

高能物理与核物理领域面向对象软件技术的发展*

叶沿林 应 军 陈 陶

(北京大学技术物理系 北京 100871)

摘 要 介绍了高能物理与核物理领域引入面向对象软件技术的背景,报道了 CERN 和美国的几项重要计划的进展情况,对有关技术的优劣与前景进行了评价.

关 键 词 面向对象 软件技术 高能物理与核物理

分 类 号 TP316

1 引 言

面向对象(Object-Oriented)是程序设计与实现的一种新模式,它包括抽象、封装、继承与多态性等基本属性.面向对象的软件具有开发效率高、易于维护和扩展、结构接近实际事物等优点,因而已成为90年代大型软件系统设计与制作的主要方向.以程序语言为例,C++就是一种具有面向对象特性的语言,可以用它来开发面向对象软件.

过去几十年,高能物理与核物理实验数据分析普遍采用 CERN 软件系统,如 ZEBRA、GEANT、PAW 和 HEPDB 等.这些软件主要用 FORTRAN 语言写成,自成体系,长期以来虽然通过加入新的层次与界面不断升级,但其基础始终是70年代的面向过程的程序模式.所谓面向过程,就是以程序代码为主体,程序中的数据被分离处理.代码可以在任何步骤上改变数据(如 FORTRAN 中的公用数据).这在大型程序中很容易造成混乱,使程序的维护与改造变得十分复杂与困难,比如在 GEANT 中^[1],增加一种探测器几何形状就需要修改约60个子程序,很难正确把握.另外,过程化程序语言的数据形式都比较单调,难以直接反应实际事物的特征.比如对一个物理粒子的描写,同时需要字符、整数和浮点数等,用传统的 FORTRAN 语言处理起来很别扭,而采用 C++中的结构或类

(Class),就能够自然地解决问题.

未来几年中,高能物理与核物理领域将有若干台大型实验装置投入运行,如欧洲的 LHC,美国的 CEBAF、RHIC、PEP-II,日本的 KEK-B 工厂和50 GeV PS,中国的 τ -Charm 工厂以及各国的若干台放射性核束装置等.由于能量的提高和流强(亮度)的成百倍增,实验数据的量(每年 10^{15} Bytes 量级)和复杂性都会达到前所未有的程度,这对数据模拟、分析和管理等提出了全新的要求.另外,理论计算中由于大量采用类似于物理模拟的方法,也要求软件技术有质的变化.这种要求与社会上面向对象软件技术的发展相吻合,促成了全国革新 CERN 软件系统的潮流.

目前这方面最具有代表性的,就是 CERN 的 GEANT4 计划(RD44)^[2]和持久性数据管理计划(RD45)^[3].美国的几个大国家实验室也在进行类似的软件革新.

2 GEANT4 计划

GRANT 是超过20万行和由1300多个子程序构成的大型软件,主要用于高能物理与核物理探测器的模拟.经过多年的扩展,该程序已变得十分庞大且难于管理.LHC 时代的模拟计算必须有更高的伸缩性和适应性.采用面向对象技术可以从根本上解决长期积累的问题,适应全新的要求.新的 GEANT 将立

* 国家自然科学基金资助课题.
1997 - 02 - 24收稿.

足于“任何用户均可以容易地将物理过程的模拟挂接到 GEANT 上来运行”,而不是传统的试图“使 GEANT 本身模拟所有已知的物理过程^[2]”。或者说,GEANT 将成为一种灵活方便的工具,而不是一个大而全的直接解决方案。举例说,用户将以一种简单的方式(利用面向对象的继承与多态性)加入新的探测器几何形状、新的物理过程、引入新的相互作用形式,以及动态地加载其他程序库等。

正是基于以上思路,GEANT4 研究计划 RD44 于 1994 年底提出并得到批准执行。有二十多个机构的几十位科学家参与此计划,其主要工作包括:

1) 用面向对象程序的类的概念重新建立数据模型,分别表示探测器几何与图形、粒子径迹、物理作用形式、事件的过程与历史、粒子种类、探测器材料性质、数值化信号等。

2) 用 C++ 语言重新开发全部核心程序。数据与代码将按面向对象的特性封装,从而易于维护与扩展,安全性好。当然,传统积累的数值计算方法仍将在程序中全面实现。

3) 使用商业化的用户界面。原来的 GEANT 有两种界面:用户程序界面和基于 X11 的自行开发的图形界面。GEANT4 将尽可能采用流行的窗口或网络图形界面(如 Netscape),使用户的模拟工作与其他计算环境融为一体。采用分享库(Shared Libraries)和动态连接(Dynamic Linking)等方法,使得在包含商用软件功能后 GEANT4 的执行程序的规模不至于太大。

