第 1 章　应用介绍

千百年来，地震一直严重威胁着人类社会的安全，它具有突发性强、破坏性高、次生灾害严重等特点，可以在几秒或者几十秒内造成大量的房屋倒塌和人员伤亡，其破坏性堪比一场核战争。强烈的地震通常会产生各种次生灾害，如火灾、水灾、滑坡、泥石流甚至瘟疫，这些次生灾害进一步威胁着人类社会的生命财产安全。由于突发性强、破坏力高，地震不仅对一个地区甚至一个国家的社会生活和经济活动会造成巨大的冲击，对人们的心理也造成了重大的影响。

我国位于环太平洋地震带与亚欧地震带的交汇部位，地震断裂带十分发育。中国地震局统计表明，我国陆地地震约占全球陆地地震的 33%，地震死亡人数占全球地震死亡人数的 50% 以上，是一个震灾极其严重的国家[1]。二十世纪以来，我国共发生 6 级以上地震近 800 次，涉及近 30 个省份，灾害面积达 30 多万平方千米，房屋倒塌超过 700 万间，死亡人数超过 50 万，占全国各类灾害死亡人数的 54%。其中最严重的是 1976 年河北省唐山市发生的 7.8 级大地震，直接造成了 24 万人死亡，16 万人重伤，一座重工业城市毁于一旦，直接经济损失超过 100 亿元[1]。

地震灾害对人类社会的毁灭性破坏推动着学者们孜孜不倦地研究地球内部——地震学。地震学起源于人类抵制地震的需要。公元 132 年，东汉著名天文学家张衡设计了地动仪，这是中国历史上最早的抗震减灾科学工作之一[2]。如图 1.1所示，地动仪有八个方位，每个方位上各有龙头口含龙珠，每个龙头下方有一只蟾蜍张口等待龙珠掉落。任何一方如有地震发生，指向该方向的龙头口中的龙珠会落入蟾蜍口中，由此便可推断出发生地震的方向[3]。

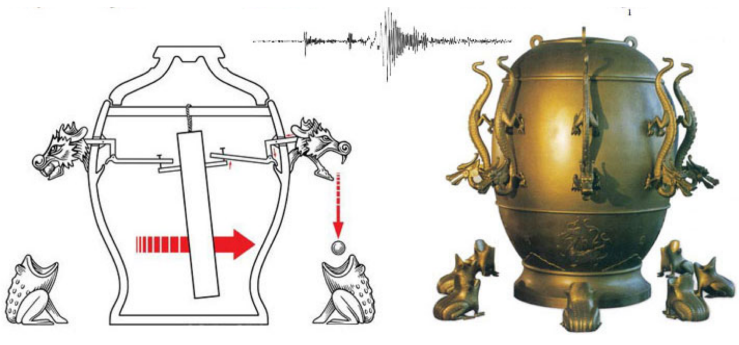

图 1.1　张衡在公元 132 年设计的地震仪的内部结构和外观[4]。发生地震时，指向该方向的龙头口中的龙珠会落入蟾蜍口中。

随着 20 世纪科技的迅猛发展，人们发明了天然地震信号采集和测量设备，能有效地记录地震波在地球内部传播的信息。基于统计的方法逐渐成为地震风险分

析和预报的主要工具。地震数值模拟工具就像一个“数字地动仪”，通过模拟地震波在地球内部的传播和地面运动，可以让科学家“模拟”甚至“预报”地震，也可以为地震相关风险提供定量评估，进一步理解地球内部结构以及地震的演变机制。此外，地震模拟工具还能够与其他科学模型（如交通、电力、水力模型）进行耦合，为地震活跃地区建立抗震救灾公用事业系统。

虽然地震数值模拟能为地震科学研究提供宝贵的“实验平台”，但它也是超级计算领域传统的“巨大挑战”。一次典型的地震模拟通常需要覆盖几百万立方米的空间范围（水平面数百公里、深度数十公里），即使计算网格空间分辨率超过 100 米，也会涉及数十亿到数万亿的未知数[5]，这在过去几乎是不可解的问题。直到最近的二三十年里，随着高性能计算的迅猛发展，基于超级计算机的地震数值模拟工具才成为了科学家研究地震的主要工具之一。

大规模地震模拟的研究工作可以追溯到 1996 年，Bielak 等人在 Cray T3D 的 256 个处理器上使用非结构化网格进行了区域大小为 $140km \times 100km \times 20km$ 的地震模拟[6]，性能达到了 8 Gflops。随后，日本和美国科学家相继在地球模拟器[7]、Jaguar[8]、Cori-II[9] 和 Titan[10] 等超级计算机上不断研制支持范围更广、分辨率更高、模拟更精确的地震模拟工具。与此同时，随着超级计算机的规模不断扩大、架构持续更新，科学家对模拟精度、时间和空间分辨率的需求也不断提升，研制面向十亿亿次异构架构超级计算机的地震模拟工具也面临着前所未有的巨大挑战。

第 2 章　算法简介

地震正演是模拟地震发生时地震波传播的数值方法。正演是地震模拟的核心，也是反演的基础。正演过程在每个时刻不断更新地震波场，因此伴随着巨大的计算量，也是地震模拟最主要的计算开销。正演算法的性能直接影响着地震模拟和反演方法的效率，提升正演算法的性能具有重要的意义，也成了许多研究员和学者孜孜不倦的研究课题[11,12]。

2.1　声波方程

声波方程法是目前使用最普遍的正演算法。与射线法不同，声波方程法抛弃了高频假设，采用更准确的双向波动方程描述波场在介质中的传播。声波方程具有计算量适中、模型描述简单和准确性高等优点。

无阻尼、无强迫力介质中的声波方程的一阶形式为：

$$\begin{cases} \frac{\partial}{\partial t}P(\mathbf{x},t) = -k(\mathbf{x})\nabla\cdot v(\mathbf{x},t) \\ \frac{\partial}{\partial t}v(\mathbf{x},t) = -\frac{1}{\rho(\mathbf{x})}\nabla\cdot P(\mathbf{x},t) \end{cases} \tag{2-1}$$

