



# Proyectos en la UNAM con altos requerimientos de ancho de banda

Alejandro Ayala, Luciano Diaz, Lukas Nellen, Guy Paic  
ICN-UNAM

CUDI Reunión Otoño  
Monterrey, 7 de octubre 2008

[lukas@nucleares.unam.mx](mailto:lukas@nucleares.unam.mx)

# Introducción

- Hemos identificados varios proyectos maduros con altos requerimientos de ancho de banda
- Continuamente nacen nuevos proyectos
- La realidad de la red de CUDI ya limita nuestros proyectos
- Redes crecen juntos:
  - ▶ de datos
  - ▶ humanas

# Colaboración internacional

- Gran ciencia se hace en colaboración
  - ▶ Altas Energías: ALICE, CMS
  - ▶ Astronomía: GTM/LMT
  - ▶ Astro-partículas: Pierre Auger, HAWC
- GRID: meta-proyectos, para **apoyar la ciencia**
- Grandes proyectos son **motores para la innovación tecnológica**

# Proyectos

- Bio-informática
- Observatorio Virtual Solar
- Observatorio sismológico
- Geografía, cartografía digital y satelital
- Optimización de redes para computo de alto rendimiento
- Laboratorios virtuales e instrumentación remota
- HEP y astro-partículas

# Proyectos

- Bio-informática
- Observatorio Virtual Solar
- Observatorio sismológico
- Geografía, cartografía digital y satelital
- Optimización de redes para computo de alto rendimiento
- Laboratorios virtuales e instrumentación remota
- HEP y astro-partículas

# HEP y astro-partículas

- Proyectos internacionales
- Participación de varias instituciones en México
- Requerimientos documentados
- Necesidades
  - ▶ continuas
  - ▶ de pico: QOS o ancho de banda

# ALICE @ LHC

- Estudio de colisiones de iones pesados
  - ▶ Re-crear el estado de materia en el universo temprano
  - ▶ Estudio del plasma de cuarks y gluones
- Programa de física proton-proton
- Dos detectores de ALICE fueron construidos en México



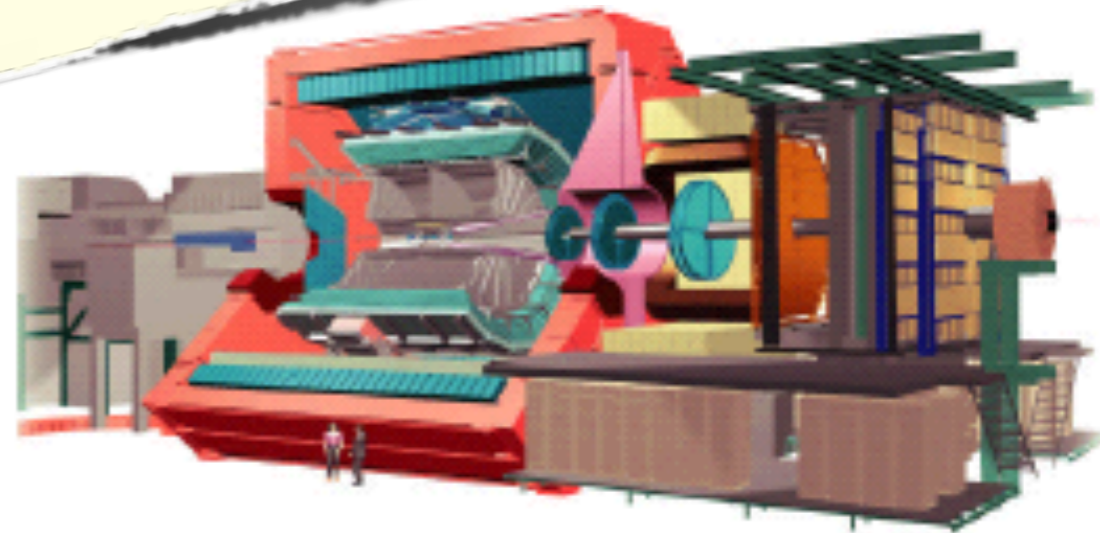
# ALICE @ LHC

- Estudio de colisiones de iones pesados

- ▶ Re-crear el estado de materia temprano

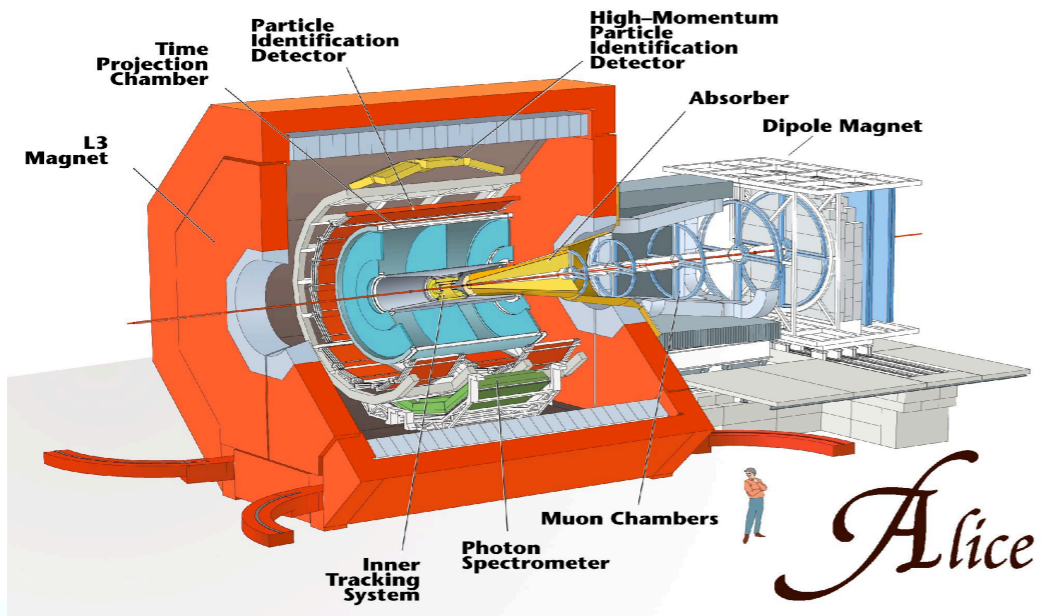
**Colaboradores mexicanos:**  
**BUAP, CINVESTAV, Culiacan,**  
**UNAM (ICN, IF), UMSNH**

- F
- D
- ALICE fueron construidos en Mexico





# Generación de datos en Alice



## Alice detector

Total weight	10,000t
Overall diameter	16.00m
Overall length	25m
Magnetic Field	0.4Tesla