4) 重新安排物理过程。传统上,GEANT 在内部处理电磁过程,而强过程需通过外部程序计算,如 GHEISHA、FLUKA 和 MICAP 等。这造成两类过程在数值方法的运用上不协调,用户难以理解和比较。同时,加入一种新的过程或修改作用截面等要牵涉到二十多个子程序,很难正确把握。在新的设计中,将把强过程的计算吸收到 GEANT 内部,按统一的面向对象模型处理,这样会使物理参数的变动和新过程的加入变得容易。另外,也可

以方便地将简化的快模拟程序植入到 GEANT 中。

按照“螺旋线模型”,GEANT4 开发两年来进展顺利,其进程大体可分为三个阶段:第一阶段,1995 年实现了所有类的定义,制作出了每一类的初步操作程序(Methods);第二阶段,1996 年产生了比较完整的工作样本并用于实验检测(LEP、HERA)。探测器几何与径迹跟踪的部分已用 C++ 语言重新制作,功能与原有 GEANT 相当。但物理过程的描写与数据的储存管理还是用旧的系统;第三阶段预计到 1998 年底完成,到时所有核心程序都将用 C++ 写成,并有完整的界面与外部的粒子产生程序、重建程序和储存管理程序等相联结。GEANT4 完成后,预计还需要 2~4 年时间在高能物理和核物理界推广使用,并逐步淘汰旧式的 GEANT 软件。

3 持久性对象的储存与管理

通常,将处理过程结束后仍然存在的对象,称为持久性对象(Persistent Object)。实验和分析产生的谱数据、刻度数据、探测器几何描写、过程控制信息、事件描述以及事件数据本身,均构成持久性对象。由于高能物理与核物理持久性数据的量大、结构复杂并需动态使用处理,传统的数据模型和文件技术达不到要求。为此,在 GERN 系统中曾经专门开发了动态管理系统 ZEBRA^[4]。用 ZEBRA 可以定义出非常复杂的数据结构(如链状和树状结构),但它毕竟是 70 年代的产物,是从外部勉强地去补充 FORTRAN 语言的不足,当然不如现代结构化语言(C、C++)处理复杂数据那样自然和方便。一方面,ZEBRA 本身只是一个较低级的系统。用户必须在相当程度上直接控制储存安排,这在涉及众多子程序和大量数据转换的情况下很容易出错,且难以比较检查;另一方面,在 ZEBRA 基础上发展的较高级的系统,如 HEPDB,又脱离了商用软件发展的标准和主流,缺乏广泛认可

的概念与层次. 为了从根本上解决这些问题, CERN 提出了“高能物理中的持久性对象管理”的研究计划(RD45)^[3], 于1995年初批准. 最初参加这项计划的有十几个单位的几十名科学家, 这两年队伍一直在不断扩大.

RD45要解决的首要问题, 是选择什么系统取代 ZEBRA 文件系统作为持久性数据管理的基础. 经过1995年的工作, RD45合作组在1996年2月得出结论: “商业化的面向对象数据库系统(ODBMS)与大容量的储存设备相结合, 是解决 LHC 时代持久性数据问题的最佳方案^[5]”. 这实质上是宣布未来的数据处理在一定程度上是“基于数据库的计算^[6]”. 目前社会上通用的是关系数据库^[7], 它具有持久性管理、并发控制、事务管理、错误恢复、查询管理、版本安排、整体自洽性、安全控制等标准功能, 但数据形式和结构比较简单, 难以用于实验数据处理. 现在正在迅速发展的面向对象数据库, 在传统数据库的标准上, 增加了面向对象的功能, 具有较接近于实际的数据抽象、继承性以及指针化的对象标识等, 适于管理复杂数据, 并且与面向对象语言(C++等)自然相容^[8], 预计2000年前后, 单个 ODBMS 的容量可达1TB(10^{12} Bytes). 将多个这样的数据库并行管理, 可达到 LHC 时代动态数据处理的要求. 根据多种评价指标, 目前 RD45选择商业化的 Objectivity/DB 和 O₂ 数据库管理系统作为开发高能物理与核物理应用系统的基础. 在此基础上已开发了初步的应用样本, 其顶层结构见图1.

