

Area Temática de Física Experimental de Partículas Elementales.

Relación con el desarrollo de la e-Ciencia en España

26 de Octubre de 2004

La Física Experimental de Partículas Elementales, muy a menudo llamada Física de Altas Energías (en adelante, FAE), es un campo de bastante actividad en España, considerado prioritario desde 1983 y actualmente parte del Programa Nacional de Física de Partículas 2004-2007.

LA COMUNIDAD:

Algunas tendencias de FAE en las últimas décadas se describen en el Anexo I. Actualmente, la comunidad trabaja organizada en "experimentos", cada uno de los cuales es una actividad colaborativa a nivel mundial para diseñar, construir, operar y realizar el análisis de datos de un detector. Los detectores están localizados en un número bastante reducido de sitios, tradicionalmente Laboratorios que operan aceleradores (esencialmente en Europa, USA, Japón y China) o en lugares idóneos para detectar rayos cósmicos o realizar otros tipos de experimentos sin acelerador (esta modalidad está teniendo un gran auge en los últimos 10 años y se espera que esta tendencia siga, con experimentos en lugares generalmente bastante remotos en Argentina, Antártida, el desierto del Sudoeste de USA, el fondo del Mar Mediterráneo, Siberia, Australia, etc. o cerca de observatorios astronómicos en Chile, Canarias, etc.).

Sin preocuparnos de factores de orden 2 y con la intención de generar una pincelada fácil de recordar, podemos hablar de los siguientes órdenes de magnitud:

Número de científicos en FAE.....	10^4
Tamaño promedio de un grupo colaborador en un experimento dado.....	10
Número de institutos con grupos en FAE.....	10^3
Número de países con grupos en FAE.....	10^2
Duración en años de un experimento (desde concepción hasta desmantelamiento).....	10

Una tendencia importante es que un experimento en particular tiene una duración total cada vez mas larga, y que viene a ser del mismo orden que la longitud de una carrera profesional. Por otro lado, tanto los detectores como los aceleradores se construyen espaciados en el tiempo, por tanto en cualquier momento se puede encontrar alguno en cualquier fase (conceptualización, diseño, construcción, toma de datos, análisis, *upgrade*, desmantelamiento). Por tanto, un físico o ingeniero o técnico particular tiende a trabajar simultáneamente en unos cuantos experimentos, con una constante de tiempo de cambio de contexto que varía entre unos pocos días y unos pocos años. Otra característica importante es que aunque los físicos de FAE se especializan, lo hacen de una manera plural cubriendo aspectos de las fases anteriormente descritas, y

por tanto participan íntimamente en muchas de dichas fases. Esto quizás sea algo diferente de lo que ocurre en otros campos, donde hay una diferenciación mas aguda entre los "científicos" y los "instrumentalistas". Por último, hay esencialmente una completa descorrelación entre las localidades geográficas de la institución de un físico particular, de las instituciones de sus colaboradores y de sus experimentos.

Los grupos de Altas Energias españoles estan participando en experimentos del LHC y en otros experimentos de otras características (por ejemplo, en experimentos de Astroparticulas – AMS, Antares, Magic- o en experimentos en Estados Unidos y Japón (CDF, BaBar, K2K). A continuación se dan los grupos que participan en experimentos LHC:

- Grupo de Altas Energias del CIEMAT (Madrid) que participa en CMS
- Grupo de Altas Energias del IFCA (Santander) que participa en CMS
- Grupo de Altas Energias de la UAM (Madrid) que participa en ATLAS
- Grupo de Altas Energias del IFAE (Barcelona) que participa en ATLAS
- Grupo de Altas Energias del IFIC (Valencia) que participa en ATLAS
- Grupo de Altas Energias de la UB (Barcelona), que participa en LHCb
- Grupo de Altas Energias de la USC (Santiago de Compostela), que participa en LHCb

Ademas esta el PIC (Puerto de Informacion Científica) de Barcelona cuya finalidad es, entre otras, el suministrar servicios GRID a los tres experimentos del LHC (ATLAS, CMS y LHCb).

Por otro lado existen otros grupos de Altas Energias en España que no participan en ningun experimento del LHC :

- Grupo de AAEE de la Universidad de Zaragoza
- Grupo de AAEE de la Universidad Complutense de Madrid
- Grupo de AAEE de la Universidad de Granada

Un aspecto de esencial importancia dentro de las Altas Energias en España es nuestra participación en el CERN del que nuestro pais es miembro desde 1984. Podemos decir que la comunidad de fisicos de Altas Energias de grupos españoles esta mayoritariamente en colaboraciones o experimentos del CERN (concretamente en experimentos de LHC: ATLAS, CMS y LHCb) , que es el Laboratorio Europeo de Altas Energias que actua, entre otras cosas, como gran facilidad de suministro de aceleradores y de física de detectores. Debido a la excelencia de este centro, el CERN es tambien el centro lider de proyectos GRID tales como el DATAGRID, el EGEE, etc.

