

托卡马克大规模数值模拟研讨会

数值计算物理数据库的 初步实现和展望

夏 凡

核工业西南物理研究院

北京 . 2013. 4. 24

Outline

- A Story
- Philosophy
- Technology
- Calculation between Discharges
- Vision

故事

□ 小刘想要自动判断H模的产生以及计算相关参数

偏滤器处的电子温度密度分布

Beta

三角形形变

Q剖面

辐射功率

孔栏半径

电流分布

电流中心

电子密度剖面

储能

加热功率

约束时间

电子温度剖面

小刘

离子密度剖面

离子温度剖面

气体种类

拉长比

内外打击点位置

压力剖面

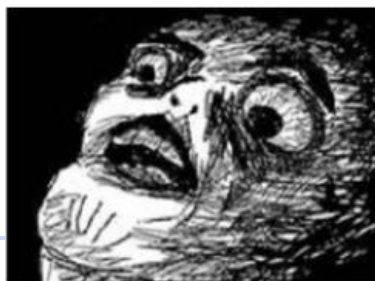
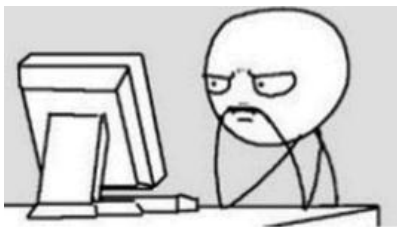
X点位置

内感

线积分密度

最外层磁面

统计分析



原因分析

- ◆ 当前HL-2A的数据库是为采集系统而设计，即一个通道对应一条随时间变化的曲线，不能处理零维和多维数据
- ◆ 这些数据有如下的特点：
 - 大部分需要在采集数据库的基础上进行处理，算法只有当事人知道
 - 二维剖面数据，随时间变化，例如温度剖面、密度剖面
 - 图形，随时间变化，例如质谱图
 - 一炮可能只有一个值，例如评判、气动探针连线图、孔栏半径、工作气体、壁面处理情况等等
 - 一天只有一条曲线，例如EPICS监视数据（气压等等）
 - 这些数据可能要根据不同的情况做相关的修正，而一般情况下一般人不知道修正算法，例如去跳点、补偿等等
- 这些因为各种原因而导致没有能进行统一入库管理，导致分散在各个地方
 - 不同的数据不同的存储格式，例如中控使用dpf格式，温度剖面matlab格式，要用数据，不知道找谁要
 - 要用数据，得去找相关的负责人
 - 要用数据，需要熟知相关的处理程序，看得懂代码，能分析系统
 - 要用数据，需要精通其数据格式
 - 要用数据，负责人已经更换，早期数据已经没有了

解决方案的战略和原则

- 跟随战略
 - ITER是SWIP的旗帜，ITER的解决方案均经过多方考虑，其技术方案能经得住推敲，值得顺势紧跟
- 统一共享战略
 - 统一的格式，统一的管理，统一的接口，专人维护，使用方便，数据授权公开共享
- 时效原则
 - 相关计算利用大规模并行计算技术在炮间或者24小时之内计算完成
- 重点原则
 - 对于典型的炮号各种数据尽量完善
- 自洽原则
 - 此物理数据库最终的目的是建设数据可以互相自洽的数据库，即要保证数据是相对正确的，错误或者无效的数据不进入到此数据库中

技术解决方案-HDF5

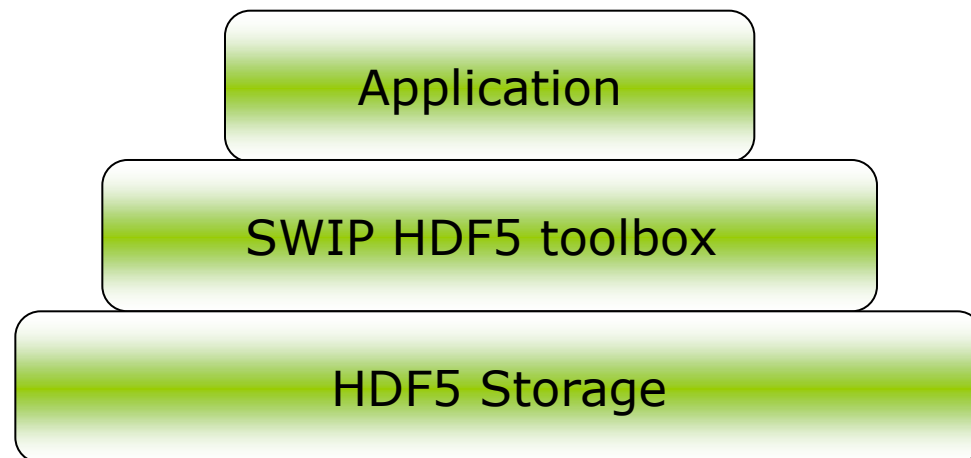
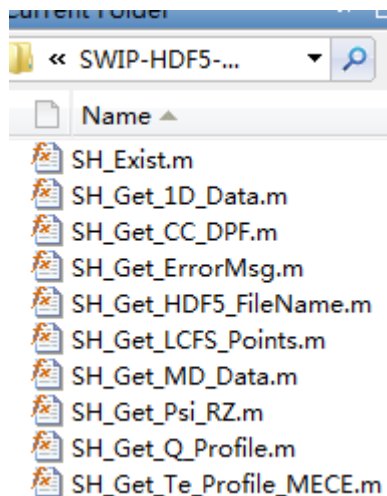
◆ 优点:

- ITER采用文件存储格式
- 将来很多工具可以从ITER得到
- 是一个文件系统标准，各行各业都在用，不局限于Tokamak
- Matlab直接提供接口，是由Mathworks公司官方维护，非常方便，Windows下和Linux下都直接使用
- 有LabView接口
- 有Windows下C，C#接口
- HDF5的数据库架构同当前HL2A的自有数据库架构类似，迁移的时候不需要推翻重来
- 可以存储多种格式的数据

因此，采用HDF5作为物理数据库的存储格式

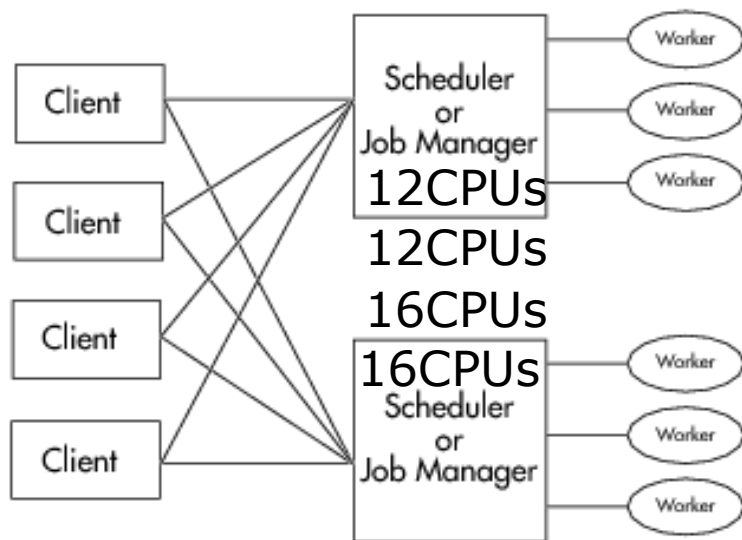
SH Matlab工具包

- 将底层具体的存储方式和内容加以屏蔽，使得相关的使用人员可以不用熟悉里面的具体格式，从而再次降低其使用门槛。
- 通过提供工具包的方式，在保证函数名（接口）不变的情况下，只更新这个工具包，就可以在不修改上层程序的情况下，达到存储格式更新的目的。



炮间（8分钟）大规模并行计算

- 目前共有4台工作站，56个cpu参与炮间大规模并行计算，实现炮间入库功能，系统由EPICS消息驱动，平台使用Matlab并行计算工具箱

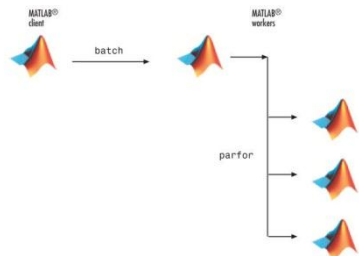
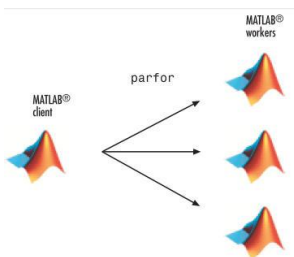


实现的功能有：

- 离线EFIT的计算，平均每炮1800个时间片
- 常规故障诊断
- 诊断系统常规初步计算
- 计算结果存入HDF5数据库
- 相关大屏幕显示

正在实现的功能有：

- TSC跟跑计算，典型的一次计算需要2个小时，通过并行计算，实现当天放电数据全部入库



炮间计算

网络 > 192.168.10.11 > hdf > 2A

编辑(E) 查看(V) 工具(T) 帮助(H)

刻录 新建文件夹

名称	修改日期
HL-2A=Structure.h5	2012/12/18 16:44
HL-2A=22813=PhysicsDB.h5	2013/4/19 17:30
HL-2A=22812=PhysicsDB.h5	2013/4/19 17:21
HL-2A=22811=PhysicsDB.h5	2013/4/19 17:10
HL-2A=22810=PhysicsDB.h5	2013/4/19 17:02
HL-2A=22809=PhysicsDB.h5	2013/4/19 16:54
HL-2A=22808=PhysicsDB.h5	2013/4/19 16:44
HL-2A=22807=PhysicsDB.h5	2013/4/19 16:33
HL-2A=22806=PhysicsDB.h5	2013/4/19 16:23
HL-2A=22805=PhysicsDB.h5	2013/4/19 16:15
HL-2A=22803=PhysicsDB.h5	2013/4/19 16:00
HL-2A=22802=PhysicsDB.h5	2013/4/19 15:43
HL-2A=22801=PhysicsDB.h5	2013/4/19 15:33
HL-2A=22800=PhysicsDB.h5	2013/4/19 15:23
HL-2A=22799=PhysicsDB.h5	2013/4/19 15:06
HL-2A=22798=PhysicsDB.h5	2013/4/19 14:57
HL-2A=22797=PhysicsDB.h5	2013/4/19 14:46
HL-2A=22796=PhysicsDB.h5	2013/4/19 14:35

