

# 应用于脑机接口系统的动态 稀疏矩阵压缩算法

高原雨<sup>1</sup>, 尤昌华<sup>2</sup>, 李朋<sup>3</sup>, 姚镭<sup>1</sup>

(1. 上海大学 微电子学院, 上海 201899;

2. 中国科学院空天信息创新研究院 传感器技术国家重点实验室, 北京 100094;

3. 山东信芯电子科技有限公司, 济南 250013)

**摘要:** 在脑机接口系统中, 高通道数神经信号采集是一个核心功能模块, 能够为外部计算机设备采集大量人脑中的神经信息; 在高通道数神经信号采集中, 因其原始数据量巨大, 直接传输和处理产生的原始数据会消耗极大的功耗并增加硬件设计上的难度; 为了解决这个问题, 一个有效的方法是在数据传输和处理前依据原始神经信号数据的特点对其进行压缩; 神经元动作电位信号具有不应期即有效信号的时域宽度与信号重复周期之比很小; 利用此特点, 能够将多通道神经信号的数字标记输出在一定时间范围内定义为一个稀疏矩阵, 并对此稀疏矩阵进行特征提取, 根据其特征动态地采用优化算法进行数据压缩; 所提出的算法在 Xilinx 平台使用 FPGA 进行设计与实现, 并且将其作为中控硬件在 32 通道神经信号采集硬件系统上通过实时验证, 实验证明提出的动态稀疏矩阵压缩算法可实现 83.4% 的数据压缩率。

**关键词:** 神经信号采集; 多通道; 稀疏矩阵; 数据压缩算法; FPGA

## Dynamic Sparse Matrix Compression Algorithms for Brain-computer Interface Systems

GAO Yuanyu<sup>1</sup>, YOU Changhua<sup>2</sup>, LI Peng<sup>3</sup>, YAO Lei<sup>1</sup>

(1. School of Microelectronics, Shanghai University, Shanghai 201899, China;

2. State Key Laboratory of Transducer Technology, Aerospace Information Research Institute,  
Chinese Academy of Sciences, Beijing 100094, China;

3. Shandong Xinxin Electronic Technology Co., Ltd., Jinan 250013, China)

**Abstract:** In brain-computer interface systems, multi-channel neural signal acquisition is a core functional module, which can collect much neural information in human brain for external computer equipment. In the multi-channel neural signal acquisition, because of the huge amount of original data, the generated original data is directly transferred and processed to take huge power consumption and increase the difficulty of hardware design. To solve this problem, an effective method is to compress the original neural signal data according to its characteristics before the data transmission and processing. The neuron action potential signal has the refractory period property, that is, the effective signal has the very low ratio of time domain width to repetition period. This paper adopts this feature to define the digital label output of the multi-channel neural signals as a sparse matrix in a certain time range, extracts the feature of this sparse matrix, and dynamically uses the optimization algorithm to compress the data according to its feature. The proposed algorithm is designed and implemented on a Xilinx platform using FPGA, and taking it as the central control hardware to pass the real-time verification on the 32 channels neural signal acquisition hardware system. Experiments show that the proposed dynamic sparse matrix compression algorithm can achieve a data compression rate of 83.4%.

**Keywords:** neural signal acquisition; multi-channel; sparse matrix; data compression algorithms; FPGA

## 0 引言

脑机接口<sup>[1-3]</sup>是一种将生物的神经系统与外部人工设备进行直接通信连接和交互的技术。脑机接口可以应用于不同场景: 如人脑等活体神经系统可以向外部人工电子设备

传输神经信号, 以实现特定功能如控制机械臂或轮椅<sup>[4-5]</sup>; 外部电子设备也可以通过电或化学的方式反馈于生物神经系统, 通过神经调控的方式对其进行直接影响和干预<sup>[6-7]</sup>。多通道神经信号采集和实时信号传输是脑机接口中的核心技术之一。

收稿日期: 2023-06-13; 修回日期: 2023-07-05。

基金项目: 国家重点研发计划资助(2021YFB3200600)

作者简介: 高原雨(1998-), 女, 在读硕士。

通讯作者: 姚镭(1982-), 男, 博士, 教授。

引用格式: 高原雨, 尤昌华, 李朋, 等. 应用于脑机接口系统的动态稀疏矩阵压缩算法[J]. 计算机测量与控制, 2024, 32(5): 238-245, 324.

就神经信号采集而言, 目前有植入式脑机接口与非侵入式脑机接口之分。非侵入性方法能够通过颅骨记录数百万个神经元的复合电信号, 但是这种信号是非特异性和扭曲的<sup>[8]</sup>。植入式脑机接口能够最准确地捕捉神经电活动, 可以从分布的、功能相关的神经元群中记录单个动作电位<sup>[9]</sup>, 对其进行放大、数字化处理, 并通过发射器将其电活动信息传输到体外。由于神经信号主要通过锋电位信息进行传递, 所以锋电位信息提取和压缩就成了需要解决的重要问题<sup>[10-11]</sup>。

最近, Neuralink 公司已经建立了小型且灵活的电极阵列 (每个阵列 3 072 个电极), 使记录数千个通道神经信号<sup>[12]</sup>成为可能。大规模多通道的神经信号记录在降低功耗和芯片面积方面提出了许多挑战<sup>[13]</sup>。由于大量的电极和片上可用于实时处理和传输的有限功率预算, 信号传输变得更具挑战性<sup>[14]</sup>。由于数据在时间和空间上的高度相关性, 直接传输原始数据会消耗巨大的功耗。为了解决这一问题, 需要在传输前对记录器芯片中的神经数据进行压缩, 这对植入式脑机接口的信息提取至关重要。

神经元动作电位信号存在不应期性, 它的有效信号的时域宽度与信号重复周期之比很小。本文运用这一特性, 提出一种基于 FPGA 的应用于多通道神经信号采集系统的动态稀疏矩阵压缩算法, 将已有的神经信号采集芯片与 FPGA 硬件结合形成一个系统级的模型对得到的结果进行分析研究, 可应用于神经信号采集系统的性能测试、神经科学研究、脑机接口等领域。