## Compare to Shuttle



Total weight	2,000t
Overall length	17.3m

8 kHz (160 GB/sec)  
level 0 - special hardware

level 1 - embedded processors

level 2 - PCs

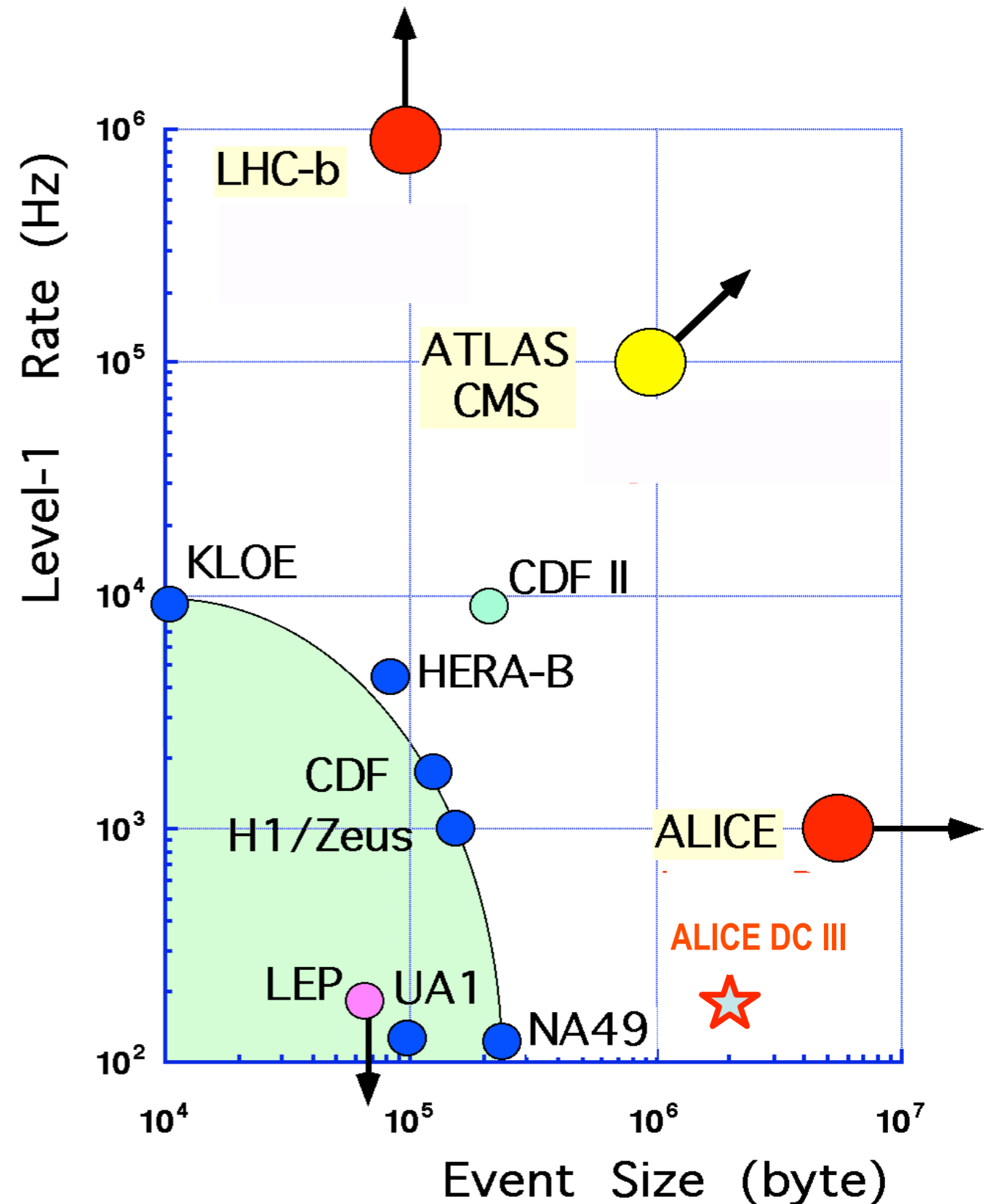
30 Hz (1.25 GB/sec)  
data recording & offline analysis

**online system**  
multi-level trigger  
filter out background  
reduce data volume

**More than 3 PB/ year!**

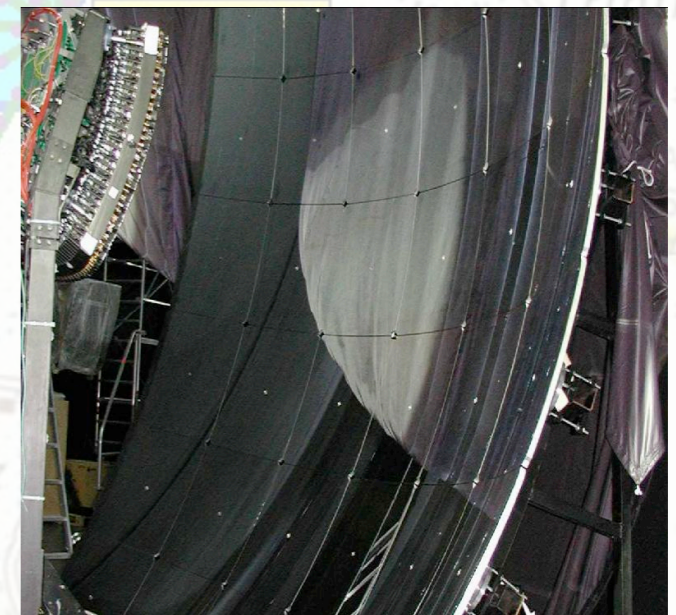
# Volumen de datos en HEP

- Aumento en cantidad de datos generados requiere nuevas estrategias para su procesamiento
- Modelos de cómputo basados en GRIDs



# Observatorio Pierre Auger (OPA)

- El observatorio más grande para rayos cósmicos
- Sitio sur: Malargüe, Argentina
- Participación mexicana
  - ▶ Tanques (ROTOPLAS)
  - ▶ Diseño óptico
  - ▶ Software para análisis de datos
  - ▶ Optimización de software para GRID