La participación española en la puesta a punto del sistema GRID internacional para el LHC (que es el llamado Proyecto LHC Computing GRID) esta estimado en el 5% del esfuerzo total

Y como se ha comentado, tambien hay grupos que estan participando en proyectos de aceleradores de EEUU y de Japón y estan interesados en utilizar los avances en GRID para el procesado de datos. Por ejemplo estan las participaciones en CDF y en BaBar que son experimentos que estan en fase de toma de datos (running experiments) que

pueden servir con banco de pruebas de algunos aspectos de procesado de datos 'a la GRID'.

EL DÍA A DÍA DE LA COLABORACIÓN:

Dado este panorama, no será ninguna sorpresa que FAE utiliza de manera generalizada el soporte digital en modo colaborativo, cubriendo toda la gama desde correo electrónico, Webs colaborativos, repositorios con acceso remoto de diseños mecánicos, electrónicos y de software, hasta la videoconferencia y la gestión y operación remota de dispositivos y sistemas de adquisición de datos. Dado el gran número de países, instituciones y dominios de administración que intervienen, la única manera práctica de realizar todas estas actividades es utilizando herramientas con estándares abiertos a través de Internet. El resultado neto es que, excluyendo el análisis colaborativo de datos, cada colaborador en FAE es un "e-Físico" que colabora continuamente a escala mundial utilizando un ordenador personal que es necesario renovar aproximadamente cada 3 años, una infraestructura de correo, Web y repositorios colaborativos que contiene unos 100s de GB y que, en condiciones ideales, consumiría individualmente el equivalente nominal de una Ethernet tradicional, es decir 10 Mbps. Es interesante notar que 10^4 personas x 10 Mbps = 100 Gbps nominales, aunque evidentemente la estructura de picos y valles de utilización es muy aguda. La falta de red académica de calidad es un factor crítico limitante a la colaboración, que sorprendentemente no sólo se manifiesta en los llamados "países menos desarrollados" sino también en Europa, USA, Canadá y Japón debido a deficiencias de "último Kilómetro" o incluso "últimos metros" en campus universitarios.

El principal problema de estos "e-Físicos" es la necesidad de acceder continuamente de un dominio de administración/seguridad a otro y su sueño es tener un "*single sign-on*" global como individuo por un lado, y por el otro el poder cambiar su *contexto o rol* de una manera fácil y relativamente dinámica para facilitar el trabajo cuasi-simultáneo en varios experimentos. Estos deseos se están haciendo realidad poco a poco, gracias al despliegue de Infraestructuras Grid como EGEE. Se espera que gracias al despliegue coordinado de un conjunto interoperable de Autoridades de Registro y de Certificación X.509 todas las personas de la comunidad FAE tengan en breve un certificado reconocido a nivel mundial. Luego, gracias a sistemas todavía en fase experimental, como el "*Virtual Organization Management System*" desarrollado en el proyecto EU DataGrid, se puede gestionar el acceso a recursos asignados a cada experimento (es decir, a cada "Organización Virtual") según la rol asignada a cada individuo.

EL CASO DISTINGUIDO: TRATAMIENTO Y ANÁLISIS DE LOS DATOS:

En el apartado anterior se excluyó el tratamiento de los datos de FAE en sí, ya que este problema requiere atención especial por su magnitud y dificultad. Es difícil generalizar, pero se puede decir que casi cualquier experimento actual está generando al menos 10-100 GB *por día* de datos a analizar, y en muchos casos 1 TB por día es habitual. Como se verá mas abajo, el Large Hadron Collider tendrá un volumen muy

superior. Una figura muy aproximada es que el tratamiento de estos datos requiere entre 100 y 10000 "instrucciones" por byte, utilizando la palabra "instrucción" en el sentido de la definición de MIPS. El número de individuos que apliquen dichos tratamientos está generalmente anti-correlacionado con los recursos necesarios, es decir que hay muchos tratamientos de pocas "instrucciones" por byte (el análisis individual de datos muy filtrados o muy reconstruidos) y pocos tratamientos de muchas "instrucciones" por byte (re-procesamientos o simulaciones organizados por la colaboración experimental). Una figura extremadamente aproximada es que en total se necesitan unas 20000 "instrucciones" por byte por año, y que el flujo integrado global de datos alcanza al menos el triple del flujo que emana del sistema de adquisición de datos.

