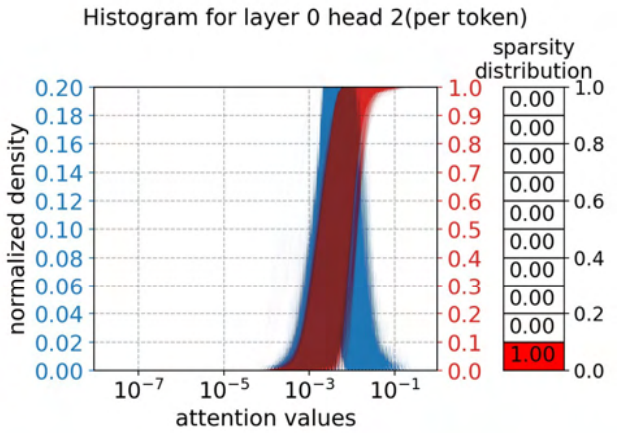
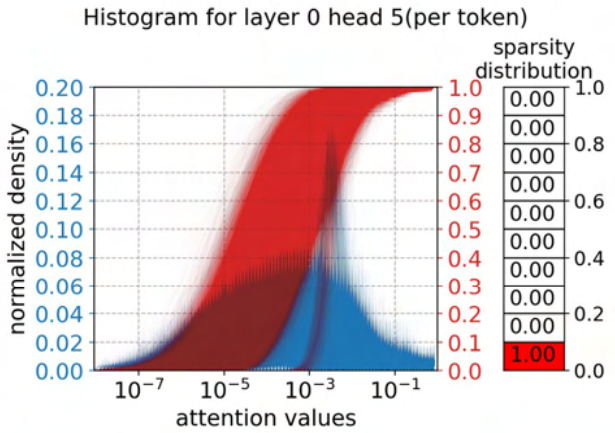
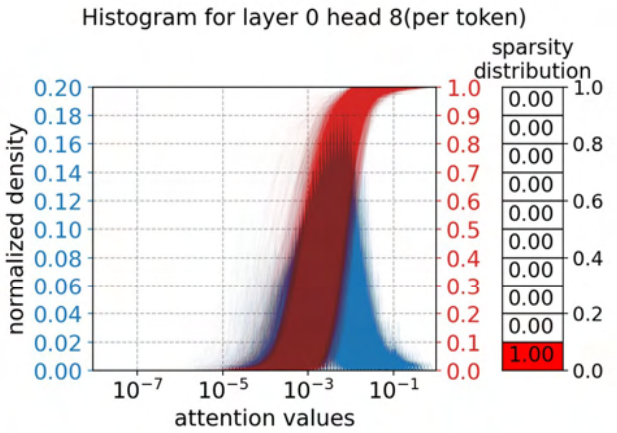
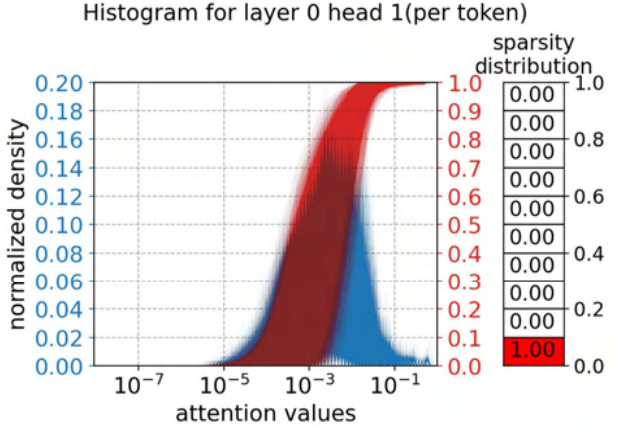
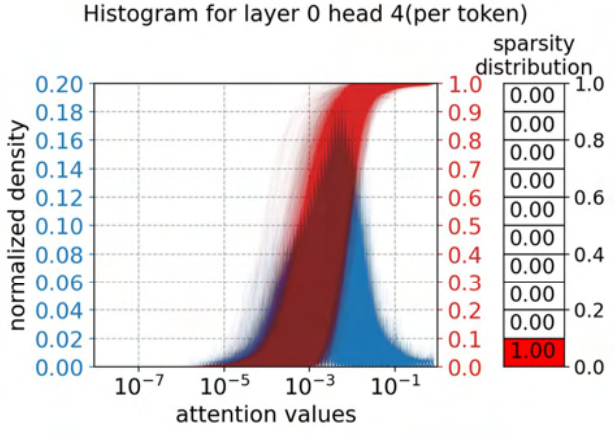
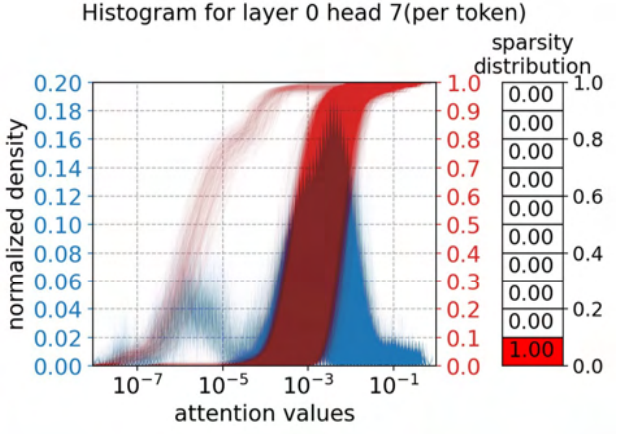

