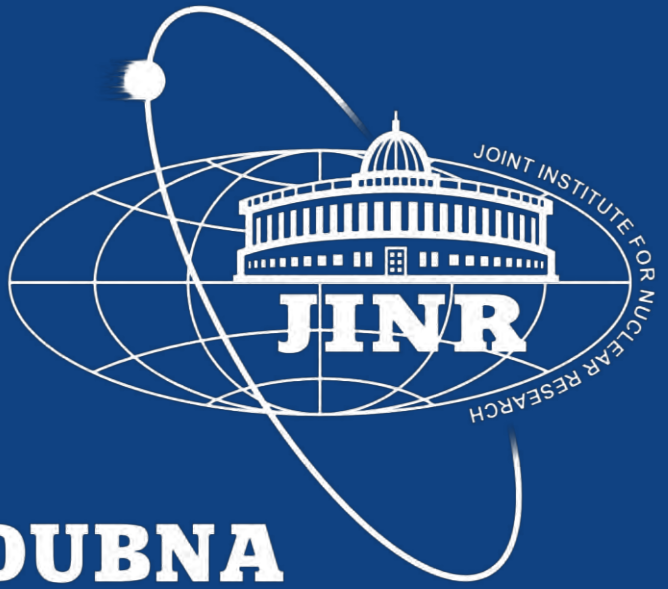


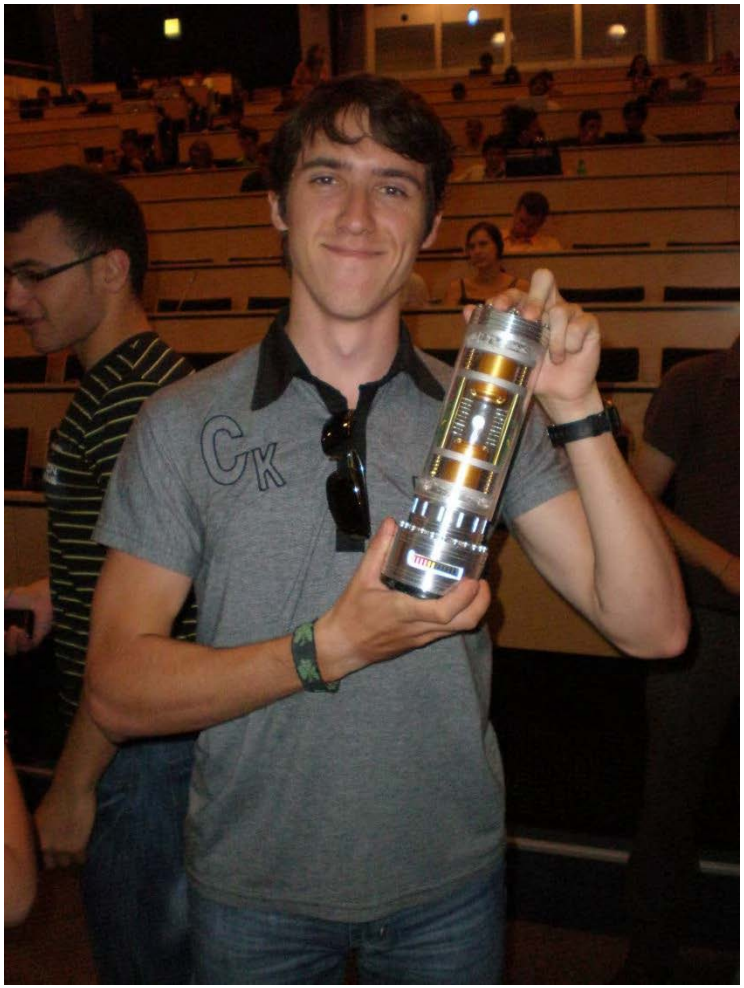
DUBNA

Объединённый
Институт
Ядерных
Исследований

BigData в большой науке о физике частиц

Рассказывает:
Игорь Пелеванюк

Обо мне



Игорь Пелеванюк (1991)

Окончил - Университет «Дубна» в 2013

Профессия – программист

В ОИЯИ – с 2013г
Разрабатываю веб сервисы
Объединяю ресурсы
Преподаю в университете
Провожу экскурсии

О Дубне

Дубна (1956)

121 км на север
от Москвы

Население – 75000

- Институт
- Университет
- Особая экономическая зона

Об ОИЯИ

Объединённый Институт Ядерных Исследований (1956)

Более 4500 сотрудников

Member States

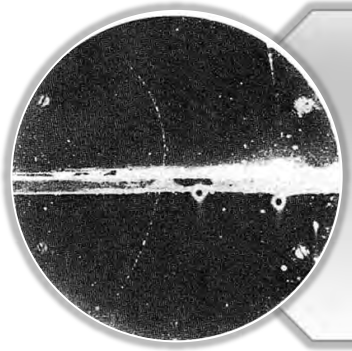
Armenia, Azerbaijan, Belarus, Bulgaria, Cuba, Czech Republic, North Korea (Membership suspended), Georgia, Kazakhstan, Moldova, Mongolia, Poland, Romania, Russia, Slovakia, Ukraine, Uzbekistan, Vietnam

Associate members

Egypt, Germany, Hungary, Italy, South Africa, Serbia

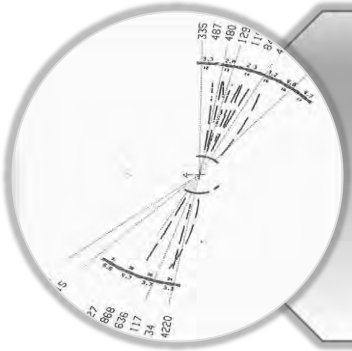
Ядерная физика, физика частиц, информационные технологии, радиобиология и др.

Как мы делаем открытия



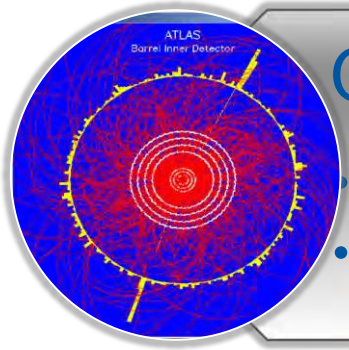
Открытие в 30х

- ~2 учёных
- Карандаш и бумага



Открытие в 70х

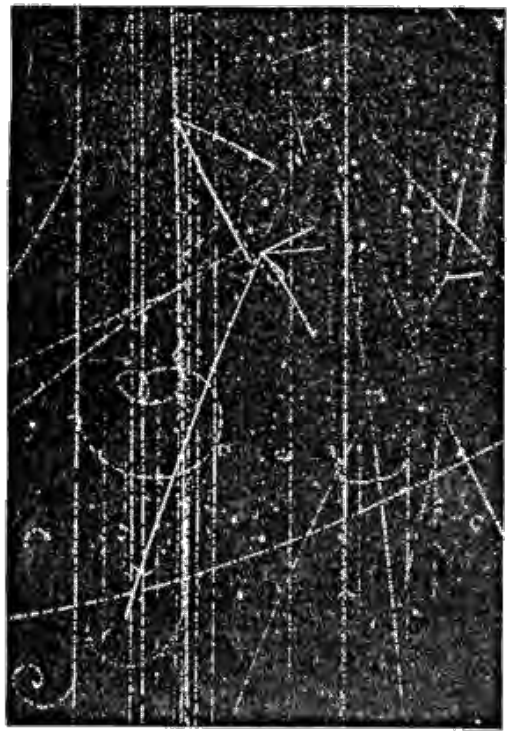
- ~200 учёных , ~10 стран
- Мэйнфрэймы

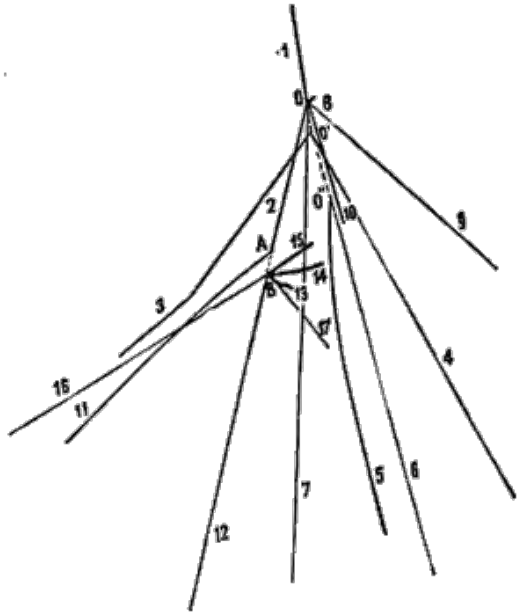


Сейчас

- ~2000 учёных, ~100 стран
- Суперкомпьютеры, GRID, распределённые вычисления

Пузырьковая камера ОИЯИ





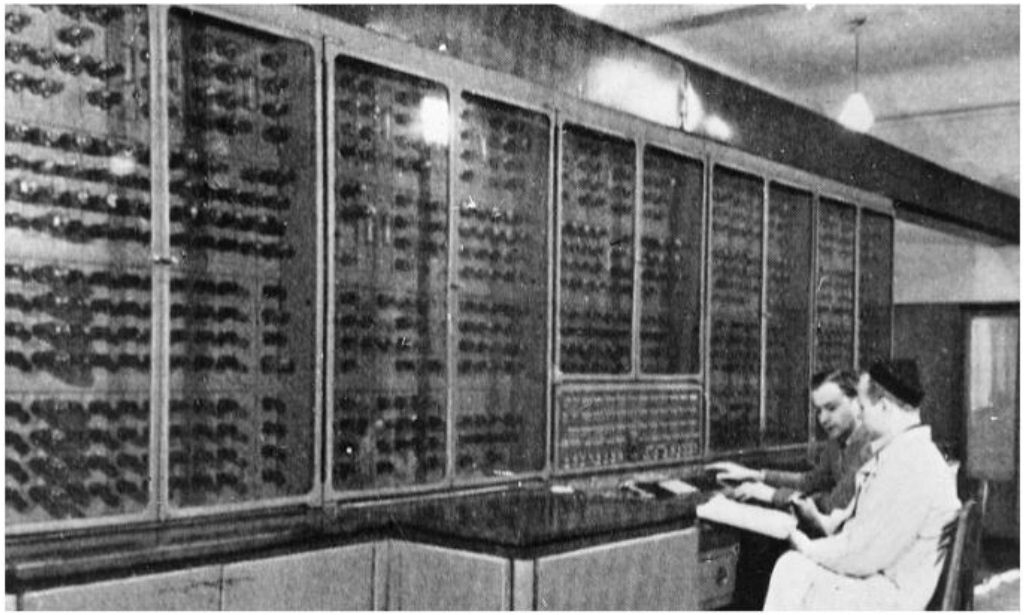
40000 стерео изображений,
просмотренны 2-3-более раз

Компьютеры 1950х

Феликс
Операций/с - ?
Цена: 15 Руб
Масса: 4-6 кг

Мерседес 37MS
6-7 операций/с
Цена: ~1000$
Масса: 17-24 кг

Компьютеры 1950х

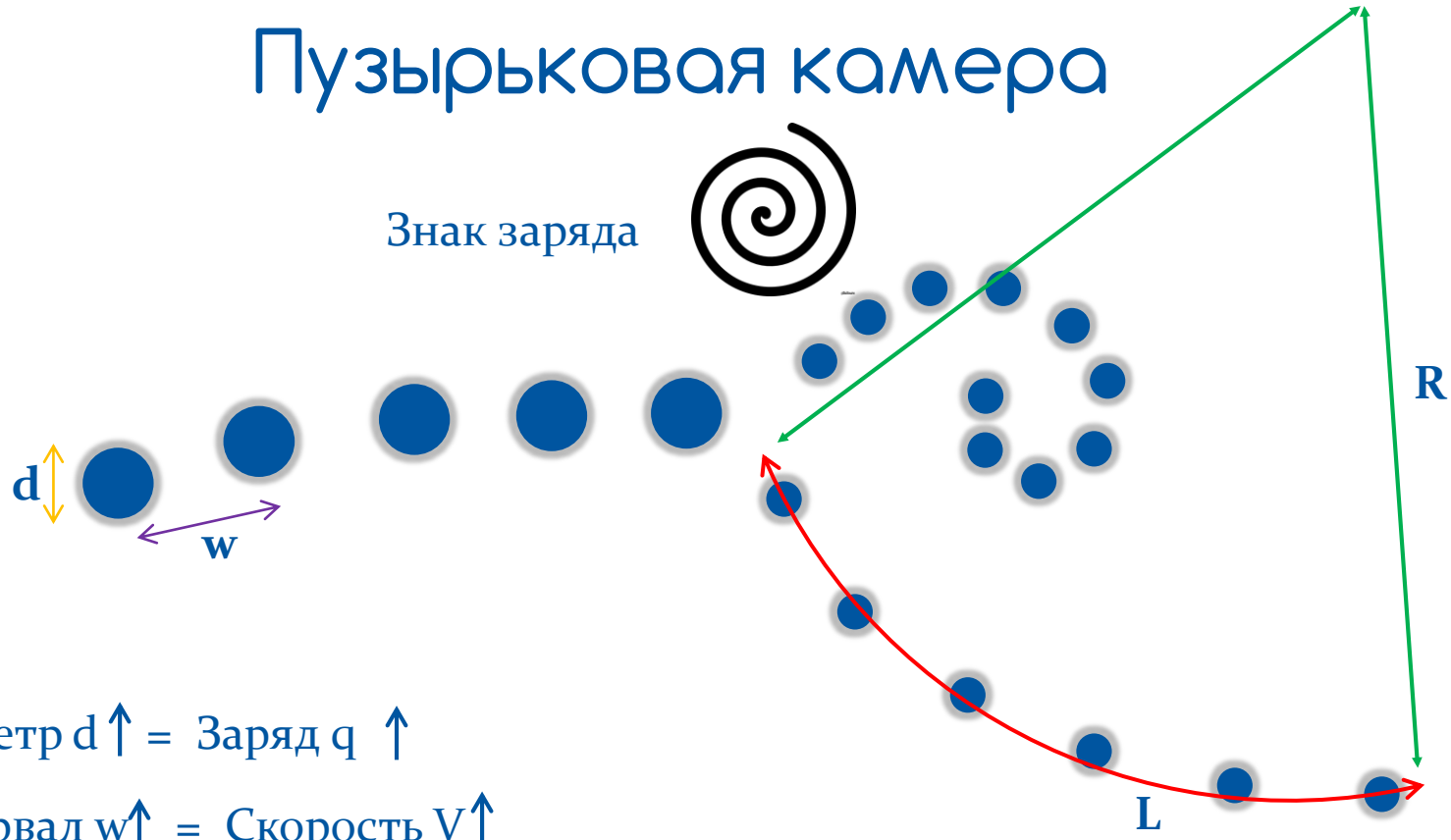


Дубна, ОИЯИ, 1958

Урал-1
Площадь: 75 м2
Количество ламп: 1000
Потребление: 7-10 кВт
Производительность: 100 оп/с