## 1 背景原理

### 1.1 神经信号不应期性

神经细胞功能的最重要特征是动作电位。该机制允许神经元在不发生传输衰减的情况下进行可靠的长距离的信息传输<sup>[15]</sup>。动作电位尖峰信号中含有丰富的外在刺激物信息, 从这些信号中提取出有效的特征可探讨动物大脑中的作用机理。

神经元的绝对不应期为 1~2 ms, 对应动作电位的持续期, 神经尖峰信号发放间隙为对数正态分布。神经尖峰信号幅值约为 500  $\mu$ V, 频率为 0.1~7 kHz<sup>[16]</sup>。

神经元动作电位信号存在不应期性, 它在时域内的有效信号宽度与信号重复周期之比很小, 即有效信号的占空比很小, 因此这段尖峰序列就具有一定的稀疏性。而在数据采集方面, 往往都是采集单个神经元的发放, 目前的 BMI 技术已经可以同时记录数千个通道, 将多个通道的神经元一次次发放的尖峰序列记录下来, 多通道神经信号的数字标记输出就会形成一个巨大的稀疏矩阵, 对于神经尖峰信号所构成的稀疏矩阵, 对其进行特征提取和压缩, 可以更高效地提取神经信号信息。

### 1.2 稀疏矩阵压缩方法

首先来回顾 4 种非常简单而又广泛使用的数据格式: Coordinate (COO), 压缩稀疏行 (CSR), ELLPACK 格式 (ELL) 以及混合型 Hybrid (HYB)。这 4 种格式是能得到

的最接近“通用”稀疏矩阵表示的格式。对于每种格式, 图 1 展示了稀疏矩阵的表示方法。下面的格式介绍只是一个简短的概述, 绝不是详尽的。有关更多细节, 请参考文献综述 [17-18]。

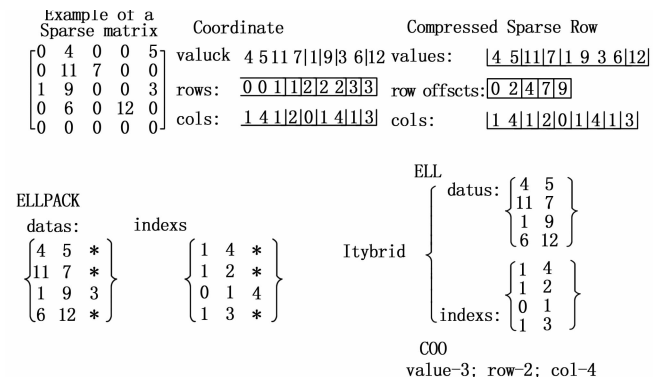


图 1 示例稀疏矩阵的三元坐标法 (COO)、压缩稀疏行 (CSR)、ELLPACK 和 Hybrid 格式的示意图表示

#### 1.2.1 COO 三元组法

COO 三元组存储格式是最为简单的一种稀疏矩阵存储格式。稀疏矩阵的每一个非零元素都要求以三元组的形式表达, 各自为其值、行号与列号, 分别与图 1 Coordinate 中 values, rows 以及 cols 相对应, 数组的长度均为 NNZ (本例中的 NNZ=9)。示例稀疏矩阵可以表示为图 1 所示, 易看出行索引为 0、列索引为 1 的位置存放元素 4, 行索引为 0、列索引为 4 的位置存放元素 5, 以此类推。

#### 1.2.2 CSR 压缩稀疏行法

CSR 属于比较规范的存储格式, 要求有三类数据表示: 值、列, 以及行偏移, 分别与图 1 Compressed Sparse Row 中的 values, cols, row offsets 对应。CSR 不属于三元组而是属于总体编码的方式。数值和列号与 COO 一致, 表示一个元素和它的列索引, 这两个数组的长度为 NNZ。行偏移是指某行第一个元素开始偏移于 values 中的位置。在存储行值时相较于 COO 方法做了压缩。

#### 1.2.3 ELLPACK 法

ELL 法用两个和原始矩阵相同行数的矩阵来存储: 其中第一个矩阵存的是元素数值, 第二个矩阵存的是列号, 行号不需要存储, 直接用元素自身所在的行来表示; 这两个矩阵每一行都是从顶格开始放元素, 如果没有元素就用个标志比如 \* 来代替。对于每行最多  $k$  个非零元素的  $M \times N$  矩阵 (这里  $k=3$ , 因为示例稀疏矩阵的第 3 行有 3 个非零元素), ELL 格式将非零值存储在稠密的  $M \times k$  的 datus 数组中, 具有少于  $k$  个非零值的行则用 \* 填充, 与 values 对应的列索引存储在 indexes 数组中, 数组大小同样是  $M \times k$ , 并具有用于填充的标记值 \*。例如 indexes [2] [2] = 4, 表示其相应元素值 datus [2] [2] = 3 在原示例稀疏矩阵中的列偏移是 4, 即 A [2] [4] = 3。

除非所有行都有完全相同的非零数, 否则二维数组中的一些元素将为 \*; 因此, 这个数据结构在内存空间和冗

余操作方面都有很大的开销。

### 1.2.4 Hybrid 法

HYB 提出的动机是, ELLPACK 和 COO 格式适用于稍微互补的情况: 因此, 这个想法是共同利用 ELLPACK 的高性能潜力和 COO 的性能不变性。混合型 Hybrid 法是 ELLPACK 和 COO 格式的混合组合, 其中大部分稀疏矩阵有效值是存储在 ELL 中。具体来说, HYB 在 ELLPACK 部分中首先分配  $K$  个非零数的零填充行, 并将剩余的非零元素存储在 COO 部分。 $K$  的值是由至少有三分之一的矩阵行包含  $K$  个或更多的非零元素的规则确定的。在对示例的稀疏矩阵压缩中, HYB 方法可以将多出来的元素, 如图 1 中第 3 行的元素“3”利用 COO 单独存储。

## 2 实验前期准备工作

### 2.1 神经信号特征

本部分将对大脑记录神经信号的特征进行简单描述。这种信号是通过植入性方法记录的, 通过这种方法将电极柄植入大脑, 对不同深度、不同部位大脑活动情况进行监控。

在这项工作中, 我们应用了不同文献中介绍的广泛使用的数据集如下。

1) 数据 1 和数据 2: 参考文献 [19] 提供了具有不同噪声水平和峰值形状的各种类型的神经信号。平均数据采样频率为 24 kHz, 发射率被定义为峰值数除以总样本数, 约为 2.4‰。本文使用了为 C\_easy1 和 C\_difficult2 这两个数据集进行信号处理和压缩, 此数据集包含来自 3 个不同神经元在噪声水平条件为 0.05 和 0.20 下的峰值。

