













Binary





















































1. FPGA LUT () LUT
 - 1.
 - 2.
2. FPGA AI
 1. PPA
 2. FPGA
 3. PPA
3. FPGA LUT AI
 - 1.
4.
 1. LUT
 2. LUT







1. LUT6
 1. 6 1 LUT 64 64bit
 2. xc7k480t 74659 slices slice 4 LUT 8 flip-flop
 3. lut 64bit*74659*4 = 2389088 byte 74659*4 / 6 * 2 * 0.5G GFlops = 50 TFlops
 4. LUT LUT
2.
 - 1.
 2. 4 6LUT 8bit lut 8 LUT8 8bit 8
 - 3.
3. LUT
 1. LUT
 1. " +weight" LUT
 2. LUT LUT

1. 
2. 
 1. 
3. 
4. 
5. 
6. 
7.  shift  scale 

LLM ☐ ☐

1.  "  "  bit 
 1. 
2. 
 1. 
 2.  "  "
 3.   
3.  8bit 
 1.  bit 
 2. 8bit  256  weight
 3. weight  bit 
 1. 
 4. 
 1.  LUT 
 2.  LUT  bit  LUT
 3. 
4. 
 1.  / 
 2.  "  "  "  " 
 3. 
 1. 
 2. 
 3. 
5. 
 1. $x \rightarrow x$
 1.  attention  qkv  bit 
 1. QK  qk  input 
 2.  LUT  reduce

--	--

1.  LLM/CNN
2. 
3.  PPA
4.  FPGA



1.
2. AI
3.
4. DNA
5.
6. LUT
 1.
 2.
 3. scale



1. BitNet b1.58 <https://mp.weixin.qq.com/s/G9ZbMnBVbeH1m45HY2JlKA>
2.



1.
2.



1. <https://arxiv.org/html/2502.19008v1>
2. Binary Neural Networks <https://zhuanlan.zhihu.com/p/117285043>
3. <https://github.com/ryuz/BinaryBrain>
4.
<https://www.slideshare.net/kentotajiri/ss-77136469>
5. BinaryConnect: Training Deep Neural Networks with binary weights during propagations
<https://arxiv.org/pdf/1511.00363.pdf>
6. Binarized Neural Networks
<https://arxiv.org/abs/1602.02505>
7. Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1
<https://arxiv.org/abs/1602.02830>
8. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks
<https://arxiv.org/abs/1603.05279>
9. Xilinx UltraScale Architecture Configurable Logic Block User Guide
https://japan.xilinx.com/support/documentation/user_guides/ug574-ultrascale-clb.pdf

Revision #59

Created 12 April 2025 08:03:12 by Colin

Updated 29 May 2025 06:52:02 by Colin