其中 $\mathbf{x}$ 表示空间位置，t 表示时间，P 和 v 是 $(\mathbf{x},t)$ 的函数，k 为介质的体积模量，ρ 为模型的密度。P 为标量，表示 $\mathbf{x}$ 处质点在 t 时刻的压强。v 为矢量，表示 $\mathbf{x}$ 处质点在 t 时刻的速度。公式2-1可以消去 v 得到关于 P 的二阶形式：

$$\frac{\partial^2 P(\mathbf{x},t)}{\partial t^2} - k(\mathbf{x})\nabla\cdot(\frac{1}{\rho(\mathbf{x})}P(\mathbf{x},t)) = 0 \tag{2-2}$$

若 ρ 为常数，则方程2-2可以进一步简化为：

$$\frac{\partial^2 P(\mathbf{x},t)}{\partial t^2} - c(\mathbf{x})^2\nabla^2\cdot P(\mathbf{x},t) = F(\mathbf{x},t) \tag{2-3}$$

其中 c 为介质的声速，$F(\mathbf{x},t)$ 为额外引入的强迫压力项。一般情况下，方程2-3没有解析解，只能利用数值方法求解。

偏微分方程的定解条件还包括初始条件和边界条件。一般情况下初始条件为 $P(\mathbf{x},0)=0$，表示初始无振动。随着时间的推移，波源以 $F(\mathbf{x},t)=\delta(\mathbf{x}-\mathbf{x}_s)f(t)$ 的形

式加入到介质空间 $\mathbf{x}_s$ 中，其中 δ 为脉冲函数。脉冲函数有不同的形式，其中最常用的为 Ricker 子波。

边界条件为另外一个非常重要的定解条件，其中最简单的边界条件为自由反射边界，但并不符合实际地震模拟中的情景。地震数值模拟只对有限的区域空间进行模拟，而实际地震波却会无限制传播，因此常使用较复杂的吸收边界条件，如海绵吸收边界[13]、完美吸收边界[14]。

2.2 弹性波方程

弹性波方程是求解速度和应力张量耦合的偏微分方程，令

$$v = (v_x, v_y, v_z) \tag{2-4}$$

表示波场中质点的速度向量，且令

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \tag{2-5}$$

为对称应力张量，则弹性动力学方程为

$$\begin{aligned} \partial_t v &= \frac{1}{\rho} \nabla \cdot \sigma \\ \partial_t \sigma &= \lambda (\nabla \cdot v) I + \mu (\nabla \cdot v + \nabla \cdot v^T) \end{aligned} \tag{2-6}$$

其中，λ 和 μ 是 *lamé* 系数，ρ 是介质密度。公式2-6可以将速度向量分解成三个标量方程，应力张量分解成六个标量方程。

地震波在地球内部传播过程中能量会逐渐衰减，这也必须体现在数值模拟中。本研究使用了品质因子 Q_s 和 Q_p 分别量化 S 波（横波）和 P 波（纵波）的滞弹性衰减。对于频率较高的情况，岩石和土壤中的非线性的响应以及盆地内浅层沉积岩的非线性行为会成为重要考虑因素。为了适应这些非线性效应，本研究结合 Drucker-Prager 塑性[15]，得到如下产量应力方程：

$$Y(\sigma) = \max(0, c \cos\varphi - (\sigma_m + P_f) \sin\varphi) \tag{2-7}$$

其中 c 为凝聚量（cohesion），φ 为摩擦角度（friction angle），P_f 为流体压力（fluid

pressure），σ_m 为平均压力（mean stress）。产量函数也用以判断是否更新应力：

$$\sigma_{ij} = \sigma_m^{\text{trial}}\delta_{ij} + rs_{ij}^{\text{trial}} \tag{2-8}$$

其中 r 是由产量应力 $Y(\sigma)$ 计算所得，s_{ij} 是应力偏导。

公式2-6分解后的九个速度和应力标量方程可以用九个交错网格有限差分方程来近似。公式2-9近似了交错网格有限差分方程的时间导数：

$$\begin{aligned} \partial_t v(t) &\approx \frac{v(t+\frac{\Delta t}{2}) - v(t-\frac{\Delta t}{2})}{\Delta t} \\ \partial_t \sigma(t+\frac{\Delta t}{2}) &\approx \frac{\sigma(t+\Delta t) - \sigma(t)}{\Delta t} \end{aligned} \tag{2-9}$$

速度和应力的空间导数可以用统一的方程描述。令 Φ 表示速度或应力分量，h 表示速度或应力网格中等距的空间分辨率，则网格点 (i,j,k) 的空间偏导 $\partial_x\Phi$ 可以用如下有限差分方案近似：

$$\partial_x\Phi(i,j,k) \approx D_x^4(\Phi)_{i,j,k} = \frac{c_1\left(\Phi_{i+\frac{1}{2},j,k} - \Phi_{i-\frac{1}{2},j,k}\right) + c_2\left(\Phi_{i+\frac{3}{2},j,k} - \Phi_{i-\frac{3}{2},j,k}\right)}{h} \tag{2-10}$$

其中 $c_1 = 9/8$，$c_2 = -1/24$。公式2-10可用以近似每个速度和应力分量中的空间导数。

2.3 正演算法

有一系列数值方法能够模拟天然地震破裂和地震波传播过程，包括有限差分（finite difference）、有限元（finite element）、谱元（spectrum element）和有限体积（finite volume）方法。每种方法有各自最适合的应用场景，但从准确度、计算效率和编程实现难易程度综合考量，有限差分方法最适合用于地震波传播的数值模拟。

此处的正演算法是波动方程的数值解法，更狭义的指在规则网格上基于有限差分方法的波动方程显示时间解法。对于声波方程2-3的数值解法，本研究使用中心有限差分方法对二阶偏微分算子 ∇^2 进行数值离散。例如，网格点 (ix, iy, iz) 在 x 方向的中心差分形式为：

$$\frac{\partial^2 P(ix,iy,iz)}{\partial x^2} = \frac{1}{dx^2}\sum_{k=-N}^{N} C_{|k|}P(ix+k,iy,iz) + O(dx^{2N+1}) \tag{2-11}$$