15,159 个对象

脱机状态: 脱机

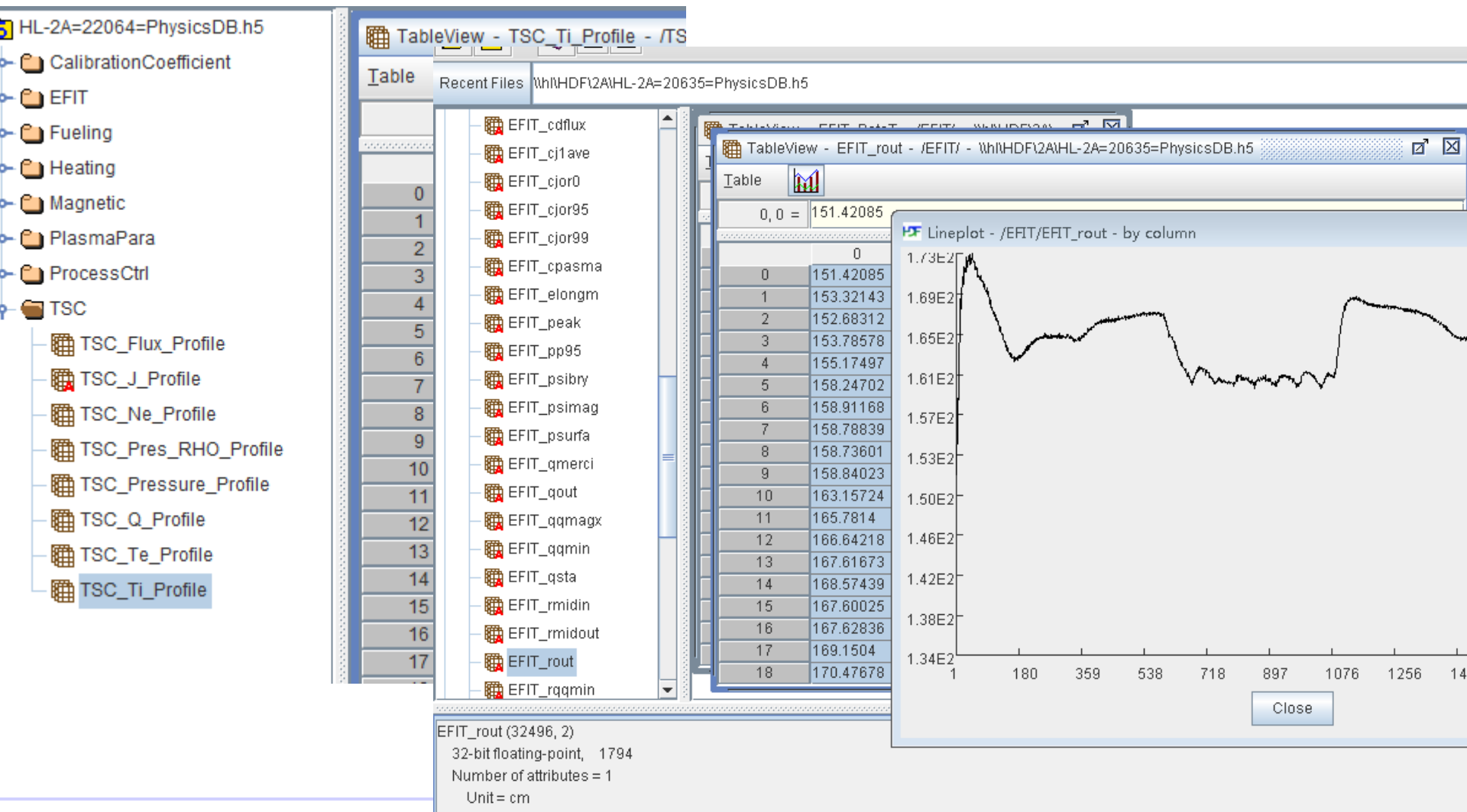
脱机可用性: 不可用

常规故障诊断

HL-2A Automatic Alarm System

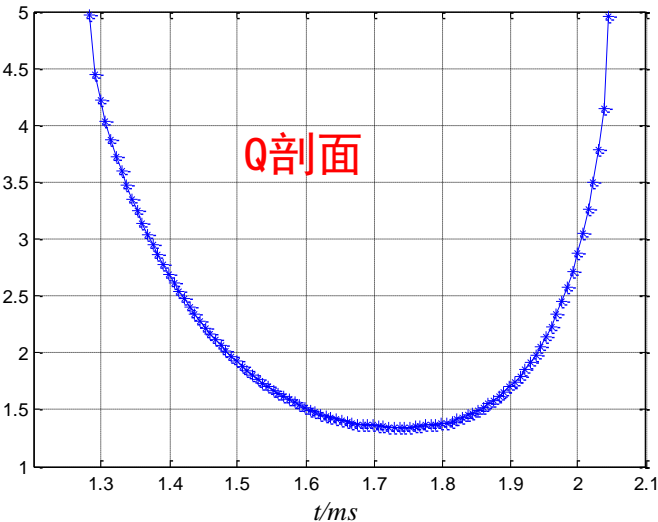
13:52:27 PLC-OPC 机器恢复.
13:47:54 M: Better Wall Condition is Good Idea
13:47:54 P:Near 366ms, Major Disruption!
13:47:50 Current Shot: 21508
13:47:50 =====
13:38:22 M: Better Wall Condition is Good Idea
13:38:22 P:Near 104ms, Major Disruption!
13:38:21 Something Wrong at 4th Channel in HCN
13:38:18 Current Shot: 21507
13:38:18 =====
13:29:10 Current Shot: 21506
13:29:10 =====
13:19:44 M: Better Wall Condition is Good Idea
13:19:44 P:Near 391ms, Major Disruption!
13:19:40 Current Shot: 21505
13:19:40 =====
13:06:26 Current Shot: 21504
13:06:26 =====
NC 14:40:52 M:Add more Gas!
NC 14:40:52 S:May Cause Major Disruption
NC 14:40:52 P:Ne near Density Limit:0.3148
NC 14:40:52 HCN第4道错误, 需要修正
NC 14:40:48 Current Shot: 21393
NC 14:40:48 =====

当前物理数据库概览

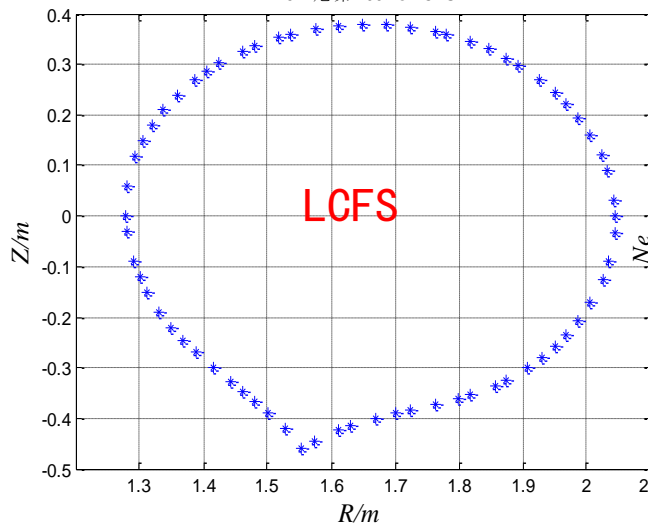


相关物理数据 (EFIT)