# Observatorio Pierre Auger (OPA)

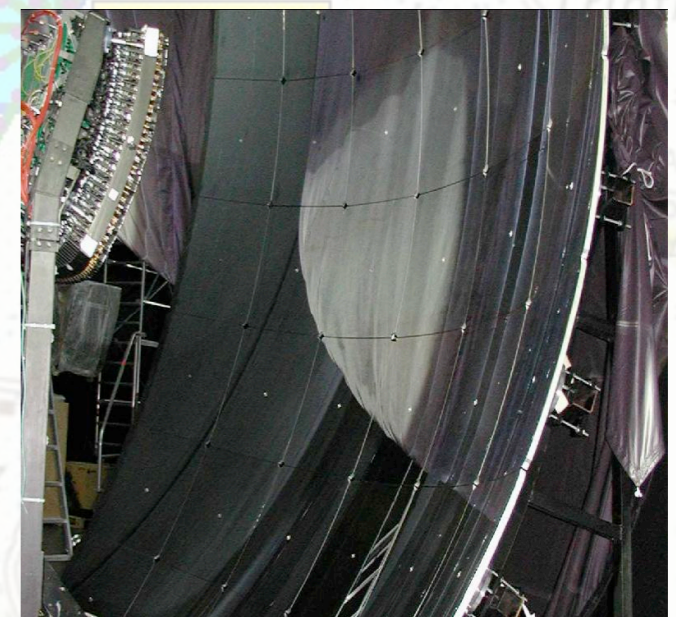
- El observatorio más grande para rayos cósmicos

- Sitio sur: Malargüe, Argentina

- Participación mexicana

**Colaboradores mexicanos:**  
**BUAP, CINVESTAV,**  
**UNAM (ICN, IGeof), UMSNH**

- ▶ Optimización de software para GRID



# Software para OPA

- Fuerte colaboración en diseño y implementación
  - ▶ Desarrollo en C++
- Servicios centrales instalados en la UNAM
  - ▶ Web + wiki
  - ▶ Bugzilla
  - ▶ SVN
  - ▶ listas de correo
- La colaboración **consideró re-ubicar** estos



**Generan visibilidad**

# HAWC

- Observatorio de rayos gama y de rayos cósmicos
- Sitio en la Sierra Negra (vecino del GTM)
- Colaboración México-EEUU



# HAWC

- Observatorio de rayos gama y de rayos cósmicos
- Sitio en la Sierra Negra (vecino del

**Colaboradores mexicanos:**

**BUAP, CINVESTAV, UG**

**UNAM (IA, ICN, IGeof, IF), UMSNH**



# Tochtli

- **Master node (CE):**  
`tochtli.nucleares.unam.mx`
- **15 Worker Nodes (WN)**
- **Storage Element (SE)**  
`tlapiacalli.nucleares.unam.mx`  
(almacen)  
~30TB disk space
- **Alice VO box**  
`calpolli.nucleares.unam.mx`  
(vecindad, tribu)





**Storage element**

**Nodo maestro**

**Nodos de trabajo**

tochtli.nuclear

15 Worker Node

Storage Element

tlapiacalli.nuclear

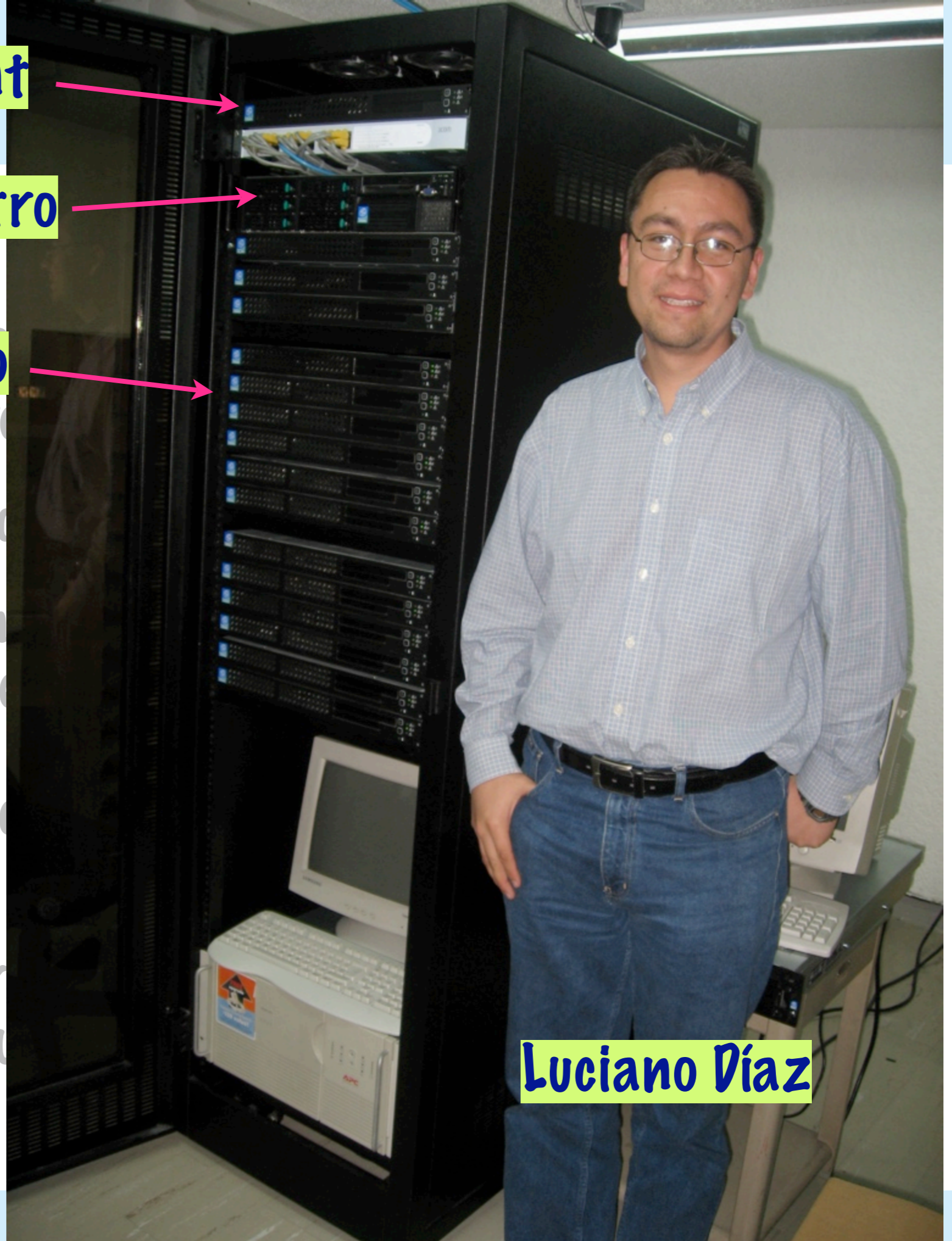
(almacen)