其中：

$$\begin{cases} C_k = \dfrac{(-1)^{k+1}\prod_{i=1,i\neq k}^{N} i^2}{k^2\prod_{i=i,i\neq k}^{N}|i^2-k^2|} \quad (k=1,2,\ldots,N) \\ C_0 = -2\sum_{k=1}^{N} C_k \end{cases} \tag{2-12}$$

该格式是 $2N$ 阶精度差分。∇ 空间的其他分量也可以同理进行差分离散化。中心差分格式展开后为星形的计算，空间中一点的二阶微分值由该点上下左右前后各 N 个点以及他本身的值共同计算决定（如图2.1所示）。这种计算被称为 *Stencil* 运算，记为 ∇_s。

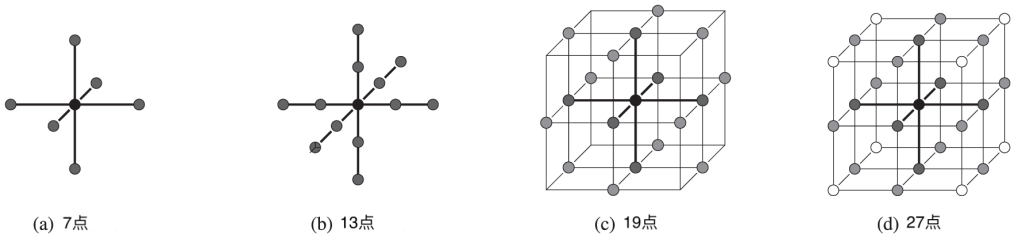

图 2.1　不同 stencil 算子的空间结构[16]。（a）7 点三维结构；（b）13 点三维结构；（c）19 点三维结构；（d）27 点三维结构。

在时间维度上，时间偏微分算子同样可用二阶形式的中心差分格式：

$$\frac{\partial^2 P(it)}{\partial t^2} = \frac{P(it+1)-2P(it)+P(it-1)}{dt^2} \tag{2-13}$$

于是方程2-3在空间和时间经过二阶形式的中心差分离散化后可写成：

$$P(it+1) = 2P(it) - P(it-1) + \frac{c^2}{dt^2}\nabla_s P(it) + F \tag{2-14}$$

公式2-14为经典的基于二阶时间精度有限差分的波动方程数值解法。其中 $P(it)$、$P(it+1)P(it-1)$ 分别表示当前时刻、下一时刻和前一时刻的波场。

算法1是均匀介质三维声波正演伪代码，为了方便描述，使用的是基于 2 阶的有限差分算子[17]。但是 2 阶有限差分算子会带来严重的频散，实际生产和研究中，常常使用 10 阶或者 12 阶差分算子。

算法1仅描述了在一个速度模型样本，单个震源或者震源编码情况下的地震波正演。在传统的多震源正演中，不同震源之间是相互独立的，可以完全使用不同的节点分别对不同的震源进行地震波模拟，相互之间没有通信开销。因此，提升

Algorithm 1 均匀介质三维声波正演伪代码

```
1:  for ( it = 0; it < nt; it++ ) {
2:      for ( ix = 0; ix < nx; ix++ )
3:          for ( iy = 0; iy < ny; iy++ ) {
4:              for ( iz = 0; iz < nz; iz++ ){
5:                  p(it,ix,iy,iz)=2*p(it-1,ix,iy,iz)-p(it-2,ix,iy,iz)+v(ix,iy,iz)*v(ix,iy,iz)*dt*dt*(
6:                                  p(it-1,ix,iy,iz)*(2/dx/dx+2/dy/dy+2/dz/dz)+
7:                                  (-p(it-1,ix,iy,ix-1)-p(it-1,ix,iy,ix+1))/dx/dx+
8:                                  (-p(it-1,ix,iy-1,iz)-p(it-1,ix,iy+1,iz))/dy/dy+
9:                                  (-p(it-1,iz-1,iy,iz)-p(it-1,iz+1,iy,iz))/dz/dz);
10:
11:                 if ( is_boundary(ix, iy, iz) ) //吸收边界处理
12:                         apply_absorb_boundary_condition(p);
13:             }
14:             if (is_record_seismo) //输出合成地震记录
15:                 save_seismo(p(it, ix, iy, :);
16:         }
17:
18:     if ( is_record_wavefield ) //输出波场快照
19:         output_wavefield(p(it, :, :, :));
20:
21:     for ( is = 0; is < ns; is++ ) //在正传波场中添加震源信号
22:         p(it, src[is].x, src[is].y, src[is].z) += wavelet(is, it);
23: }
```

正演算子的性能只需要关注单个样本、单个震源下的地震波传播模拟即可。地震波正演过程中，核心计算是在 $t = 0$ 到 $t = n - 1$ 时刻不停往波场中注射震源激励信号、更新波场（第1到13行），并根据需要输出合成地震记录或者正传波场。

正演数值算法也需要在模拟区域的边界处添加吸收边界条件（第11到13行），消除地震波在边界处的反射现象。

第 3 章　框架介绍

完整的大规模地震模拟工作流程十分复杂，本研究以模块化的方式开发了不同的组件，并以良好的接口将不同的组件耦合成统一的地震模拟软件框架（如图3.1所示）。它包括动态破裂震源生成器模块、地震波波场传播模块、三维模型插值与划分模块、震源划分模块、波场快照输出和重启模块。不同模块的设计原则如下：