2) 数据 3: 中国科学院神经科学研究所赵郑拓老师研究组提供的由 Intan 芯片从大鼠海马体记录的 128 通道胞外神经信号数据。

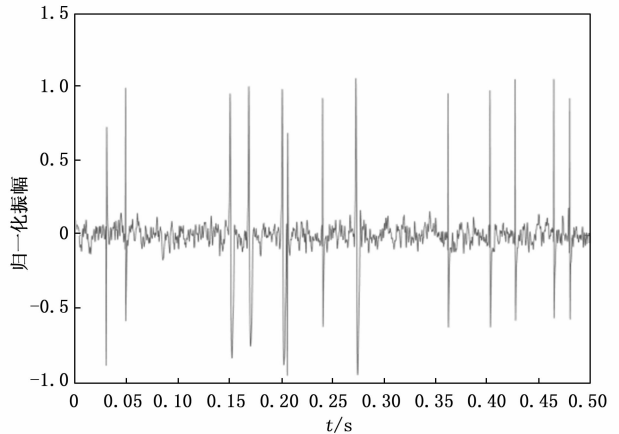
图 2 显示了来自于参考文献 [19] 和中国科学院神经科学研究所赵郑拓老师研究组提供的 Intan 芯片的神经信号数据样本。

### 2.2 神经信号预处理

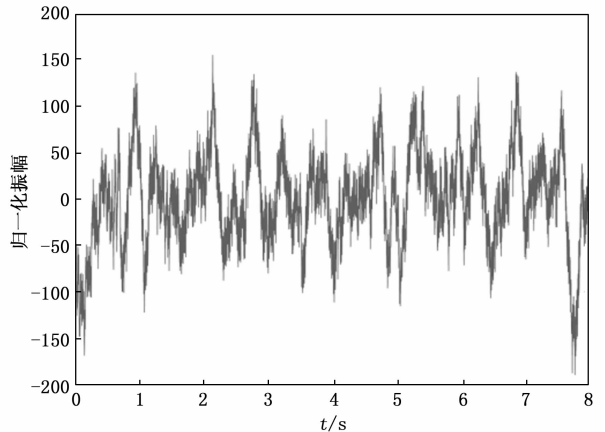
以数据 1 为例, 本文通过 MATLAB2018a 软件对神经信号数据集进行预处理, 具体步骤如图 3 所示。首先通过带通滤波信号 (300~6 000 Hz, 四阶巴特沃斯滤波) 来检测信号的振幅阈值<sup>[20]</sup>。阈值 ( $Thr$ ) 设定为

$$Thr = 4\sigma_n \quad (1)$$

其中:  $\sigma_n$  是  $\frac{|x|}{0.6745}$  的中位数;  $x$  是带通滤波的信号;  $\sigma_n$  是背景噪声的标准偏差的估计值<sup>[21]</sup>。通过滤波得到神经信号动作电位后, 需要对神经信号峰电位进行检测。峰电位的检测可以通过阈值检测的方法, 将滤波后的信号与初设的阈值进行比较, 当信号超过阈值时, 则认为此处出现一个峰电位的波形。将神经信号的输出时数字标记在信号值超过阈值时定义为“1”, 其他时刻定义为“0”; 在多通道的神经信号采集系统中, 一定时间范围内的多通道神经信号的时间戳标记就会形成一个巨大的稀疏矩阵。



(a) 从参考文献[19]中得到的噪声水平为0.05的样本神经信号, 它只有背景噪声和尖峰



(b) 一个真实的单通道全带宽神经信号

图 2 来自不同数据集的样本数据

在峰电位检测的同时, 利用检测到的峰电位进行对齐以减少由于采样抖动引起的影响。对齐通常需要先对信号进行采样 (插值), 然后以某种方式在时间上对齐, 对齐后需要再降采样到原来的频率 (抽取)。本文通过对峰电位的峰值进行对齐, 随后选取峰电位最大值所提取的样本点来对神经信号峰电位进行分类。

### 2.3 神经信号压缩

本文首先利用第 1 节介绍的稀疏矩阵压缩方法分别对 3 种来自不同数据集的神经信号进行数据压缩, 得到结果如图 4 所示。

压缩率是评估压缩方法性能的最重要的标准之一, 它表明了数据压缩的比例。压缩率越小, 即压缩结果越好。压缩率的定义为:

$$\text{压缩率} = \frac{\text{文件压缩后的大小}}{\text{原始文件的大小}} \quad (2)$$

可以看出在信号完整的情况下, CSR 方法对于 4 种来自不同数据集的神经信号的压缩效果最好。但是在极短时间 1 ms 内, COO 方法的压缩效果更好。

本文分析原因是在信号完整的情况下, 稀疏矩阵过于庞大, 由于数据样本数比较大, 因此数据中有效尖峰数量

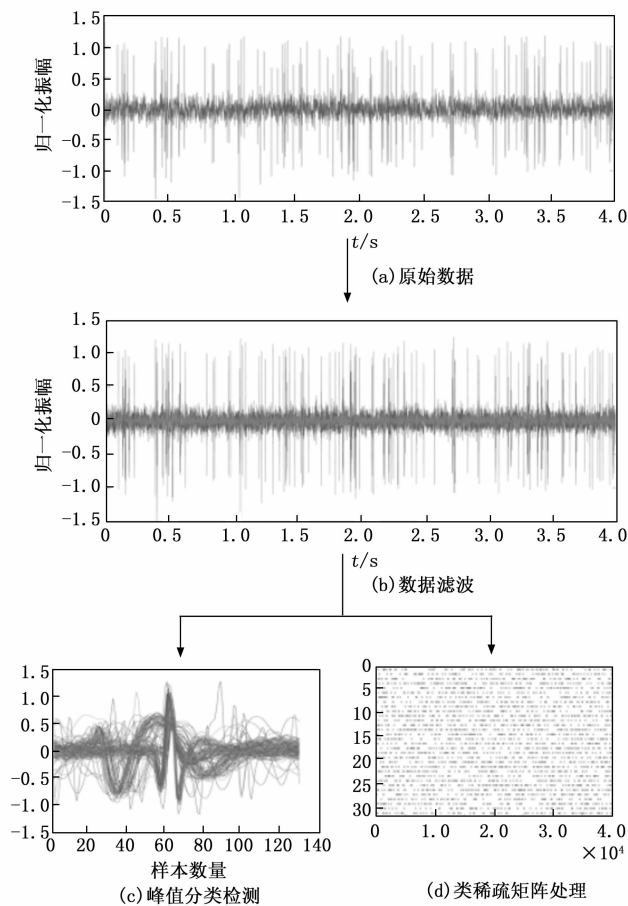


图 3 信号预处理过程概述

较多, 相同的稀疏度下有效信号数目也会相对增加, 此时采用 CSR 方法可以对完整信号下的所有有效信号进行很好的压缩。而在毫秒级的时间范围内, 由于神经信号的不应期性, 在较短的时间中采集到的尖峰信号数相对较少, 此时可以由图 4 (b) 中看出, ELL 格式与 HYB 格式已经不适用于在较短时间内的神经信号, 面对研究时间间隔很小的神经信号矩阵时, COO 方法的压缩效率最高。

### 3 Meanswin 算法描述

在第 2 节的讨论中, 我们得出了当使用典型的稀疏矩阵压缩方法来压缩由神经信号组成的稀疏矩阵时, 其最优的压缩形式是不一样的结论。也就是说, 当多通道神经信号完备时, 使用 CSR 方法对多通道神经元信号构成的稀疏性矩阵进行最优压缩; 在较短时间尺度上, 利用 COO 方法将神经信号稀疏矩阵中小规模子矩阵进行存储, 能获得较好的压缩效果。

对于不同的时间尺度下, 由神经信号构成的稀疏矩阵所适用的压缩格式有所差异的情况, 本文提出一个动态的稀疏压缩算法名为 Meanswin, 拟寻找出一个适当的时间尺度值  $Tt$ , 通过分析在  $Tt$  的时间范围内的稀疏矩阵的矩阵特征, 针对矩阵的特点, 动态优化经典压缩算法, 实现神经信号稀疏矩阵的数据压缩。具体的算法描述由图 5 的决策

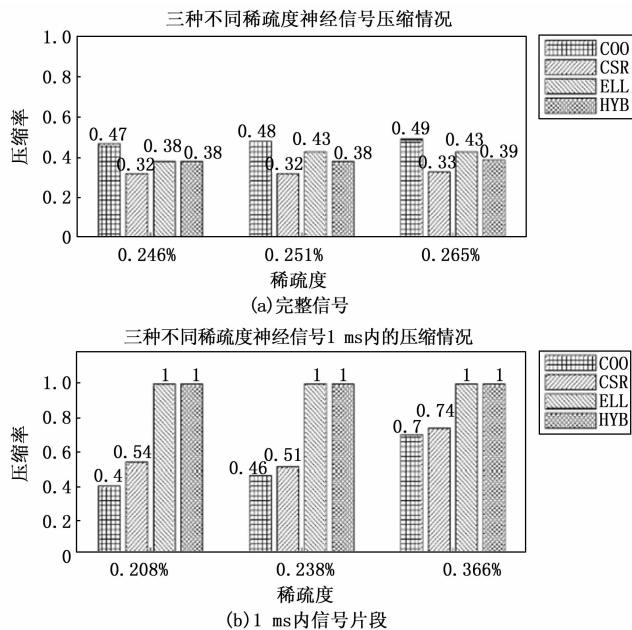


图 4 不同数据集神经信号压缩情况

树所示。

首先在对多通道神经信号进行采集后, 将其按照 2.2 节所述方法进行滤波、尖峰信号检测等预处理, 将多通道的神经信号处理为只含有“0”和“1”的稀疏矩阵, 其中数值为“1”的点即为原始神经信号数据中的尖峰信号点。接下来统计矩阵中的非零元素个数即有效信号的数目, 依据公式 (3) 计算此矩阵的稀疏度, 紧接着判断矩阵稀疏度是否大于 5%, 若多次处理后仍大于 5%, 则判定此矩阵为非多通道神经信号的稀疏矩阵, 不适于再做后续处理。随后经过实验得到适用于神经信号稀疏矩阵压缩方法选择的适配时间窗口尺度值, 最后通过比较选择好的时间尺度子矩阵内的有效尖峰信号数与采样通道数的大小关系, 决定采用何种压缩方法进行压缩。

### 3.1 决策点的阈值选择

#### 3.1.1 稀疏度的阈值选择

由于不同神经信号数据集的采样率不同, 因此得到的神经信号稀疏矩阵的稀疏度也不同, 从得到的数据集可以看到矩阵的稀疏度分别为 2.08%、2.38%和 3.66%。

由于稀疏度被定义为:

$$\text{稀疏度} = \frac{\text{有效数目}}{\text{样本总数}} \quad (3)$$

考虑到极端情况, 本文将神经信号稀疏矩阵的稀疏度阈值放宽为 5%。若矩阵的稀疏度过大达到百分之十的情况, 则将其判断为非神经信号稀疏矩阵。此时用 Meanswin 法压缩时的内存资源将大于此矩阵不做任何处理的原始所占资源, 本文在后面的仿真实验中模拟了矩阵稀疏度过大的情况。