~30TB disk space

Alice VO box

calpolli.nuclear

(vecindad, tribu)

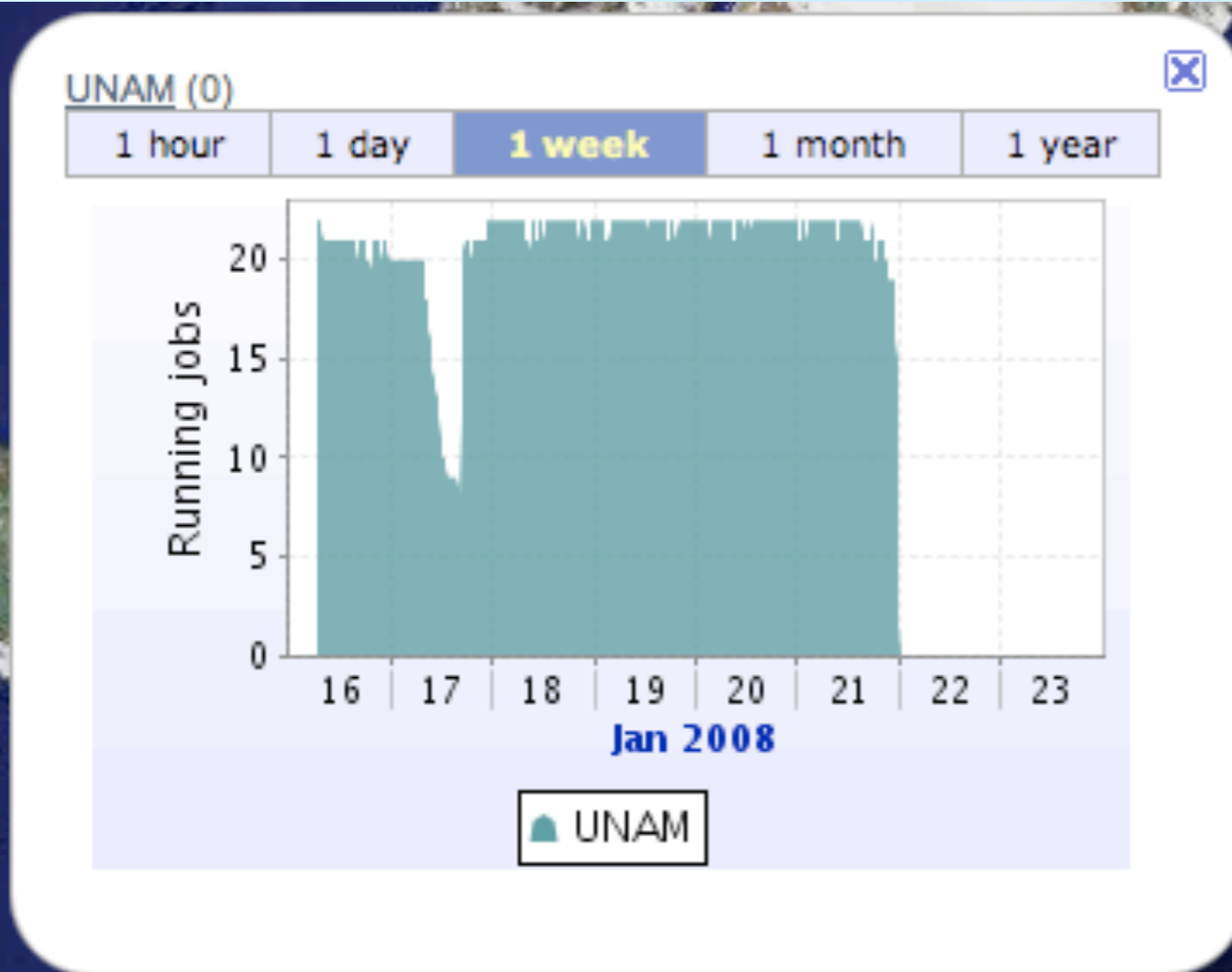


**Luciano Díaz**

# Producción de simulaciones

- Xeon@2.5GHz: ~0.2 MB/s
- Opteron@2GHz: ~0.5 MB/s
  - ▶ Por el momento para trabajo interactivo
- Cores instalados: 30 Xeon, 16 Opteron
  - ▶ ~15 MBit/s para transmisión de resultados
- Los jobs en las últimas corridas eran menos demandantes