- 计算最密集的模块：动态破裂震源生成和地震波波场传播模块。这两个模块的设计首要考虑该模块代码的可扩展性，使其能够高效扩展到神威超算百万甚至上千万核心，其次要考虑每个模块核心代码的运算效率，主要是提升有限差分运算的计算效率；
- 预处理模块：三维震源划分、三维模型插值/划分模块。这些模块根据动态破裂震源生成和地震波波场传播模块在不同平台的实现方式和并行规模对震源和模型进行不同的划分和插值，通过预处理来耦合动态破裂震源器和地震波波场传播模块，并以高效的计算、通信、IO 处理为地震波传播模块提供输出；
- 后处理模块：波场快照输出模块与重启模块。为了提高地震波场传播模块的效率，地震波传播模块以最高效的方式将波场快照和重启参数输出到磁盘中，然后调用波场快照输出模块与重启模块对输出的数据进行后处理，形成最终结果。

虽然高效的有限差分运算代码是提升地震波模拟性能的关键，但当计算规模扩展到成千上万甚至十几万进程时，通信和 IO 变得同样重要甚至更重要。因此本研究将涉及不同资源（计算、通信、IO）的任务进行模块化处理，并集成到统一的软件框架中。这是大规模地震模拟的基础。

3.1　动态破裂震源生成与地震波波场传播模块

软件框架中的动态破裂生成模块是基于 CG-FDM 软件[18] 进行二次开发而成。该模块具有初始化断层应力、执行摩擦定律控制以及通过波传播生成震源等功能。动态破裂震源生成器的输出是地震破裂震源，这可作为地震波传播模块的输入。地震波传播模块是基于 AWP-ODC[5] 软件并针对太湖之光的特殊体系结构进行了全新的设计。这个模块占据了主要的计算时间，也是并行优化和创新的重点。地震波传播的核心计算是求解速度-应力张量方程，主要流程包括速度更新、应力更新、

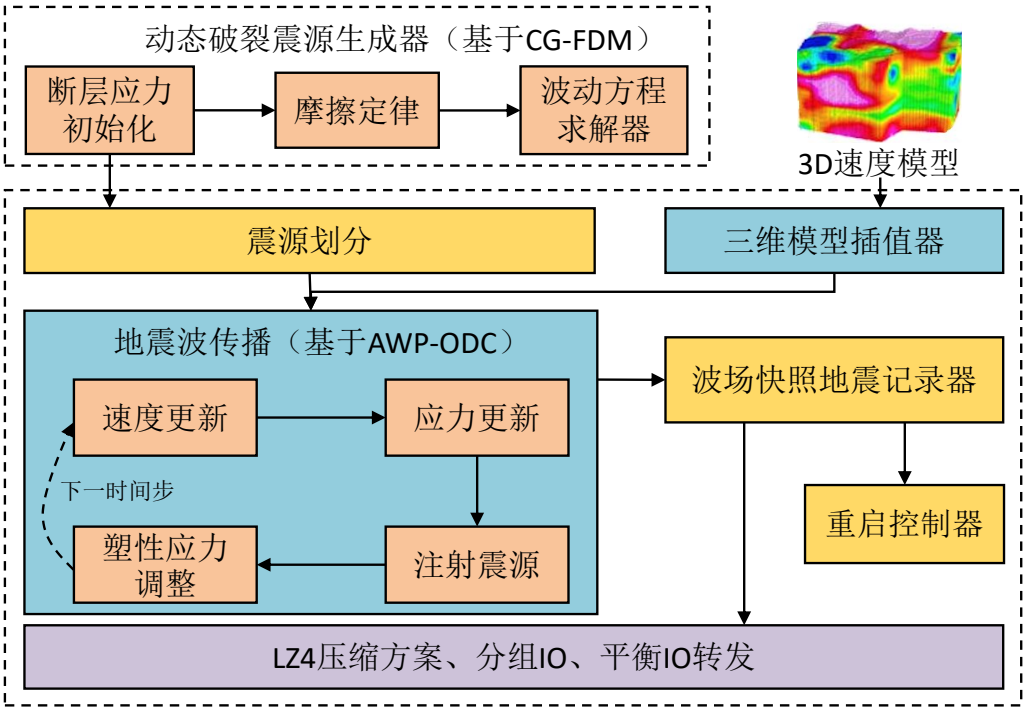

图 3.1　基于太湖之光的统一地震模拟软件框架。该框架由动态破裂震源生成器、数值计算网格的生成器、地震波传播模块以及其他辅助模块组成。

震源注入和塑性应力调整。

图 3.2描述了地震模拟的核心计算流程，每个核函数（kernels）都涵盖了 20-40 个三维网格。为了支持复杂的工作流程和促进未来可能的算法调整，本研究在重新设计和移植 AWP-ODC 的过程中将每个核函数封装为一个标准模块以便轻松地与其他模块衔接。

在典型的地震模拟中，每一个迭代步的工作流程都由以下五个主要步骤组成：

- 速度更新：基于上一个时间步的速度和应力数组，完成速度的更新，包括 Halo 区域（由核函数 *dvelcy* 完成）和中心区域（由核函数 *dvelcx* 完成）；
- 应力更新：使用速度、应力、Lam 系数和频率相关的地震波震级衰减参数 Q 等数组来计算 Halo 区域和中心的应力更新（由核函数 *dstrqc* 完成）；
- 断层应力调整：检查屈服应力是否越界（由核函数 *drprecprc_calc* 完成），并对断层区域进行相应的调整（由核函数 *drprecprc_app* 完成）；
- 震源注入：注入震源来更新当前时间步的波场（由核函数 *addsrc* 完成）。
- 自由表面应力调整：对自由边界进行应力调整（由核函数 *fstr* 完成）。