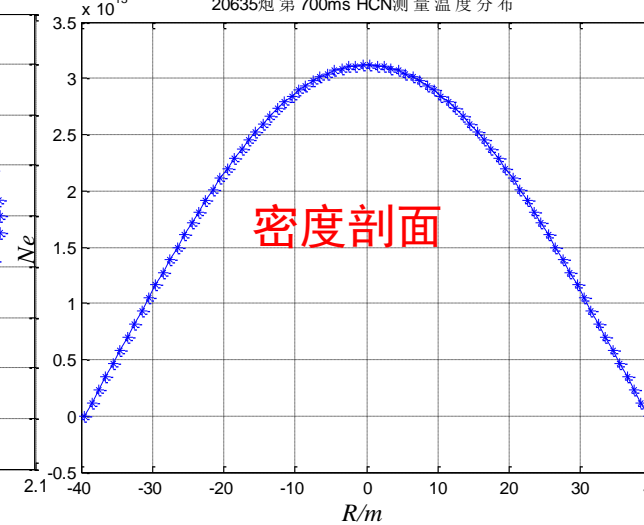
11617炮第700ms Q剖面



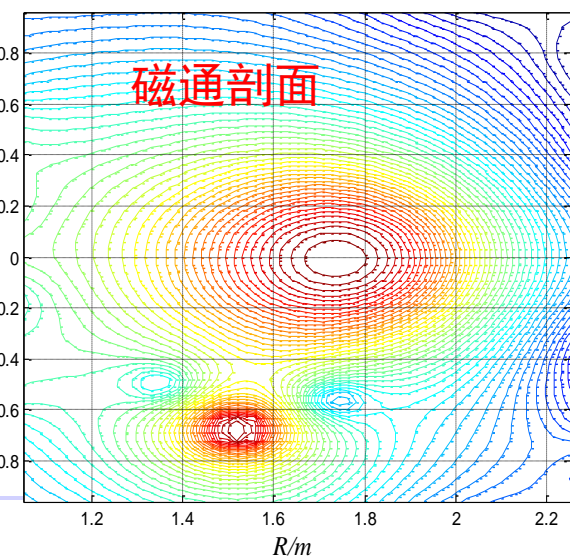
11617炮第700ms LCFS



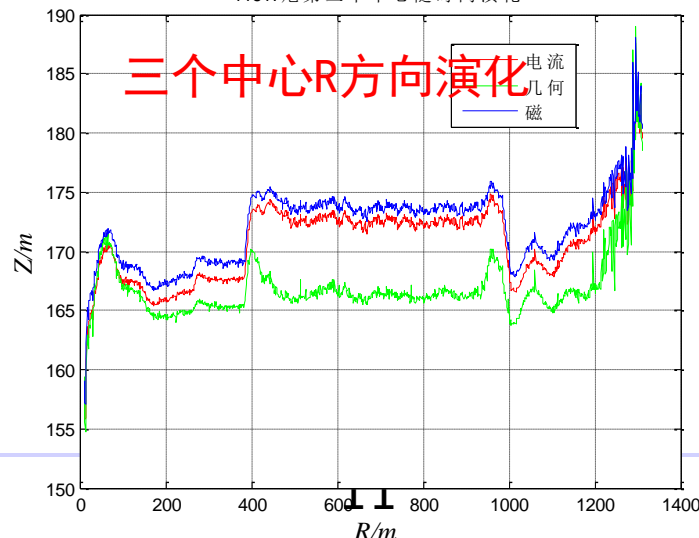
20635炮第700ms HCN测量温度分布



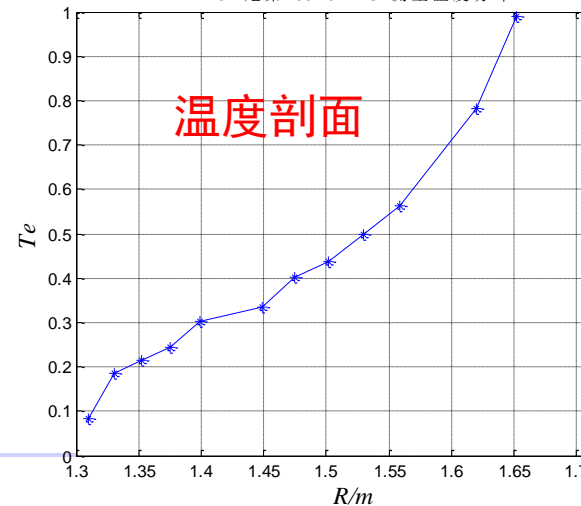
11617炮第700ms 磁通分布



11617炮第三个中心随时间演化



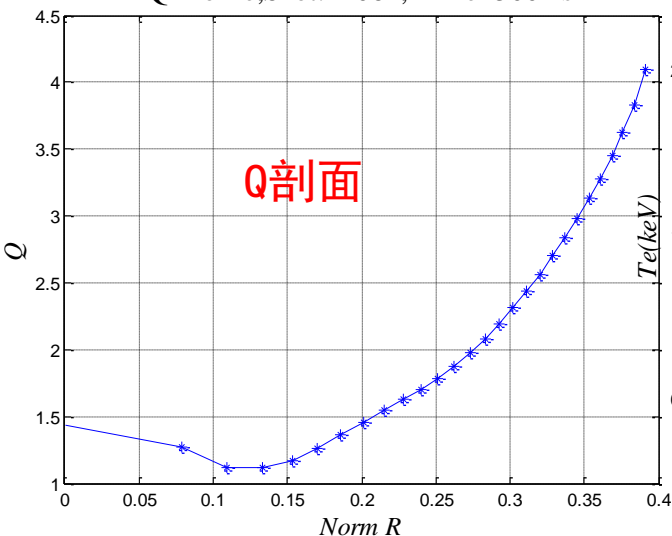
11617炮第700ms MECE测量温度分布



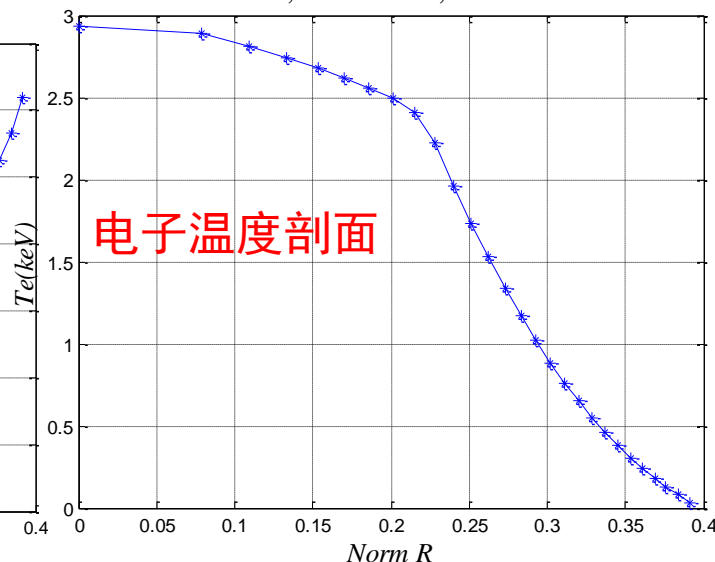
相关物理数据 (TSC)

反剪切

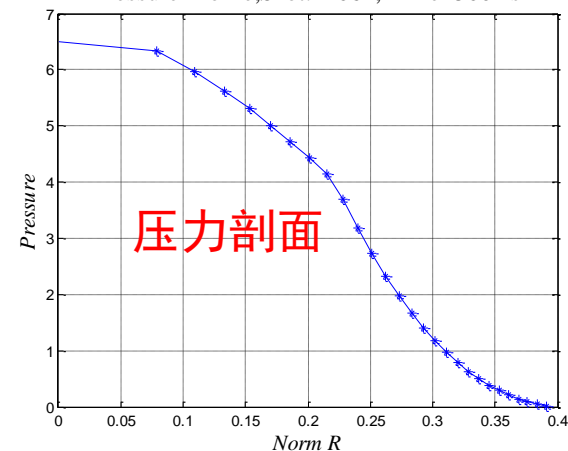
Q Profile, Shot: 22064, Time=500ms



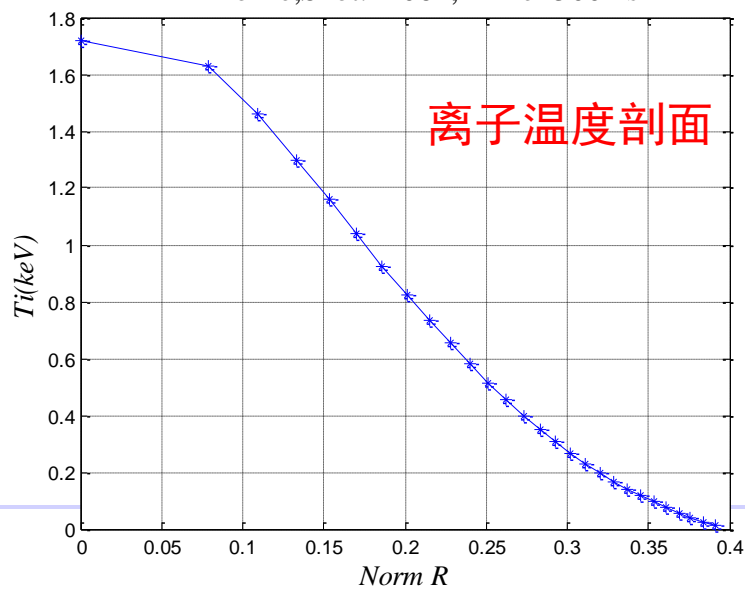
Te Profile, Shot: 22064, Time=500ms



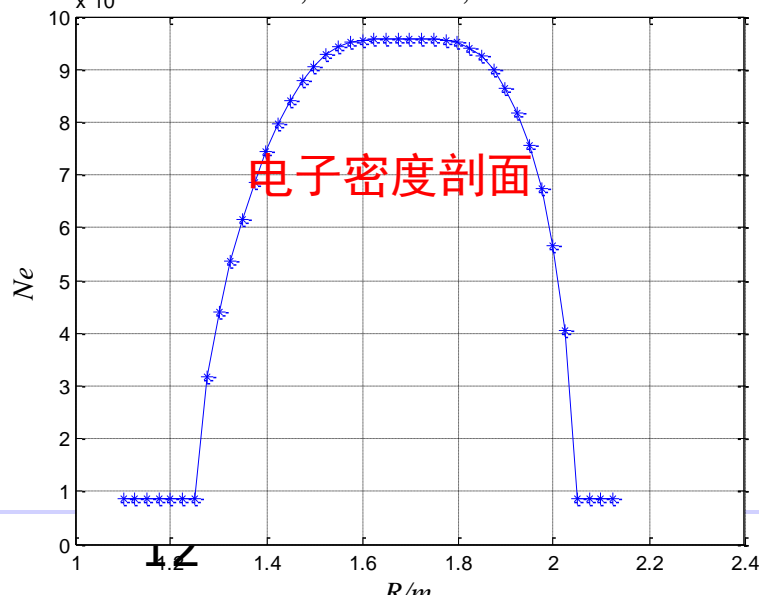
Pressure Profile, Shot: 22064, Time=500ms



Ti Profile, Shot: 22064, Time=500ms



Ne Profile, Shot: 22064, Time=500ms



展 望

- 从历史和现在的情况来看，物理数据库的工作开展更依赖于非技术问题的解决，包括：行政支持、数据来源、数据认证、版权和出图等等，这些值得综合各方面的资源进行解决
- 数据的准确性非常依赖于诊断系统的提升，例如MSE投入使用后，在EFIT中加入，将会大幅度提高对等离子体芯部参数计算结果的精度
- 与实验相关的模拟计算结果加入到数据库中充实数据
- 从更大的范围来讲，可以考虑将SWIP、ASIPP、HUST的物理数据资源在一定的协议框架下进行共享，将会对中国的核聚变理论以及数值模拟研究提供极大便利，当然，这需要从更高层面的协调以及三个机构与各院校的博弈。

谢谢！

物理数据库建设的基本原则

- 对于HDF5数据库，并不是针对于某个人而创建，而是针对于所有与SWIP两个装置的物理数据产生联系的人而创建，因此，必须创建一套规则，在这个规则下，以统一地界面进行数据的读取、写入和维护
- 此物理数据库最终的目的是建设数据可以互相自洽的数据库，即要保证数据是相对正确的，错误或者无效的数据不进入到此数据库中
- 具体内容请参看<\\192.168.10.11>下面的文档《基于HDF5的物理数据库.doc》