Estas cifras son bastante superiores a las de décadas anteriores y desbordan el modelo habitual en la década de los 90 basado en realizar gran parte del tratamiento de datos en estaciones de trabajo RISC y ordenadores personales conectados por Ethernet. Afortunadamente, el hecho de que los datos de FAE son mayoritariamente un enorme conjunto de registros independientes cuyo tratamiento y análisis constituye un ejemplo de computación "vergonzosamente paralela", sigue vigente, y por tanto soluciones basadas en clusters de, por ejemplo, PCs con Linux y Gigabit Ethernet son posibles. La única peculiaridad, que afecta mas a gestores que a científicos, es el hecho de que el número de ordenadores pasa a ser superior al número de físicos.

Para dar números mas concretos, aunque con la advertencia de que están en proceso de revisión y que pueden haber cambios de factores de 2 o mas, se citan en la Tabla 1 algunas de las estimaciones recientes para la computación de los cuatro experimentos del Large Hadron Collider (LHC). Estas necesidades probablemente representen del orden de la mitad de las necesidades globales de FAE en 2006-2010.

Si las capacidades de almacenamiento, proceso y transferencia de datos necesarias para el LHC son de por sí bastante altas, es importante mencionar también que dichos datos se corresponden a unas 10^{11} a 10^{12} colisiones, entre las cuales los 6.000 físicos del LHC distribuidos en cientos de institutos buscarán subconjuntos quizás tan pequeños como 10^3 o 10^2 colisiones, utilizando características iterativamente derivadas de los datos y no conocidas *a priori*. Es evidente que un nivel importante de gestión automática y fiable de subconjuntos de datos será muy necesario. Dicho sistema de gestión aún queda por desarrollar y deberá de ser el objeto de intenso desarrollo en un futuro inmediato.

RELACIÓN AL DESARROLLO DE LA E-CIENCIA EN ESPAÑA

La comunidad de FAE ha invertido mucho esfuerzo en desarrollos de técnicas Grid. La razón es análoga a la adopción universal de redes compartidas TCP/IP: Al poder "infraestructurizar" parte de los servicios por un lado, y al "universalizarlos" por el otro, se pueden obtener grandes economías de escala. Por ejemplo, no tiene sentido que FAE mantenga su propio sistema de Autoridades de Registro y Certificación, cuando puede ser compartida con otros campos científicos por un lado, y además la mayoría de Universidades y Centros de Investigación también desean dicho sistema para sus propias aplicaciones. En este sentido una política coherente hacia el desarrollo de Infraestructura Grid en España es muy importante para FAE.

Localización	Parámetro	Σcapacidad 2006-2008	Incremento anual >2008
CERN	Almacenamiento en disco	5.000 TB	2.500 TB
	Almacenamiento en cinta	20.000 TB	10.000 TB
	Transferencia de cinta	4.000 MB/s	2.000 MB/s
	CPU en cluster	20.000 KSI2000	10.000 KSI2000
	Red agregada del cluster	1.000 Gbps	500 Gbps
	Conectividad externa	80 Gbps	40 Gbps
Conjunto de centros de transformación de datos	<i>Número de centros</i>	10	-
	Almacenamiento en disco	20.000 TB	10.000 TB
	Almacenamiento en cinta	20.000 TB	10.000 TB
	Transferencia de cinta	1.000 MB/s	500 MB/s
	CPU en cluster	40.000 KSI2000	20.000 KSI2000
	Red agregada del cluster	2.000 Gbps	1.000 Gbps
	Conectividad externa	200 Gbps	100 Gbps
Conjunto de centros de análisis de datos y simulación	<i>Número de centros</i>	25-40	?
	Almacenamiento en disco	10.000 TB	5.000 TB
	Almacenamiento en cinta	5.000 TB	3.000 TB
	Transferencia de cinta	No evaluada	No evaluada
	CPU en cluster	40.000 KSI2000	20.000 KSI2000
	Red agregada del cluster	No evaluada	No evaluada
	Conectividad externa	100 Gbps	50 Gbps

Tabla 1

Además, podría ser el caso (aunque esto queda por probarse), que los recursos computacionales y de almacenamiento se logren utilizar de una manera mas eficiente si se incorporan a Infraestructuras Grid. De hecho, es la automatización de la replicación y gestión automática de complejos subconjuntos de datos lo que es realmente crucial para FAE (y lo que desafortunadamente no ha recibido énfasis y comprensión adecuadas por parte de la prensa o por autoridades de política científica, que tienden a enfatizar los aspectos de "supercomputación" del Grid).