Histogram for layer 0 head 2(per token)
sparsity distribution
normalized density
attention values

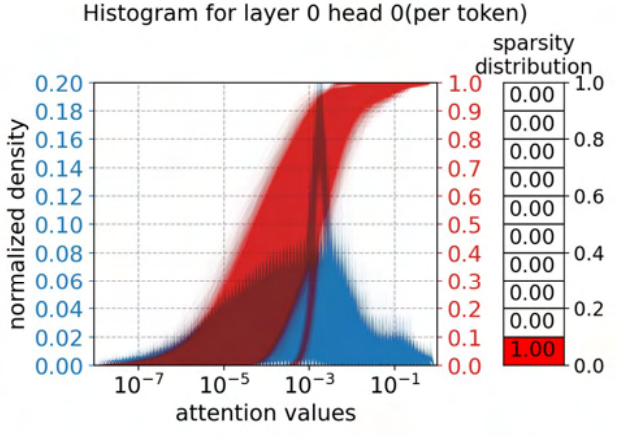

Histogram for layer 0 head 5(per token)
sparsity distribution
normalized density
attention values

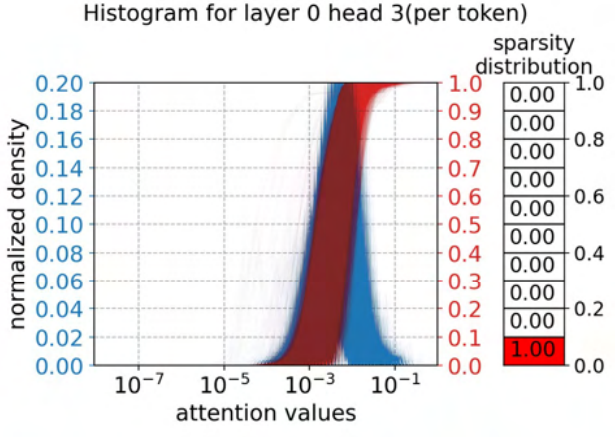

Histogram for layer 0 head 8(per token)
sparsity distribution
normalized density
attention values

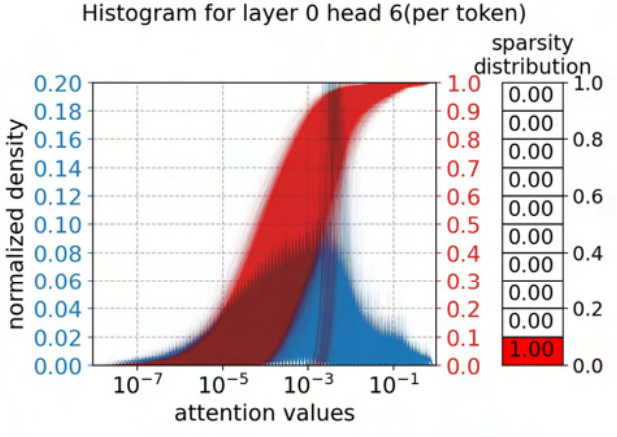

Histogram for layer 0 head 1(per token)
sparsity distribution
normalized density
attention values

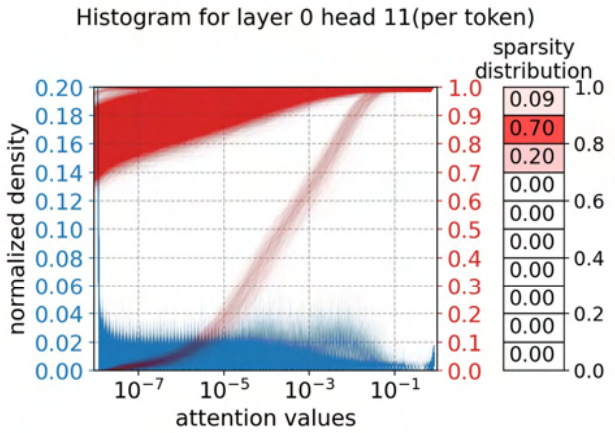

Histogram for layer 0 head 4(per token)
sparsity distribution
normalized density
attention values

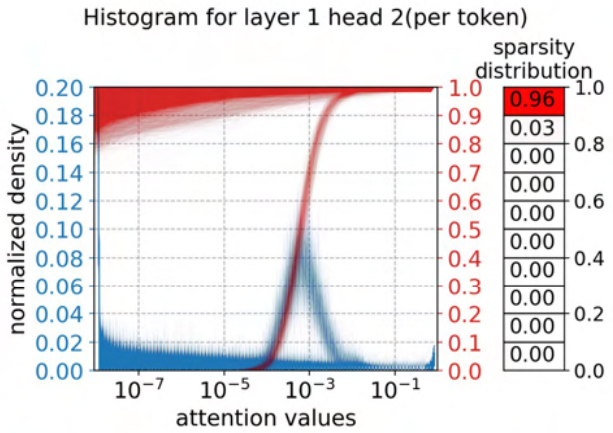

Histogram for layer 0 head 7(per token)
sparsity distribution
normalized density
attention values