Задачи тех времён

Пузырьковая камера



Диаметр d ↑ = Заряд q ↑

Интервал w↑ = Скорость V↑

Радиус R ↑= m и V↑ , q↓

Длина L ↑= Энергия W↑

Рождение лаборатории

Именно производительность средств обработки экспериментальной информации будет, в конечном счёте, определять “производительность” физических исследований

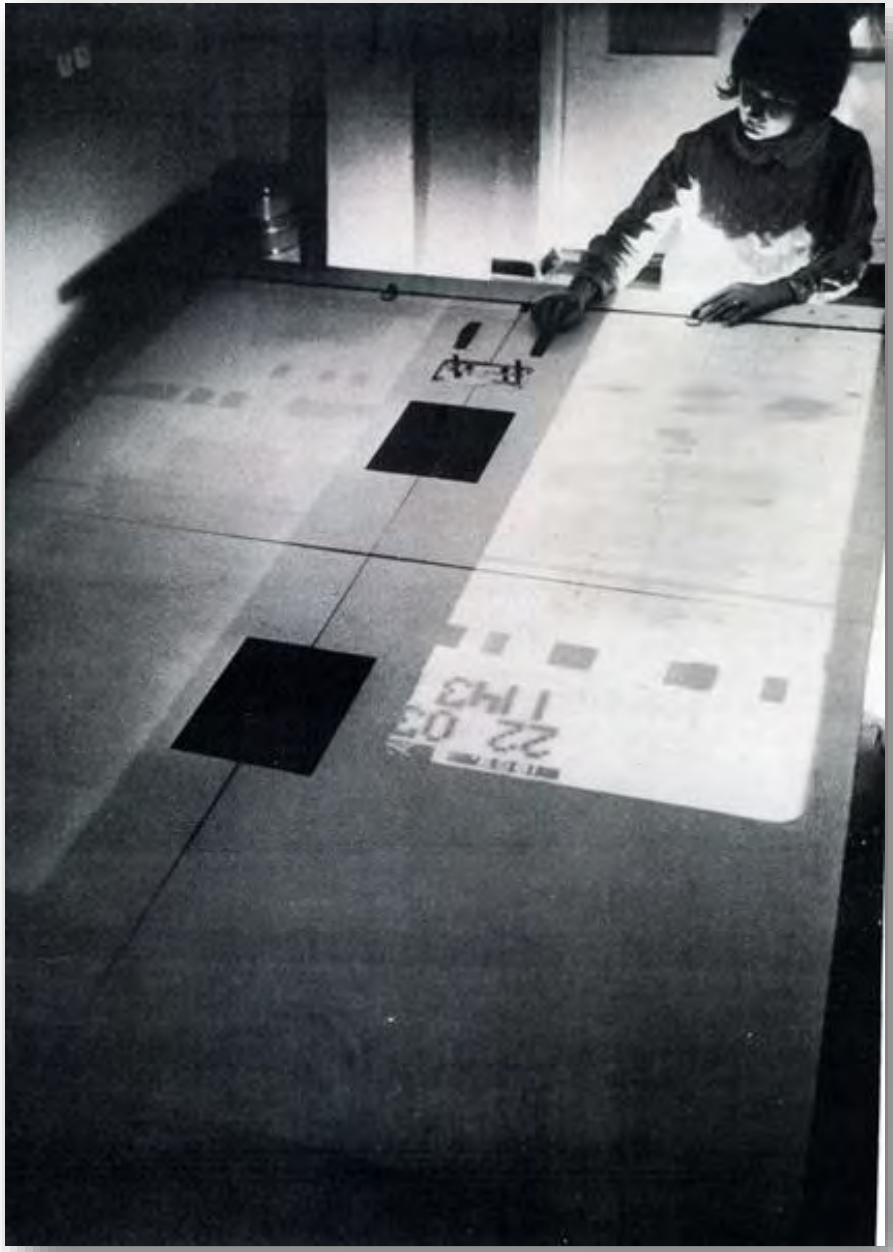
6 августа 1966

Михаил Григорьевич Мещеряков

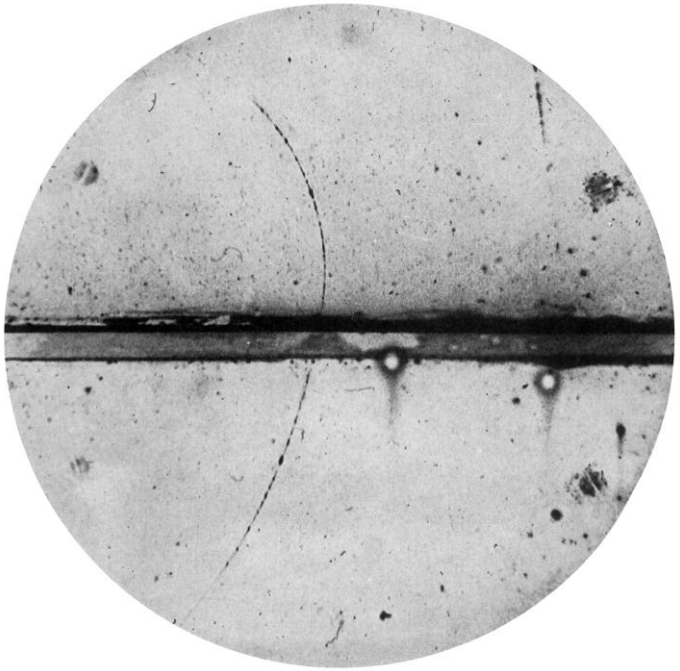
Николай Николаевич Говорун

Лаборатория Вычислительной Техники и Автоматизации

Рождение ЛВТА



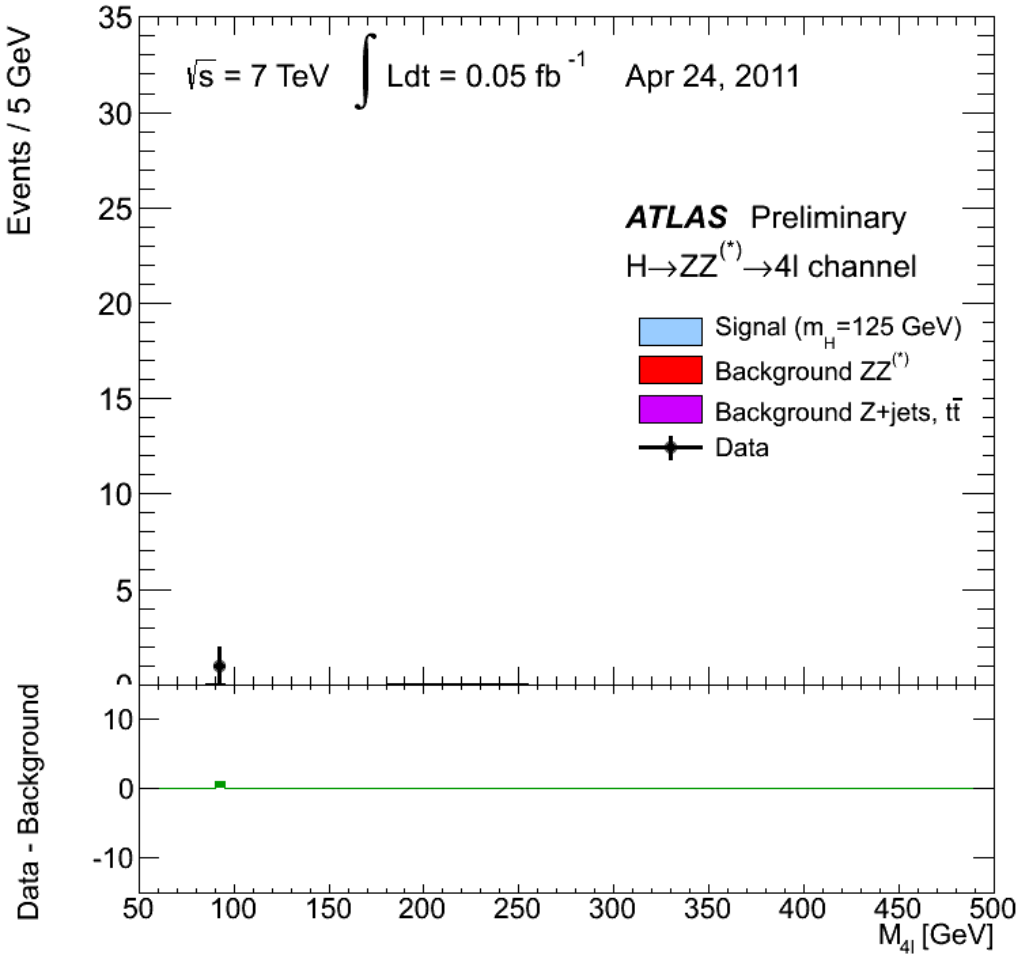
Как делаются открытия?



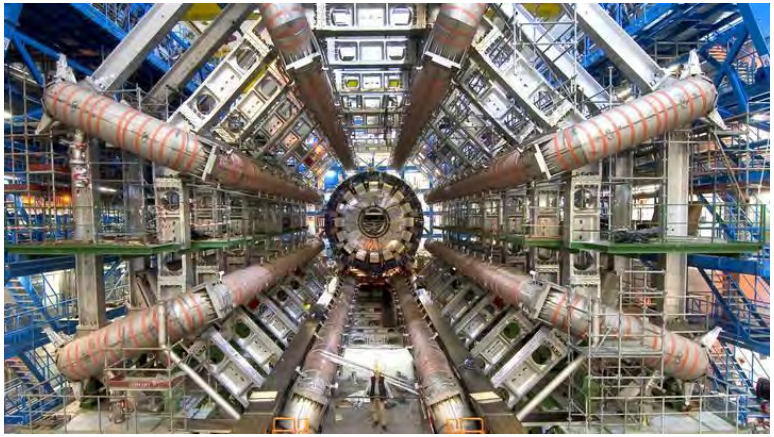
1300 фото,
15 похожи на позитроны

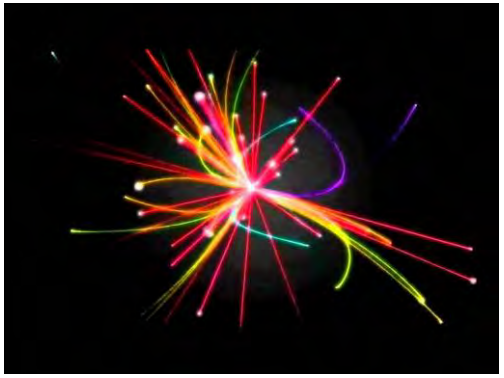
A magnetic field deflect the particle trajectory and the insertion of a lead plate allow to determine the direction of its movement. The deflection indicates a positive particle which could not be a proton whose path would be much shorter. *C.D. Anderson, Physical Review 43, 491 (1933)*.

Как делаются открытия?



Эксперимент



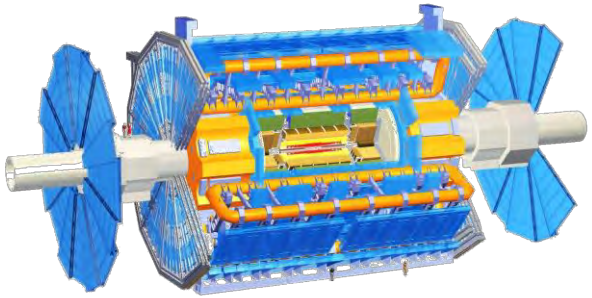


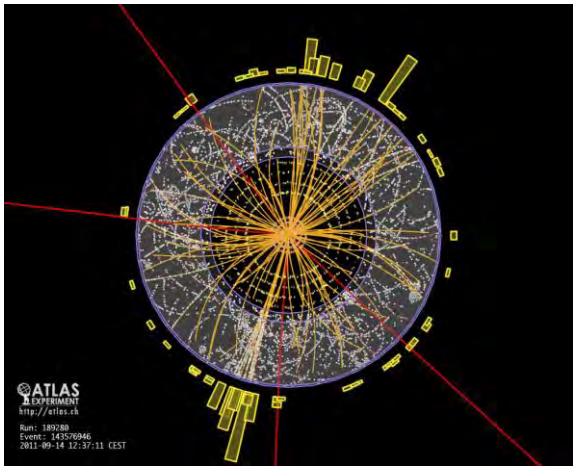
Сырые данные

```
CH0:0.001;
CH2:0.14;
CH4:0.34;
…
CH98039232:0.08;
```

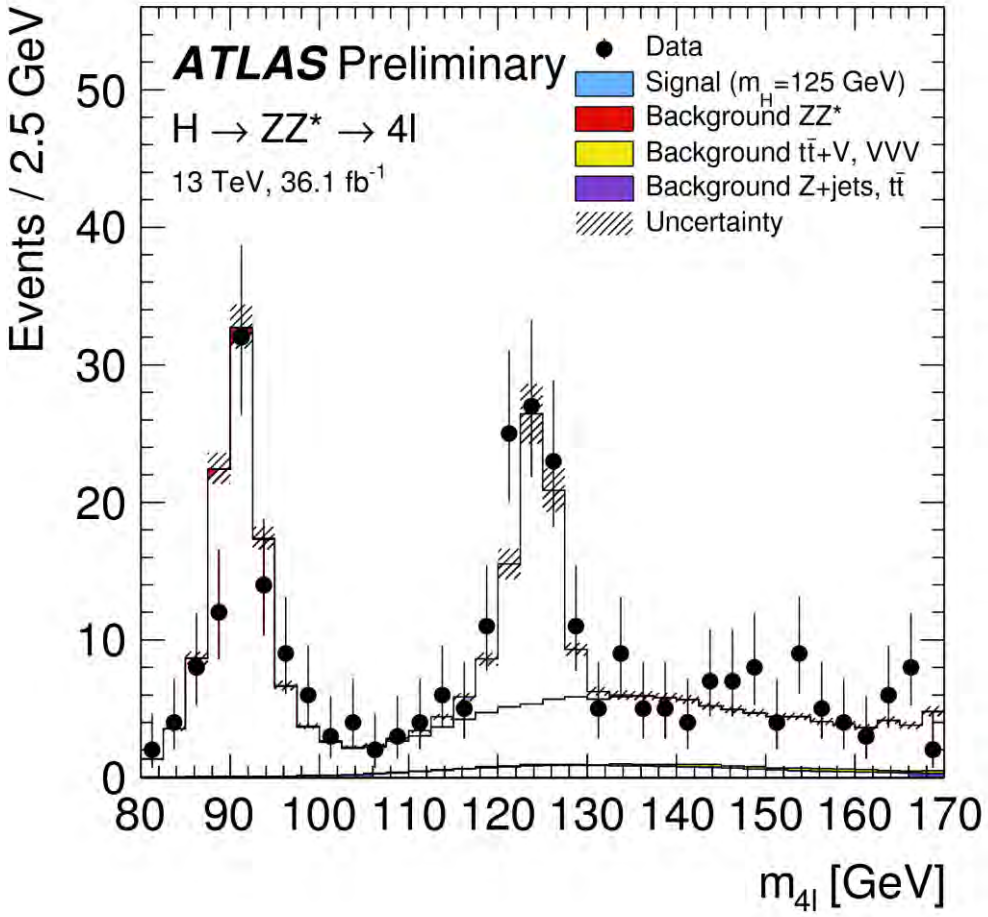
Реконструкция

```
CH0:0.001;
CH2:0.14;
CH4:0.34;
…
CH98039232:0.08;
```