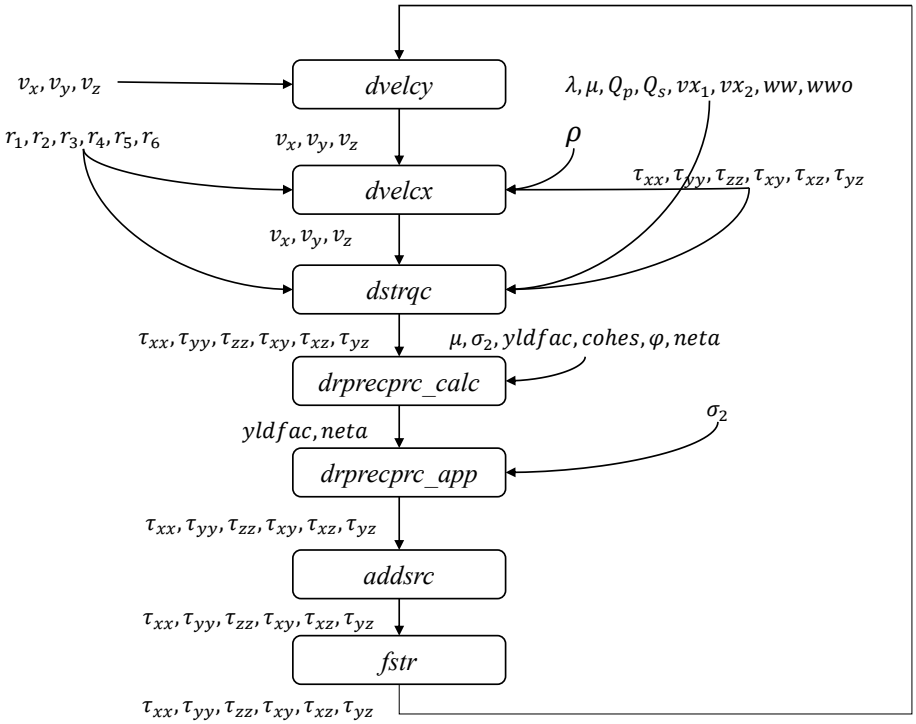

图 3.2　地震模拟的核心计算流程，其中变量以数组表示，计算以 kernel 表示。

3.2　并行震源划分与介质模型插值模块

在大规模地震模拟中，当震源和介质模型的文件规模逐渐上升到 TB 量级，地震模拟的预处理也逐渐成为整个软件的瓶颈，耗时不断上升，甚至因为内存不足导致程序无法运行。因此设计高效和高可扩展性的震源、模型划分软件是大规模地震模拟必要过程。本研究采用两种划分方式对震源进行划分：多进程串行划分方法和并行 MPI-IO 划分方法。

多进程串行划分方法按照既定的地震波传播并行划分规则，每个进程按照所处理的区域将单一的大型震源划分成若干小型震源，并输出到不同的文件中。这种方法能够充分利用神威超算系统的 IO 带宽。但由于动态破裂震源通常为模拟区域内的一个小范围曲面，震源位置非常集中，在划分的过程中会出现严重负载不均衡问题，降低了运行效率。此外，震源震动时间长、局部性高，容易导致位于震源处的进程所需内存超过系统最大内存。本研究在按空间划分的基础上再按照时间进行划分，有效的避免了划分时部分进程内存不足问题。

并行 MPI-IO 方法并不对大震源物理切分成若干个小文件，而是在地震传播模块中采用并行 MPI-IO 接口读取大震源的不同部分。理论上，MPI-IO 库在调用良好的情况下能够提供可观的带宽和扩展性，但在神威超算系统中，并行 MPI-IO 接口却展现出非常糟糕的带宽和扩展性。本文作者猜测原因是移植的 MPI-IO 标准库并未针对神威超算的特殊架构进行深度优化。

因此从效率和可扩展性出发，并行震源划分采用了多进程串行划分方法。虽然该划分方法会导致了负载不均衡，但划分过程的效率主要取决了 IO 带宽，处理密集震源的核心几乎能够利用全部 IO 带宽，因此并不会因为负载不均衡而严重影响效率。

实际介质模型通常不大，在进行多尺度地震模拟时，介质模型需按照地震模拟的网格分辨率进行插值，得到新的介质模型。在大规模地震模拟中，再使用并行划分方法为每个进程构建输入。为了简化流程，提高效率，本研究直接在地震波传播模块中集成插值模块，将实际介质模块插值得到目标模型，可省去将插值后的介质模型存储到磁盘的中间环节。

3.3 波场快照输出与重启模块

地震波场快照是地震模拟结果之一，是可视化的必要输入数据。本研究提供了灵活的波场快照输出方式，用户可以指定任意范围任何分辨率的波场快照输出，且同时支持不同分辨率的快照输出。低分辨率的波场快照可用于地震模拟可视化，高分辨率的波场和地震记录可精细分析地面运动和烈度。

大规模的地震模拟伴随着巨大的计算量，即便拥有大型超算的支持，计算时间仍然可能需要十几个小时甚至几十个小时。更多计算核心的参与也意味着发生硬件和软件错误的概率更大。这种情况下，软件的容错能力则显得非常重要。例如，在神威超算全机模拟中，有超过一千万核心参与运算，四小时内发生几乎会发生一次硬件或软件错误，如果没有容错机制，整个地震模拟过程需要重新计算，严重地浪费了之前四小时的计算。不具备容错机制几乎无法完成需要长达十几个小时的地震模拟。本研究设计了重启模块，在地震波传播过程中，定期将每个进程的内部状态变量存储到硬盘中（检查点）。当程序异常退出时，地震模拟可从最近的检查点启动以便继续模拟。

在波场快照输出与重启模块中，IO 是影响性能最大的瓶颈。例如在 16 米分辨率下，每个重启检查点需输出的内部状态变量超过 108 TB，这对超算系统的磁盘带宽和容量都是巨大的挑战。本文使用与第3.2节类似的多进程串行输出方法，每个进程将各自的串行输出到硬盘中，并调用后处理程序对数据进行合并。此外，我们还采用了 LZ4 压缩方法以降低输出数据量，降低存储开销，并采用诸如“组 I/O”和“平衡 I/O 转发”等技术，实现了 120GB/s 的峰值 I/O 带宽，达到了该文件系统峰值带宽的 92.3‰