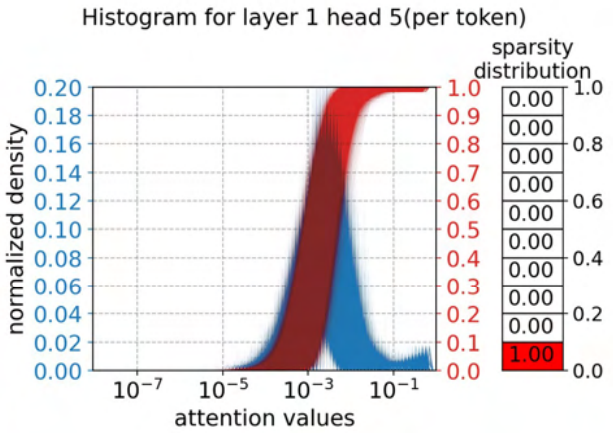

Histogram for layer 0 head 0(per token)
sparsity distribution
normalized density
attention values

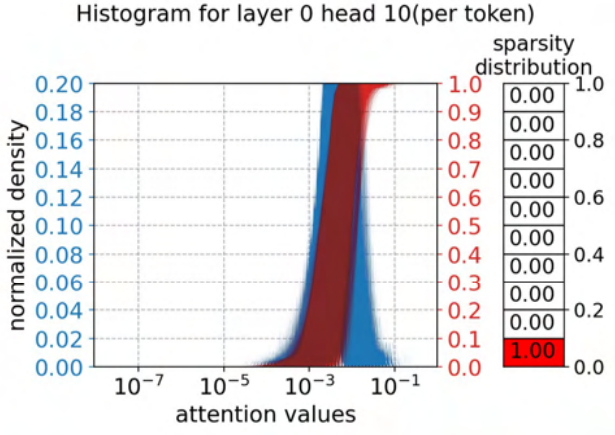

Histogram for layer 0 head 3(per token)
sparsity distribution
normalized density
attention values

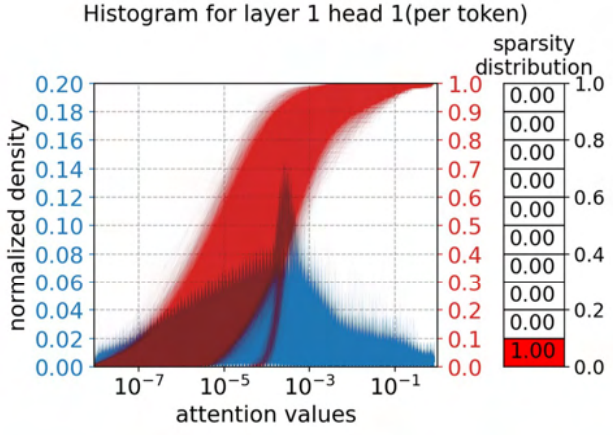

Histogram for layer 0 head 6(per token)
sparsity distribution
normalized density
attention values

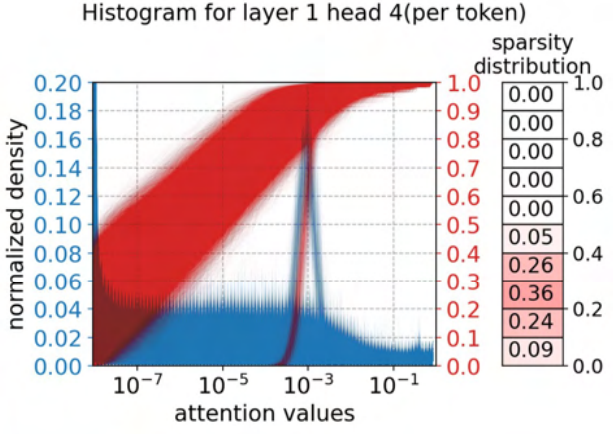

Histogram for layer 0 head 11(per token)
sparsity distribution
0.09
0.70
0.20
0.00
0.00
0.00
0.00
0.00
0.00
0.00
normalized density
attention values

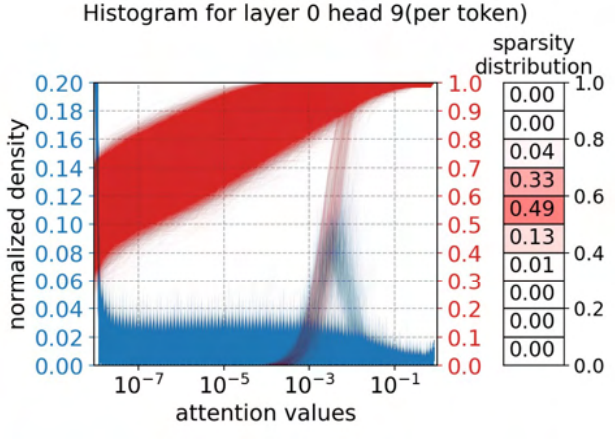

Histogram for layer 1 head 2(per token)
sparsity distribution
0.96
0.03
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
normalized density
attention values