原因分析

- 当前核聚变的研究的一个重点仍然是实验研究，需要综合各种物理量，试图找出其中的各种物理现象和规律；
- 因此，可以认为，能够用来做上述等离子体物理分析之用的数据就是物理数据，
- 从这个范畴上讲，部分工程数据也是一种物理数据，例如加热功率等等
- 并且根据分析对象的不同，可以将新的数据纳入到其范围内，其范围是动态的；
- 物理数据常见的获取方式是在采集数据库的基础上，经过计算得到

❑ 目前的HL2A的公共数据库是建立在采集系统之上，其优点是：

- 统一的格式
- 统一的管理
- 统一的接口
- 专人维护
- 使用方便
- 数据公开

❑ 也存在缺点

- 最原始的工程数据和诊断系统数据，一个通道对应一条随时间变化的曲线，对多维数据的处理困难

❑ 当前的很多物理数据由于没有得到有效的整理和收集，均放置在系统负责人那里，其优点是：

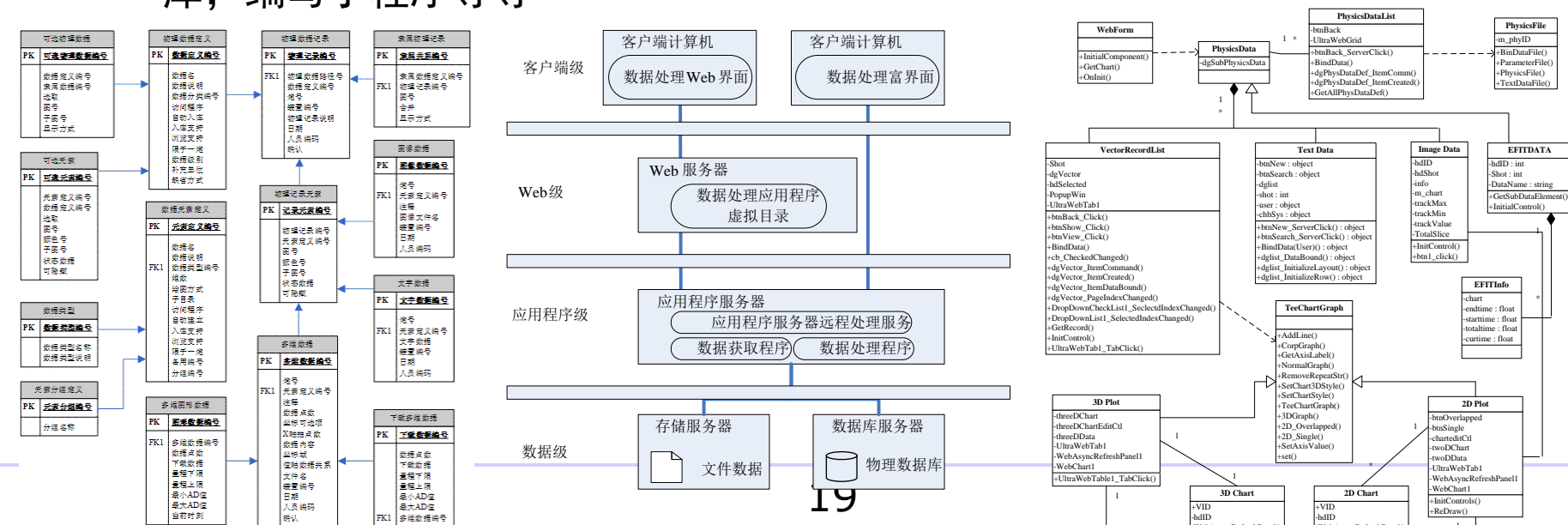
➤ 数据保密性好

❑ 但也存在缺点，

- 不同的数据不同的存储格式，例如中控使用的是自定义的dpf格式，温度剖面使用的是matlab格式，早期的ECE使用的是txt格式
- 要用数据，不知道找谁要
- 要用数据，得去找相关的负责人
- 要用数据，需要熟知相关的处理程序，看得懂代码，能分析系统
- 要用数据，需要精通其数据格式
- 要用数据，负责人已经更换，早期数据已经没有了

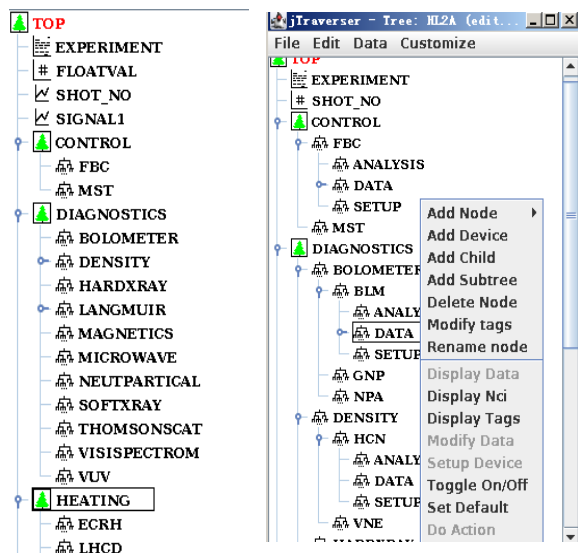


- HL2A建设初期，106制定针对于采集系统的数据库存储规范，十余年的使用证明此数据库充分满足采集系统的使用、扩充的需求
- 针对于此物理数据库的配套软件的开发了Proda用于通用数据的浏览分析
- 从2008年开始，宋博士持续开发了matlab版本的Dataproc，其中的hl2adb程序为使用最广泛的读取此数据库的matlab代码。
- 2006年左右，106开始考虑多维数据的入库，并提出物理数据库的概念
- 106室对此作了大量的工作，设计了数据存储规范，规定了接口，设计了数据库，编写了程序等等



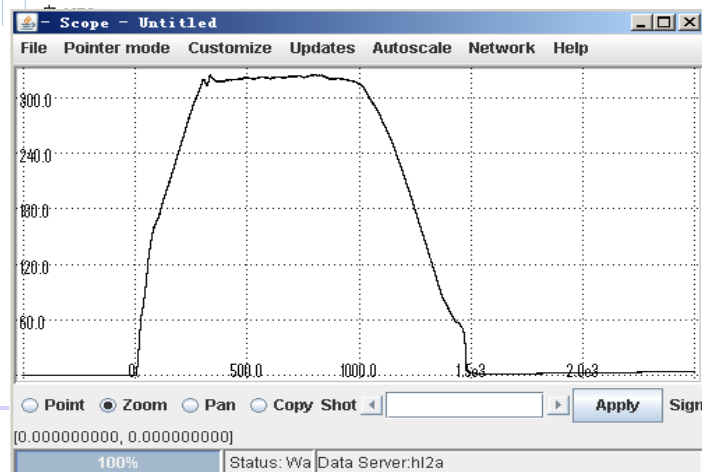
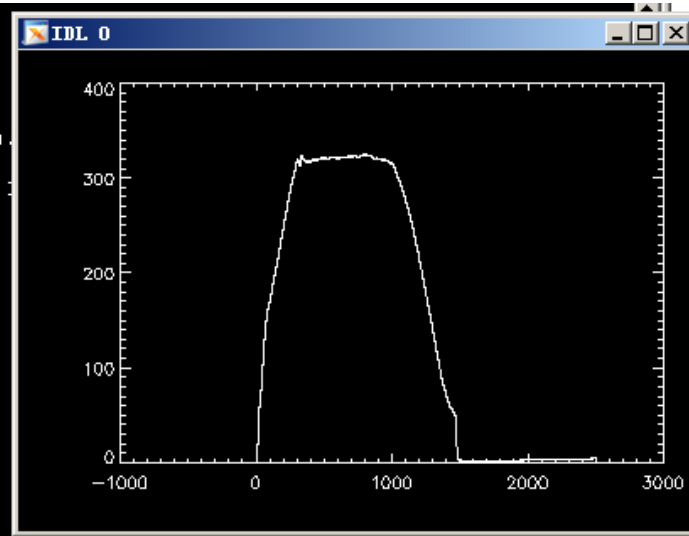
历史渊源

- 2008年，考虑到外访人员使用数据库上的困难，106开始将采集数据转移到MDSPlus上



```
% Syntax error.
IDL> plot,x,y
IDL> exit
[root@localhost mdsplus]# idl
IDL Version 6.3 (linux x86_64 m64)
Installation number: 220094.
Licensed for use by: Southwestern I

IDL> mdsconnect,'192.168.10.47'
% Compiled module: MDSCONNECT.
% Compiled module: MDSDISCONNECT.
IDL> mdsopen,'hl2a',13300
% Compiled module: MDSOPEN.
% Compiled module: MDSVALUE.
% Compiled module: MDSCHECKARG.
% Compiled module: MDSISCLIENT.
IDL> y=mdsvalue('\ip')
IDL> x=mdsvalue('dim_of(\ip)')
IDL> plot,x,y
IDL>
```




MDSPlus 入库

运行状态

DAS目录: G:\2adas\
 目标IP: 192.168.10.47:8000 端口: 12300
 自动入库状态: ON

手动入库

开始炮号:
 结束炮号:
 入库

日志

Received UDP message:+PLS_14452 (2010-6-13 14:32:33)
 Start to put data in MDSPlus! (2010-6-13 14:37:33)
 Shot:14452 trees creation complete!
 SXR Data Input Completion!
 HCN Data Input Completion!
 TWP Data Input Completion!
 EPC Data Input Completion!
 EPD Data Input Completion!
 EPE Data Input Completion!
 SX2 Data Input Completion!
 MST Data Input Completion!
 TA1 Data Input Completion!
 TA2 Data Input Completion!
 MIR Data Input Completion!
 OFA Data Input Completion!
 OFC Data Input Completion!
 FBC Data Input Completion!
 OFB Data Input Completion!
 EMD Data Input Completion!
 OFD Data Input Completion!