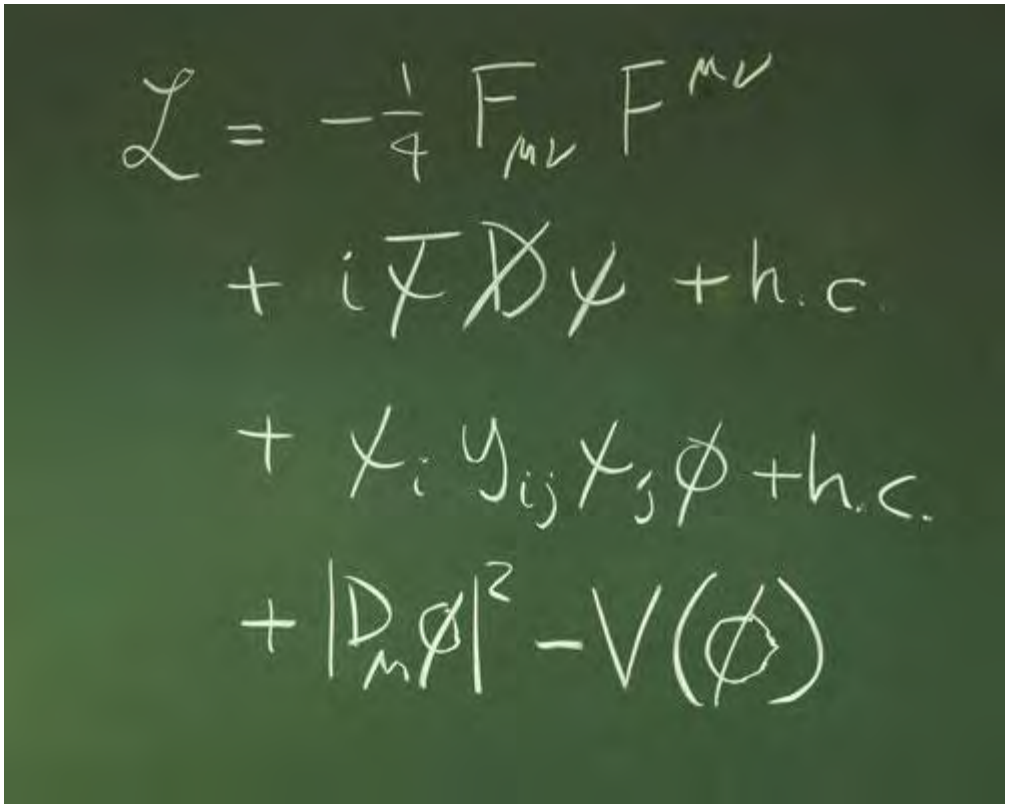
+  + Алгоритм реконструкции

= 

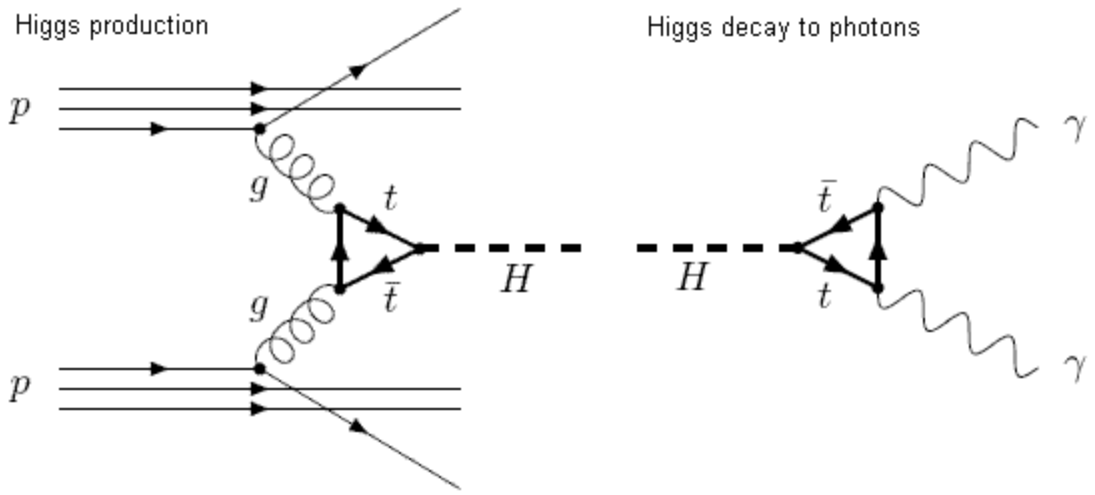
Этого не достаточно



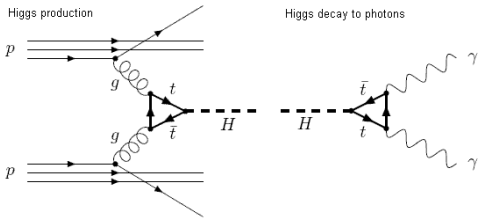
Модель


$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$
$+ i \bar{\chi} \not{D} \chi + h.c.$
$+ \chi_i y_{ij} \chi_j \phi + h.c.$
$+ |D_\mu \phi|^2 - V(\phi)$

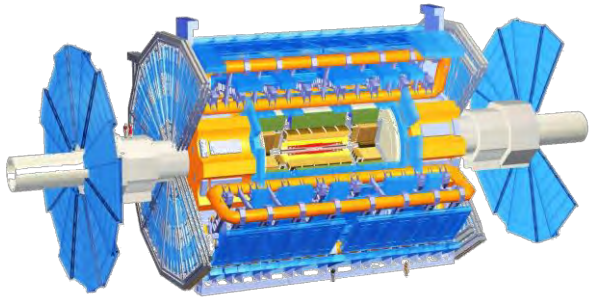
Событие


Higgs production
Higgs decay to photons
p
p
g
g
t
$\bar{t}$
H
H
$\bar{t}$
t
γ
γ

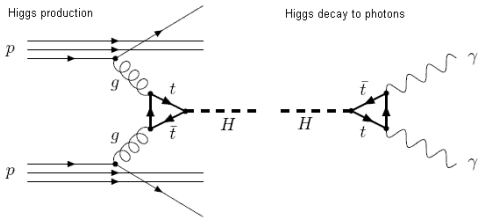
Симуляция



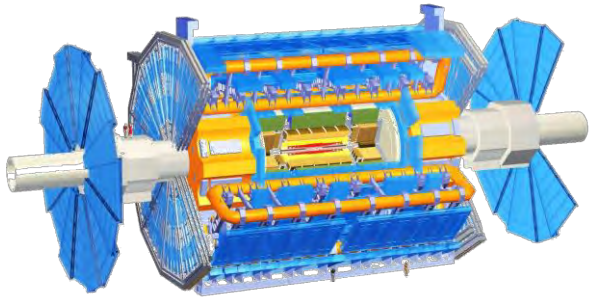
+



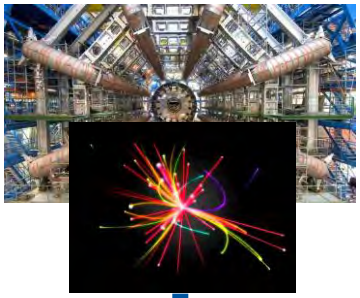
Симуляция



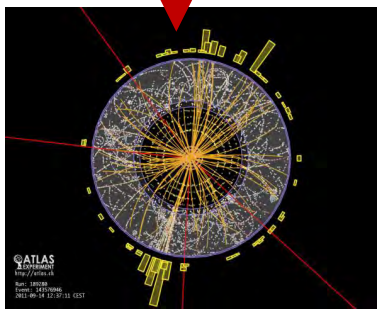
+

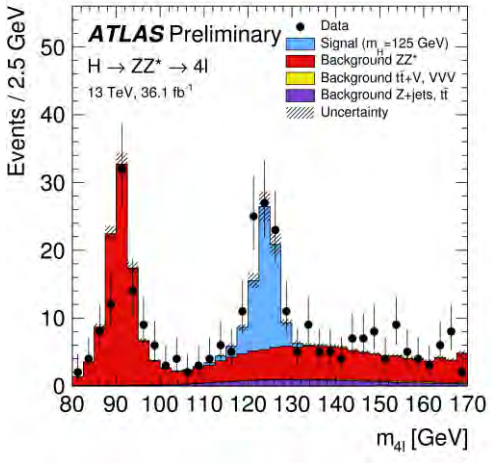


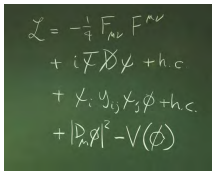
```
CH0:0.001;
CH2:0.14;
CH4:0.34;
…
CH98039232:0.08;
```

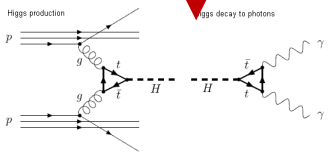


CH0:0.001;
CH2:0.14;
CH4:0.34;
…
CH98039232:0.08;

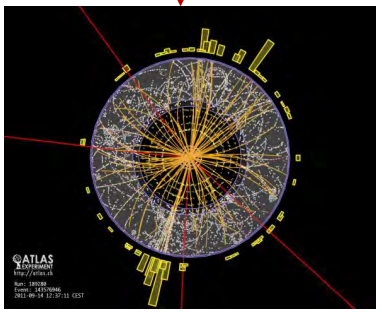








CH0:0.001;
CH2:0.14;
CH4:0.34;
…
CH98039232:0.08;



BigData

Volume	Velocity	Variability
Более 1 ExaByte	До 4 GigaByte/s	Разные энергии, Разные детекторы, Разные состояния

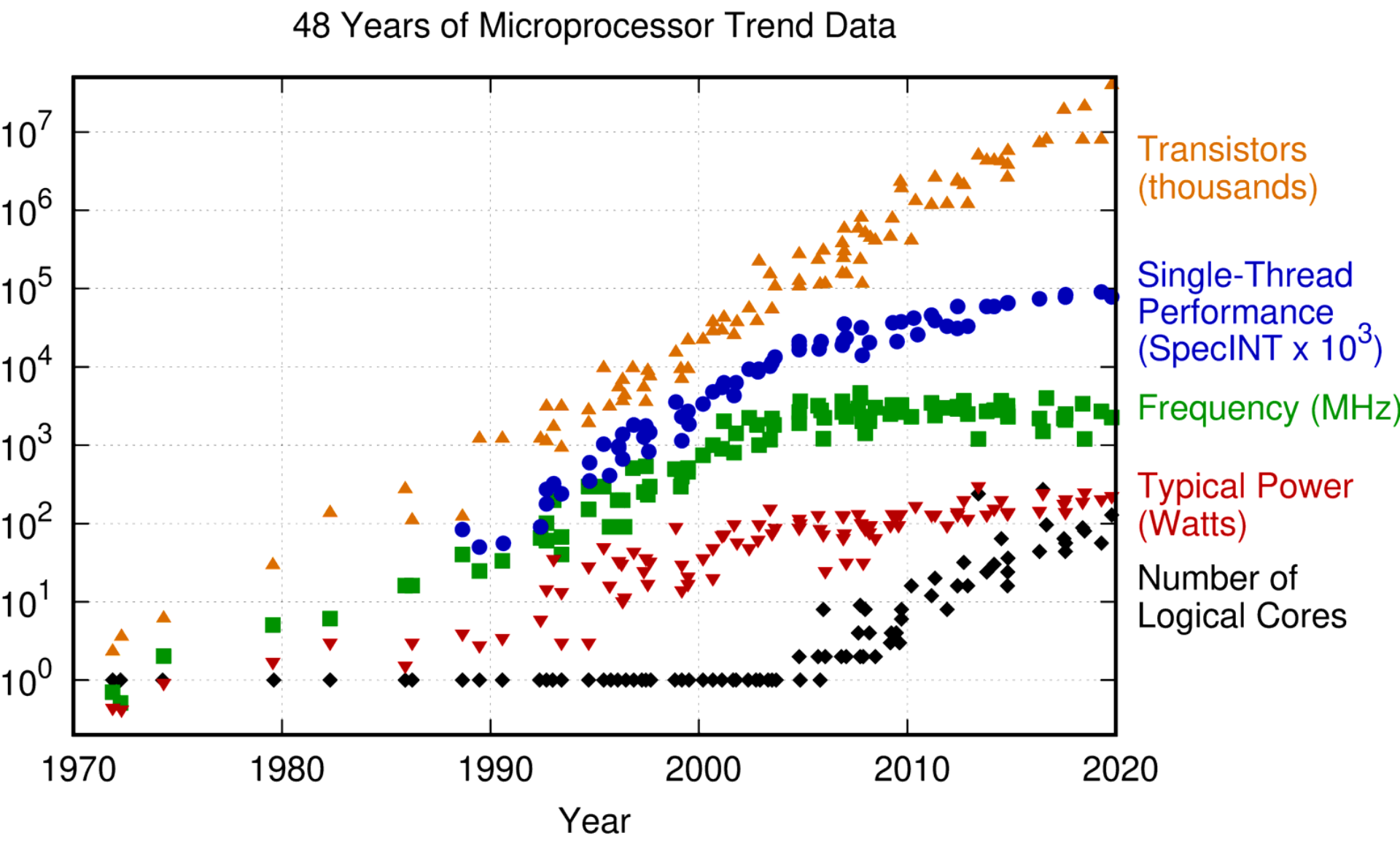
1 Megabyte – фотография

1 Gigabyte – фильм

1 Terabyte – ёмкость стандартного ПК

1 Petabyte – 1000 Terabyte

1 Exabyte – 1000 Petabyte



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2019 by K. Rupp

Параллельность

Суперкомпьютеры

Распределённость

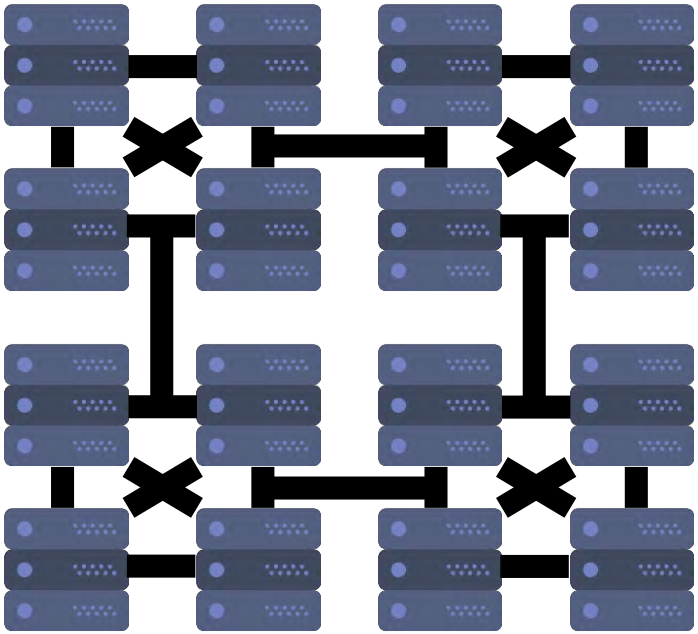
Грид,
кластера

Эффективность

Новые алгоритмы

Суперкомпьютер

Одна большая задача



Результат

Что такое суперкомпьютер?