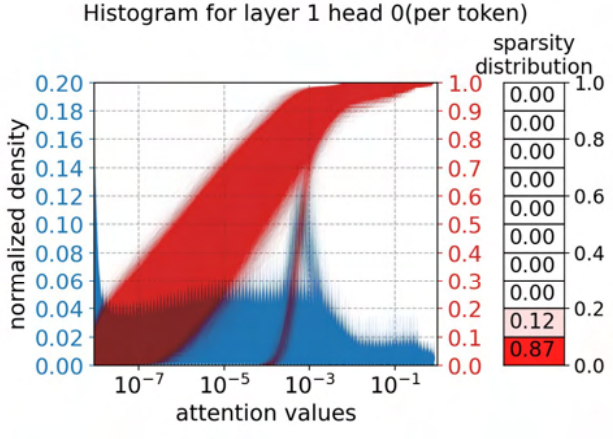

Histogram for layer 1 head 5(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

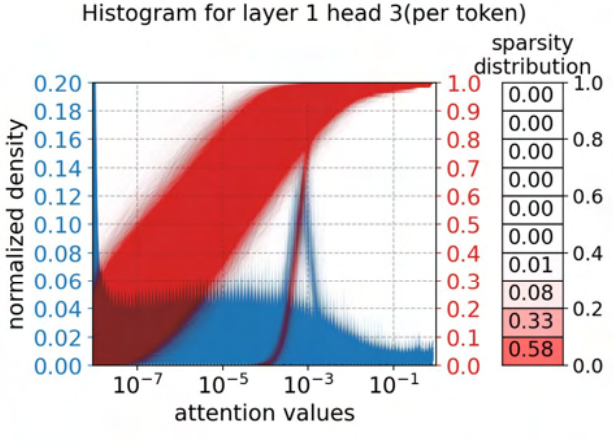

Histogram for layer 0 head 10(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

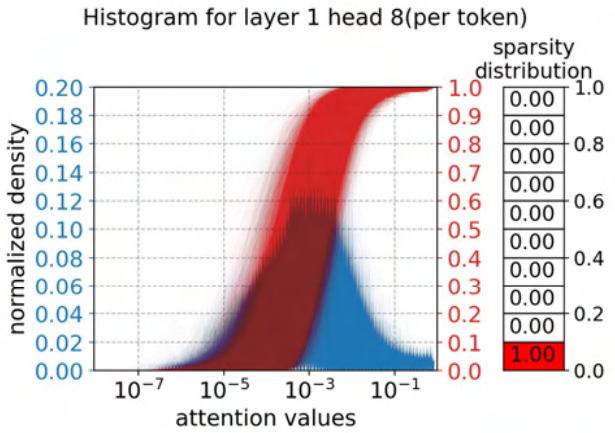

Histogram for layer 1 head 1(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

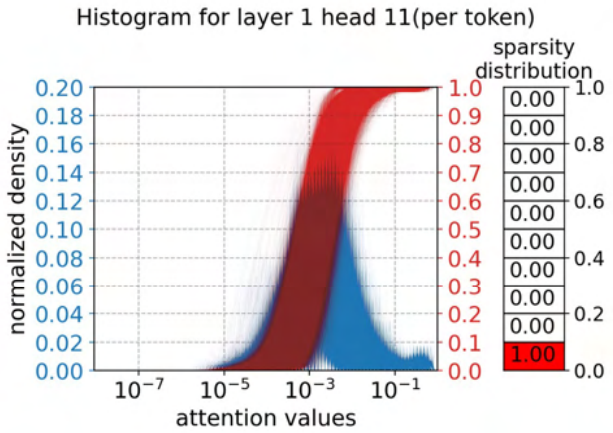

Histogram for layer 1 head 4(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.05
0.26
0.36
0.24
0.09
normalized density
attention values

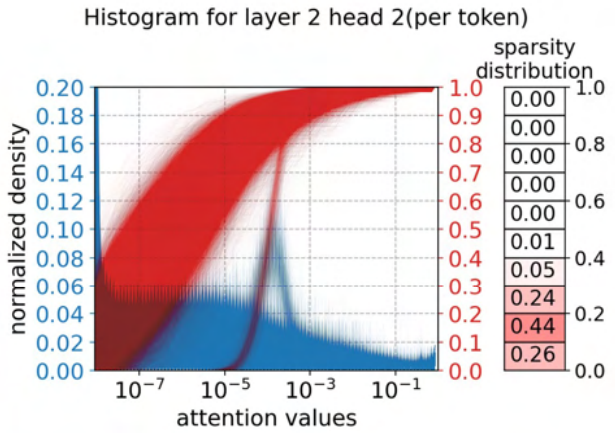

Histogram for layer 0 head 9(per token)
sparsity distribution
0.00
0.00
0.04
0.33
0.49
0.13
0.01
0.00
0.00
0.00
normalized density
attention values