#### 3.1.2 适配时间窗口尺度值的选择

由于不同时间尺度下神经信号稀疏矩阵所适用的稀疏

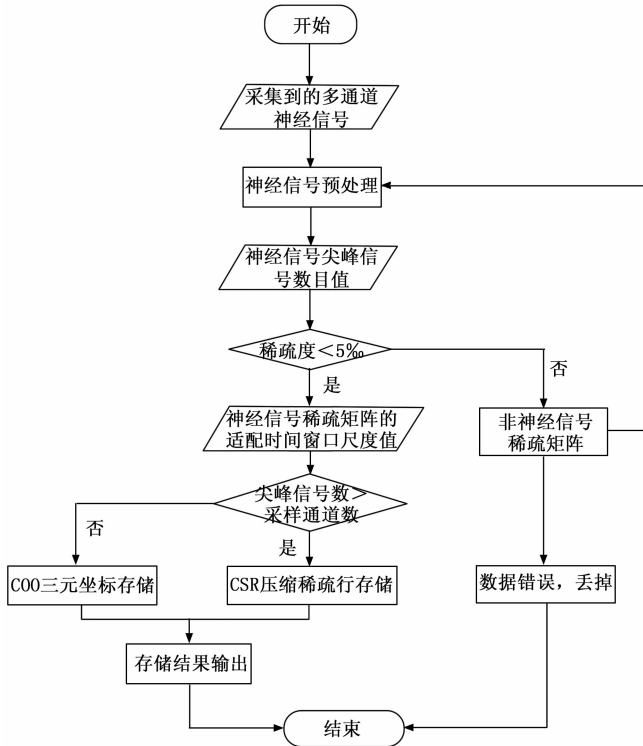


图 5 神经信号稀疏矩阵压缩存储格式的动态选择算法。

其中决策的阈值是基于对数据集的分析而选择的。

矩阵压缩方法不同，因此需要分析出一个适合的时间窗口尺度值对完整神经信号矩阵进行压缩处理。

如图 6 所示，我们通过对三组神经信号数据构成的稀疏矩阵进行不同时间尺度的划窗处理，观察发现其有效点数据分布随着时间窗口尺度在变化：在时间尺度为 240 即 1 ms 时，easy01 和 diff01 的数据集有效点数目集中在 15~20 之间，利用 COO 方法进行压缩最有效，在时间尺度为 600 即 2.5 ms 时，有效点数目分布在 40~50 之间，利用 CSR 方法进行压缩最有效；在 128 通道的数据中，这两种

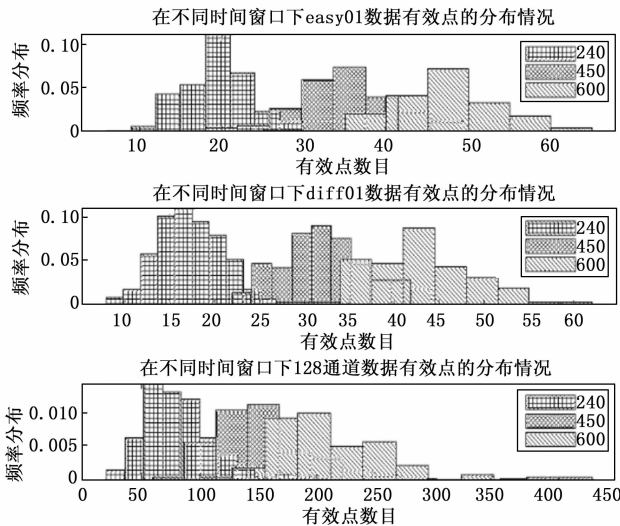


图 6 在不同时间窗口尺度下三组神经数据有效点的分布

时间尺度也会存在数据只能使用一种方法最优压缩的情况。

为了有利于完整神经信号稀疏矩阵的整体压缩，节省更多的空间与资源，本文通过实验后将时间尺度即稀疏矩阵的划窗值定为 450，在此时间窗口尺度下，数据有效点的中位值集中在 3.1.3 节的第三个决策点处，更有利于稀疏矩阵的动态压缩。

### 3.1.3 压缩方法的选择

稀疏矩阵压缩方法的选择很大程度上决定于稀疏矩阵的特征，本文为这些矩阵推导出了表 1 中所描述的特征。这组特征是 SpMV 自动调优文献中通常使用的特征的子集<sup>[22-23]</sup>。特别地，参考文献 [22] 的工作表明，用于检测格式的一组特征取决于格式本身。

本文定义一个评估标准，此标准在评估最佳压缩格式时往往会有帮助。假设我们想评估一种叫做“ANS”的新格式。设  $A$  是一个有  $NNZ$  个非零元素的  $M \times N$  稀疏矩阵，需要存储在计算机内存中。 $A$  的稀疏度为  $\frac{NNZ}{M \times N}$ 。

表 1 用来描述稀疏矩阵的一组特征

特征	描述
$M$	矩阵中的行数
$N$	矩阵中的列数
$NNZ$	矩阵中的非零数目
$DES$	矩阵的密度
$AVGNZR$	每行非零数的平均数
$MINNZR$	每行的最小非零数
$MAXNZR$	每行的最大非零数

设  $S_{ANS}$  表示在 ANS 中存储关于给定稀疏矩阵的非零元素结构的信息所需的字节数，本文将称之为矩阵结构内存占用。

对于 COO 格式存储， $S_{COO}$  可以表示为：

$$S_{COO} = NNZ \times \log_2(M \cdot N) = NNZ \times \log_2 M + NNZ \times \log_2 N \quad (4)$$

当使用 CSR 格式存储时， $S_{CSR}$  可表示为

$$S_{CSR} = NNZ \times \log_2 N + (M + 1) \times \log_2 NNZ \quad (5)$$

通过公式 (4) 和公式 (5) 可知：若稀疏矩阵中的有效点数目，即矩阵中的非零数目  $NNZ$  大于矩阵的行数  $M$ ，采用 CSR 格式对其进行压缩时压缩效果更好；当矩阵中的非零数目小于或等于矩阵的行数时，应该采用 COO 格式对矩阵进行压缩。在本文中，针对于多通道神经信号所形成的稀疏矩阵，其行数即为神经信号的采集通道数 32。

### 3.2 MATLAB 仿真结果

针对本文中所提出的 Meanswin 方法，我们使用 MATLAB R2018a 软件对 3 种不同数据集的神经信号进行仿真验证，实验得到的压缩结果如图 7 所示。实验共模拟和检验了 3 种不同的算法，即采用 450 时间尺度的划窗对初始矩阵进行动态划窗处理，Meanswin 方法是根据每一个划窗内子矩阵的特征动态地选择使用哪种方法进行压缩；COOwin