1985 – Cray 2

1.9 GFLOPS
2 GB RAM
Мощность 200kW
Цена– 32.000.000$
NASA, US Defense …

2010 – iPhone 4

1.6 GFLOPS
512 MB RAM
Мощность 5.4 W
Цена– 700 $
Миллионы были проданы

Цена vs производительность

System	MegaFLOP/s	Inflation Adjusted Cost (2010 $)	Cost per MegaFLOP/s
CDC 6600	1 (Megaflop)	$49 million	$49 million
Cray-2	1,000 (1 Gigaflop)	$32 million	$32,000
iPad-2	1,650 (1.65 GFLOP/s)	$699 (64GB storage, no optional 3G plan, no cover)	$0.42
Lenovo W510 laptop	24,239 (24.23 GFLOP/s)	$2,100 (i7 920, 2.0GHz quad core, 16GB RAM)	$0.086
Generic business desktop	39,675 (39.67 GFLOP/s)	$1700 (i5, 2.66GHz, quad core, 8GB RAM)	$.0428
Hydra-1 (personal project)	122,680 (122.68 GFLOP/s)	$10,000 est. (2x Xeon 5690, 3.46 GHz, 6-core, 24GB RAM)	$.0815
ASCI Red	1,000,000 (1 TeraFLOP/s)	$76 million	$76
Roadrunner	1 billion (1 PetaFLOP/s)	$101 million	$0.10
K Computer	10 billion (10 PFLOPS/s)	$1.25 billion (to design and build)	$0.13

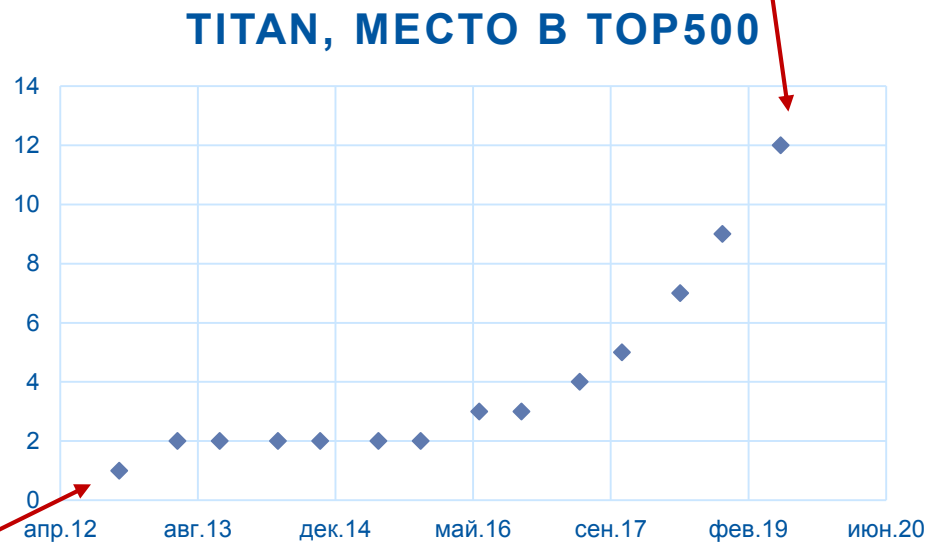
Что такое Top500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,299,072	415,530.0	513,854.7	28,335
2	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371

Суперкомпьютер Titan

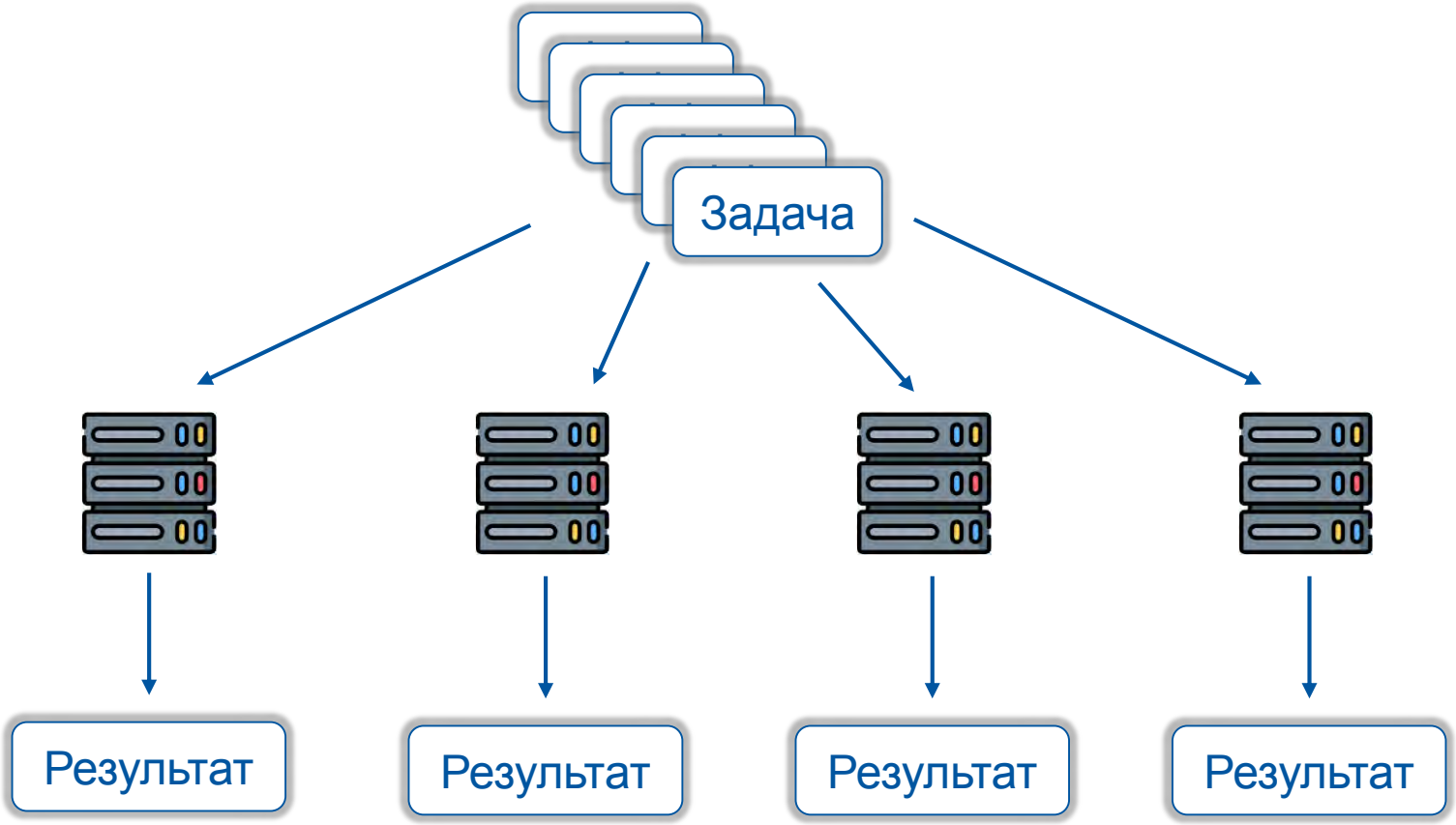
Производительность – 17,6 PFlops
Пиковая произв-ть – 27 Pflops
Ядра - 560,640
RAM - 710,144 GB
Мощность 8.2 MW
Цена – 100M $ + 9M $ в год

Вывод из эксплуатации



Ввод в эксплуатацию

Распределённые вычисления

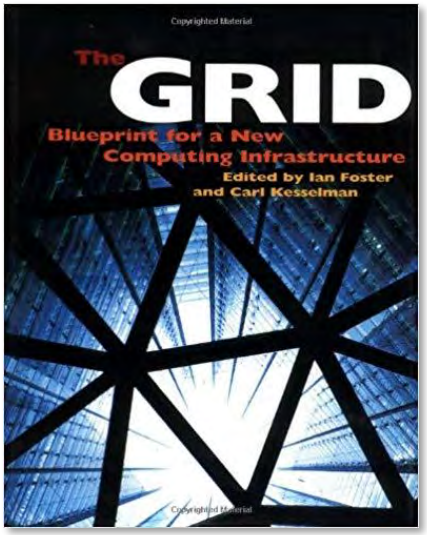


Грид

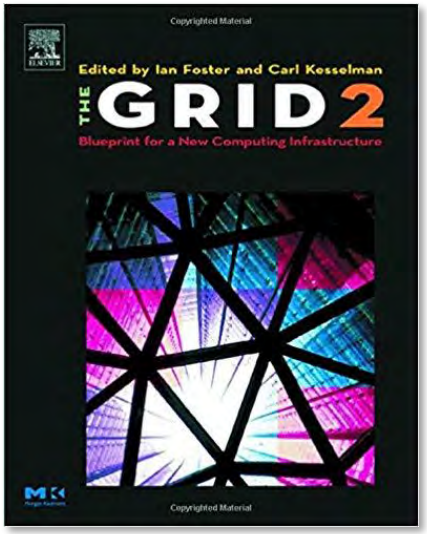
Ян Фостер

Карл Кессельман

1998



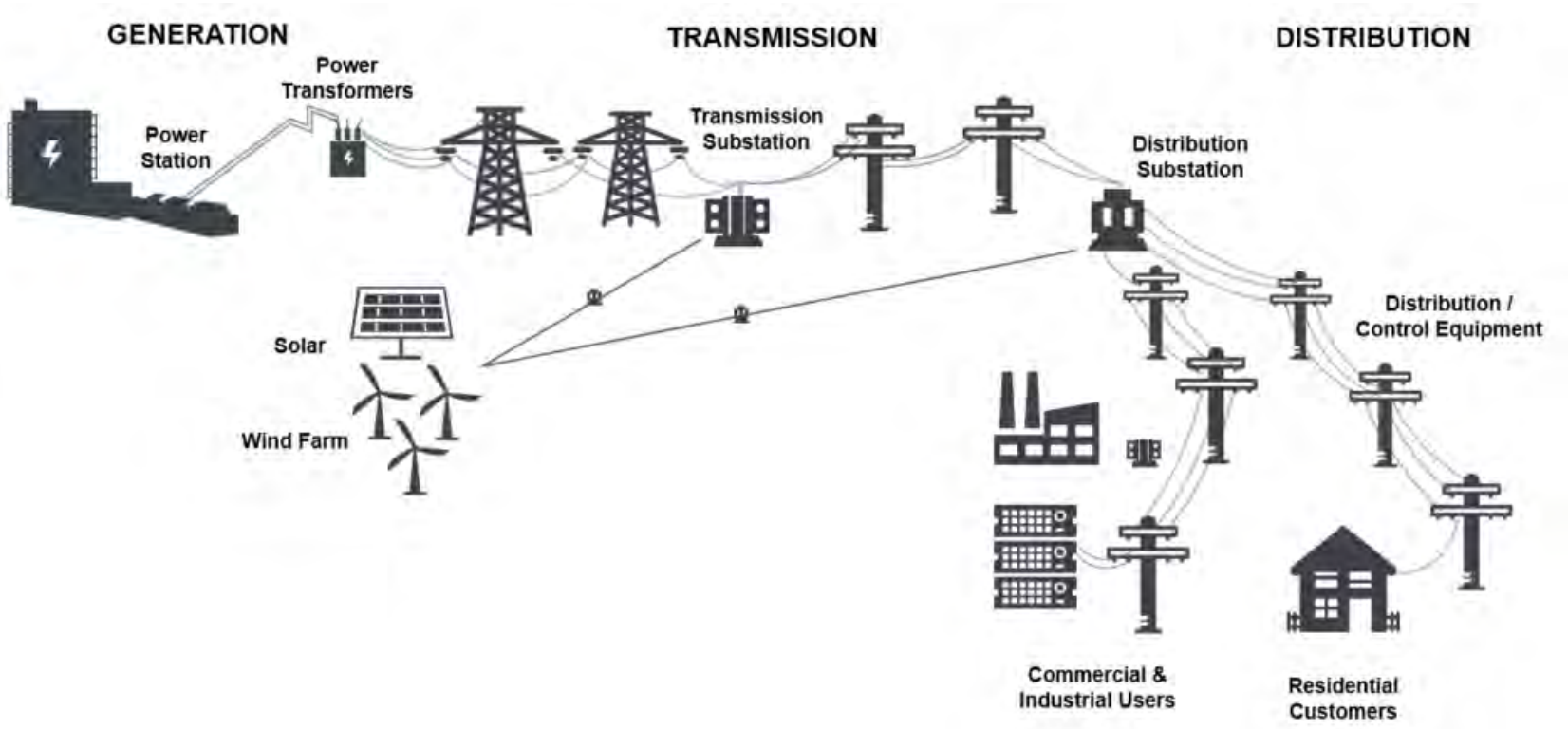
2003



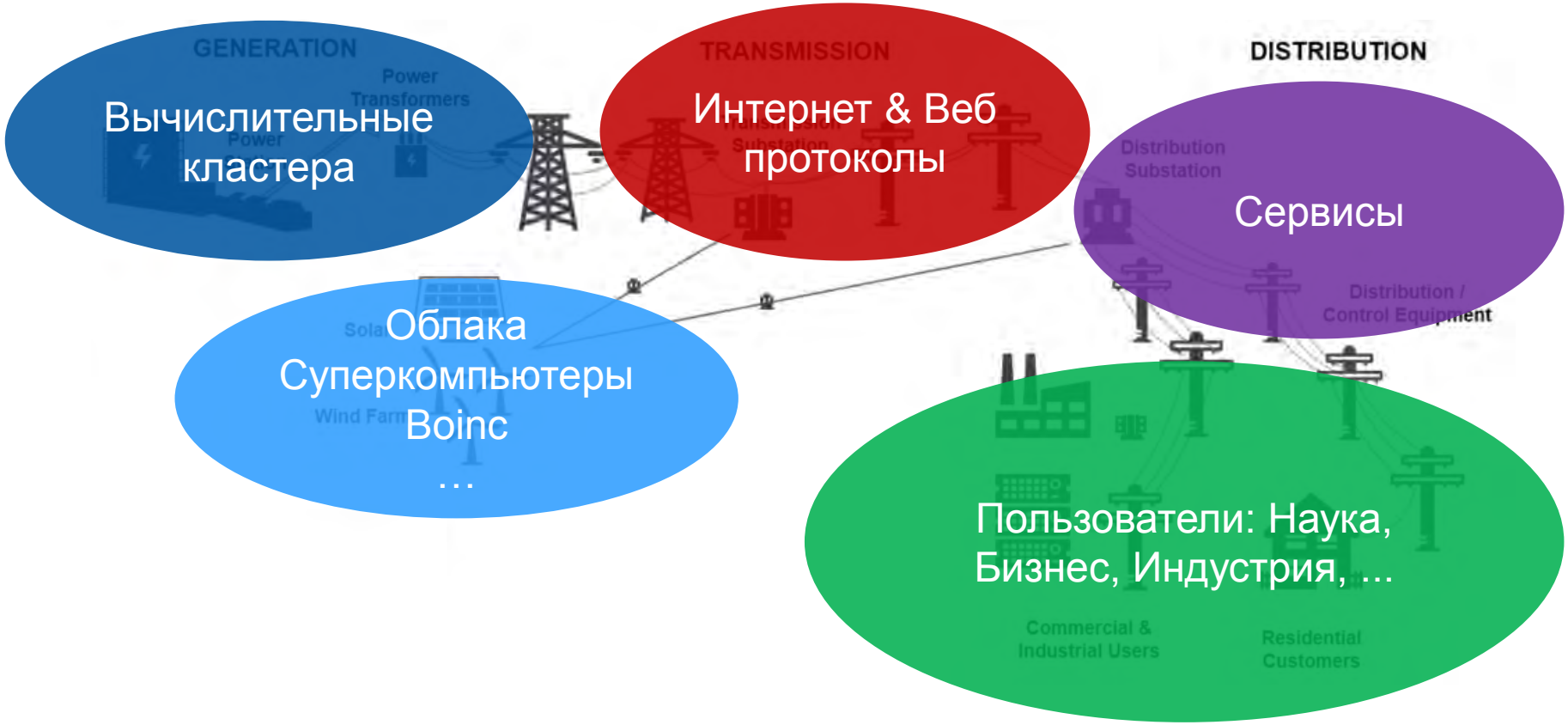
Грид подходит для:

- Передача данных
- Реконструкция
- Долговременное хранилище
- Анализ
- Симуляция

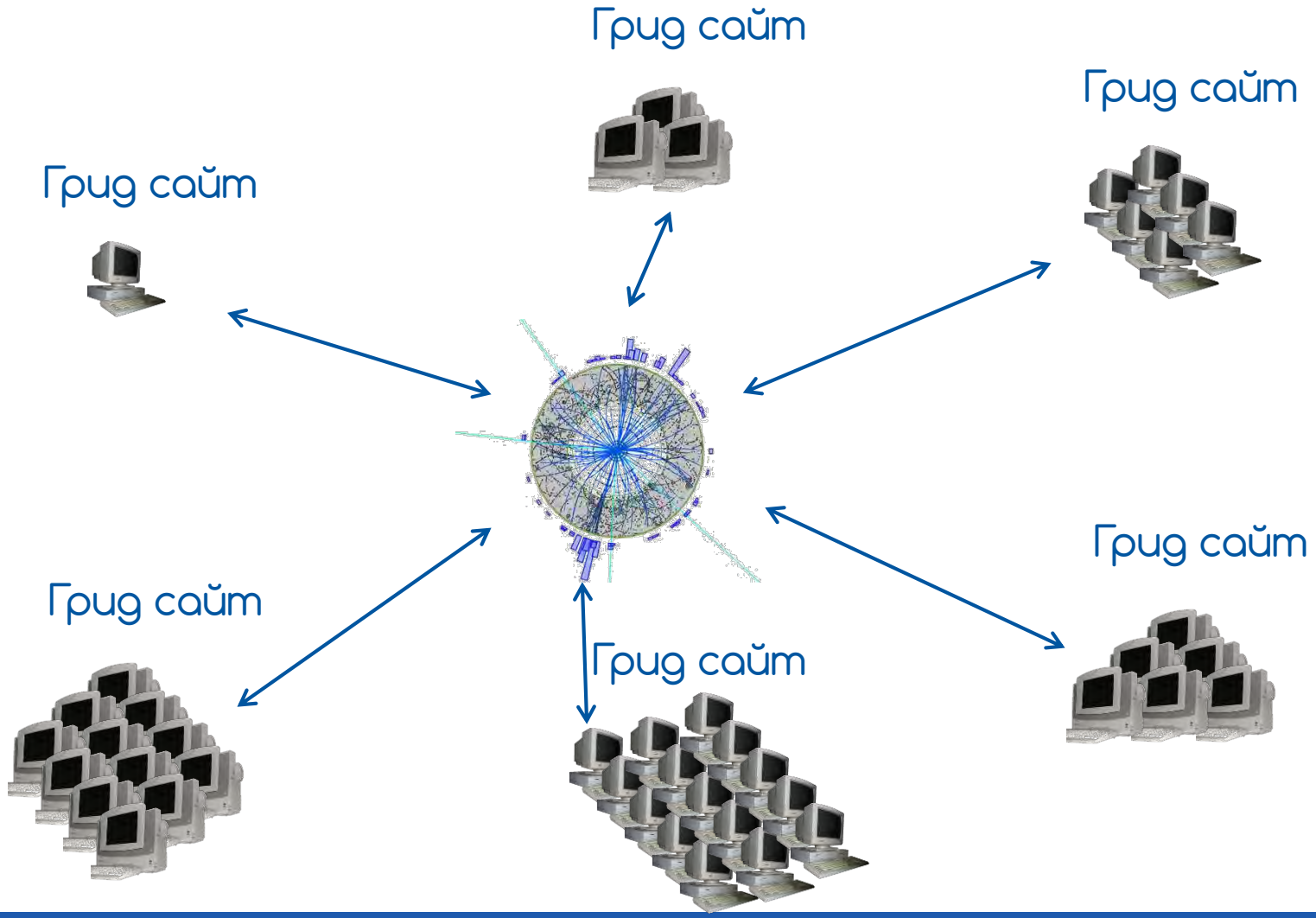
Power Grid


GENERATION
TRANSMISSION
DISTRIBUTION
Power Transformers
Power Station
Transmission Substation
Distribution Substation
Solar
Wind Farm
Distribution / Control Equipment
Commercial & Industrial Users
Residential Customers

