申请人：之江实验室，浙江大学

发明人：黄科杰 刘润 沈海斌

第一发明人身份证号：330222198007126952

专利负责老师：黄科杰17706443800；huangkejie@zju.edu.cn

联系人电话及邮箱：刘润17865195158 [liurun@zju.edu.cn](mailto:liurun@zju.edu.cn)

寄信地址及邮编：浙江大学玉泉校区老生仪楼3楼 邮编310027

**专利申请提交的技术交底材料**

**申请发明或者实用新型专利应提交交底材料应包括以下内容：**一、名称  
一种基于存内计算的混合基FFT计算电路。

二、技术领域

本发明属于神经形态计算领域，涉及一种基于存内计算的混合基FFT的神经形态计算电路。

三、背景技术

傅里叶分析是一种广泛应用于工程，科学和数学等领域的将信号从原始域变换到频率域的分析方法，快速傅里叶变换（FFT）是其在实际中常用到的快速计算方法。同时FFT计算是一种高并行度，大数据量的计算方式，随之而来的是计算中产生的巨大功耗，这在越来越多的低功耗终端应用中显现出巨大的挑战。与此同时，在传统的冯洛伊曼计算架构中，由于存储和计算模块的分离，导致这种计算架构产生严重的“内存墙”效应，这种效应的本质使得数据在存储和计算模块中不断进行传输和计算。在大数据量和大计算量的智能时代，这种计算模式使得计算系统出现了成本高，能耗高，能耗效率低和延迟高等各种问题，让智能时代的计算问题变成了一个巨大的挑战。并且，随着智能时代终端数量的爆炸式增长和摩尔定律的即将失效，通过半导体工艺升级的办法来提升处理器的性能已经难以维持，低功耗和大数据量的矛盾日益突出，我们急需要新的硬件架构和软件算法来提升计算和存储能力的同时，减少计算所产生的功耗。

存内计算技术是一种采用新型存储单元同时实现存储和计算的一种技术，是神经形态计算电路的一种实现方式。存内计算技术通过给存储器赋予计算功能，从而解决了冯洛伊曼结构中存在的内存墙问题，是一种非常适合用于进行低功耗计算的神经网络处理芯片设计的异构计算方式。存内计算中并行的输入输出方式适合于大数据量的并行计算，大大降低了计算延迟。并且，存内计算并行乘累加的计算方式很适合FFT计算中所使用到的矩阵乘法。这样采用存内计算的FFT计算就可以获得高并行，低功耗，低延迟的效果，从而提升电路对于FFT计算的处理性能。

现有的FFT处理器设计主要分为两种，一种是流水线式，一种是存储器中心式。流水线式FFT处理器通过对输入原始数据进行分类，每次计算其中一部分数据从而产生流水效果；以存储器为中心的设计模式则是将优化数据在存储器和计算单元的流动方式作为重点，设计不同的数据流模式以达到计算的目的。但是这两种方式仍然没有跳出传统的冯诺依曼设计结构，因此都会产生相当的功耗和一定的延迟。本发明则是采用新兴的存内计算技术，结合FFT矩阵计算和存内计算乘累加的特点，设计混合基FFT的处理电路，该电路可以在降低功耗的同时，实现高并行和低延迟的信号FFT处理。

四、发明内容

本发明针对于传统FFT设计中高功耗的特点，结合低功耗的存内计算技术和FFT计算的本质特点，设计了一种低功耗的FFT计算电路，并且可以实现高并行度和低延迟时间。

存内计算模式的本质是数据的乘累加和，输入数值转化为二进制形式，每次输入单比特的二进制数，控制相应的存储单元的计算（图1）。而使用大规模的阵列式存内计算单元，即可实现在阵列的一个方向控制乘法计算数值，在阵列的另一个方向实现乘法计算数值的累加。这种计算模式类似于矩阵计算的模式。在基础的离散傅里叶变换（DFT）中，原始数据和旋转因子乘累加的模式也是一种矩阵计算，这就很自然地可以将存内计算和FFT结合起来（图2）。然而，如果直接使用DFT矩阵进行存内计算，则需要大量的存内计算单元，并且不同点数的FFT也需要不同数量的存内计算单元，使得这样的计算电路缺少灵活性。为了兼具存内计算的低功耗和FFT计算的灵活性，本发明采用了基于混合基分解的FFT计算方式，将DFT计算中一个大的矩阵分解成为多个小型矩阵计算，将一次完成的DFT矩阵计算分解为多级的DFT计算，这种混合基分解的计算方式可以使得不同点数的FFT计算变得非常灵活。本发明的创新点具体如下所述：