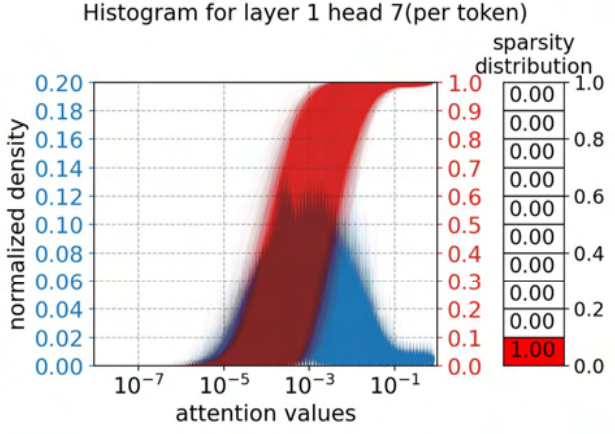

Histogram for layer 1 head 0(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.12
0.87
normalized density
attention values

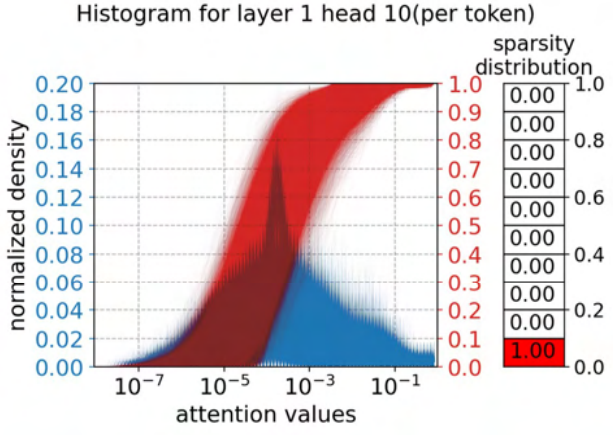

Histogram for layer 1 head 3(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.08
0.33
0.58
normalized density
attention values

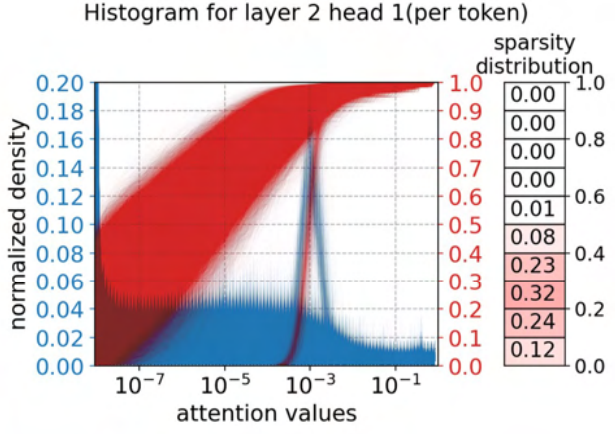

Histogram for layer 1 head 8(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

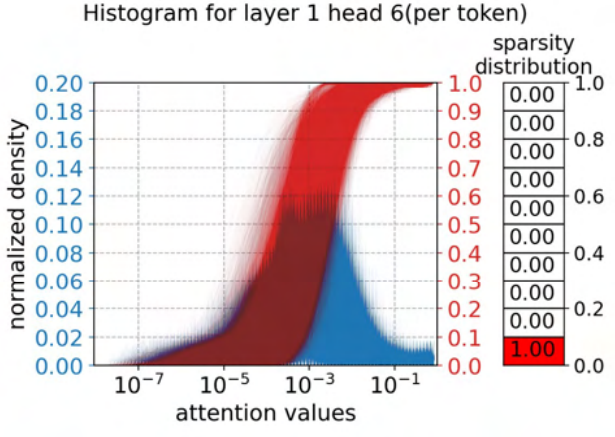

Histogram for layer 1 head 11(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

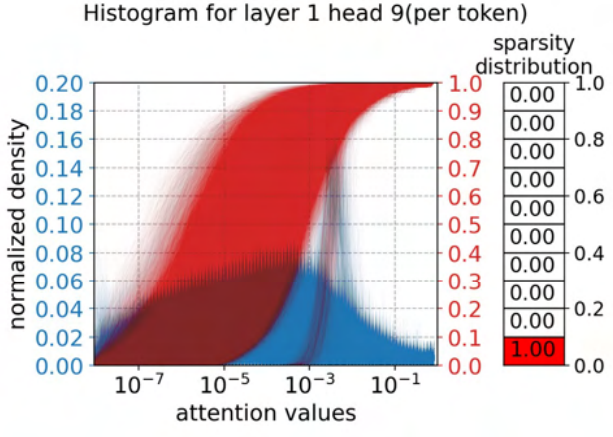

Histogram for layer 2 head 2(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.01
0.05
0.24
0.44
0.26