历史渊源

- 但是由于种种原因，上述两种数据库使用的效果不佳，并未推展开来
- 2012年，ITER释放出来的文档中决定不采用mdsplus作为其数据库，而采用HDF数据库
- 2012年，在开发实时位形重建代码（PARC）的过程中，用C重写EFIT，为保证正确性，需要对离线efit做出修改，并要保证一次计算可以计算一炮中所有的时间点（1kHz），由此，计算一炮就产生大量的数据（100M左右一炮），特别是二维的磁通剖面数据，需要得到存储方便以后的使用。
- Efit的数据于是直接使用HDF5存储，作为前期HL2M数据库的尝试
- 鉴于HDF5数据库的种种优点，106决定重启物理数据库的项目，并开展相关的尝试
- 目前入库的数据仅为本人所关心的数据

原因分析

- 从前面的叙述可以看到，106很早以前就意识到这个问题，但是几次尝试的没有进行下去的原因很大程度在于太着重于数据库技术层面的研究，与物理实验的结合度不够
- 物理数据库的实施问题其实可以归结到两个问题，一个是技术问题，即数据库技术问题，一个非技术的问题，即如何找到物理数据来源并验证入库的问题。
- 对106来讲，数据库技术问题（入库、查询、存储、维护）其实不是问题。
- 对于101, 102以及其他的室来讲，只要心态开放，第二个非技术问题其实也不是问题
- 通力配合，扬长避短，团队协作，才是物理数据库能成功实施的根本保障

MDSPlus的优缺点

◆优点:

- 可以直接通过网络访问
- DIIID和EAST都在用，有些年头了，技术相对成熟

◆缺点:

- 必须注意到的是，MDSPlus重点在于Linux平台，而非Windows平台，可以从其源代码看出
- 在Windows下，重新编译成dll较为困难，因为用到了Linux下的网络库在Windows下，应用在Matlab2008版本以上的需要重新编译其dll，需要对应的Linux网络函数库（C语言），较难编译成功
- MDSPlus主要应用在Tokamak，应用面窄，外围工具支持较少
- 大规模并发访问的时候可能会出现问题，这也是ITER的考虑因素之一

HDF5的优缺点

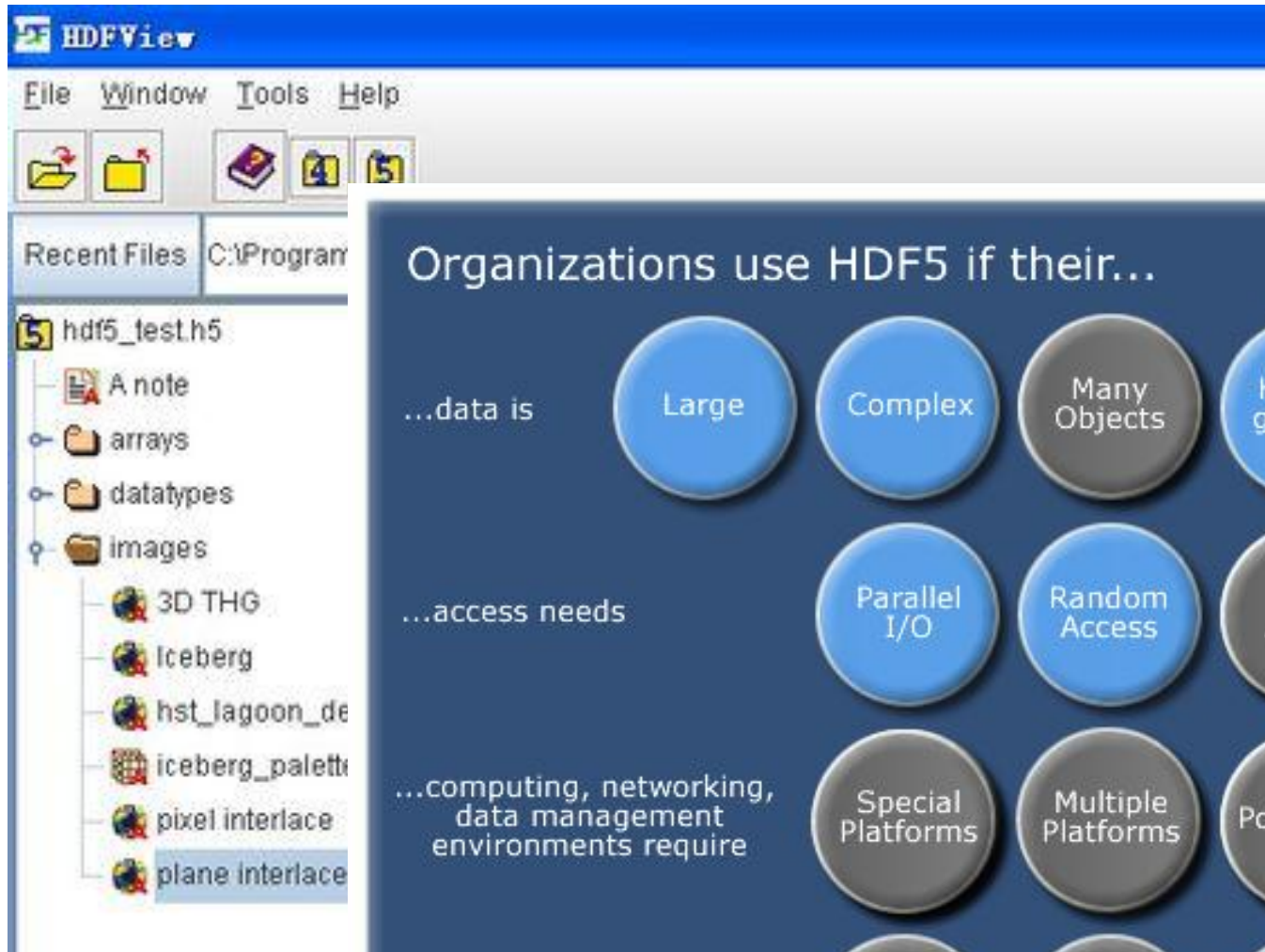
◆ 缺点:

- 同HL2A一样，是文件系统，不可以直接通过网络访问，需要通过中间的数据库

◆ 优点:

- ITER采用文件存储格式
- 将来很多工具可以从ITER得到
- 是一个文件系统标准，各行各业都在用，不局限于Tokamak
- Matlab直接提供接口，是由Mathworks公司官方维护，非常方便，Windows下和Linux下都直接使用
- 有LabView接口
- 有Windows下C，C#接口
- HDF5的数据库架构同当前HL2A的自有数据库架构类似，迁移的时候不需要推翻重来
- 可以存储多种格式的数据

因此，采用HDF5作为物理数据库的存储格式



Organizations use HDF5 if their...

...data is

Large

Complex

Many
Objects

Hetero-
geneous

Esoteric

...access needs

Parallel
I/O

Random
Access

Fast
Access

Partial
I/O

...computing, networking,
data management
environments require

Special
Platforms

Multiple
Platforms

Portability

Efficient
Storage

...or if they need

A Standard
Format

Available
Tools

Low
Cost

Image or
Table API

物理数据库建设的基本原则

- 对于HDF5数据库，并不是针对于某个人而创建，而是针对于所有与SWIP两个装置的物理数据产生联系的人而创建，因此，必须创建一套规则，在这个规则下，以统一地界面进行数据的读取、写入和维护
- 此物理数据库最终的目的是建设数据可以互相自洽的数据库，即要保证数据是相对正确的，错误或者无效的数据不进入到此数据库中
- 具体内容请参看<\\192.168.10.11>下面的文档《基于HDF5的物理数据库.doc》

HDF5物理数据库的使用前的说明

- 为了保证数据的安全性，HDF5的数据全部都放在 \\192.168.10.11 服务器里面，这个服务器同时也存储有HL2A的采集数据库，具有最高级别的数据安全保障。
- 但是，使用这个服务器上的数据之前，需要进行登录认证，认证的用户名和密码就是SWIP的域账号和密码。登录成功之后，Windows会记录这个用户名和账户，下次再使用这个数据库上的数据，就不需要进行登录。
- \\192.168.10.11与\\HL为同一个服务器，但考虑到解析的时候可能会出现问題，所有的路径均使用IP地址的方式
- 考虑到物理分析的因素，目前进入到物理数据库中的数据不包括放电失败的炮。
- 目前已经入库13600多炮的数据

HDF5物理数据库的使用-Win7下的登录认证



相关工具-Matlab与HDF5

- HDF5的数据格式广泛应用于各研究领域，有二十多年的发展历史，技术非常成熟，并且广泛的被各种语言官方支持，其中包括Matlab官方支持。
- 从使用角度来讲，Matlab是使用HDF5最方便的语言之一。
- Matlab的版本要求：Matlab 2011以上
- Matlab2012a的下载地址：[\\192.168.10.11\hdf\](http://192.168.10.11/hdf/)
相关软件

相关工具- Hdfview

- Hdfview是HDF组织官方提供的查看HDF数据库结构的工具，此软件基于Java开发，Hdfview在安装的时候会自动安装Java运行环境。

- 下载地址与Matlab2012a的地址一样：

\\192.168.10.11\hdf\相关软件

此目录下，有32位和64位的安装版本，可以根据机器的配置选择合适的版本。

hdfview_install_win64.exe

hdfview_install_win32.exe

- 安装完毕后，可以直接打开后缀名是.h5的文件，查看里面的数据组织结构和查看曲线，例如，第20635炮的数据组织如下：

当前物理数据库的内容-EFIT平衡计算数据

- 中平面上的q剖面-EFIT_Q_at_MidPlane
- LCFS上的点- EFIT_LCFS_Points
- 二维磁通剖面- EFIT_PsiRZ
- 计算区域坐标- EFIT_Grid_R, EFIT_Grid_Z
- 磁轴磁通和边界磁通-EFIT_psimag, EFIT_psibry
- 磁轴中心坐标-EFIT_Mag_R, EFIT_Mag_Z
- 电流中心坐标-EFIT_Cur_R, EFIT_Cur_Z
- 几何中心坐标（大半径，Z）-EFIT_rout, EFIT_zout
- 等等

当前物理数据库的内容-EFIT平衡计算数据

□ 说明:

- 因为在目前的efit数据库中，并没有包含内部剖面的约束，只使用到了外部的磁探针、磁通环的约束，因此电流剖面并未得到有效的修正，得到的芯部的Q并不一定是正确的
- 目前正在考虑使用压强剖面条件对EFIT的计算进行进一步地约束计算，但是首先需要得到密度剖面，温度剖面等相关参数