方法是将子矩阵都以 COO 格式压缩; CSRwin 方法就是对所有的子矩阵都采用 CSR 方法进行压缩。

图 7 清晰地表明对于 3 种神经信号, 使用同等大小的时间滑窗窗口分别用 COOwin 格式、CSRwin 格式和 Meanswin 方法对其进行压缩, Meanswin 方法不仅可以在一定程度上降低压缩率, 取得更好的压缩效果, 并且可以动态地调整最优压缩方法的选择去匹配神经信号稀疏矩阵的压缩。

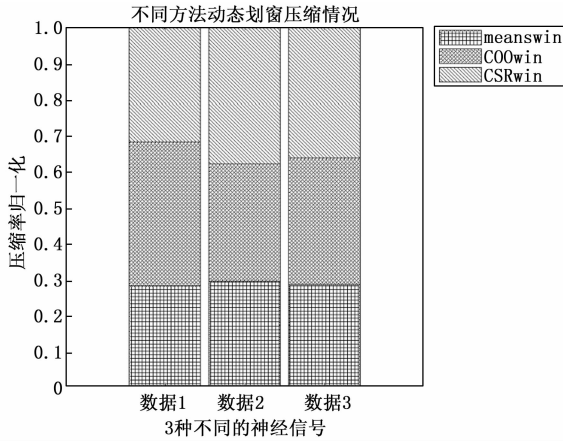


图 7 3 种不同神经信号在不同方法动态划窗下的压缩情况

### 4 Meanswin 算法描述

#### 4.1 FPGA 硬件架构

在 Vivado2018.2 的开发环境下, 基于 Verilog 语言,

设计并实现了 Meanswin 算法的硬件化描述, 并对该算法进行仿真测试与功能验证。在 XILINX 的 ZYNQ7000 开发平台 AX7020 硬件平台上对系统性能进行测试。

Meanswin 算法的 FPGA\_信号采集 IC 系统架构示意图如图 8 所示: 在 FPGA 中由时钟生成模块产生前端信号采集模块和 Meanswin 算法所需时钟; 采集芯片控制模块输出采集 IC 控制信号以使能前端 ADC 信号采集芯片开始工作; 芯片将采集到的神经信号数据经过内部的模数转换处理后输入到 FPGA 中, 随后 FPGA 信号预处理模块中的子模块通过 3.1 节得出的决策点阈值对从芯片中采集到的神经信号数据进行多线程并行的预处理, 将其转化为对应的稀疏矩阵, 并在此模块中得到矩阵中的非零数目 NNZ (图中为 nonzerosum) 输出到压缩方法选择模块; 压缩方法选择模块判断后输出存储格式算法模块的使能信号, 将数据输入进入到 Meanswin 算法模块中进行压缩, 压缩后的数据存储在对应格式的 RAM 块中; 最后利用 Vivado2018.2 软件自带的逻辑分析仪在 PC 上位机中查看实验结果。

#### 4.2 Meanswin 算法功能仿真测试

本文以数据库中已采集的神经信号集模拟前端神经信号采集 IC 采集的多通道神经信号输入到 FPGA 中。

首先通过数据库中已有的神经信号数据改造, 模拟不同稀疏度的神经信号矩阵对 Meanswin 算法进行功能验证测试, 输出波形如图 9 (a) 所示。随后将数据库中采集到的真实神经信号数据作为输入, 对 Meanswin 算法进行功能测试, 输出波形如图 9 (b) 所示。

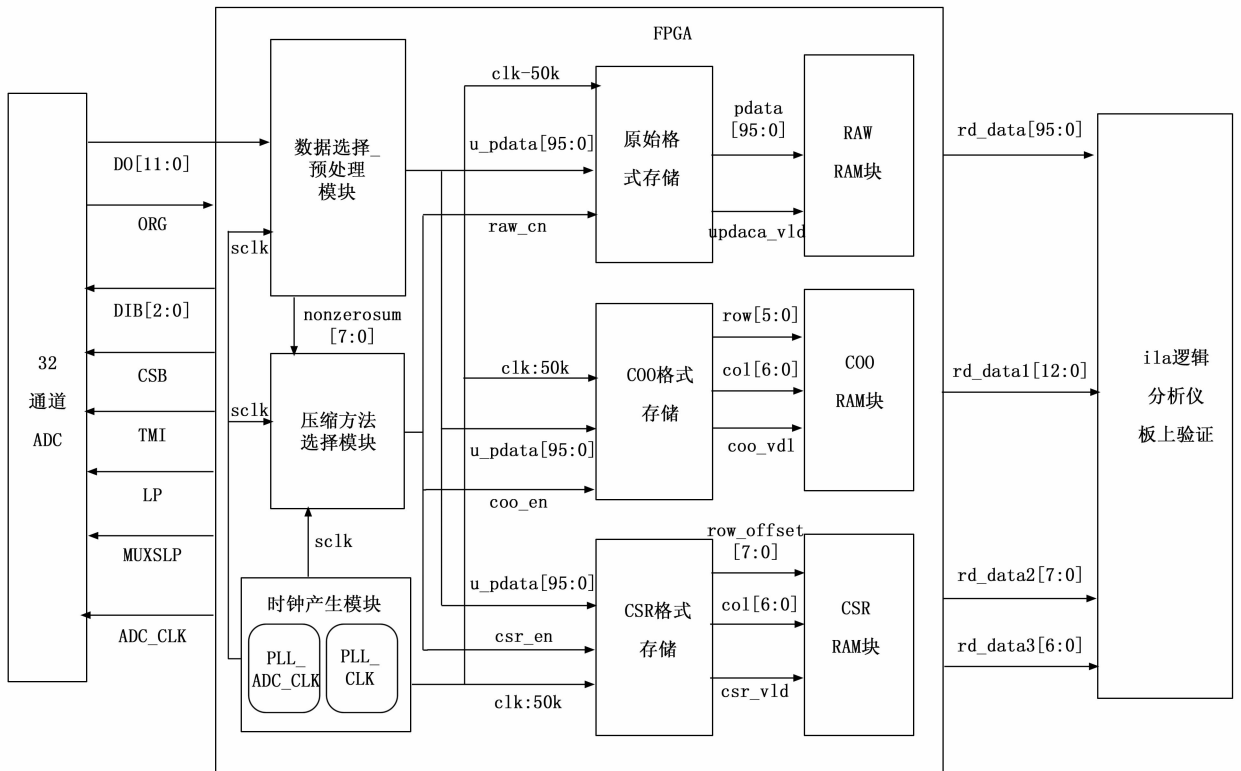
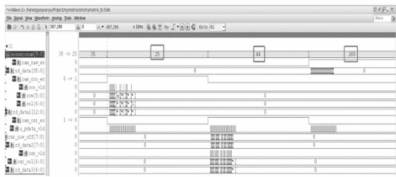
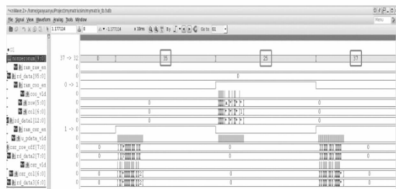