# Jobs corriendo en la UNAM



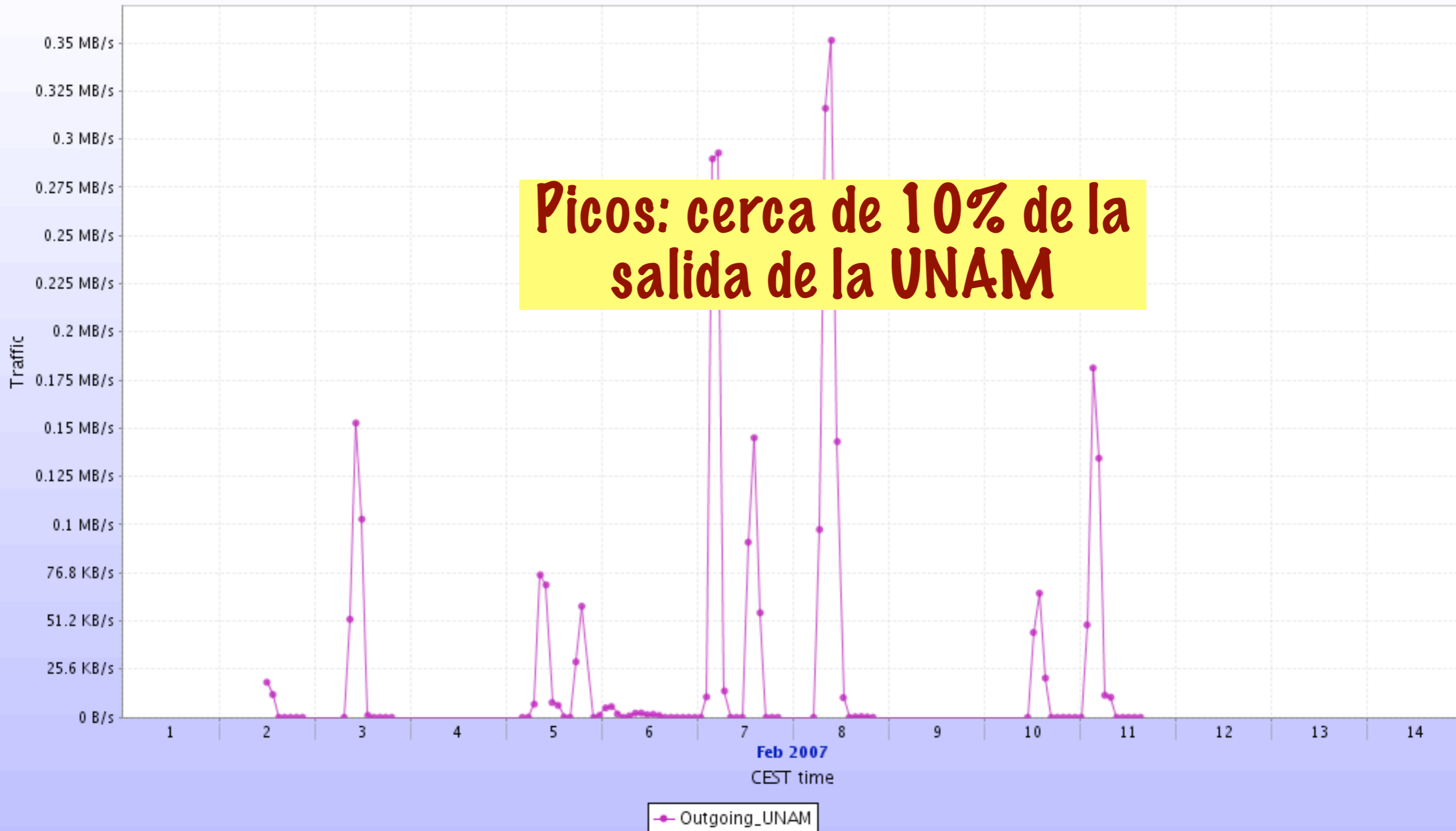
Ejecutando ~22 jobs

Interrupción del servicio por reconfiguración

- Almacenamiento
- Nodos de trabajo

# Trafico generado por ALICE

Site Outgoing Traffic



# Simulaciones par OPA

- Actualmente: 0.1 MBit/s/core
- Con mejoras en el software: 0.5 MBit/s/core
- Parecidos a los requerimientos de ALICE

# HAWC - datos reales

- Observatorio ubicado en México
- 6000 cascadas/s, 1.7 KB por evento
  - ▶ 11 Mbit/s continuas
- Se requiere x5 para operar el sistema de replica de datos
  - ▶ transferencia a la UNAM
  - ▶ transferencia a EEUU

# Evaluación de la red actual

- Transferencias CERN  $\Leftrightarrow$  UNAM
  - ▶ Marzo 2008
  - ▶ September 2008: jumbo frames activados
- Salida de la UNAM saturada
- No es posible evaluar la capacidad en CUDI

# Detalles técnicos (Marzo)

- Herramienta: Iperf
- TCP Window size: 256, 512, 1024 KB
- Número de hilos (conexiones simultáneas):  
1,4,8,16
- Pruebas realizadas: 24
- Duración total: 1 día
- Pruebas en ambas direcciones simultáneamente,  
no sincronizadas



# Detalles técnicos (Marzo)

- Herramienta: Iperf

- TCP Window

- Números de hilos (1, 4, 8, 16)

- Pruebas realizadas: 24

- Duración

- Pruebas en ambas direcciones simultáneamente, no sincronizadas

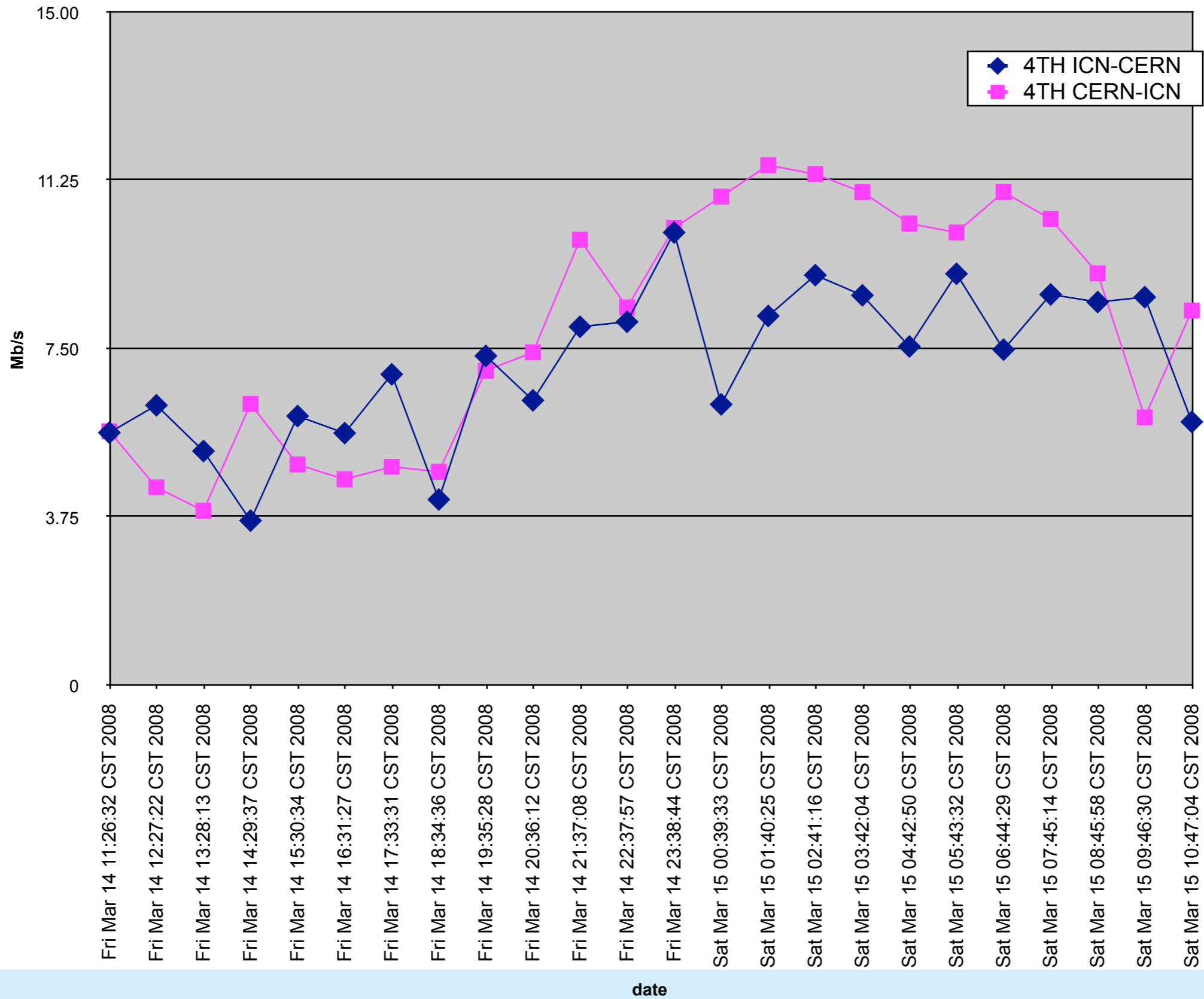
**Presentación completa  
de resultados**