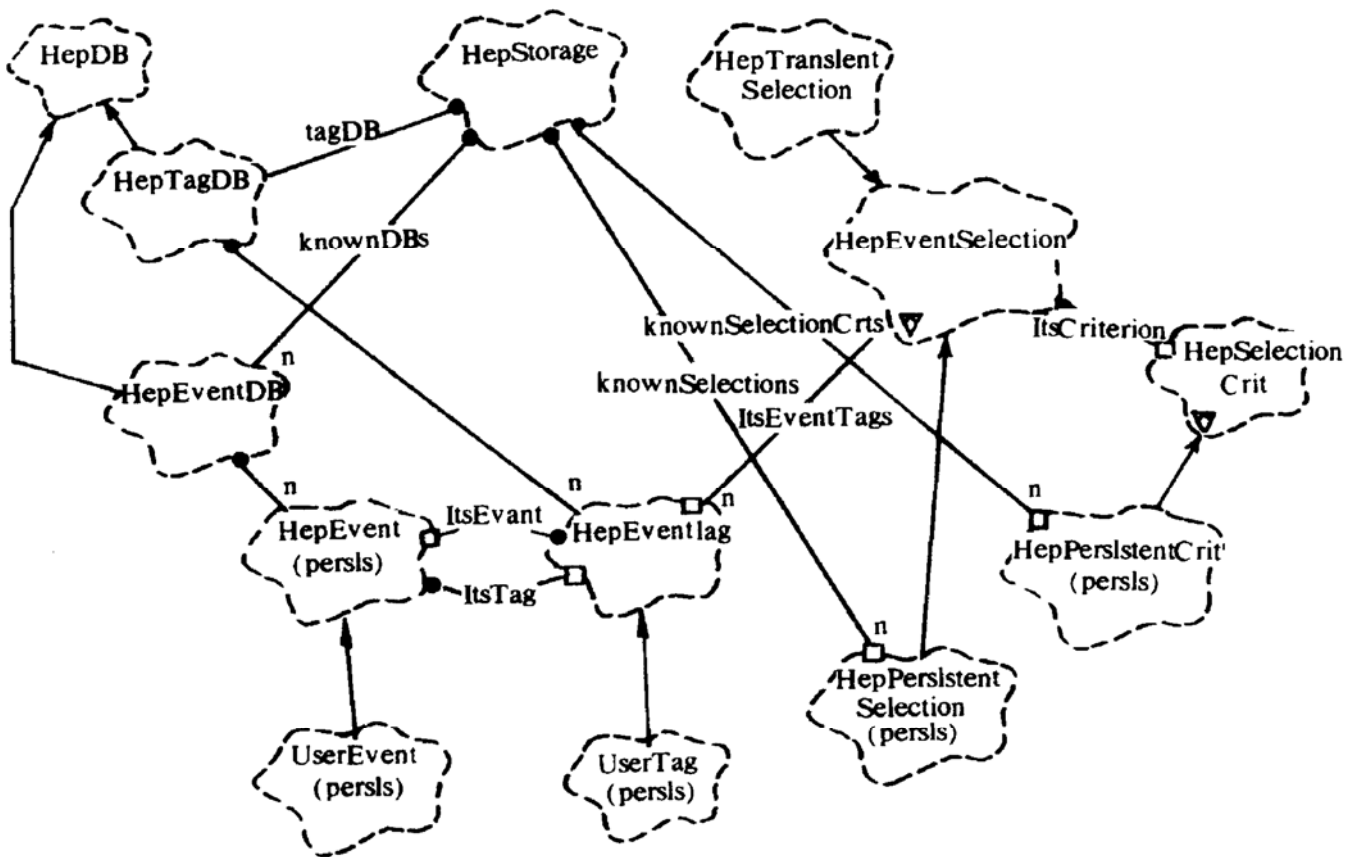
对于大容量储存设备, RD45期望使用美国的“高性能储存系统研究计划”(HPSS)的成果. HPSS 是由政府资助的, 由若干个国家实验室与大公司联合承担的, 其目的是发展下一代的符合 IEEE 标准的、分布式的以及在容量与性能上易于扩展的分级存储系统. HPSS 的主要指标为: 存储量达到若干 PB (10^{15} Bytes); I/O 速度达若干 GB/秒; Bit 文件容量达 2^{64} ; 多用户操作; 分布式多过程服务器管理等.

RD45在1996年已开始将 ODBMS 运用到具体实验环境下(NA45), 并开发了多种与其他系统的界面, 开始为 GEANT4 提供支持. RD45每年有两次以上的会议, 在高能界影响很大. 预计到2000年前后将完成完整的系统, 为 LHC 时代的持久性数据存储与管理提供长期的支持.