图 8 Meanswin 算法 FPGA 架构示意图

以图 9 (b) 为例：在 nonzero-sum 信号表明神经信号稀疏矩阵中的非零数目分别为 35 和 25 时，由于触发的使能信号不同，神经信号被存储的压缩方法也有区别。在 nonzero-sum 为 35 时 ram\_csr\_en 使能信号置高，使用 CSR 存储，rd\_data2 和 rd\_data3 分别表示非零元素的所在列数和它在 values 里面的起始偏移位置；在 nonzero-sum 为 25 时信号 ram\_coo\_en 置高使用 COO 格式存储，rd\_data1 中的前六位比特 [12: 7] 表示它的所在行，后七位 [6: 0] 表示它的所在列数。在图 9 (a) 中，当 nonzero-sum 为 268 时，此时矩阵的稀疏度已经远远大于 5%，因此 ram\_raw\_en 信号置高，神经信号不做处理直接输出。



(a) 模拟不同稀疏度神经信号对 Meanswin 算法功能测试



(b) 数据库中真实神经信号对 Meanswin 算法功能测试

图 9 Meanswin 算法功能测试仿真图

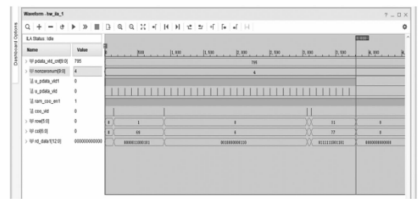
### 4.3 基于片上系统的算法性能测试

本文以 Xilinx 公司的 Zynq7000 系列的芯片，型号为 XC7Z020-2CLG400I 的 FPGA 作为硬件平台，前端连接一款 32 通道低功耗低噪声信号采集芯片 MTXR32P 作为 FPGA 输入信号，设计并实现了本文提出的动态稀疏矩阵压缩算法 Meanswin。

MTXR32P 是一款 32 通道低功耗低噪声信号采集 IC，其内部每个通道集成了低噪声放大器 (LNA, low noise amplifier)、可配置带通滤波器 (BPF, band-pass filter)、12 位低功耗 SAR ADC、以及数字信号处理和接口模块，可应用于各类对功耗和设备尺寸有要求的多通道电生理信号采集仪器和设备。芯片使用一个 1.024 MSPS 的 SAR ADC 轮流对 32 个输入通道进行采样，每个通道的最高采样率是 32 kHz。实验中前端的神经信号输入为采集到的小鼠神经信号音频文件。

FPGA-信号采集 IC 片上系统的测试结果如图 10 所示。实验表明，系统实物测试与仿真结果一致。如图 10 (a) 中，nonzero-sum 输出为 4，小于采集通道数 32，此时 ram\_coo\_en 信号置高，压缩后的数据为 rd\_data1，每个 13 bit 的数据表明了此段信号中有效神经信号的位置信息。例如第一个有效信号的坐标值为 1 和 69，则 rd\_data1 输出 13 bit 信号 0000011000101，前 6 bit 表示此尖峰信号的所在

通道值 1，后 7 bit 表示此尖峰信号出现的时间戳位置为 69。图 10 (b) 中 nonzero-sum 的输出为 36，此时 ram\_csr\_en 信号置高有效，压缩后神经尖峰信号在此稀疏矩阵中的行偏移索引值存储为 rd\_data2，信号出现的时间戳位置即其列索引值存储为 rd\_data3，同样利用 CSR 格式的存储方法可以锁定有效信号的位置信息。



(a) Meanswin 算法片上系统测试的 COO 结果



(b) Meanswin 算法片上系统测试的 CSR 结果

图 10 Meanswin 算法片上测试图

### 4.4 测试结果和对比分析

节省空间的比率 SSR (space saving ratio)<sup>[24]</sup> 是评估压缩方法性能的最重要的标准之一，它表明了数据减少的比例。压缩值越大，压缩结果越好。SSR 的定义为：

$$SSR \% = 1 - \frac{\text{压缩数据的位数}}{\text{原始数据的位数}} \quad (6)$$

在本文中，使用所提出的 Meanswin 方法压缩了来自文献 [19] 的 12 个数据集，平均 SSR 为 83.4%。

表 2 展示了所提出的算法架构与最先进的神经信号压缩算法的比较。由于一些先进的技术主要集中在有损压缩方面，所以很难直接进行比较。表 2 中所有在特定应用集成电路 (ASIC) 上实现的 SSR 都是无损模式下的压缩结果。本文使用 Synopsys 公司的 DC 综合工具，利用 180 nm 技术和 1.8 V 电源电压完成对本文压缩算法的硬件综合实现，此设计的面积与功耗均优于<sup>[27]</sup>，在数据压缩方面明显优于<sup>[24]</sup>，对于多通道神经信号采集系统的实时动态压缩有显著的效果。

表 2 与最先进的设计进行比较

参考文献	压缩方法	SSR / % <sup>1)</sup>	功率 / Ch	面积 / μm <sup>2</sup>	工艺 / nm
[25]	减少预测器顺序	64.8	N/A	N/A	N/A
[20]	预测方法压缩	65.6	N/A	N/A	N/A
[26]	Haar 小波	43.8	N/A	3 700	130
[27]	离散小波变换	30	95 μW	5 750 000	500
[28]	离散余弦变换	~50	N/A	N/A	N/A
[24]	自适应硬件加速器	61.84	1.05 μW	792.44	28
本文	动态稀疏矩阵压缩算法	83.4	0.78 μW	1 272 872	180