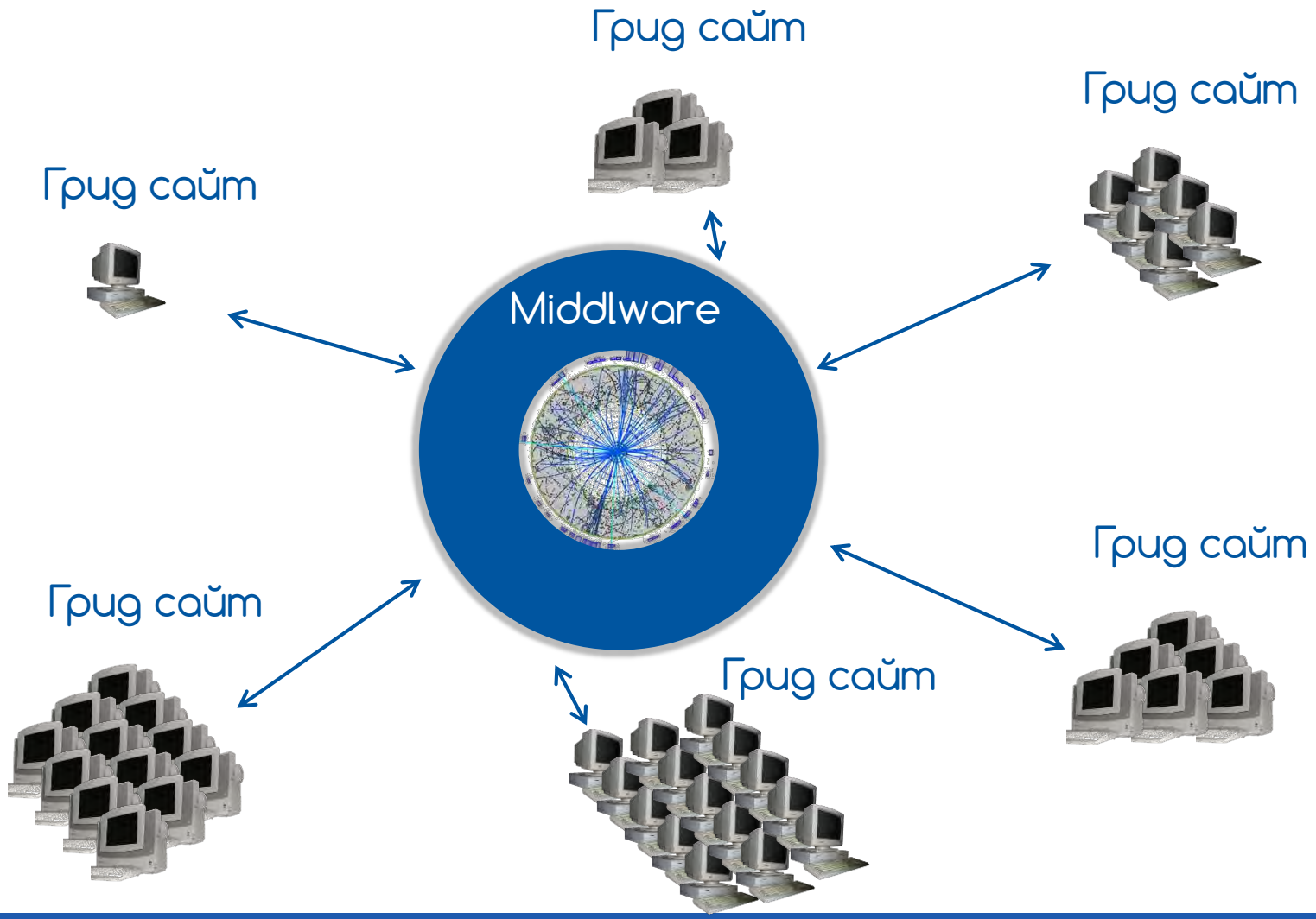
Грид компьютинг



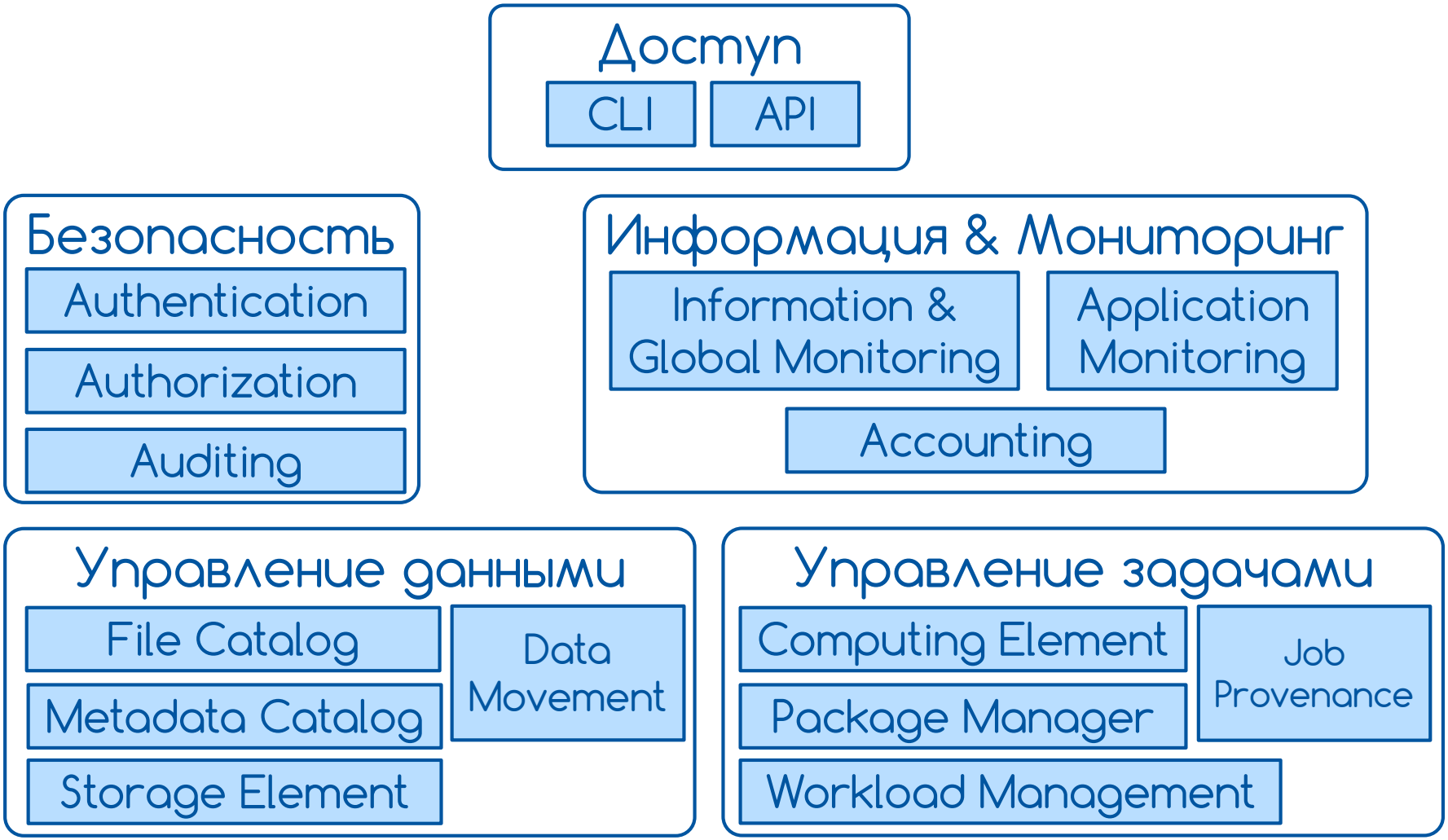
Простейший грид



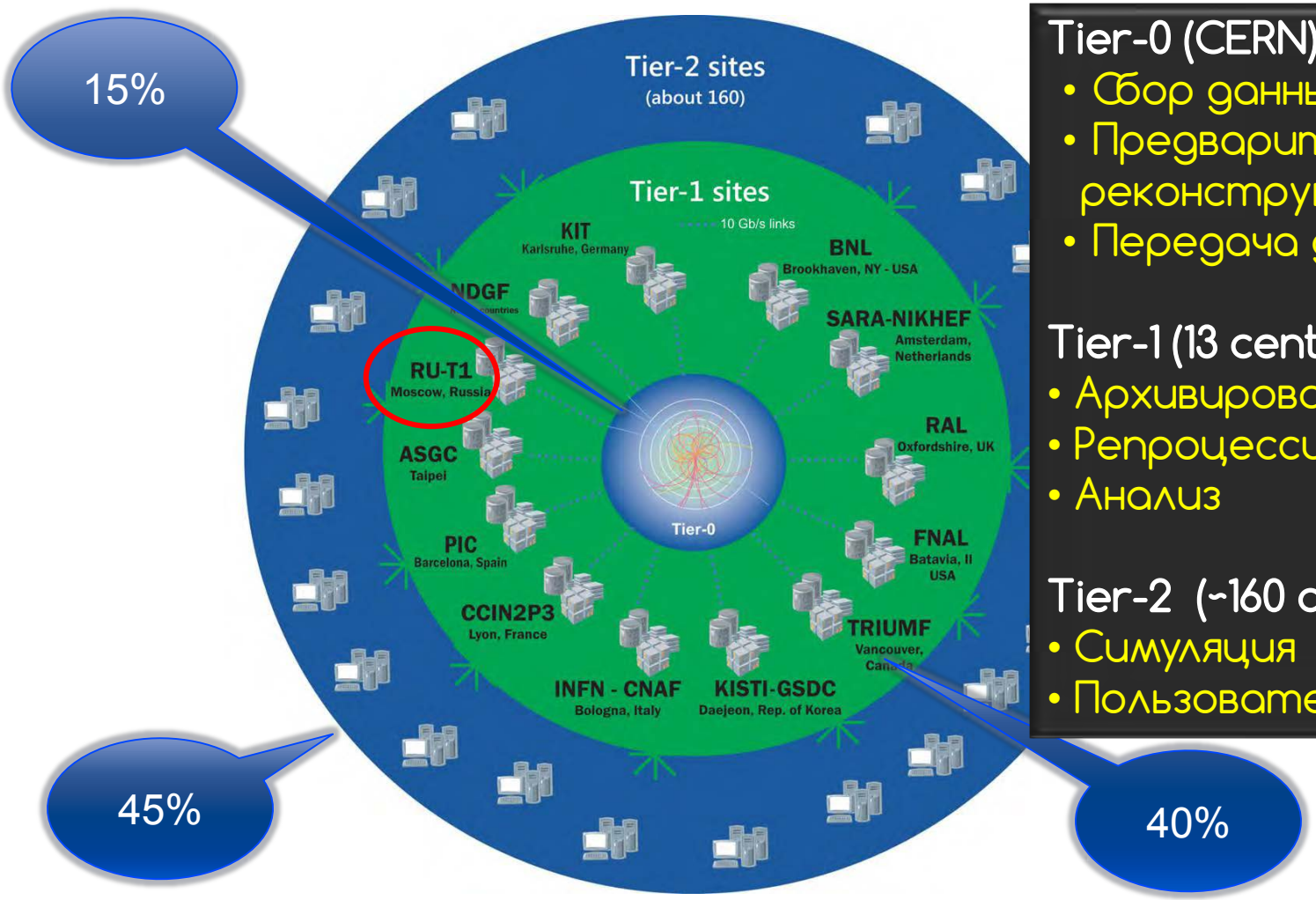
Middleware


Грид сайт
Грид сайт
Грид сайт
Middlware
Грид сайт
Грид сайт
Грид сайт

Grid Middleware



Иерархя в грид



Tier-0 (CERN): (15%)
- Сбор данных
- Предварительная реконструкция
- Передача данных

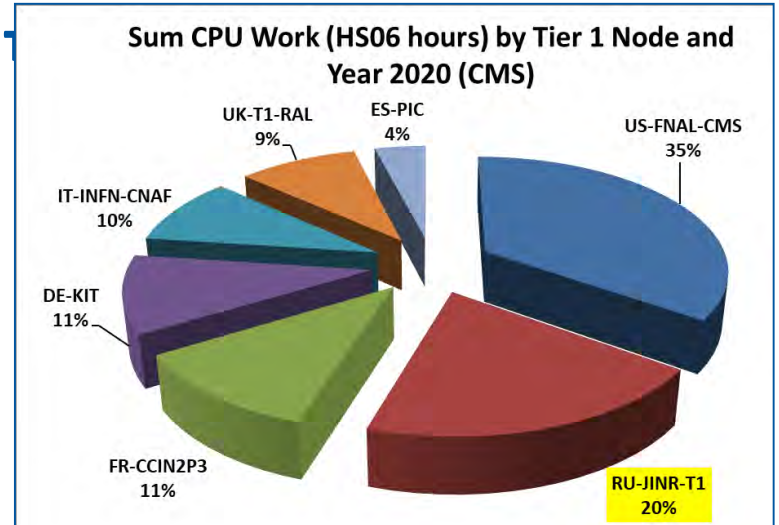
Tier-1 (13 centers): (40%)
- Архивирование
- Репроцессинг
- Анализ

Tier-2 (~160 centers): (45%)
- Симуляция
- Пользовательский анализ

Grid в ОИЯИ

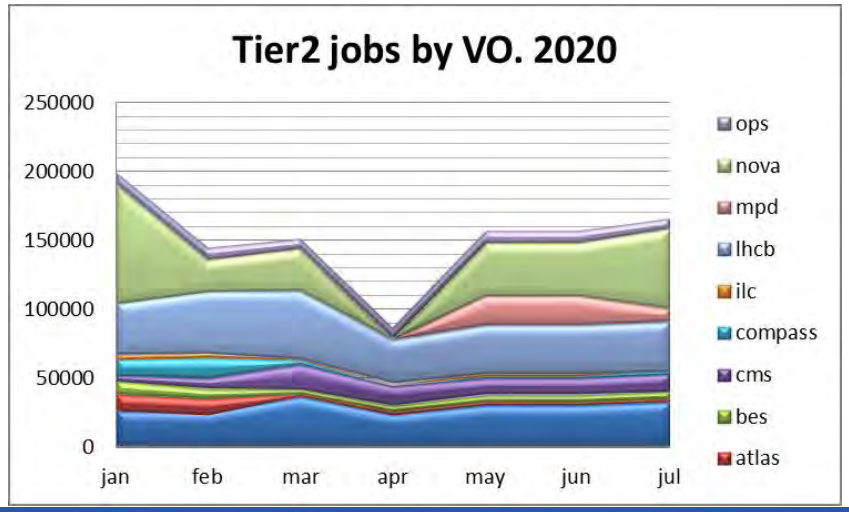
Эксперимент

BM@N,
MPD,
CMS,
ATLAS,
ALICE,
LHCb,
COMPASS,
PANDA,
CBM,
STAR,
NOvA,
BESIII,
DIRAC,
OPERA
NEMO
Mu2e,
NUCLON,
HONE,
BIOMED



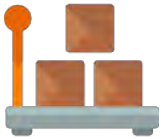
JINR Tier1

- **второе** место среди Tier1 центров для эксперимента CMS
- **20% от CPU времени всех выполненых задачь** среди всех Tier1 центров CMS
- **13,23 PB** данных передано на Tier1 и **19,58 PB** загружено с него



Сложности

Безопасность



Передача данных

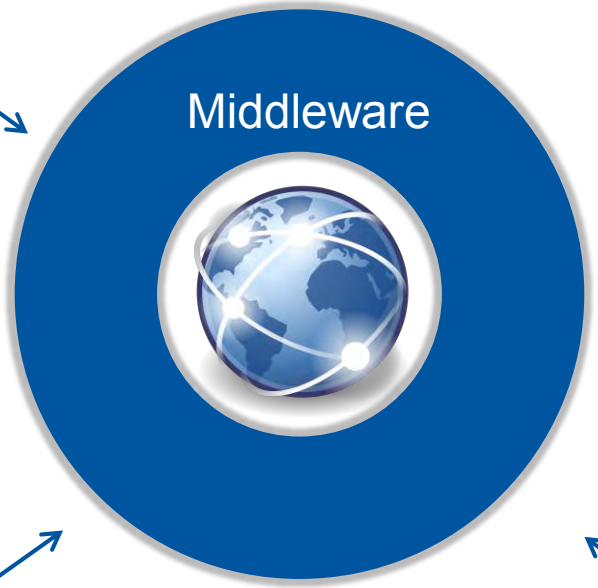
Хранение данных

Аккаунтинг

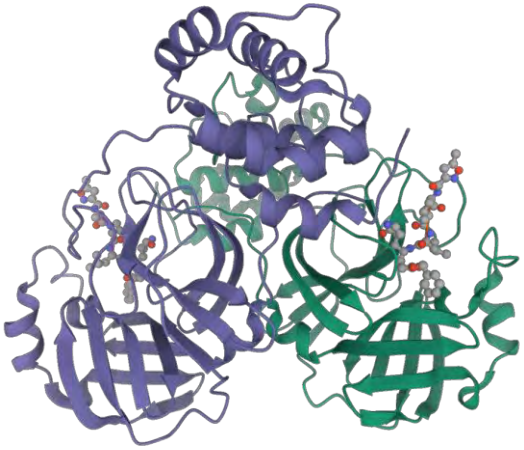
Мониторинг

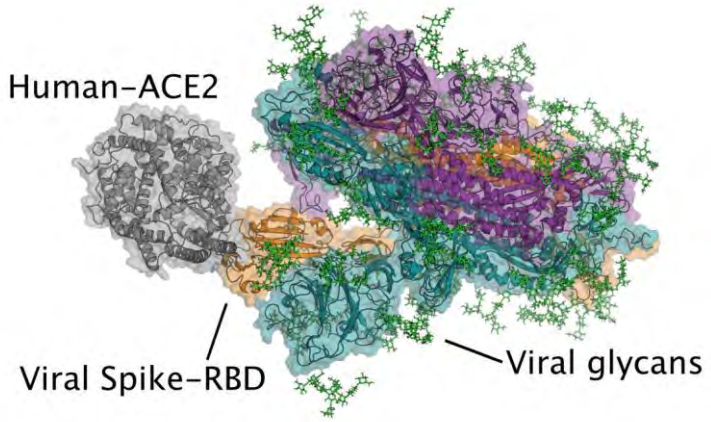
Упраление нагрузкой

Грид сегодня



Folding@HOME





Team: Joint Institute for Nuclear Research

Date of last work unit	2020-09-13 07:02:52
Active CPUs within 50 days	4,257
Team Id	265602
Grand Score	27,222,185
Work Unit Count	11,863
Team Ranking	6473 of 255001
Homepage	http://www.jinr.ru/main-en/

Team members

Rank	Name	Credit	WUs
64,603	CLOUD.JINR.ru	12,645,224	5,355
77,918	CLOUD.PRUE.ru	9,453,851	4,175
194,662	CLOUD.IPANAS.az	1,542,618	910
198,800	CLOUD.INP.by	1,465,167	599
224,832	CLOUD.NOSU.ru	1,083,663	395
230,931	CLOUD.INP.kz	1,012,919	413
N/A	CLOUD.INRNE.bg	18,743	16

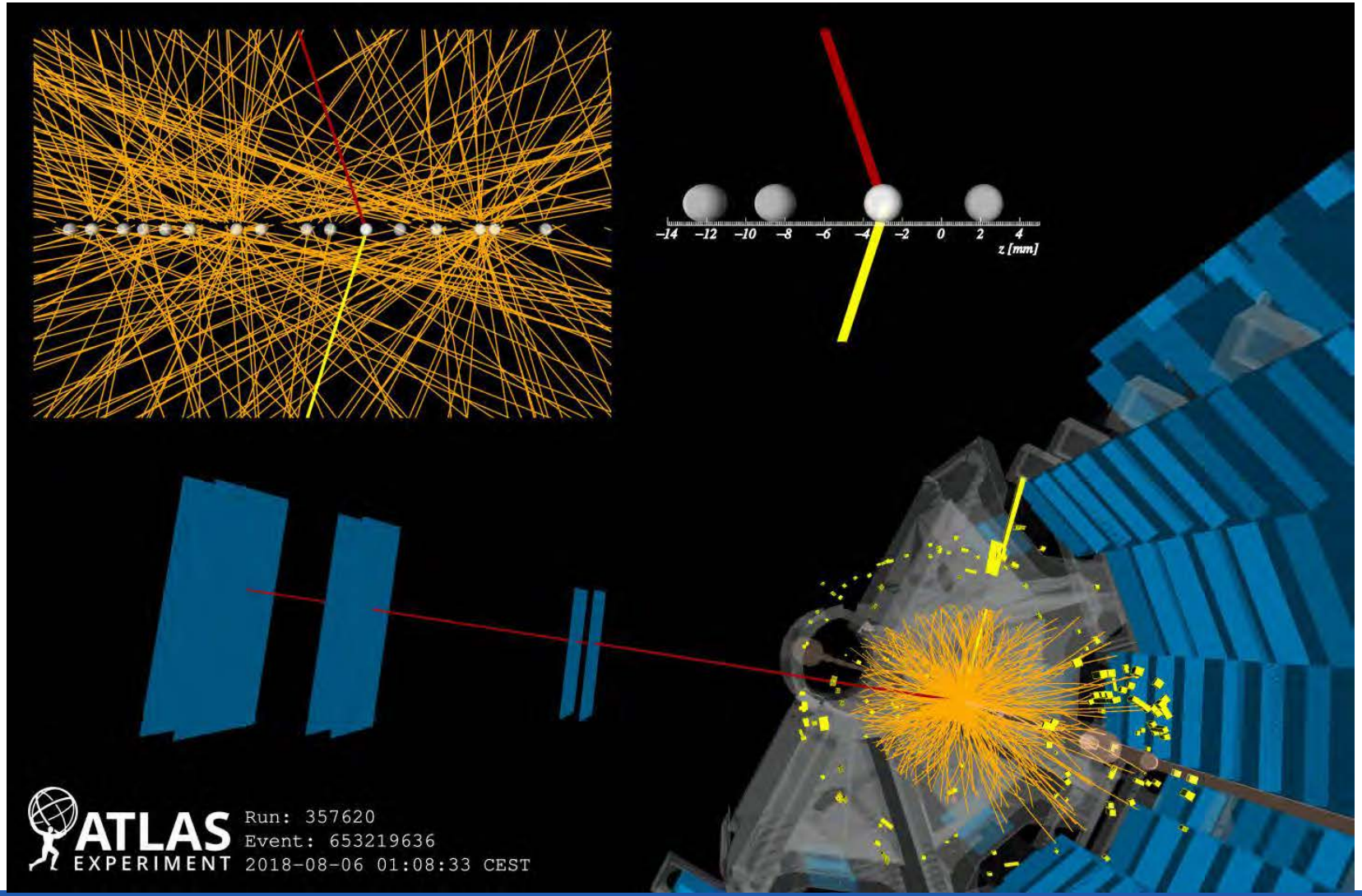
Что с новыми алгоритмами?

