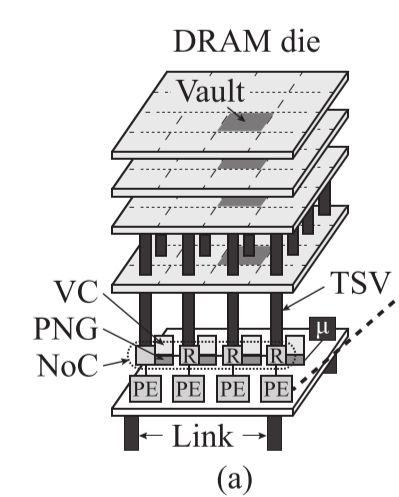
论文——《Neurocube: A Programmable Digital Neuromorphic Architecture with High-Density 3D Memory》

# Neurocube架构

Neurocube可以分为DRAM die和Logic die。

## DRAM Die

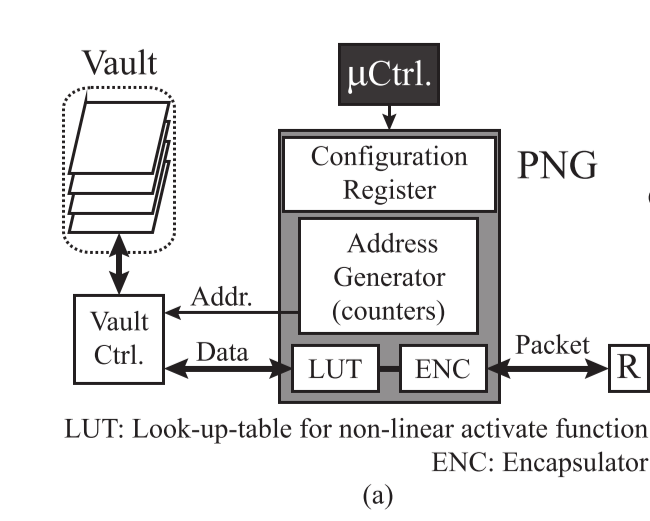
使用HMC架构。



将多层的DRAM分成16个vault，每个vault表示竖直方向一组内存分割块，内存分割块之间利用高速传输介质TSV进行传输。

每个Vault在Logic die都有唯一一个Vault控制器与之对应。

## Logic Die



每个Vault控制器对应唯一一个PNG，PNG与通过2D网格网络连接的Router连接，与PE通信。

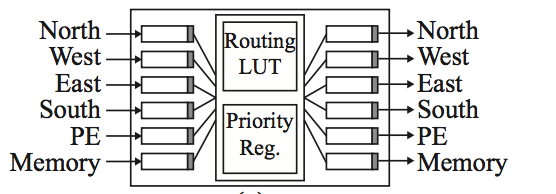
### PNG

PNG的架构如上图。

PNG负责控制神经计算所需的数据流动，生成在前一层连接的神经计算在内存中的地址和权值。

需要向Router发送的数据包在PNG中封装了源ID（Vault ID）和目的ID（PE ID），通过在NoC中的Router发送给对应的PE。

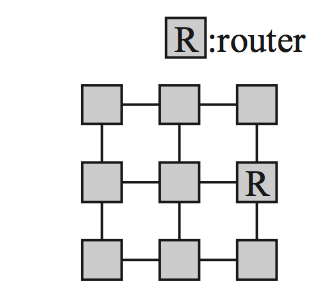
### Router



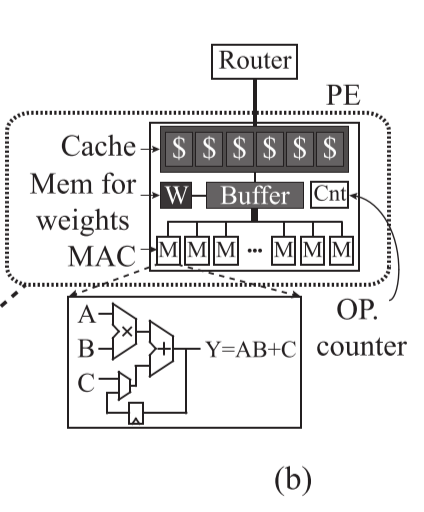
Router，可称为路由器。

路由器共有6个输入和6个输出（其中4个输入输出为连接临近的路由器，2个输入输出为连接PE和Memory）。

2D网格网络（Mesh Grid NoC）如下图，



### PE



PE是主要的计算单元，包含许多乘法累加器（MAC），一个cache内存，一个存放数据包的buffer，一个存放权值的buffer。

# 数据传输

不同的Vault和不同的PE之间的数据传输都需要经过Router，在Router所在的2D网格网络上进行传输。

当对应MAC的包到达PE时，

* 若它的OP-ID大于当前OP计数器的计数值时，它将被存入cache
* 否则直接放入Buffer，以供MAC使用。

当OP计数器变化时，先将cache中的包再放入Buffer。