第 4 章　性能与分析

4.1　计算核心优化结果

正演算法的计算热点比较集中，主要为速度更新和应力更新两个核函数。对于速度物理量的更新，计算中心区域和 Halo 区域的核函数 *dvelcxdvelcy* 是计算量最大的两个计算核心。对于应力物理量的更新，*dstrqc* 是计算量最大的计算核心。对于塑性部分，*drprecprc_calc* 是整个程序中最耗时的部分。剩下的内核包括 $fstr$，*drprecprc_app*，MPI 的预处理和后处理（*unpack_VY*，*gather_VX* 和 *unpack_VX*），这消耗了总运行时间的 1-2‰。本研究对上述的核函数都进行了极致的优化，以便实现最高的性能。

图 4.1演示了不同方法下核函数的性能和带宽优化结果。可以看到，几乎所有的核函数经过优化后都获得了 30 倍左右的性能提升，并且 DMA 传输带宽达到了总带宽的 70％到 80％左右。唯一的例外是 $fstr$ 核函数，$fstr$ 核函数的计算密度很低，内存访问不规整，只实现 4 倍至 5 倍的加速。从图 4.1还可看到，经过了一系列的优化后，不同核函数优化前和优化后在总执行时间内的耗时分布并无太大变化。

4.2　弱扩展性结果

弱扩展性表示当计算规模和计算核心同时增大时计算效率的变化。图 4.2描述了线性和非线性情况下的地震模拟程序的弱扩展性结果。在这两种情况中每个核组计算的网格大小均为 $160\times160\times512$，然后逐渐扩展到整个机器。本研究的大规模唐山大地震模拟从 8,000 个进程到 160,000 个进程下实现了 97.9% 的并行效率，几乎实现了完美的线性加速效果。这意味着该应用的通信策略非常高效，几乎能将通信完美隐藏在计算中。

无压缩的线性地震模拟在 160,000 进程下的峰值性能达到了 10.7 PFlops。非线性地震模拟的计算密度更高，相同进程下的峰值性达到了 15.2 PFlops。相同的内存带宽在采用实时压缩方案后能够处理更多的数据，将线性和非线性地震模拟的性能分别进一步提高至 14.2 Pflops 和 18.9 Pflops。

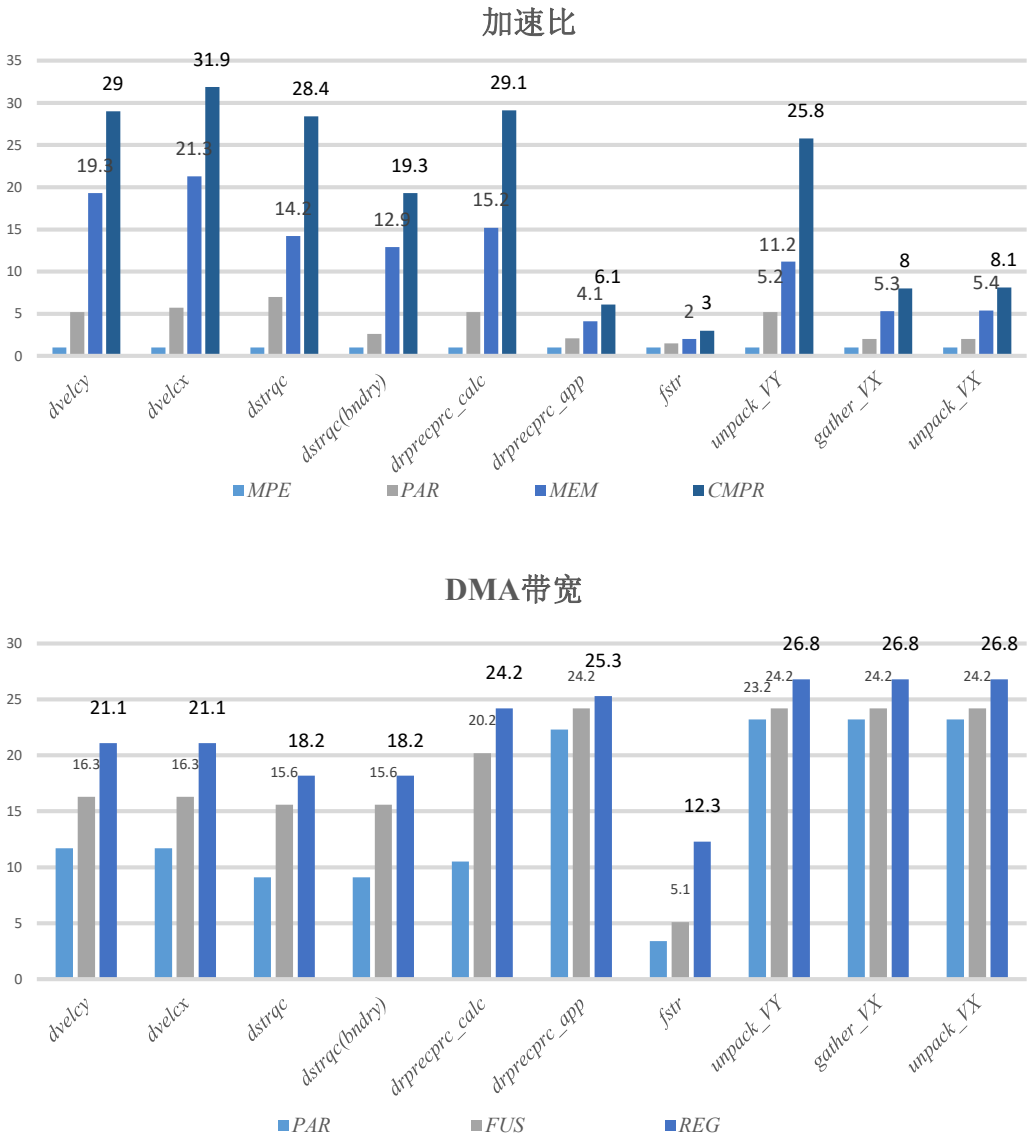

图 4.1　地震模拟中不同核函数采用了不同优化技术后取得的性能和内存带宽提升。‘MPE’代表仅使用主核的原始版本。‘PAR’ 是指使用了多层级并行划分方案并使用了 64 个从核并行计算的版本。‘MEM’ 是指采用所有与内存相关的优化的版本。“CMPR”是指进一步应用实时压缩方案的版本。