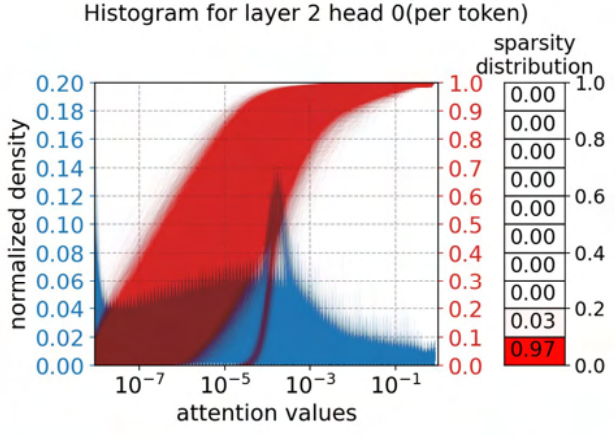

Histogram for layer 1 head 7(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

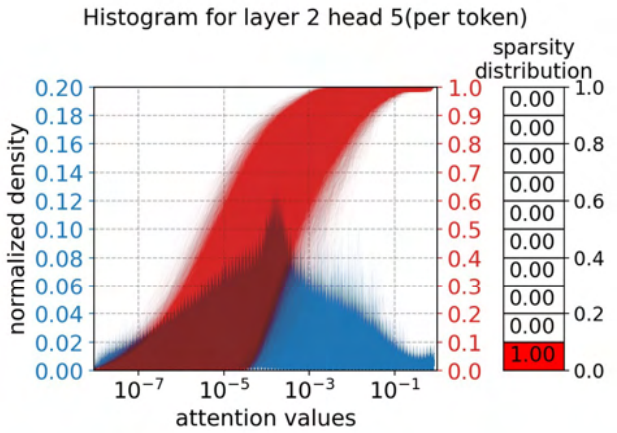

Histogram for layer 1 head 10(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

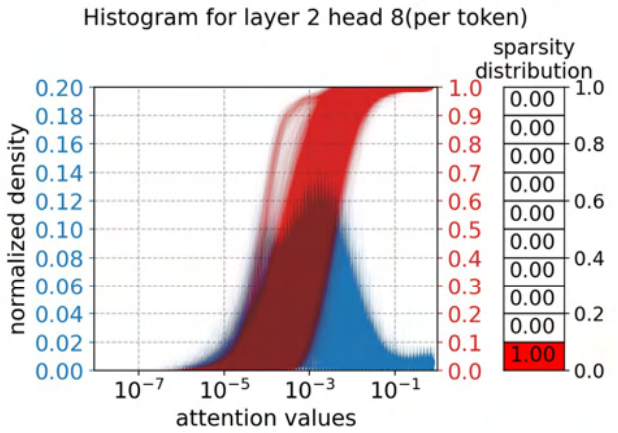

Histogram for layer 2 head 1(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.01
0.08
0.23
0.32
0.24
0.12

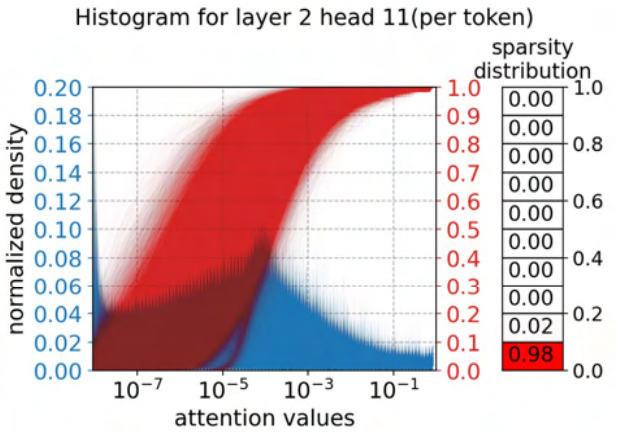

Histogram for layer 1 head 6(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

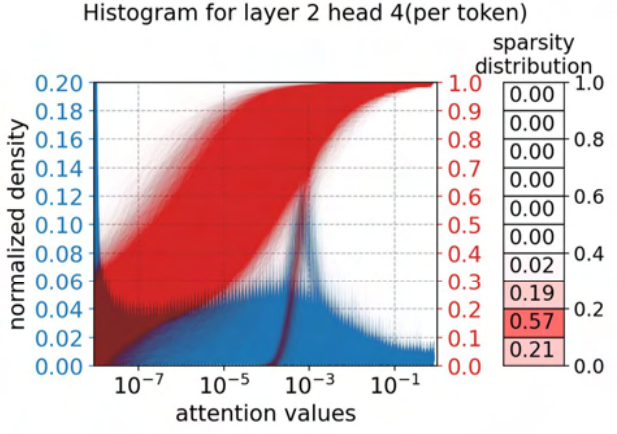

Histogram for layer 1 head 9(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