Огромное количество кода

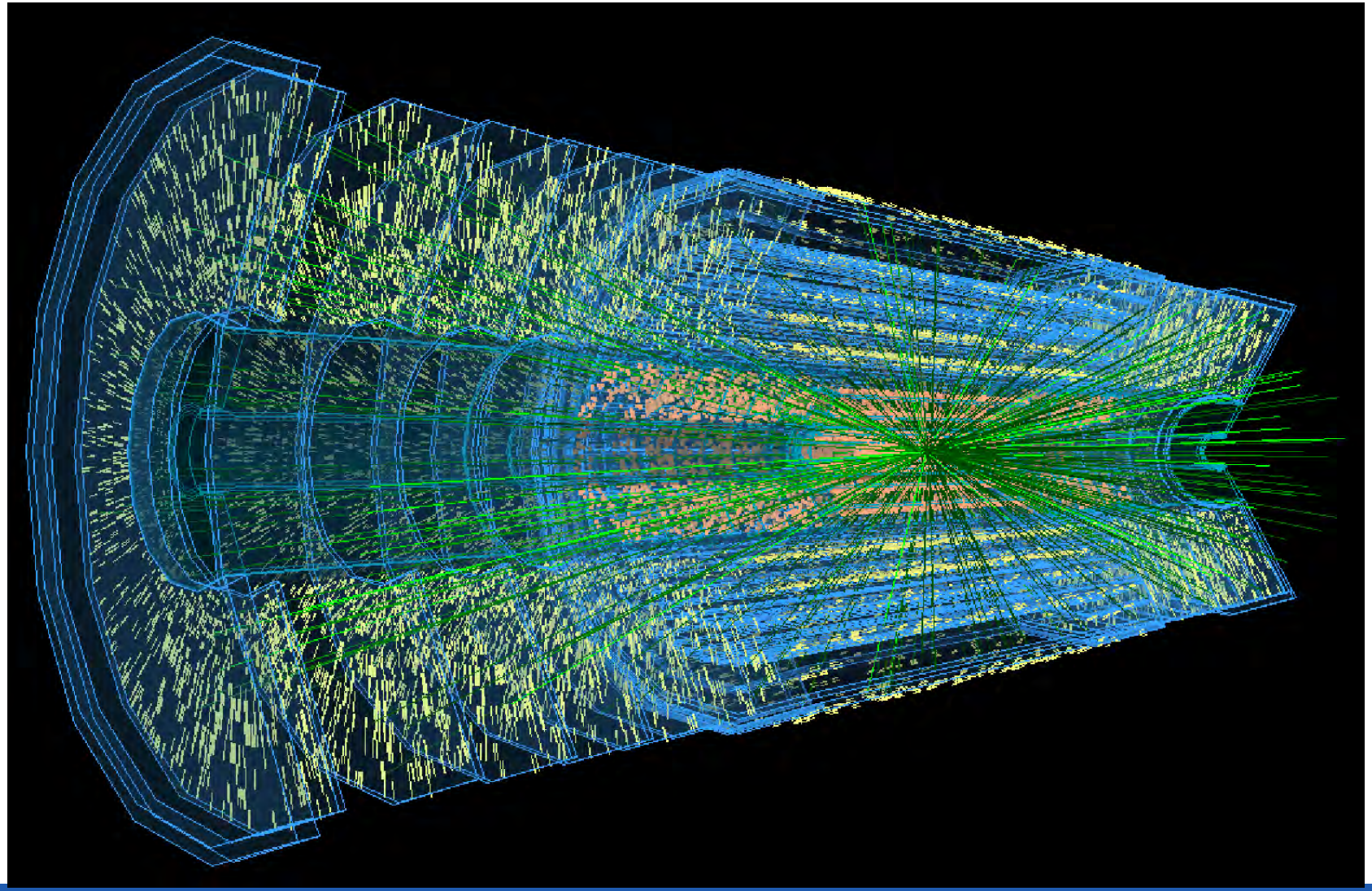
Микро оптимизация не спасёт

Старые подходы будут работать по-старому

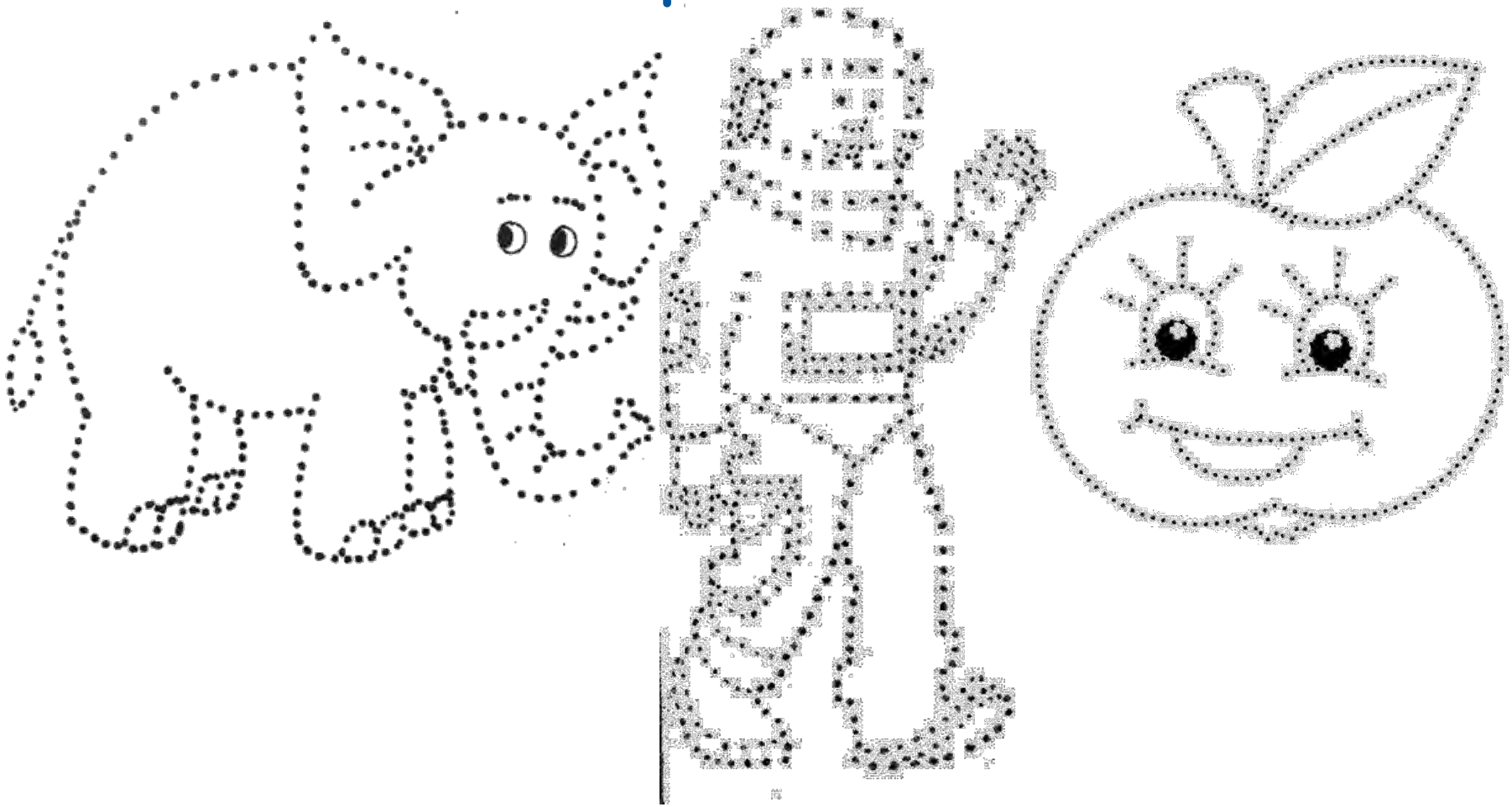
События сейчас



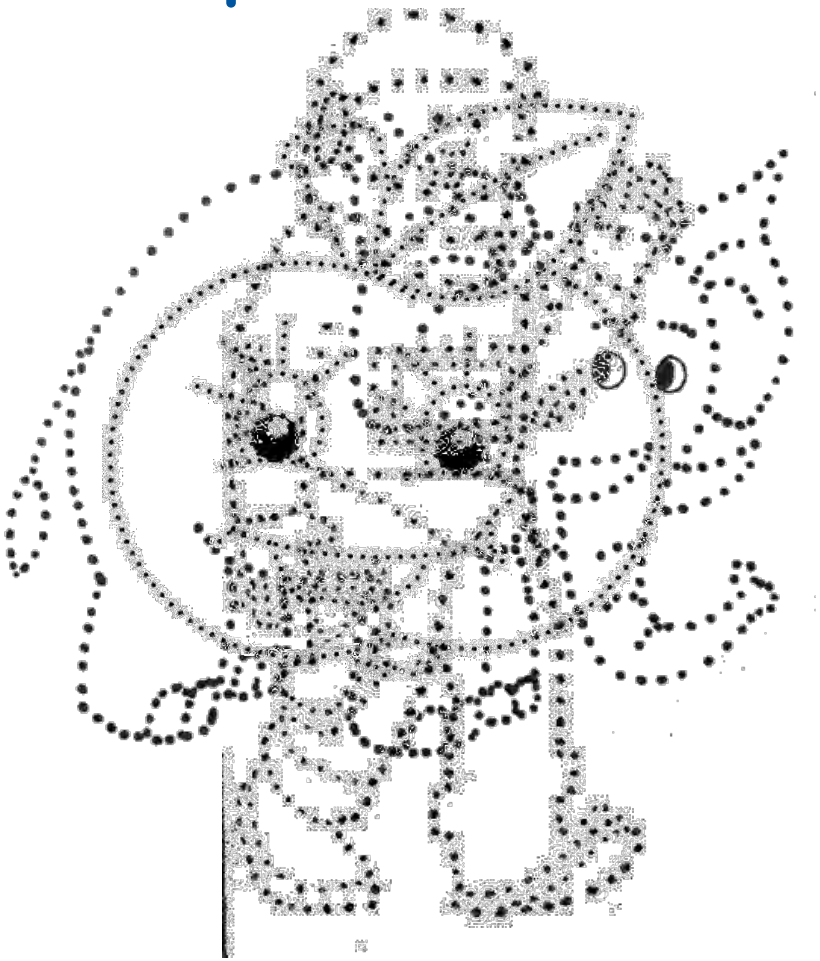
Событие посложнее



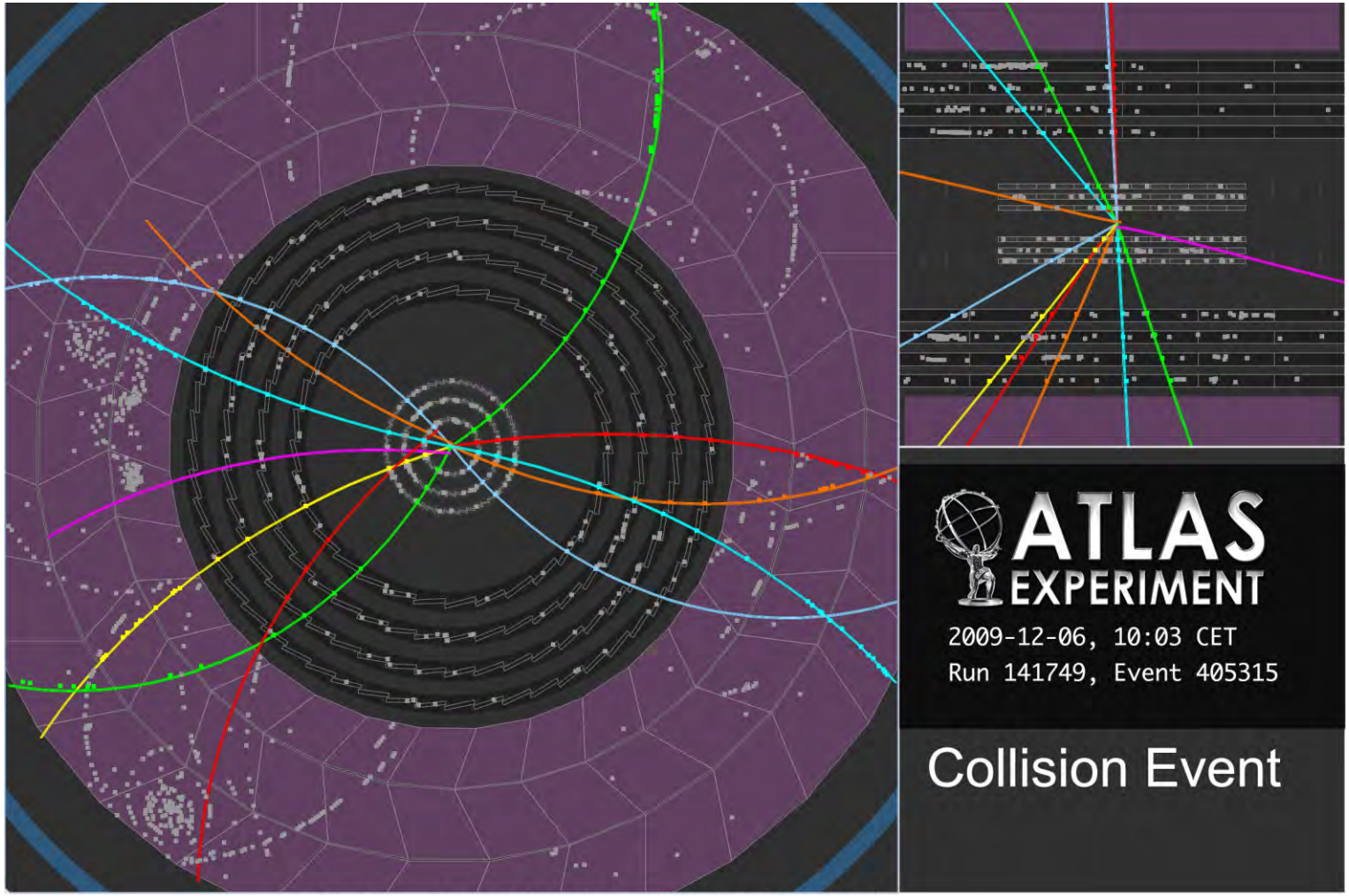
Трэкинг



Трэкинг

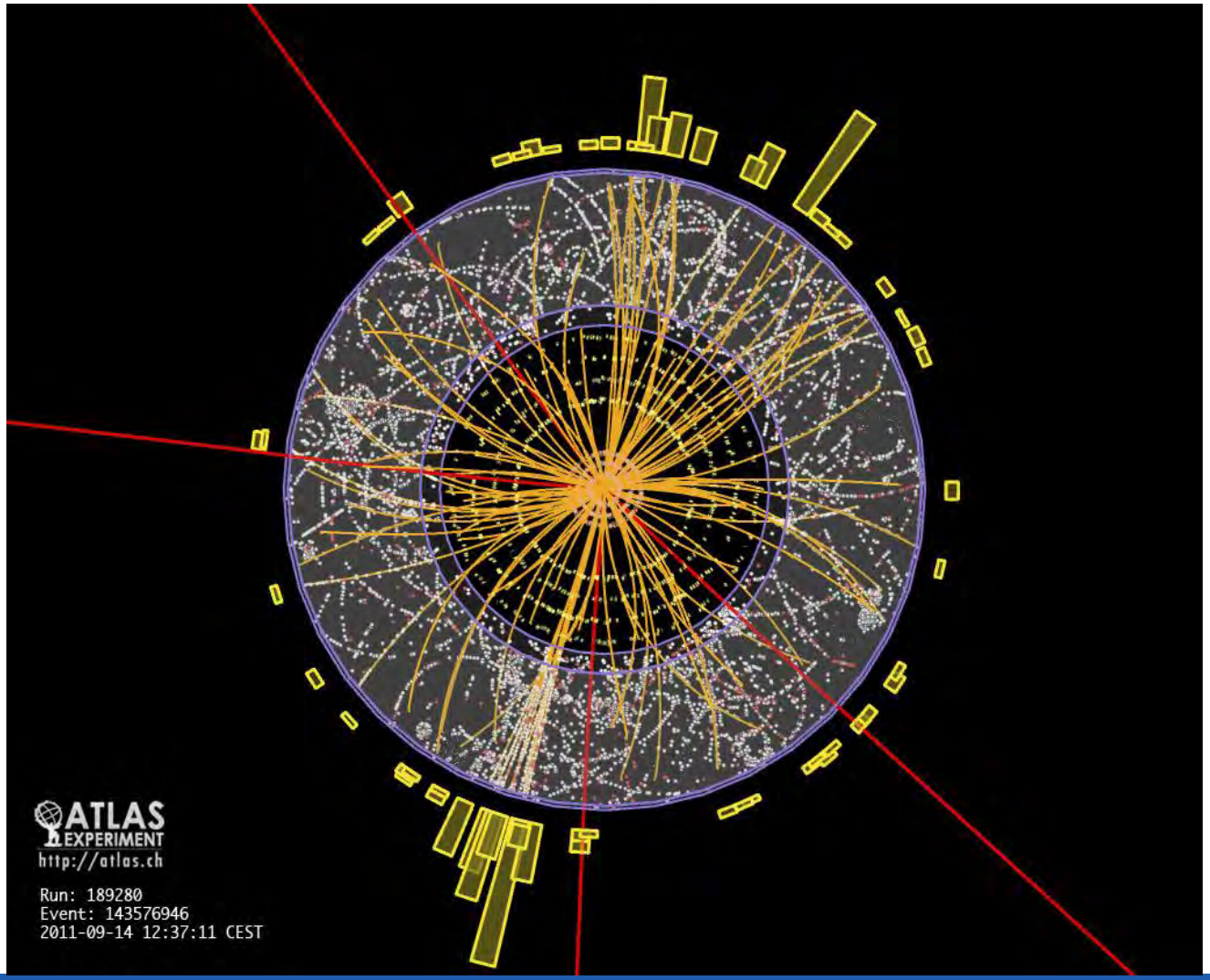


Трэкинг

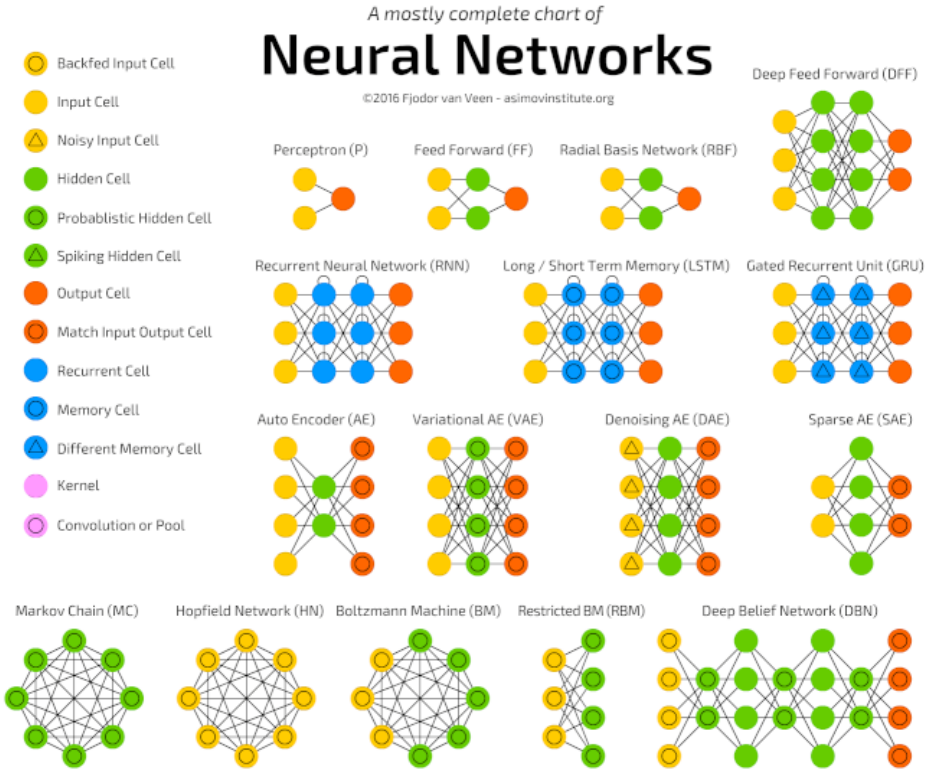


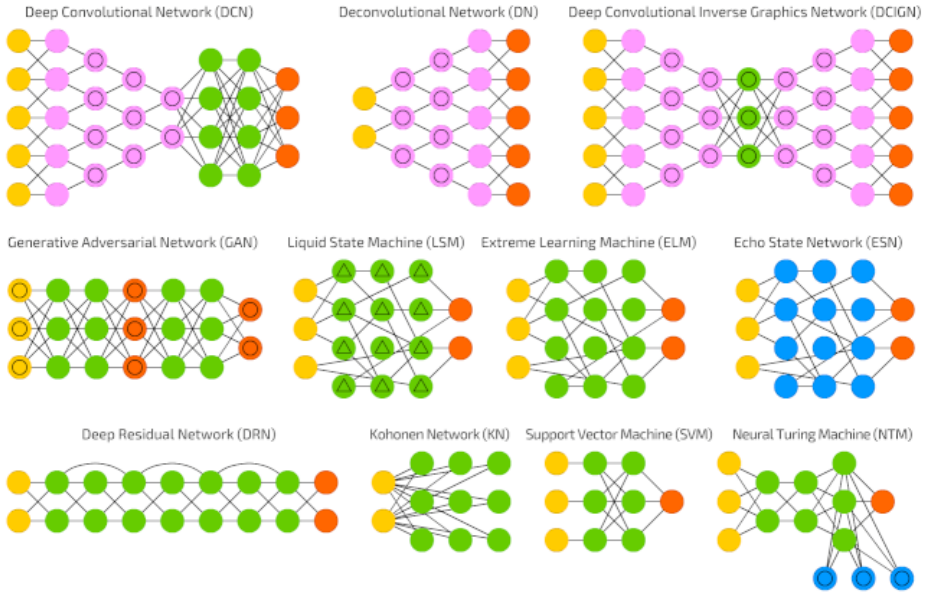
http://atlas.web.cern.ch/Atlas/public/EVTDISPLAY/events.html

Трэкинг


ATLAS
EXPERIMENT
http://atlas.ch
Run: 189280
Event: 143576946
2011-09-14 12:37:11 CEST

Нейронные сети





Нейронные сети

До 7.8 TFLOPS при вычислениях с двойной точности

До 125 TFLOPS при работе с нейронными сетями

С нейронными сетями, никаких серьёзных изменений в коде не требуется (при переходе с CPU на GPU)

Параллельность