当前物理数据库的内容-标定系数

□ 标定系数在
CalibrationCoefficient
下面

□ FluxLoop

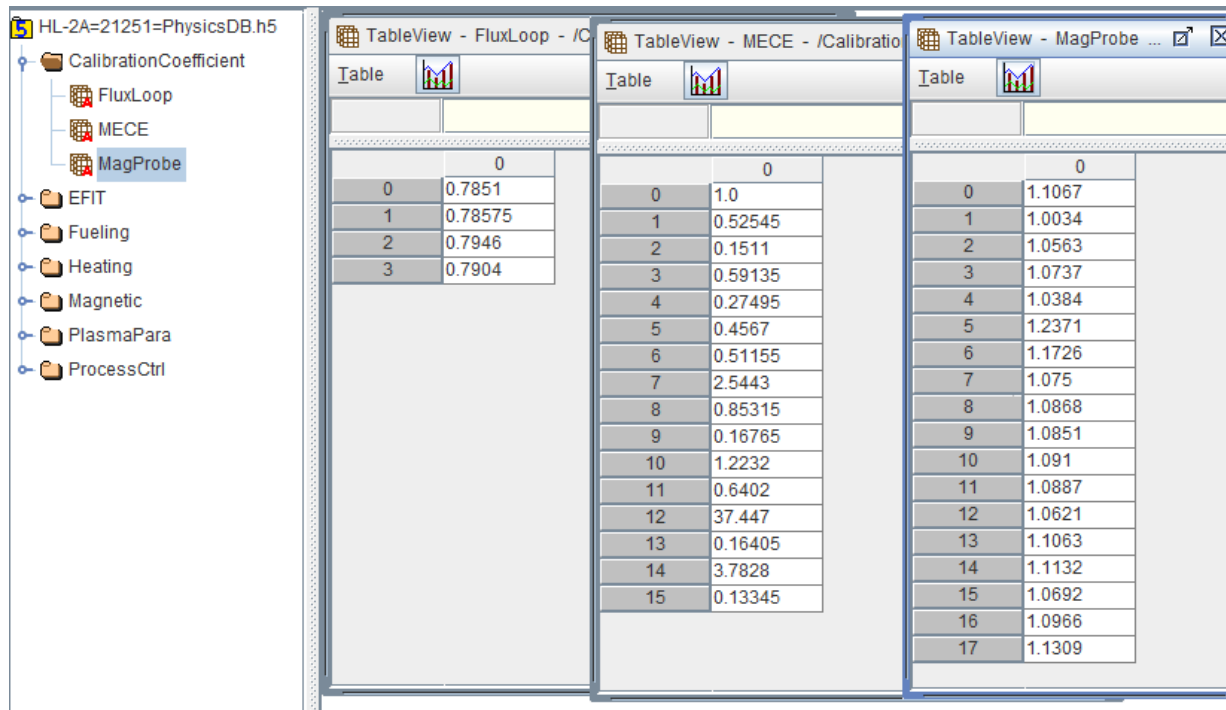
- 四个磁通环的标定系数，
顺序为：
- 内上，内下，外上，外下

□ MagProbe

- 18个极向磁探针的标定系数，
顺序为磁探针序号的
升序

□ MECE

- 16个MECE的标定系数



The screenshot displays three tables of calibration coefficients from the HL-2A=21251=PhysicsDB.h5 database. The left panel shows the database structure with 'CalibrationCoefficient' expanded, highlighting 'FluxLoop', 'MECE', and 'MagProbe'. The middle panel shows the 'FluxLoop' table with 4 rows. The right panel shows the 'MECE' table with 16 rows. The far right panel shows the 'MagProbe' table with 18 rows.

	0
0	0.7851
1	0.78575
2	0.7946
3	0.7904

	0
0	1.0
1	0.52545
2	0.1511
3	0.59135
4	0.27495
5	0.4567
6	0.51155
7	2.5443
8	0.85315
9	0.16765
10	1.2232
11	0.6402
12	37.447
13	0.16405
14	3.7828
15	0.13345

	0
0	1.1067
1	1.0034
2	1.0563
3	1.0737
4	1.0384
5	1.2371
6	1.1726
7	1.075
8	1.0868
9	1.0851
10	1.091
11	1.0887
12	1.0621
13	1.1063
14	1.1132
15	1.0692
16	1.0966
17	1.1309

过程控制参数- ProcessCtrl

□ Configuration

- Configuration是HL2A中央控制的配置，为一个struct结构。

□ Limter_Radius

- 本炮孔栏半径，因为在有活动孔栏的装置中，这个孔栏半径会随时间变化，而这种变化一般情况下，并不会被记录，但是对于形状控制是非常重要的。

□ DischargeSummary

- 为一个数据结构，记录放电概要信息，内容正在增加中

□ Bolometer_State

- 表示16个中央Bolometer的状态，此状态由为通过分析16个Bolometer的采集数据得到，为一个数据结构

等离子体参数相关-PlasmaPara

- IP
 - 采用的是反馈使用的数据
- P_Rad_Bolometer,
 - 等离子体辐射功率，一维数据，已经修正因为采集通道失效造成的错误
- Ne_Profile_HCN
 - 电子密度剖面，二维数据，因为一些未解决问题，已经入库，但是不建议引用
- Te_Profile_MECE
 - 电子温度剖面，二维数据，2012年12月后的数据没有修正，最近准备修正

等离子体参数相关-PlasmaPara

□ ElectronDensity

- 中平面上的电子平均密度随时间演化，一维数据，数据由HCN测量得到。存在跳变的炮，跳变没有修正，使用时注意，将来会对此进行处理

□ Vloop

- 双次平滑之后的等离子体环向电压，随时间演化，已经根据使用的OH电源配置不同进行了修正

二级加热相关-Heating

□ NBI_Power

- 注入到等离子体中的NBI的加热总功率随时间演化
- NBI没有加入的炮中，HDF5中此项不存在。
- 根据近似公式计算得到，最近将要进行相关修正

□ ECRH_Power

- 注入到等离子体中的ECRH的加热总功率随时间演化
- ECRH没有加入的时候，HDF5中此项不存在。
- 直接使用使用的是ECRH组提供的数据

加料相关-Fueling

□ GasPressure_at_Divertor

- 在偏滤器位置测量得到的气压，一维数据

□ GasPressure_in_VV

- 在真空室测得的气压，一维数据

□ GasPuffing

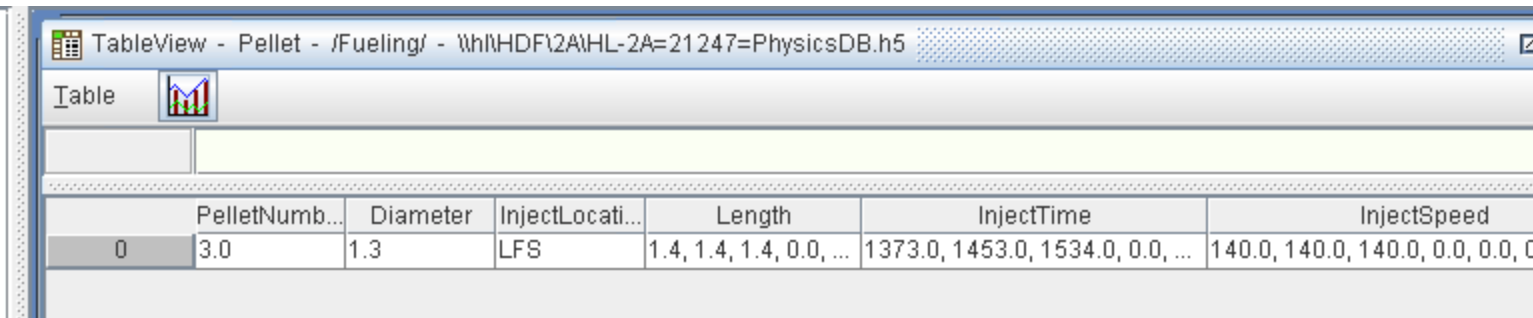
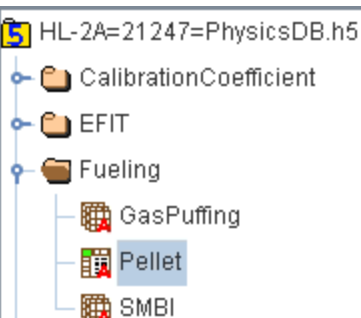
- 普通送气的波形，一维数据，此数据必定存在

□ SMBI

- 分子束送气记录，一维数据

□ Pellet

- 最近准备加入实时记录模块，并协商争取将历史数据加上



	PelletNumb...	Diameter	InjectLocati...	Length	InjectTime	InjectSpeed
0	3.0	1.3	LFS	1.4, 1.4, 1.4, 0.0, ...	1373.0, 1453.0, 1534.0, 0.0, ...	140.0, 140.0, 140.0, 0.0, 0.0, 0.0, ...