1. 设计了混合基分解和存内计算结合的电路（图5）。主要包括原始数据和各级中间数据存储模块，每一级计算时地址生成模块，数据流产生模块和存内计算阵列。对于全新的阵列结构进行了设计。
2. 对于原始数据和各级中间数据进行了量化。为了兼顾数据的精度和每一级计算的延迟，对于原始数据和每一级中间计算数据都采用有符号数的8比特量化方式。在这种量化方式下，每一个数据都会量化到-128到127的范围中，这样可以让每一级的计算周期都变成8个时钟周期，同时保持相当的计算精度。
3. 设计不同大小的存内DFT计算阵列，包括8点计算阵列，16点计算阵列，32点计算阵列和64点计算阵列（图5）。不同点数的计算阵列可以满足不同点FFT的分级要求，也可以满足相同点FFT不同分级方式的计算要求。这种不同点的阵列布置，可以使得FFT计算电路完成不同点的计算，从而大大提升整体电路的灵活性。

五、附图说明  
图1是一般存内计算电路结构图；

图2是一般DFT计算矩阵；

图3是512点FFT 16\*16\*2混合基分级模式下数据流图；

图4是16点DFT存内计算示意图；

图5是本发明提出的混合基FFT存内计算结构示意图；

六、具体实施方式

下面结合附图对发明的具体实施方式做进一步的说明。

传统DFT计算，当原始数据量较大时，使用矩阵乘法形式计算十分不方便，尤其是使用存内计算技术。对于大点的FFT计算，可以采用混合基分解形式将一个大型的旋转因子矩阵分解为多个小型旋转因子矩阵。其计算原理是，对于N点DFT，如果N是一个复合数，它可以分级为一些因子的乘积，则可以使用FFT的一般算法，即混合基FFT算法，而常用的基-2算法是这种一般算法的特例。

若N可以表示为复合数，则对于的任何一个正整数n，都可按照L基，，…, 表示为多基多进制形式,这一多基多进制所代表的数值为

其倒位序形式为=，它所代表的数值为：

在这一多基多进制的表示中：

设DFT点数是一个复合数，将n用下面公式表达：

并且有：

这样可以吧x(n)分为个点的序列。令，将频率变量k可以用下面公式表示：

并且有：

则DFT公式可以化为：

这样可以得到的混合基FFT算法，首先将n表示为按进制表示的形式,即将x(n)分为个点的序列，然后再做个点的DFT得到；将

乘以旋转因子后，再做个点的DFT得到，最后进行进制整序，得到。

图5显示的是整体系统的结构图，图中除存内计算阵列之外，都采用数字电路实现。存储模块用于对原始数据和每一级产生的中间数据进行存储。由于进行了混合基的分解，每一级中需要抽取不同位序的输入数据送入到存内计算阵列中进行计算，这就需要地址生成模块产生每一级计算的输入数据在存储模块中的地址，通过正确地产生输入数据地址，从而将正确的数据流送入到存内计算阵列中。总体的数据流控制模块则控制着整个系统中的所有控制信号，包括地址的产生，输入数据在存内计算模块中的计算，输出数据在存储模块的再存储等等。存内计算模块则采用低功耗的存内计算核，让FFT计算达到低功耗的目的。对于不同的分级方式，存内计算核提供了8点，16点，32点和64点不同的存内计算阵列，可以计算不同点数的DFT，这样满足不同点的计算需求，体现了整体设计的灵活性。

图3作为示例，显示的是512点FFT在16\*16\*2的混合基分级模式下的数据流图，

体现出混合基分解下每一级抽取不同位序数据进行计算的特点。图4是一个16点DFT在存内计算中的示意图。存内单元存储的是旋转因子矩阵中的元素，也就是权重，输入数据量化成为8bit之后按比特从横向进入存内计算阵列，实现乘法运算，在纵向实现累加和计算，并输出最后结果。