**GUY Paic,**

**reunión primavera 2008**

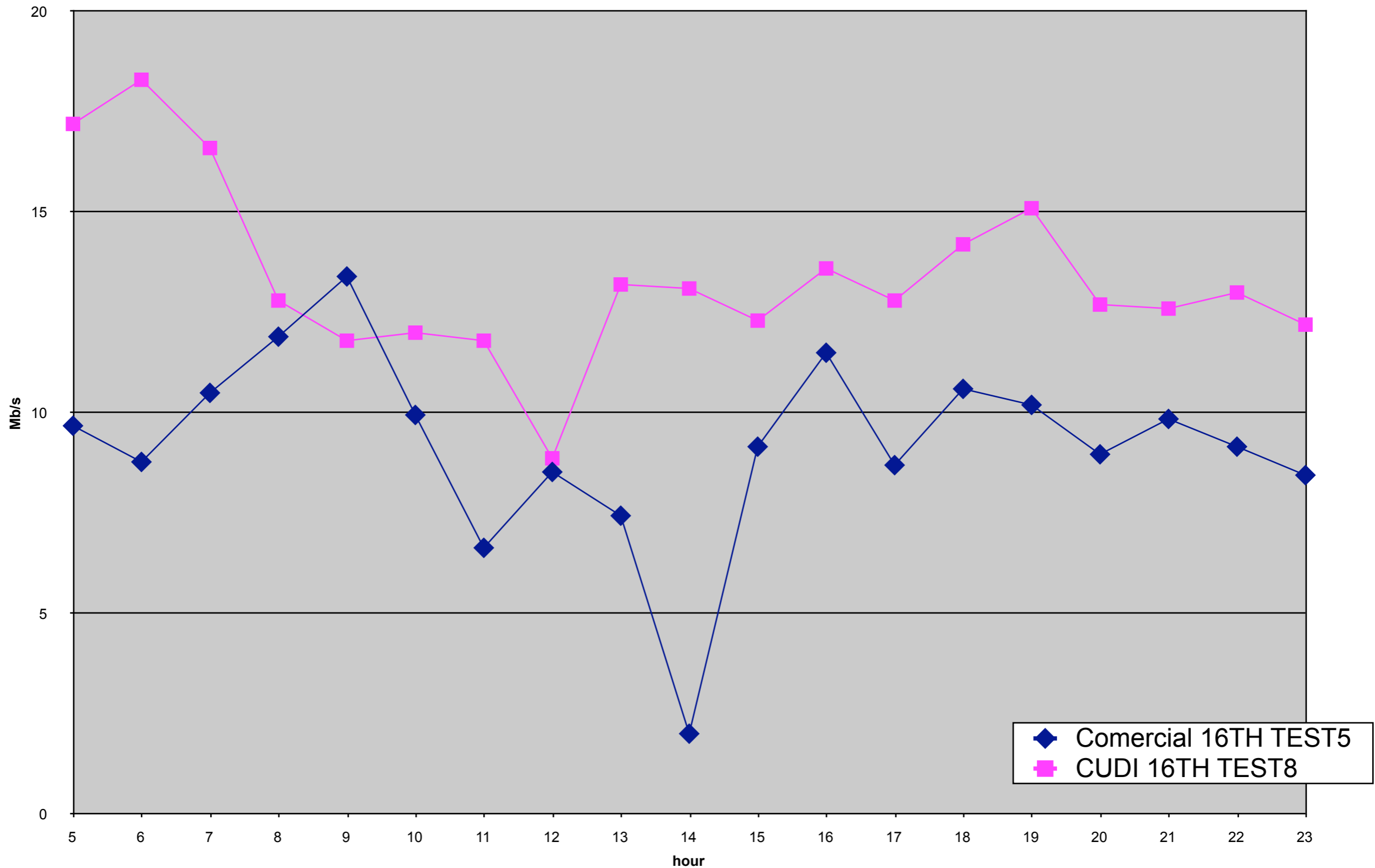
# CERN ↔ ICN-UNAM

Network test ICN-CERN <>CERN-ICN

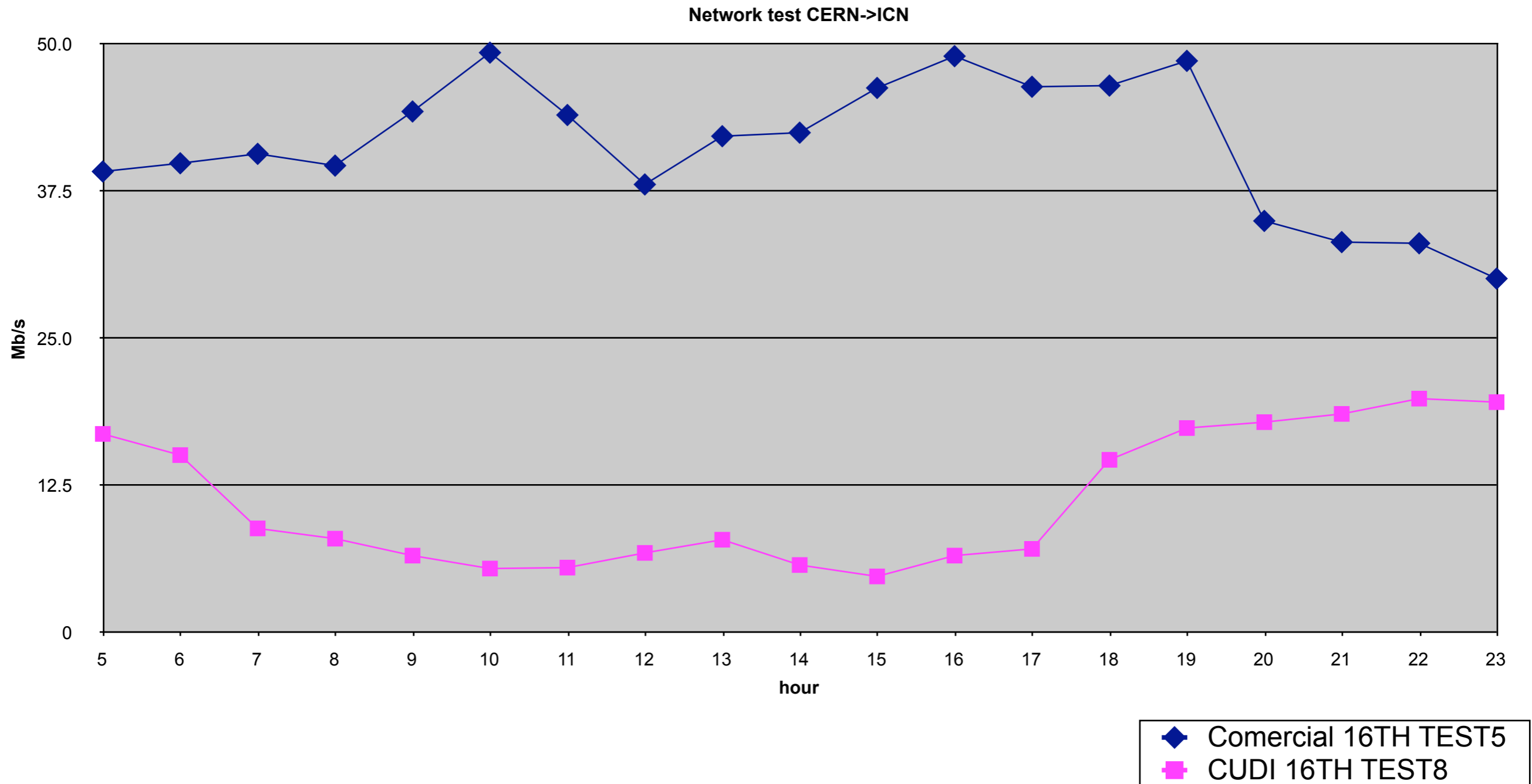


# Internet comercial vs CUDI

Network test ICN->CERN

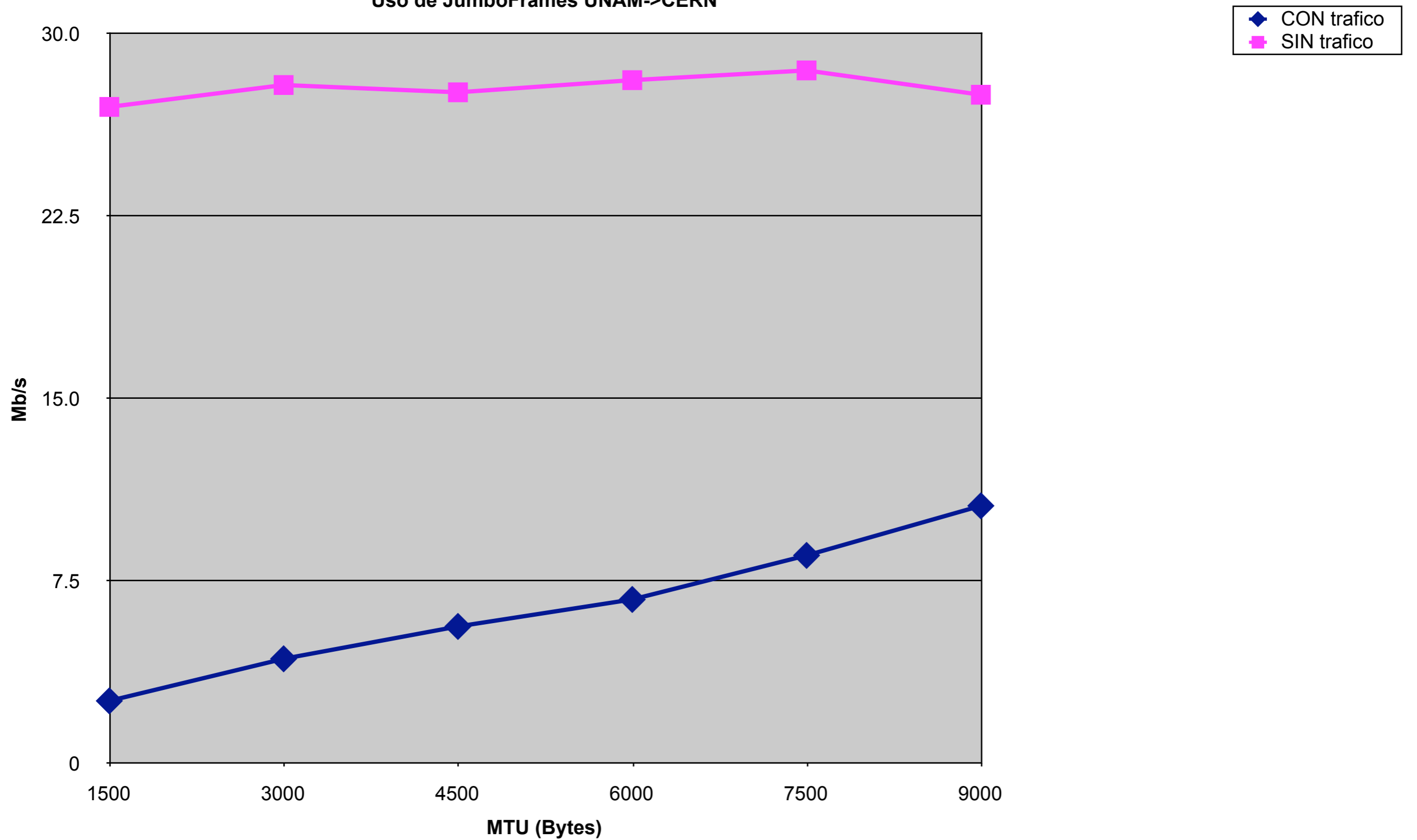


# Internet comercial vs CUDI



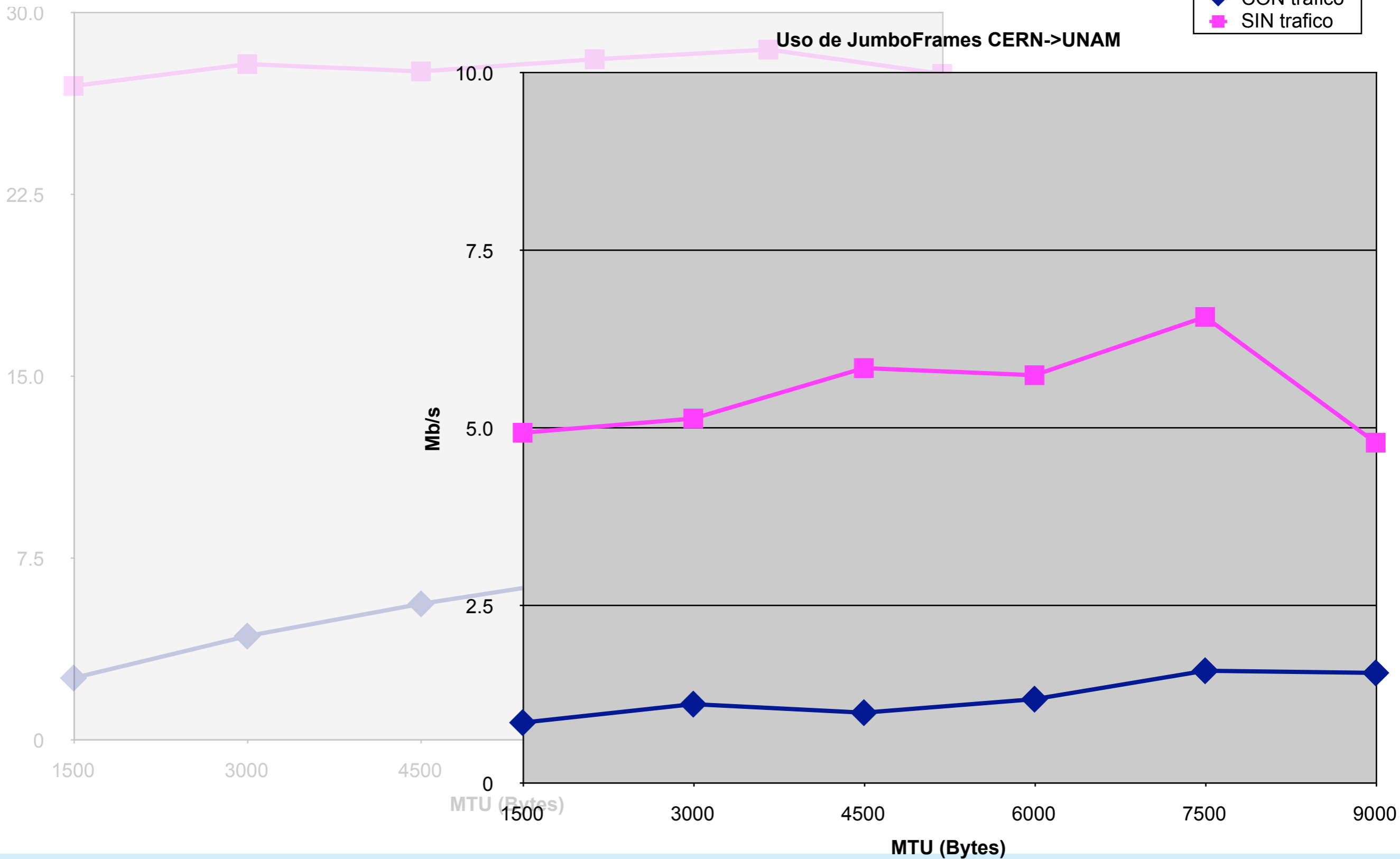
# Jumbo frames

Uso de JumboFrames UNAM->CERN



# Jumbo frames

Uso de JumboFrames UNAM->CERN



# Observaciones (I)

- En marzo, hubo cambios en la configuración del ruteo México-Europa sin anuncios
  - ▶ Motivado por la opera abierta
  - ▶ duración ~1 semana, mucho más que el evento
- CUDI y GEANT-2 ayudaron en arreglar el ruteo
  - ▶ Colaboración entre usuarios y redes es esencial  
La red **no se puede tratar como caja negra**
- Necesidad de un grupo de “**usuarios avanzados**”.

# Observaciones (2)

- Ruteo: Mexcio-EEUU-Europa
  - ▶ No usamos CLARA
- Problema con CLARA:
  - ▶ trafico viaja: DF-Tijuana-Panama-Floria-Europa
  - ▶ Latencia perjudica transferencias (“bandwidth-latency product”)