4.3　强扩展性结果

强扩展性指当问题规模不变时，增大计算核心数量带来的加速比。神威超算有 40,000 个节点，测试强扩展性的问题规模太小，则大规模并行时每个节点的任务量太小；问题规模太大，则小规模并行时节点内存可能无法容纳子问题。因此本研究针对不同的计算规模设计了三种不同网格大小算例。图 4.3显示了基于三种不同网格大小的线性，非线性，含压缩以及不含压缩情况下的强扩展性测试结果。

由图 4.3所知，不管是线性或非线性地震模拟，含压缩或者是不含压缩，地震模拟都达到了类似的强扩展性结果。但随着进程数量的不断增加，性能出现了下

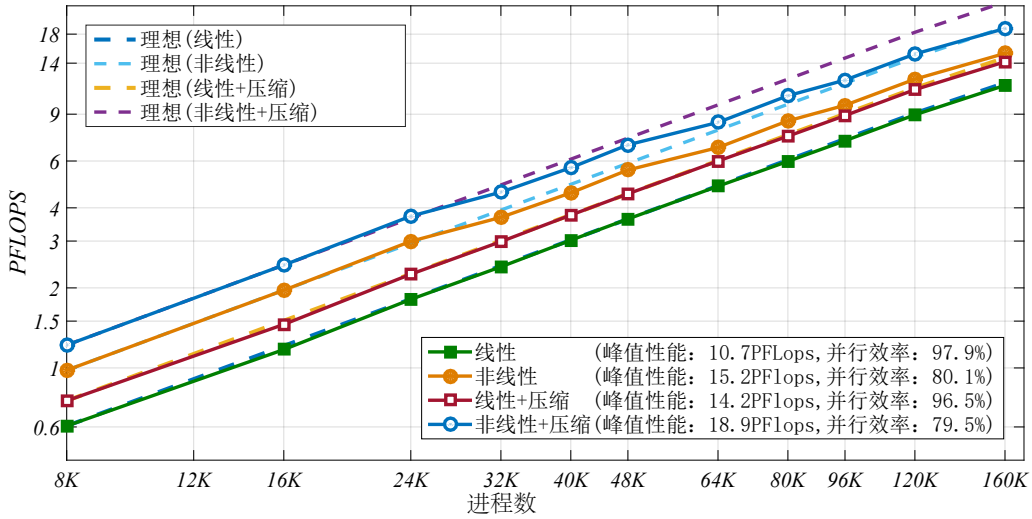

图 4.2　线性和非线性地震模拟从 8,000 个进程扩展到 160,000 个进程的弱扩展性结果，其中每个 CG 对应于一个 MPI 进程。

降。本文分析性能的下降是由两个方面造成的：（1）计算与通信的比例减小；（2）外部 Halo 区域与每个进程内计算的网格体积比例的减小，这降低了 AWP 软件的计算和通信重叠的效果，使得通信无法完全被计算隐藏。

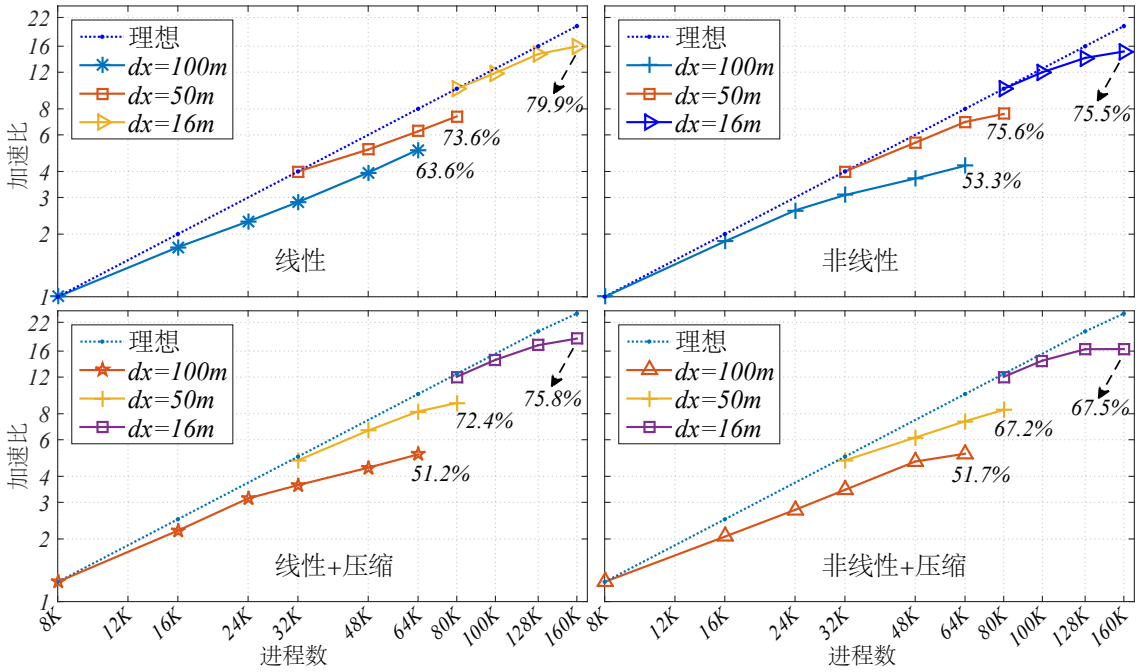

图 4.3　线性和非线性地震模拟在三种不同的问题规模中的强扩展性结果。测量的 MPI 进程从 8,000 个扩展到 16,000 个，其中每个 CG 上对应一个 MPI 进程。

第 5 章　具体需求

5.1　计算力需求

本组大地震模拟工作使用的数值解法是基于交错网格的有限差分方法，其核心是 *Stencil* 运算，这是典型的访存受限问题。因此计算并不是主要的瓶颈。

5.2　内存需求

内存的需求体现为三方面：内存容量需求、内存带宽需求和 LDM 大小需求。

5.2.1　内存容量需求

科学运算尤其是地震模拟对内存的需求是永无止境的。目前神威超算单节点内存只有 32GB，只有常规服务器内存的 1/4。由于单节点内存低，只能将任务划分到更多的节点进行计算，直接增大了节点间通信总量，在大规模并行模拟中通信会成为瓶颈。