磁相关-Magnetic

□ BT

➤ 一维数组，环向磁场

□ Tau_E_ Antimagnetic

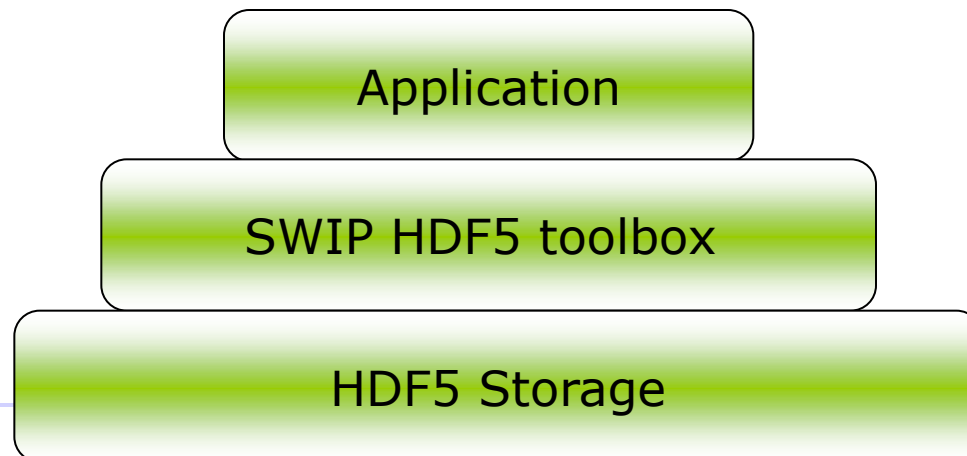
➤ 反磁测量得到的约束时间

直接用Matlab使用HDF5

```
1 - HDF_File = '\\192.168.10.11\hdf\2A\HL-2A=11617=PhysicsDB.h5'
2
3 %%
4 - DesiredTime = 710; %ms
5 - time = h5read(HDF_File, '/EFII/EFII_TimeAxis');
6 - index = find(time == DesiredTime) %
7
8 %%
9 - data = h5read(HDF_File, '/EFII/EFII_Q_at_MidPlane');
10 - qprofile = reshape(data(700, :, :), 2, 100);
11 - plot(qprofile(1, :), qprofile(2, :), '*-')
12 - xlabel('t/ms', 'fontsize', 14, 'fontname', 'Times New Roman', 'FontAngle', 'italic');
13 - ylabel(' ', 'fontsize', 14, 'fontname', 'Times New Roman', 'FontAngle', 'italic');
14 - title('11617炮第700msQ剖面');
15 - grid on
16
17 %%
18 - tt = h5read(HDF_File, '/EFII/EFII_LCFS_Points');
19 - xx = tt(500, :, :);
20 - curlen = size(xx);
21 - curlen = curlen(3);
22 - xx = reshape(xx, 2, curlen)
23 - len = xx(1, end);
24 - xx(:, len:curlen) = [];
25 - plot(xx(1, :), xx(2, :), '*-')
26 - xlabel('R/m', 'fontsize', 14, 'fontname', 'Times New Roman', 'FontAngle', 'italic');
27 - ylabel('Z/m', 'fontsize', 14, 'fontname', 'Times New Roman', 'FontAngle', 'italic');
28 - title('11617炮第700ms LCFS');
29 - grid on
```


SH Matlab工具包的开发原因

- 虽然在matlab中使用HDF5非常方便，门槛非常低，但是在物理数据库的使用过程中，发现有必要提供一些通用的matlab函数以实现常规的数据读取，将底层具体的存储方式和内容加以屏蔽，使得相关的使用人员可以不用熟悉里面的具体格式，从而再次降低其使用门槛。
- 同时，因为目前的数据存储格式并没有完全地定型，将来可能根据需要进行相关的修改，一旦进行修改，就需要通知所有的使用的人员，并让他们去修改对应的程序，会给相关的工作带来困扰。而通过提供工具包的方式，在保证函数名（接口）不变的情况下，只更新这个工具包，就可以在不修改上层程序的情况下，达到更新的目的。



SH Matlab工具包

□ 工具包中函数的命名方法

- 为了和matlab中其他的工具包的函数区分开来，在这个工具包中，所有的函数前面都加以前缀SH_，SH是SWIP HDF5这两个单词的首字母，表示此工具包只针对于目前SWIP中的HDF5数据库。

□ 关于用户验证

- 在前面已经提及如何直接在SWIP局域网内进行用户验证，有几个方面在使用这个工具包之前需要加以说明：
- 1、本工具包默认在使用之前已经进行了服务器的登陆验证，并且默认的地址是192.168.10.11；
- 2、本工具包不提供缺省的认证，此目的是为了保证只有拥有swip授权的人员才能够访问此数据库；

□ 其他说明

- 1、如果将hdf5文件拷贝到本地，不通过读取服务器文件的方式，可以修改工具包中的SH_Get_HDF5_FileName文件，将读取文件的位置转到本地硬盘上。
- 2、函数名区分大小写。
- 3、下载地址：\\192.168.10.11\hdf\SWIP-HDF5-Matlab

SH Matlab工具包

□ SH_Get_CC_DPF

- 得到指定炮号的中央控制的配置信息，DPF为HL2A中的中控配置文件中控人员称呼配置文件

□ SH_Get_1D_Data

- 如果信号是一维数据，则得到的是指定炮号的任意一个时间点的值，一维数据一般情况下指的是随时间变化的曲线

□ SH_Get_LCFS_Points

- 得到LCFS的上的点的坐标。

□ SH_Get_Psi_RZ

- 得到指定炮号，指定时间的磁通剖面

□ SH_Get_Q_Profile

- 得到指定炮号，指定时间点的中平面上的Q剖面

□ SH_Get_Te_Profile_MECE

- 得到指定炮号，指定时间点的利用MECE测量得到的电子温度剖面。

- 说明：此工具包随着工作的开展将不断进行更新

炮间入库

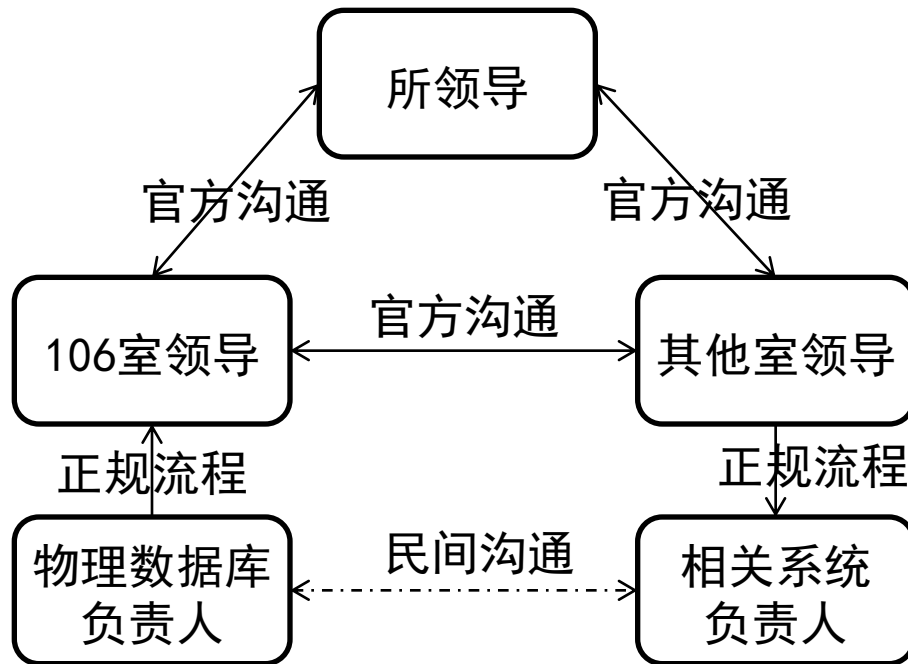
- 已经实现了炮间入库功能
- 目前已经实现在放电后四分钟内，现有数据全部入库
- 具有极强的扩展性
- 相关实现方法在将来的报告中会做介绍

非技术层面的几个考虑

- 从历史和现在的情况来看，物理数据库的工作开展更依赖于非技术问题的解决，非技术层面的问题主要有以下几个
 - 行政支持问题
 - 数据来源问题，这是目前最核心的问题
 - 数据的认证问题
 - 数据的版权问题
 - 出图问题

行政支持问题

- 此工作已经得到所长等相关领导的肯定和支持
- 物理数据库涉及室与室之间的相互配合，目前的所内合作模式是这样的：



- 希望各级领导能大力支持，对物理数据入库问题的解决给以充分的重视，并在相关的事务上给予绿色通道

数据来源问题

- 目前入库的内容是本人感兴趣以及能够提供的内容，同时作为相关技术的探索，覆盖面太窄
- 希望各位同事、科研骨干、课题组长、室领导能够将工作中想得到的物理数据、希望保存的数据、有价值的信息等等能够提到台面上来，心态开放，群策群力，互利互惠，将此数据库做大，方便其他人的同时也方便自己
- HDF5可以存储多个格式的数据
- 方法总比问题多

几种想法

❑ 你又用不到这个数据，要这个东西干嘛

- 长尾理论（The Long Tail）是网络时代兴起的一种新理论，由美国人克里斯·安德森提出。长尾理论认为，由于成本和效率的因素，当商品储存流通展示的场地和渠道足够宽广，商品生产成本急剧下降以至于个人都可以进行生产，并且商品的销售成本急剧降低时，几乎任何以前看似需求极低的产品，只要有卖，都会有人买。这些需求和销量不高的产品所占据的共同市场份额，可以和主流产品的市场份额相比，甚至更大。

❑ 数据公开了，系统的缺陷就暴露了，好丢人哦

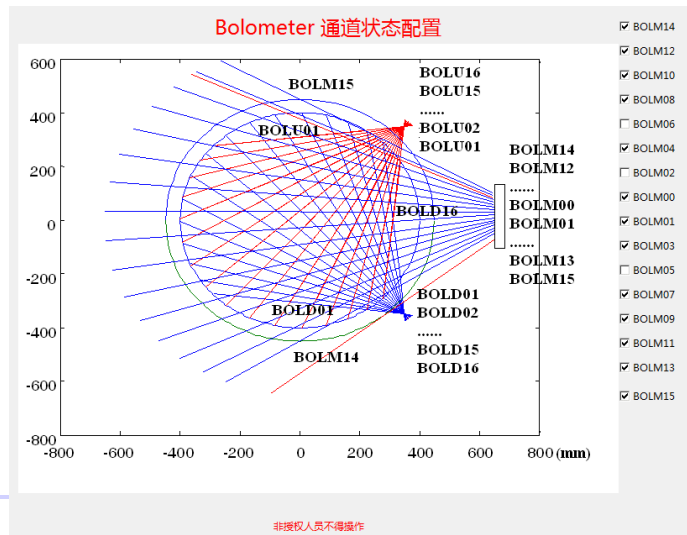
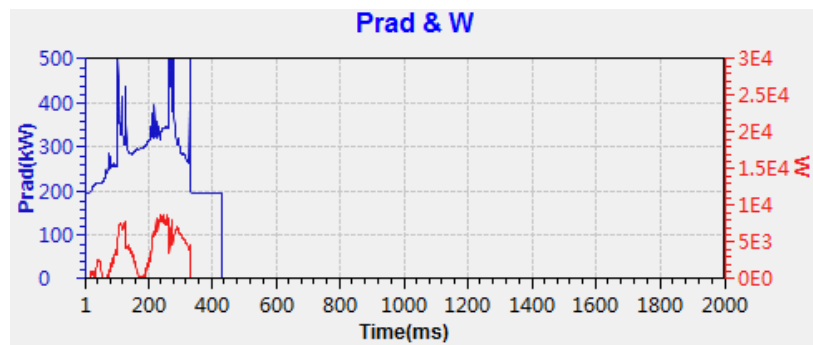
- 畏惧错误就是毁灭进步，遮掩错误就是躲避真理-周恩来
- 目前公开的采集数据物理数据里面都错误不断（包括离线efit），改正就ok了
- 心态开放，百事可成