- CLARA en su configuración actual
  - ▶ Bueno para tráfico entre America Latina
  - ▶ Problemático para Mexico  $\Leftrightarrow$  Europa



# Novedades

- UNAM duplica ancho de banda hacia CUDI
- Se adquiere un nuevo ruteador con capacidad de jumbo frames
  - ▶ Requerimiento para proyecto de transferencia de datos para Alice: ICN-UNAM  $\Leftrightarrow$  Houston
- Resultado de iniciativa de investigadores
  - ▶ Ayudo a establecer un **diálogo entre usuarios (científicos) y proveedores** de servicios

# Otras aplicaciones

- Existen también aplicaciones con necesidades variantes
  - ▶ Ejemplo: Imagenología médica
    - imágenes en patología: ~ 1 GB
    - no se pueden reducir (pérdida de información!)
  - ▶ Deseable: Acceso interactivo a archivos
- Se requiere ancho de banda de GB/s
- Que es más fácil: QOS o ancho de banda extra

# Importancia de la red

- Crecimiento limitado por falta de ancho de banda
  - ▶ Con la capacidad de cómputo instalado podemos ocupar una fracción grande de la conexión de la UNAM
- Nuevos proyectos implican más demanda
- Tenemos que pensar ya en una red de 10Gbit/s

# Escenario GRID

- Cluster de ~500 cores
  - ▶ 2 racks, ~3-4M\$ M.N.
- 0.5 Mbit/s/core (típico ALICE, Auger)
- Requiere 250MBit/s para entregar resultados
  - ▶ Una fracción notable de 1GBit/s
- Algunas colaboraciones aceptan **cubrir gastos de operación** con contribuciones en cómputo
  - ▶ Inversión en México en vez de transferir fondos

# Resumen

- Buena conectividad esencial para proyectos grandes
- Los grupos mexicanos han demostrado su capacidad de jugar papeles importantes en proyectos de frontera
- Sin incremento substancial del ancho de banda, el desarrollo será limitado
- Se requiere un esfuerzo para cubrir todo el país
- Proyectos fuertes son el motor, los **beneficiarios son todos**

# Resumen

- Buena conectividad

**Redes son  
de humanos  
de datos  
crecen juntos**

- Sin incremento substancial del ancho de banda el desarrollo será limitado
- Se requiere un esfuerzo por parte del país
- Proyectos futuros con el motor, los beneficiarios son todos