注：1) 每个参考文献都使用不同的数据集，本文与 [24] 为同一数据集。

特别注意的是, 为研究本文所用压缩算法与其他论文压缩方案的直观比较, 因此在计算 SSR 的结果时, 本文使用与文献 [24] 中相同的数据集作为压缩前原始数据, 最终得到的平均 SSR 为 83.4%, 优于其他文献压缩效果。且本文使用的硬件平台 Zynq7000 FPGA 开发板的价格为 1 699 元, 因此本文设计的方案在实际应用场景中拥有更好的性价比, 在实现高压压缩性能的同时也满足了低功耗低成本的要求。

## 5 结束语

随着科学技术的发展, 数千通道神经信号的记录为人类带来了大脑中更多有用的信息, 为了减少传输的数据量而又不影响实时性的要求, 需要在传输信号之前对记录器芯片中的神经数据进行压缩, 以达到优化数据传输速度和处理效率的目的。本文提出了一个应用于多通道神经采集系统的信号压缩算法。提出的算法基于神经信号不应期性特点, 结合多个稀疏矩阵压缩算法, 实时根据所采集神经信号的特征动态进行数据压缩, 并且搭建了基于 FPGA 和信号采集 IC 的硬件测试系统进行神经信号压缩算法的功能验证和性能测试, 验证了所提出的动态神经信号稀疏矩阵压缩算法在本文的硬件实现下可以获得较好的压缩性能。仿真和实验结果均表明, 与传统的数据压缩技术相比, 本文提出的动态稀疏矩阵压缩算法有效提高了 SSR, 使用多通道神经采集动物实验数据取得了平均 83.4% 的压缩率, 且与其他设计进行比较, 单通道的功率为  $0.78 \mu\text{W}$ , 达到低功耗低成本设计要求。

本文中所采用的动态神经信号稀疏矩阵压缩算法还有一定的优化空间, 一方面可以考虑是否可以研究出其他的神经信号稀疏矩阵特征去细化更多算法决策点的选择, 另一方面可以继续研究适用于神经信号稀疏矩阵的压缩算法, 进行进一步的优化, 提高压缩率。

## 参考文献:

- [1] ANBARASAN R, CARMONA D G, MAHENDRAN R. Human taste-perception; Brain computer interface (BCI) and its application as an engineering tool for taste-driven sensory studies [J]. *Food Eng Rev*, 2022, 14 (3): 408-434.
- [2] LUO J, XUE N, CHEN J. A review; Research progress of neural probes for brain research and brain-computer interface [J]. *Biosensors (Basel)*, 2022, 12 (12): 1167.
- [3] ABENNA S, NAHID M, BOUYGHF H, et al. EEG-based BCI: A novel improvement for EEG signals classification based on real-time preprocessing [J]. *Comput Biol Med*, 2022, 3 (148): 105931.
- [4] LIU W H. Application of EEG in wearable brain-computer interfaces [C] // Bellingham; SPIE, 2022; 12165Q.
- [5] ZENG H, WANG Y X, WU C C, et al. Closed-loop hybrid gaze brain-machine interface based robotic arm control with augmented reality feedback [J]. *Front Neurorobotics*, 2017, 11: 60.
- [6] CHIKHI S, MATTON N, SANNA M, et al. Mental strategies and resting state EEG: Effect on high alpha amplitude modulation by neurofeedback in healthy young adults [J]. *Biological psychology*, 2023, 178: 108521.
- [7] WANG Y, LIU S J, WANG H, et al. Neuron devices: Emerging prospects in neural interfaces and recognition [J]. *Microsyst Nanoeng*, 2022, 8 (1): 13.
- [8] PESARAN B, VINCK M, EINEVOLL G T, et al. Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation [J]. *Nat Neurosci*, 2018, 21 (7): 903-919.
- [9] YUSTE R. From the neuron doctrine to neural networks [J]. *Nat Rev Neurosci*, 2015, 16 (8): 487-497.
- [10] 汪 芳. 基于强化学习的植入式脑机接口神经信息解码: 算法设计及应用 [D]. 杭州: 浙江大学, 2015.
- [11] PAGIN M, ORTMANN M. Study of compressed sensing and predictor techniques for the compression of neural signals under the influence of noise [C] // New York; IEEE, 2018; 1102-1105.
- [12] PISARCHIK A N, MAKSIMENKO V A, HRAMOV A E. From novel technology to novel applications: Comment on “an integrated brain-machine interface platform with thousands of channels” by elon musk and neuralink [J]. *J Med Internet Res*, 2019, 21 (10): e16356.
- [13] NOSHAHR F H, NABAVI M, SAWAN M. Multi-channel neural recording implants: A review [J]. *Sensors*, 2020, 20 (3): 904.
- [14] DO A T, ZEINOLABEDIN S M A, JEON D, et al. An area-efficient 128-channel spike sorting processor for real-time neural recording with  $0.175 \mu\text{W}/\text{channel}$  in 65-nm cmos [J]. *IEEE T Vlsi Syst*, 2019, 27 (1): 126-137.
- [15] EUGENE M. Dynamical systems in neuroscience [M]. Cambridge, MA: MIT press, 2007: 57-66.
- [16] 郑艳瑜, 王敏敏, 高 翔, 等. 基于 fpga 和真实数据集的 128 通道神经信号模拟器设计与实现 [J]. *中国生物医学工程学报*, 2020, 39 (5): 631-635.
- [17] FILIPPONE S, CARDELLINI V, BARBIERI D, et al. Sparse matrix-vector multiplication on gpgpus [J]. *ACM Transactions on Mathematical Software*, 2017, 43 (4): 1-49.
- [18] LANGR D, TVRDIK P. Evaluation criteria for sparse matrix storage formats [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2016, 27 (2): 428-440.
- [19] QUIROGA R Q, NADASDY Z, BEN-SHAUL Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering [J]. *Neural Comput*, 2004, 16 (8): 1661-1687.
- [20] CAPURRO I, LECUMBERRY F, MARTIN A, et al. Efficient sequential compression of multichannel biomedical signals [J]. *IEEE J Biomed Health*, 2017, 21 (4): 904-916.
- [21] JOHNSTONE D I M. Ideal spatial adaptation by wavelet shrinkage [J]. *Biometrika*, 1994, 81 (3): 425-455.

(下转第 324 页)