关于数据的认证

- 谁都希望能够使用到正确的数据
- 我们希望对目前的数据进行认证，使得打了认证标签的数据可以判定是正确的。就像360来认证的软件，确定里面没有木马，垃圾广告之类，如同打了QS标签的食品可以放心食用，微博加V的人可以相信一样
- 数据完整性 是说数据的正确性、有效性、与相容性。所谓正确性是指数据的合法性；有效性是指数据是否在定义的有效范围内；相容性指同一事实的两个数据应相同。
- 数据的完整性由谁来判定？
 - 数据提供者？
 - 第三方机构？
 - 数据使用者？
- 当前的问题
 - 庞大的历史数据，13000余炮
 - 诊断算法并未成熟，
 - 诊断的通道并不是所有炮都是好的

- 系统负责人提供第一手数据和算法
- 入库系统根据数据和算法进行入库
- 程序按照一定的规则进行验证
- 使用者根据使用情况提出异议，数据提供者加以修正，同时判断的方式加入程序
- 判别算法不断得到提升

- 1、根据系统负责人提供的算法进行入库
- 2、发现数据存在问题
- 3、系统负责人分析为通道坏
- 4、提出解决方案



- 只有程序算法加以判断和自动修正

6、得到最新的解决方案和结果
==》程序判断，给出提示，相关修正

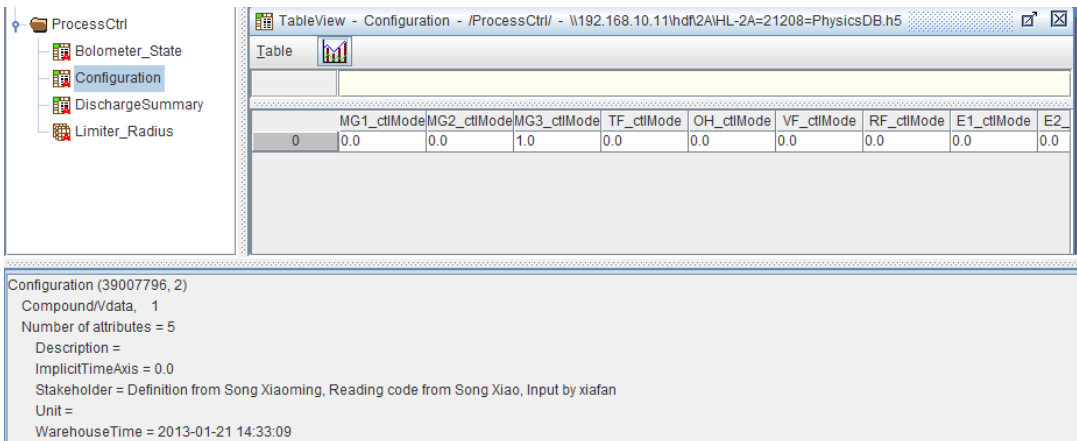
```

12: Monitor
HL 2A Automatic Alarm System
16:35:27 当前炮号 : 20870
16:35:27 =====
16:35:26 Bolometer第13道故障
16:35:26 FBC错误-没有投入或运行错误!
16:35:24 当前炮号 : 20869
16:35:24 =====
16:35:24 HCN第4道错误,需要修正
16:35:23 Bolometer第13道故障
16:35:22 FBC错误-没有投入或运行错误!
16:35:20 当前炮号 : 20868
16:35:20 =====
16:35:20 Bolometer第13道故障
16:35:20 FBC错误-没有投入或运行错误!
16:35:18 当前炮号 : 20867
16:35:18 =====
16:35:17 Bolometer第13道故障
16:35:17 FBC错误-没有投入或运行错误!
16:35:15 当前炮号 : 20866
16:35:15 =====
16:35:14 Bolometer第13道故障

```

数据的版权问题

- 物理数据来之不易，一套系统需要几年的建设
- 吃水不忘挖井人
- 在引用数据的同时，请署上数据来源相关人的名字
- HDF5物理数据库中每个物理数据均存在一个属性：Stakeholder，用以记录与数据相关的人的名字供参考，如有遗漏，请联系



The screenshot displays a software interface with a table view and a details pane. The table, titled 'Table', shows configuration data for a 'Configuration' object. The details pane below the table provides specific information about the selected configuration.

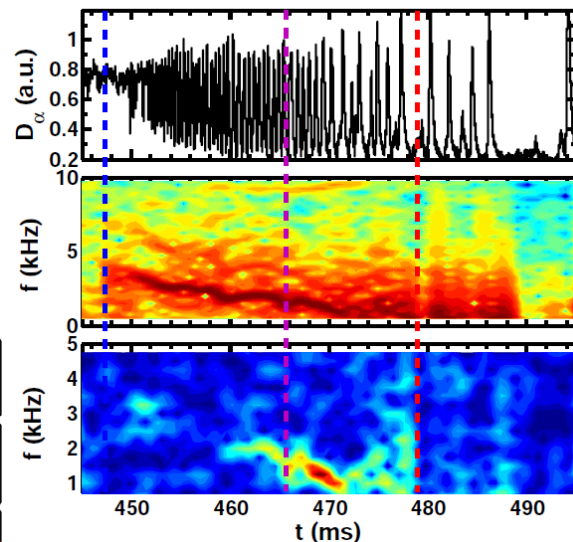
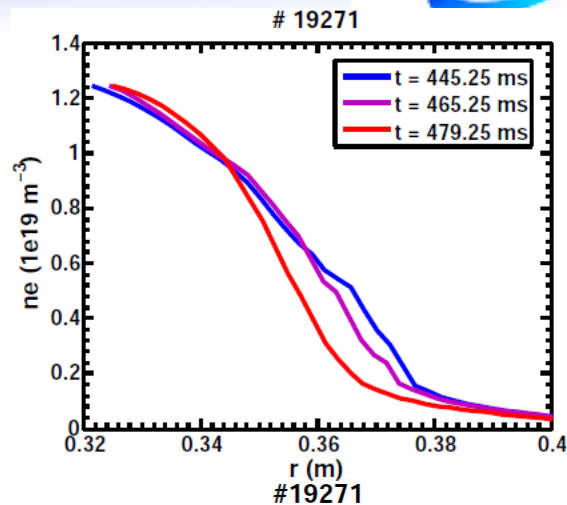
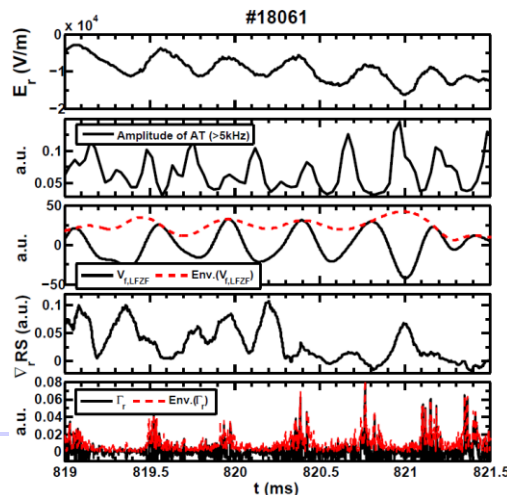
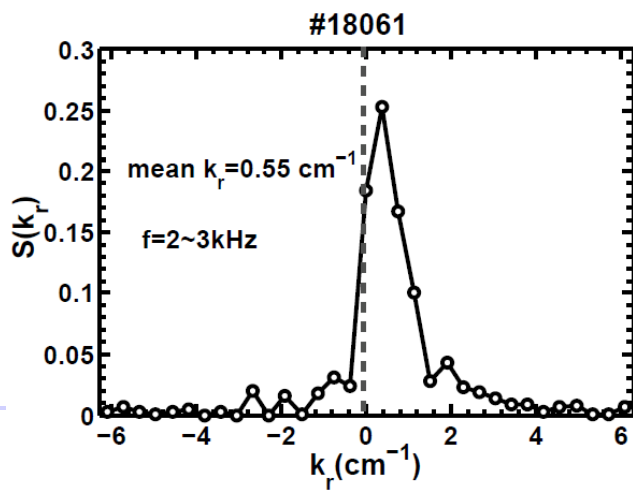
	MG1_ctlMode	MG2_ctlMode	MG3_ctlMode	TF_ctlMode	OH_ctlMode	VF_ctlMode	RF_ctlMode	E1_ctlMode	E2
0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

Configuration (39007796, 2)
Compound/Vdata, 1
Number of attributes = 5
Description =
ImplicitTimeAxis = 0.0
Stakeholder = Definition from Song Xiaoming, Reading code from Song Xiao, Input by xiafan
Unit =
WarehouseTime = 2013-01-21 14:33:09

- 这是一个纯粹的学术道德要求

关于出图

- 出图很重要
- 有数据就可以出图，没数据一切都是空谈，因此，现阶段的主要精力集中在数据的收集和整理上
- 出图需要根据各研究领域特殊的显示方式
- 希望各领域的同事能够贡献相关的出图代码，特别是以Matlab语言编写的代码，成为工具箱的一部分，使得出图部分能够得到有效共享



(这些图为从102同事的ppt中截取)

下一步工作

- 目前的工作只是一个开头，得到相关同事的大力支持，在此表示感谢；
- 下一步工作的将集中在数据的收集、验证的算法、工具箱的优化、相关应用的开发等方面
- 希望各位同事能够提出在数据方面的需求，从而不断完善，促进此数据库的进化
- QQ群：118950836，相关最新消息将在此发布，随后以文档格式保存
- 希望本报告能起到抛砖引玉的作用

谢谢！