, Mexico

Illinois Institute of Technology, Chicago, IL 60616, USA

N. Leros and J.-P. Perroud
Université de Lausanne, Lausanne, Switzerland

H.R. Gustafson, M.J. Longo, F. Lopez, and H. Park
University of Michigan, Ann Arbor, MI 48109, USA

M. Jenkins and K. Clark
University of South Alabama, Mobile, AL 36688, USA

E.C. Dukes, C. Durandet, R. Godang, T. Holmstrom, M. Huang, L.C. Lu, and K.S. Nelson
University of Virginia, Charlottesville, VA 22901, USA
Why Search for CP Violation in Hyperon Decays?

- After 35 years of experimental effort we still know very little about $CP$ violation:
  - The origin of $CP$ violation remains unknown. There is little hard evidence that it is explained by the standard model.
  - There is no compelling evidence of $CP$ violation outside of the decays of the $K_L$ and $B$ systems.
- The asymmetry is potentially large: up to several $\times 10^{-3}$.
- Hyperon $CP$ violation is not the same as $CP$ violation in kaon or $B$ decays!
- $CP$ violation is too important, and experimental evidence is too meagre, not to examine every possible manifestation of the effect.

“To extract useful information and constraints on new physics, results from hyperon decays, $K$-decays, and $B$-decays will have to be pooled and confronted with models on a case by case basis.”

S. Pakvasa
**HyperCP Data Acquisition System**

- Highest rate DAQ in the world.
- Maximum trigger rate of about 100,000 events per second.
- Sustained data logging rate of 27MB/s onto 27 Exabyte 8705 tapes.
- All custom front ends: no CAMAC, Fastbus, or VME.
**HyperCP Yields**

- In two runs, 1997 and 1999, we took the largest data sample ever by a particle physics experiment: 231 billion events, 29,401 tapes, and 119.5 TB of data.

<table>
<thead>
<tr>
<th>Events</th>
<th>Channeled beam polarity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cascade</td>
<td>$80 \times 10^9$</td>
<td>$40 \times 10^9$</td>
</tr>
<tr>
<td>All</td>
<td>$157 \times 10^9$</td>
<td>$74 \times 10^9$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Reconstructed Events</th>
<th>Channeled beam polarity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Xi \to \Lambda \pi$</td>
<td>$0.46 \times 10^9$</td>
<td>$2.0 \times 10^9$</td>
</tr>
<tr>
<td>$K \to \pi \pi \pi$</td>
<td>$0.39 \times 10^9$</td>
<td>$0.16 \times 10^9$</td>
</tr>
<tr>
<td>$\Omega \to \Lambda K$</td>
<td>$5.0 \times 10^6$</td>
<td>$14 \times 10^6$</td>
</tr>
</tbody>
</table>

- We expect a statistical precision of:

$$\delta A_{\Xi \Lambda} \approx 2 \times 10^{-4}$$
The *HyperCP* Data Processing Farm
• **I/O nodes:** Distribute the data in parallel to two independent processing systems.

• **Farm Control System:**
  
  – Submitting jobs and monitors their output.
  – Using HyperCP Farm Software System (CPFSS) to track jobs status and history.
  – Maintaining databases.
  
  – Farmer Server running on Beauty7: submit jobs, log the progress of them, direct the end of job tasks and manages the stages with input and output. Communicate to the server about the status of the performed tasks from both I/O node and PC workers.

• **PC Farms/Worker Nodes:** 55 dual-processor Linux-based PCs that the data analyzing code is running on.

• **Stages:** 70 I/O stages on FNSFH and 27 on FNSFO. Each stage is 5 GB and used to store data temporarily. A stage is cleaned up and unlocked after the data is copied to tape.
• **Tape Drives:**
  
  – Exabyte tape drives: load input raw data tapes and output data tapes for some output streams.
  – Two AIT2 tape drives: used on FNSFH and FNSFO to copy output data to 50-GB AIT2 tapes.

• **HyperCP Farm Software:**
● **Job Structure:**

- Raw data is read from an Exabyte drive and split into 10 equal files on a staging disk on the I/O node.
- The 10 input files are copied to local disks on 10 worker nodes (PCs).
- The analysis job is executed on the 10 PCs.
- Results from the analysis job are moved from the local disk to output stage on the I/O nodes.
- When an output stage is full, it is flushed to output tape and the stage is released for next job.

![Diagram](image)

**Fig.4.** The flow of farm data.
Scripts Phase0, Phase1, Phase2, and Phase3 are responsible for the tasks of a job.

- **Phase0:** A csh script launched from Beauty7(by `rsh`) but runs on the I/O node.
  * Build a short script used by FBS.
  * Launch the farm batch job using farms submit [script name]. Script name is built by the phase0 script.

- **Phase1:** A csh script which runs on the I/O node:
  * Allocating an input tape drive on the I/O nodes.
  * Requesting the tape passed by Beauty7 to be placed in the tape drive.
  * Mounting a tape on the tape drive.
  * Running the code rawSplit which is responsible for
    - Reading data from the input tape drive.
    - Splitting the data into 10 worker files.
    - Writing the data onto a stage on the I/O node.
– **Phase2:** A script which is launched onto 10 worker nodes by FBS:
  * Copying a file from the input stage on the I/O node to the local disk on the PC.
  * Setting up the environment needed by the E871 analysis code.
  * Running the analysis code.
  * Copying the stream output to the appropriate staging disks on the I/O node.
– **Phase3:** A csh script launched on the I/O node and It’s responsible for cleaning up the job.

Fig.5. The structure of a job.
• **Database:**

PostgreSQL Database is used on the farm to store job information, log files, and statistics.

- **e871farmdb:** Contains the detailed messages of a job for the three phases: for example, the time the job was submitted, when it started on which I/O node, the tape label for this job, and fail or succeed for each phase.

- **e871tapedb:** Contains the tables of input tapes and output tapes.

- **phase2sumdb:** Contains a summary for each job submitted.
Fig. 6. E871 farm Database.
• Farm Monitoring Web Page:

  – Built by Perl CGI scripts.
  – Self-refreshing and display the status of farming, job history, and statistics.
  – Error messages and resubmitting 'fail' job.
  – Allowing non-expert farmers to monitor farm processes.

Fig. 7. Running jobs on farm monitoring web page.
Fig. 8. Completed jobs on farm monitoring web page.
Data Processing

- **Tasks for Shifters:**
  - Monitor the jobs status through the web page.
  - Debug the fail job based on the information on the web page. Resubmit it if the tape is good.

![Web page showing failed jobs to be resubmitted](image)

*Fig.9. Failed jobs on farm monitoring web page.*
• **Tasks for Experts:**
  
  – Debug the failed jobs that a shifter can’t fix.
  
  – Check the data in the output tapes is consistent with the database.
Farm Results

A total of 29,843 tapes were processed and finished during

June 2000 → May 2001

• Statistics:

![Fig.10. Farm statistics.](image-url)
• **Output Tapes:**

Fig. 11. Output tapes from farm.

• **Histograms from the Farm:**

Fig. 12. Mass distribution of $\Xi$ and $\tilde{\Xi}^+$. 
Conclusion

• HyperCP farm is very efficient, well-organized, and easy to monitor.

• All the log information from the farm processing is easily accessible.

• The reconstructed DST data and 36 million histograms are now ready for the \( CP \)-violation study as well as a number of other physics studies.