4 其他研究计划

在 RD44 和 RD45 的带动下, 高能物理与核物理领域中面向对象软件技术的应用广泛展开. 实际上, RD44 和 RD45 牵涉的问题非常普遍, 不是两个组本身可以解决的. 比如 CASE 工具、图形工具、网络工具, 还有分析设计模型(OOA、OOD)、用户界面、L/O 技术等. 为此, 在 CERN 发展了 LHC++ 整体计划, 除上述 RD44 和 RD45 外, 还包括了用 C++ 语言改写所有 CERN 库基本程序, 提供新的图示工具(HEPEXplorer)来取代 HBOOK 和 HPLOT, 以及建立通用的库来提供高能物理中常用的类及其操作函数(CLHEP)等.

在美国, SLAC 的 B 工厂的探测器 BARBAR 在建造中已全面采用 C++ 语言和面向对象数据库等来开发数据管理与分析软件. CEBAF 的 CLAS 合作组, RHIC 的 STAR 和 PHENIX 合作组, 很大程度都采用了面向对象技术. 特别是美国的几个主要的核物理国家实验室最近联合向美国能源部提出了“高能物理与核物理大容量数据管理与分析”的研究计划, 被称为是 Grand Challenge Application (GCA) 计划^[9]. 此计划认为: “把物理数据作为 C++ 中的对象处理比在 FORTRAN 中作为非结构化数据处理有效的多. 为此, 目前基于 CERN 库软件将不再发展甚至终止使用. 为 LHC 时代使用的软件都将采用 C++.” GCA 将研究如下问题: (1) 结构化的大容量存储系统; (2) 广域的数据流动; (3) 持久性对象管理系统; (4) 分布式与平行计算; (5) 用户界面与图形显示.



图示 RD45数据库管理系统应用样本顶层结构图

由于美国是计算机网络最发达的地方，因此 GCA 特别强调将地区的甚至桌面的系统与中心系统相联结，形成遍布全球的新一代分布式工作环境。这将大大改变未来实验数据的工作方式。

方法新技术开发应用软件，尽快使用外来的新的软件系统，以便顺利地过渡到 LHC 时代的新环境。

参 考 文 献

5 结束语

在 高 能 物 理 与 核 物 理 领 域 中 面 向 对 象 技 术 的 发 展 非 常 迅 速，几 年 之 后，人 们 将 面 对 全 新 的 计 算 环 境 与 应 用 软 件。传 统 上，我 国 的 高 能 物 理 与 核 物 理 基 础 软 件 均 是 从 国 外 引 入 的，实 验 工 作 者 曾 经 花 了 很 长 的 时 间 来 学 习 掌 握 这 些 软 件。一 旦 发 生 彻 底 的 变 化，我 们 在 短 期 内 可 能 很 难 适 应。为 此，在 建 设 或 改 造 实 验 设 施 时，一 定 要 充 分 考 虑 到 计 算 环 境 的 先 进 性，尽 可 能 采 用 新 的 国 际 标 准。另 一 方 面，实 验 工 作 者，特 别 是 年 青 一 代，要 尽 早 采 用 新

- 1 CERN Program Library Long Writup W5013. GEANT, 1993
- 2 GEANT4: An Object-Oriented Toolkit for Simulation in HEP. CERN/RD44/DRDC/1994,58
- 3 A Perisitent Object Manager for HEP. CERN/RD45/DRDC/1994,59
- 4 CERN Program Library Entries Q100/Q101. The ZEBRA System,1995
- 5 RD45 Status Report. CERN/LHCC 96-15, 1996
- 6 Object Databases and Mass Storage System: The Prognosis. CERN/LHCC 96-17,1996
- 7 王能斌. 数据库系统. 电子工业出版社,1995
- 8 Khoshatian S. Object-Oriented Databases, J. Wiley &

Sons Inc, 1993

for High-Energy and Nuclear Physics. DoE Proposal,

9 Olson D. Data Access and Analysis of Massive Datasets

1996

Development of Object-oriented Software Technique in Field of High Energy and Nuclear Physics

YE Yanlin YING Jun CHEN Tao

(Department of Technical Physics, Peking University, Beijing, 100871)

Abstract The background for developing object-oriented software technique in high energy and nuclear physics has been introduced. The progress made at CERN and US has been outlined. The merit and future of various software techniques have been commented.

Key Words object-oriented software technique high energy and nuclear physics

(上接第105页)

are briefly described in the paper, especially, the trends in United States, Japan, Germany and some European countries in recent years. Furthermore, some brief reviews for the trends are presented. According to the international trends and the potential of our country, our several considerations are proposed.

Key Words radiotherapy heavy ion beam clinical treatment