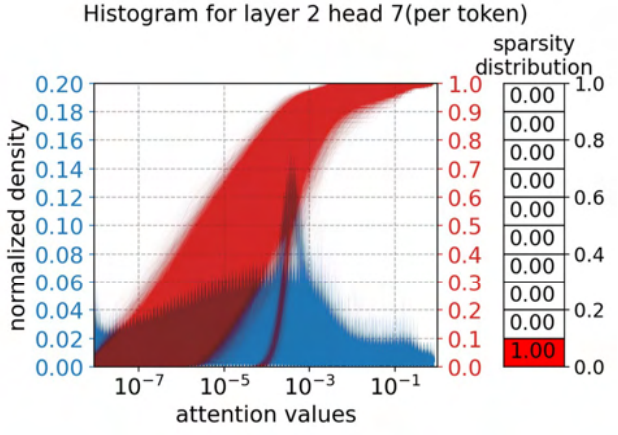

Histogram for layer 2 head 0(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.03
0.97

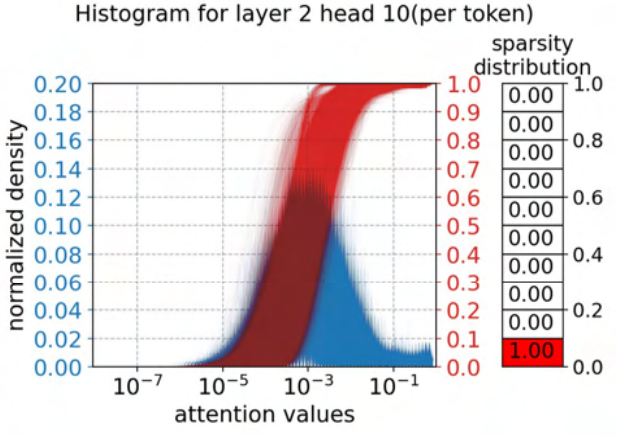

Histogram for layer 2 head 5(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

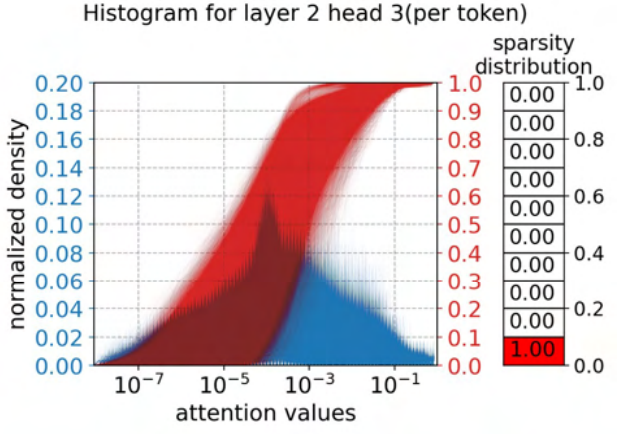

Histogram for layer 2 head 8(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

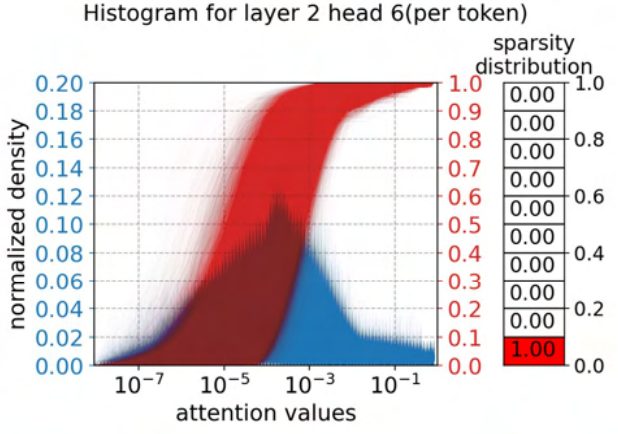

Histogram for layer 2 head 11(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.02
0.98

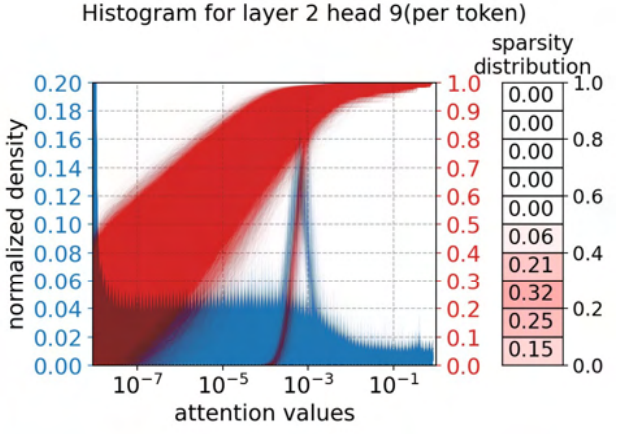

Histogram for layer 2 head 4(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.02
0.19
0.57
0.21

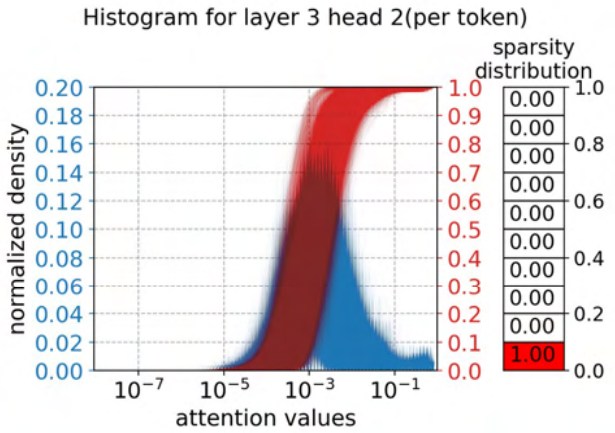

Histogram for layer 2 head 7(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