除了单节点内存问题，全机内存也有进一步提升的空间。目前最大规模的地震模拟能模拟区域为 $320 \times 320 \times 40km$，空间分辨率为 $8m$，如果想要模拟更大区域或者更高的分辨率，只能依靠更高的全机内存总量。

5.2.2　内存带宽需求

申威 26010 处理器 4 核组的峰值内存带宽为 136 GB/s，受从核 LDM 大小和 DMA 批量读取块（block size）的限制，真实应用在极限情况下的内存带宽只有不足 100GB/s。这大约只有其他超算系统内存带宽的 1/4 左右。增大内存带宽是提升 *Stencil* 运算等访存受限问题的最关键手段。

5.2.3　LDM 大小需求

Stencil 运算中每个格点的计算需要访问该格点的邻居格点。*Stencil* 的形状不同、有限差分算子长度不同，所需要访问的格点数量也不同，在 LDM 中需要额外存储的数据也不同。对于复杂的 Stencil 运算，LDM 需要存储大量 Halo 区域元素，有效计算元素比率低，会导致 Halo 区域重复读取严重，严重影响效率。因此对于 Stencil 运算而言，LDM 空间越大越好。

5.3 通信需求

Stencil 运算中每个进程在每个迭代步都需要与邻居进程通信。当进程数量达到上万量级是，通信会成为主要的瓶颈。目前地震模拟使用了空间 4 阶的有限差分运算，如果想提高模拟准确率，则需要使用更高阶的有限差分算子，这会给通信带来严重的负担。更高的网络带宽和更低的网络延迟是高效大规模并行模拟的关键。

5.4 存储需求

唐山大地震模拟中极限规模下的地震记录、波场快照输出 450TB。地震模拟软件还具备断点重启功能，断点重启功能需要将全机内存中的内部变量存储到磁盘中，每个断点的大小约为 100TB，这给 IO 带宽和存储空间造成了巨大的挑战。

此外，神威超算的 MPI-IO 性能非常糟糕，几乎无法支持任何使用标准 MPI-IO 的程序，这给移植工作引入了许多工作量。

参考文献

[1] 中国地震局. 我国地震灾情有什么特点[EB/OL]. http://www.cea.gov.cn.

[2] Stein S, Wysession M. An introduction to seismology, earthquakes, and earth structure[M]. [S.l.]: John Wiley & Sons, 2009

[3] 百度百科. 地动仪[EB/OL]. https://baike.baidu.com/item/%E5%9C%B0%E5%8A%A8%E4%BB%AA.

[4] Hsiao K H, Hong-Sen Y. The review of reconstruction designs of zhang heng's seismoscope[J]. Journal of Japan Association for Earthquake Engineering, 2009, 9(4): 4_1–4_10.

[5] Cui Y, Olsen K B, Jordan T H, et al. Scalable earthquake simulation on petascale supercomputers [C]//High Performance Computing, Networking, Storage and Analysis (SC), 2010 International Conference for. [S.l.]: IEEE, 2010: 1–20.

[6] Bao H, Bielak J, Ghattas O, et al. Earthquake ground motion modeling on parallel computers [C]//Proceedings of the 1996 ACM/IEEE conference on Supercomputing. [S.l.]: IEEE Computer Society, 1996: 13.

[7] Chen Y, Alexandru A, Dong S J, et al. Glueball spectrum and matrix elements on anisotropic lattices[J]. Physical Review D, 2006, 73(1): 014516.

[8] Carrington L, Komatitsch D, Laurenzano M, et al. High-frequency simulations of global seismic wave propagation using specfem3d_globe on 62k processors[C]//Proceedings of the 2008 ACM/IEEE conference on Supercomputing. [S.l.]: IEEE Press, 2008: 60.

[9] Breuer A, Heinecke A, Cui Y. EDGE: Extreme Scale Fused Seismic Simulations with the Discontinuous Galerkin Method[C]//International Supercomputing Conference. [S.l.]: Springer, 2017: 41–60.

[10] Cui Y, Poyraz E, Olsen K B, et al. Physics-based seismic hazard analysis on petascale heterogeneous supercomputers[C]//Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. [S.l.]: ACM, 2013: 70.

[11] Bednar J, Neale G, et al. Limited vs. full frequency wave-equation imaging[C]//2002 SEG Annual Meeting. [S.l.]: Society of Exploration Geophysicists, 2002.

[12] Stork C. Eliminating nearly all dispersion error from fd modeling and rtm with minimal cost increase[C]//75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013. [S.l.: s.n.], 2013.

[13] Cerjan C, Kosloff D, Kosloff R, et al. A nonreflecting boundary condition for discrete acoustic and elastic wave equations[J]. Geophysics, 1985, 50(4): 705–708.

[14] Berenger J P. A perfectly matched layer for the absorption of electromagnetic waves[J]. Journal of computational physics, 1994, 114(2): 185–200.

[15] Roten D, Cui Y, Olsen K B, et al. High-frequency nonlinear earthquake simulations on petascale heterogeneous supercomputers[C]//Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. [S.l.]: IEEE Press, 2016: 82.

[16] Zhang Y, Mueller F. Autogeneration and autotuning of 3d stencil codes on homogeneous and heterogeneous gpu clusters[J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 24 (3): 417–427.

[17] Fu H, Clapp R G. Eliminating the memory bottleneck: an fpga-based solution for 3d reverse time migration[C]//Proceedings of the 19th ACM/SIGDA international symposium on Field programmable gate arrays. [S.l.]: ACM, 2011: 65–74.

[18] Zhang Z, Zhang W, Chen X. Three-dimensional curved grid finite-difference modelling for non-planar rupture dynamics[J]. Geophysical Journal International, 2014, 199(2): 860–879.