Суперкомпьютеры

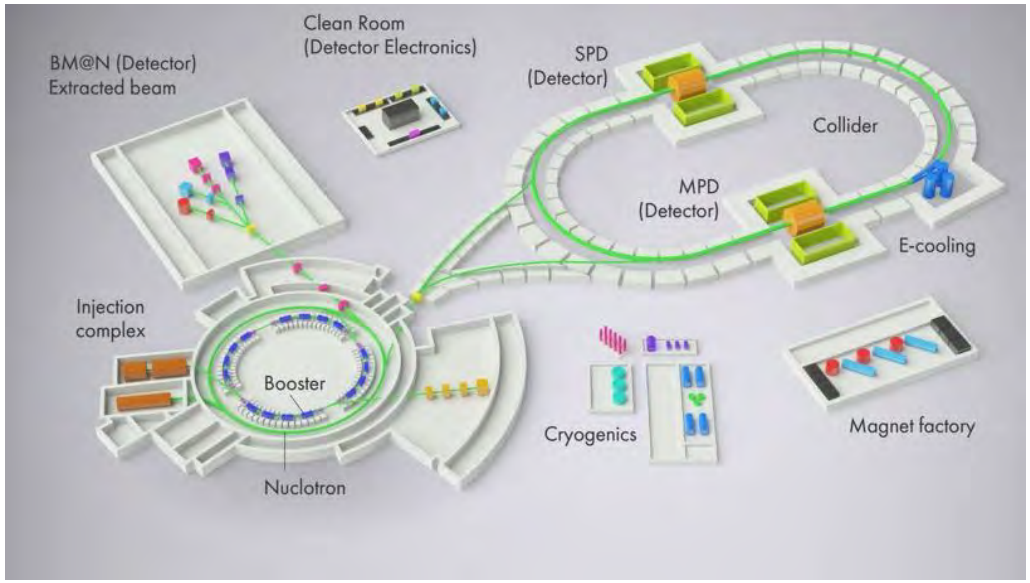
Распределённость

Грид,
кластера

Эффективность

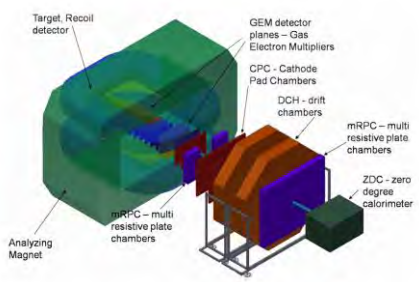
Новые алгоритмы

Проекты мегасайенс

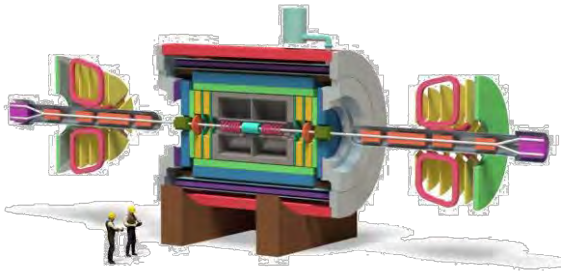


NICA

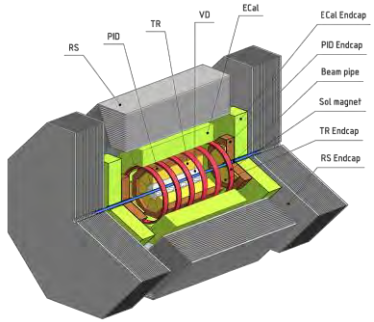
BM@N



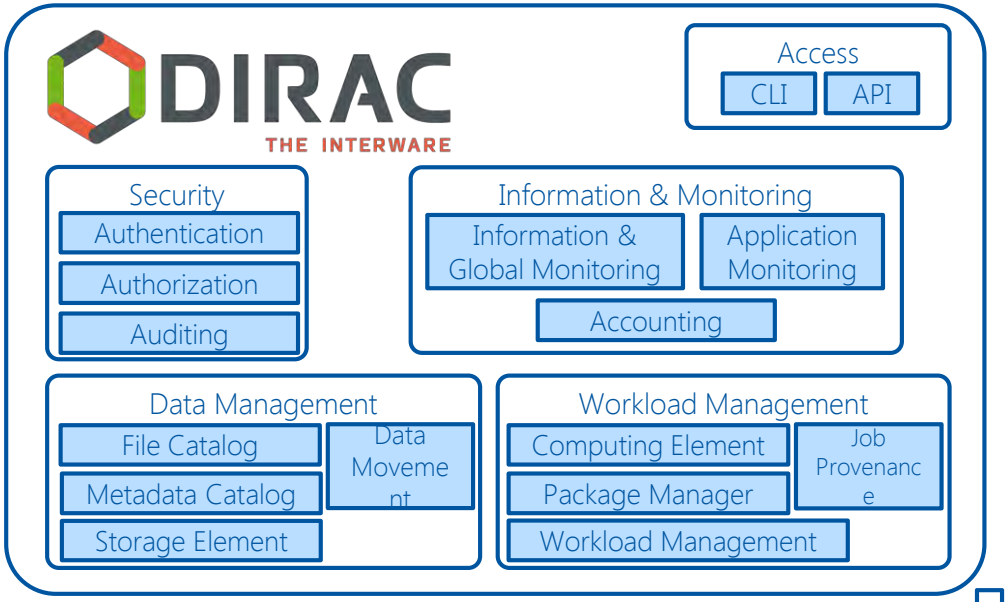
MPD



SPD



Что такое DIRAC?



+

Интеграция

+

Интерфейсы

Почему DIRAC:

- Производительность
- Единый инструмент для управления задачами, данными пользователями ...
- Расширяемость
- Активное комьюнити

Принципы интеграции

Пользователи

Ресурсы

Принципы интеграции

Пользователи

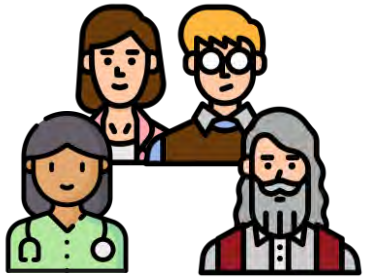
Ресурсы

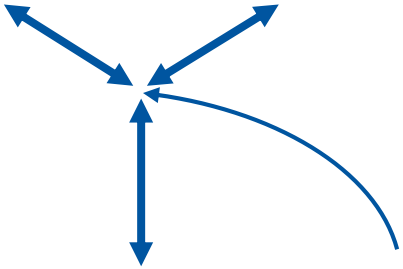
Это не волшебный инструмент

Ресурсы

Пользователи

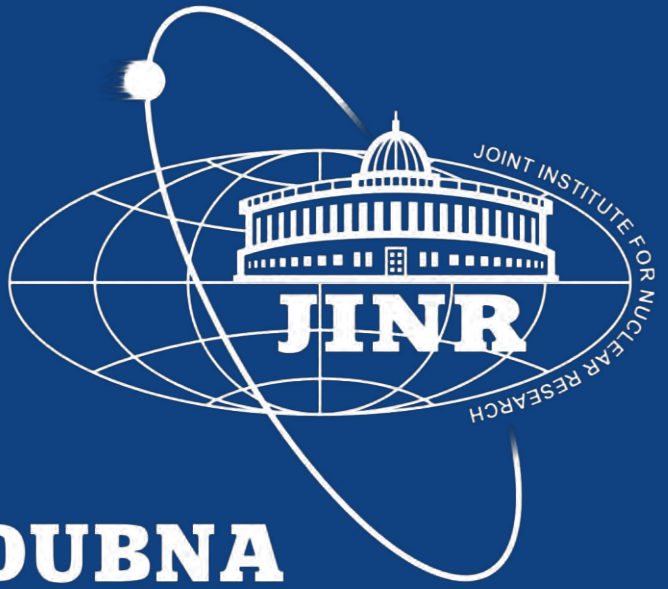
Принципы интеграции





Сотрудничество*

DIRAC
THE INTERWARE



DUBNA