Aunque para un Físico de Altas Energías de casi cualquier edad el término "e-Ciencia" resulte un tanto curioso, ya que da por hecho el ser un "e-Físico", es muy probable que se pueda realizar una beneficiosa transferencia de conocimientos de la comunidad FAE a otras comunidades científicas. En el caso español, la única Infraestructura Grid que opera en modo producción 24x7 es la derivada del banco de pruebas para el LHC, que ahora se está extendiendo a otras ciencias a través del proyecto europeo EGEE. Se considera muy conveniente que acciones de fomento de la "e-Ciencia" en España re-utilicen estas experiencias para maximizar los beneficios a corto y medio plazo.

Es justamente en esta línea donde también nuestra comunidad se ha involucrado en nuestro país: los físicos de Altas Energías están liderando y participando muy activamente en la llamada Iniciativa IRISGRID cuya finalidad es ser el embrión de una infraestructura de GRID Nacional y de dar origen al sistema de e-Ciencia a nivel español.

ANEXO I: La Física de Altas Energías en las últimas décadas

La evolución de FAE en las últimas décadas ha estado dominada por el diseño, construcción y explotación de detectores de partículas (en adelante, experimentos) cada vez más grandes y complejos, los cuales operan en aceleradores que colisionan haces de partículas cada vez más energéticos y/o intensos. Esta tendencia se puede explicar a nivel divulgativo a través de tres observaciones relacionadas a la física fundamental: a) la relación inversa entre longitud de onda y energía (mecánica cuántica), b) el incremento del número de partículas producidas cuando se incrementa la densidad de energía (similitudes entre la teoría de campos y la termodinámica) y c) el que la existencia de partículas con masa mayor que la energía disponible pueda modificar ligeramente observables (teoría cuántica de campos). Mas coloquialmente, para avanzar el conocimiento en FAE o se observan colisiones con mucha energía para explorar distancias muy pequeñas o se observa un gran número de colisiones para explorar efectos muy sutiles, o las dos cosas a la vez. Esto a su vez conlleva la construcción de experimentos de gran tamaño, ya que para poder distinguir con suficiente precisión las trayectorias individuales de un gran número de partículas producidas y también medir su momento o energía, por curvatura en un campo magnético si son cargadas o por métodos calorimétricos si no lo son, se han de vencer las limitaciones de resolución de detección de posición, máximo campo magnético generable en condiciones prácticas y cantidad de material necesario para absorción calorimétrica.

Además, particularmente en la última década, ha habido un gran desarrollo de experimentos de partículas que no utilizan aceleradores, observando partículas elementales que inciden sobre la tierra, tales como rayos cósmicos, neutrinos o rayos X. Este desarrollo está desdibujando las fronteras entre FAE, astrofísica, astronomía y cosmología. Una particularidad de estos experimentos es que muy a menudo requieren localizarse en zonas bastante remotas, tales como desiertos, el fondo marino, Antártida, etc., en los cuales no hay una infraestructura de soporte similar a la de un laboratorio de aceleradores. Además muchas veces la conectividad de red puede ser inadecuada, obligando a resucitar técnicas abandonadas por FAE hace muchos años, como por ejemplo la grabación y envío de cintas magnéticas con los datos experimentales. En España, por ejemplo, este es el caso del Observatorio del Roque de los Muchachos en La Palma, Canarias.

Las colaboraciones experimentales han crecido en tamaño a medida que se construye un menor número de detectores más complejos, más costosos y con duración de proyecto más larga. La típica colaboración experimental ha crecido en tamaño de menos de una docena de físicos provenientes de dos o tres instituciones en los años 70 a cientos de físicos de docenas de instituciones en los años 90 hasta las macrocolaboraciones del Large Hadron Collider, que típicamente tienen unos 1.500 colaboradores procedentes de cientos de instituciones. Dado que en el período 1970-presente el número de universidades y centros de investigación por país no ha

cambiado significativamente, y el número de investigadores por centro tampoco, el resultado neto es una completa internacionalización de las colaboraciones experimentales. Además hay un fenómeno de descorrelación entre las localidades geográficas de la institución de un físico particular, de las instituciones de sus colaboradores y de sus experimentos.