以上实施例仅用于说明本发明的技术方案而非对其限制，尽管参照上述实施例对本发明进行了详细的说明，所属领域的普通技术人员应当理解；其依然可以对本发明的具体实施方式进行修改或者等同替换，而未脱离本发明精神和范围的任何修改或者等同替换，其均应涵盖在本发明的权利要求范围当中。

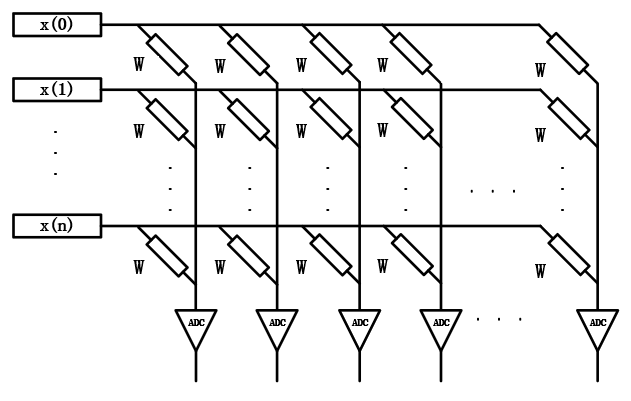


图1

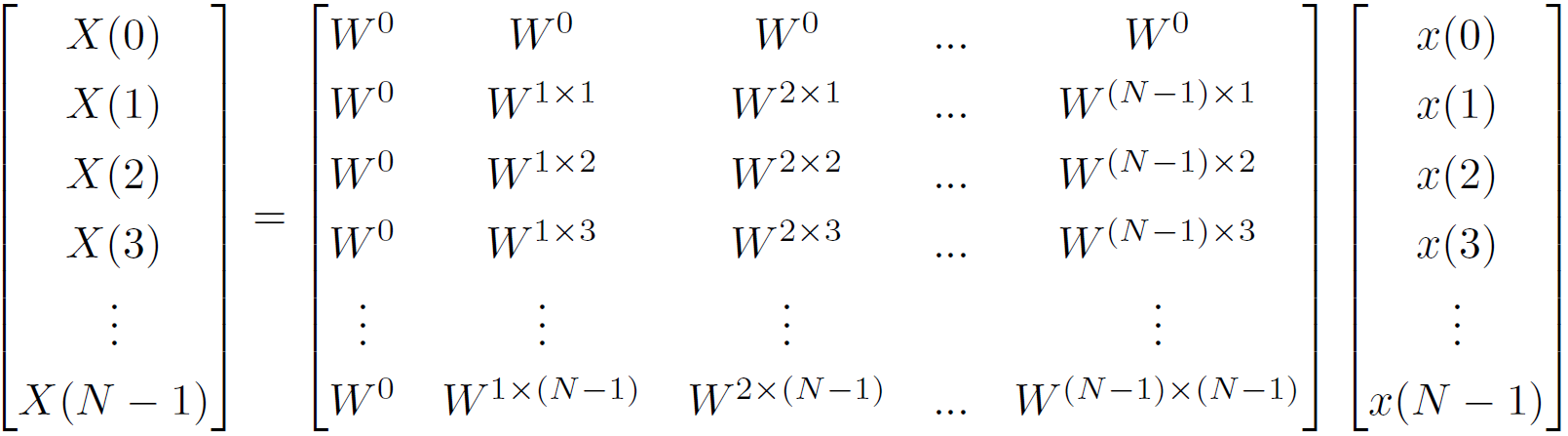


图2

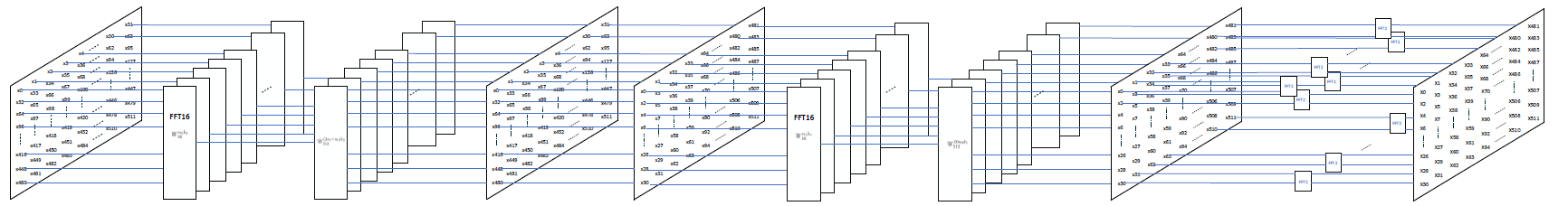


图3

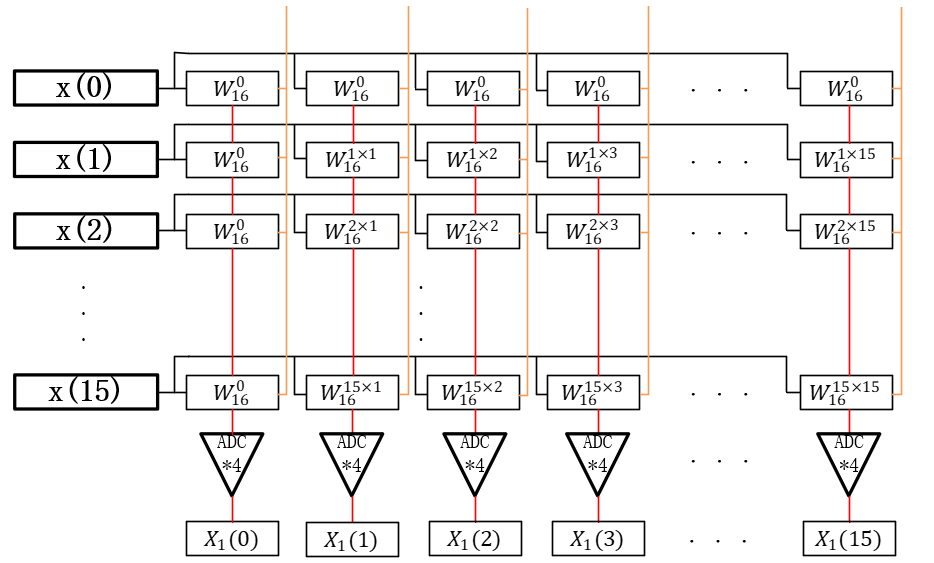


图4

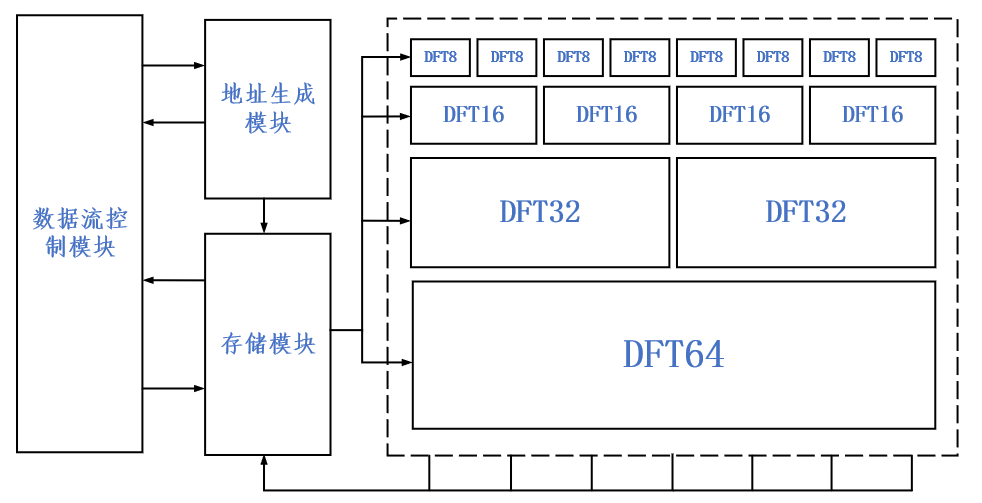


图5