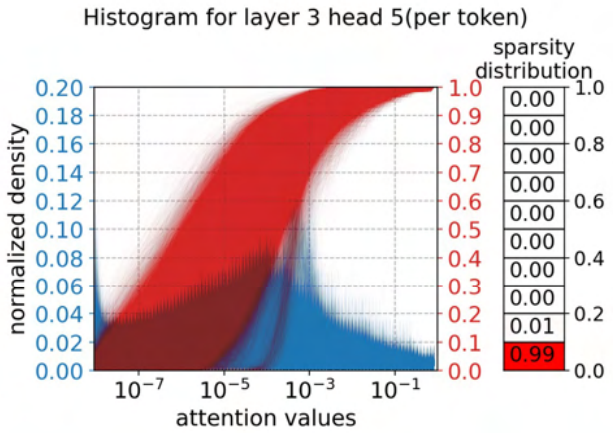

Histogram for layer 2 head 10(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

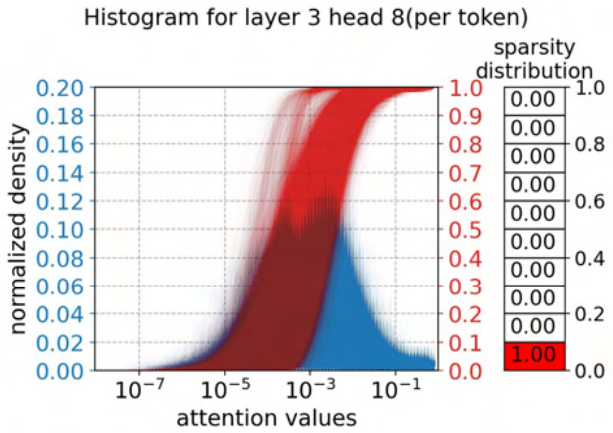

Histogram for layer 2 head 3(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

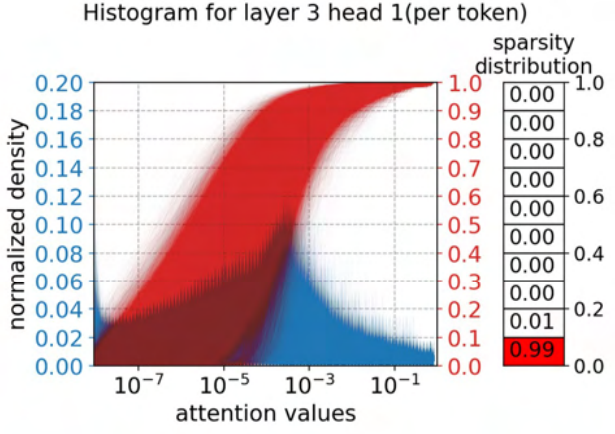

Histogram for layer 2 head 6(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

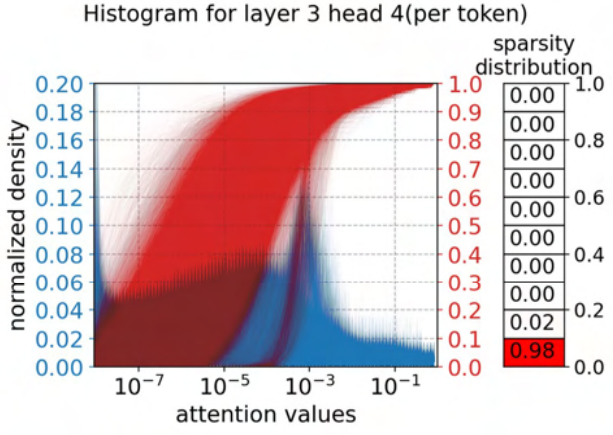

Histogram for layer 2 head 9(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.06
0.21
0.32
0.25
0.15

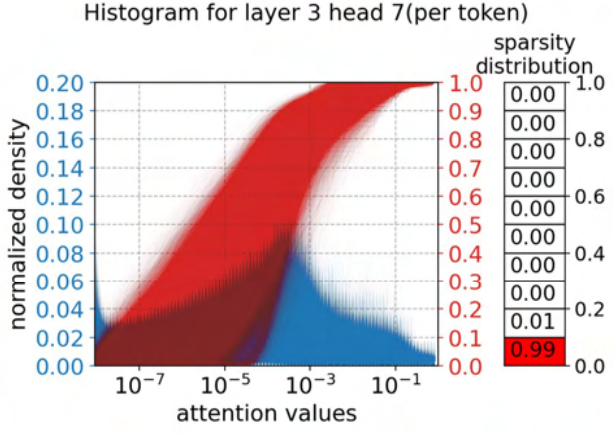

Histogram for layer 3 head 2(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

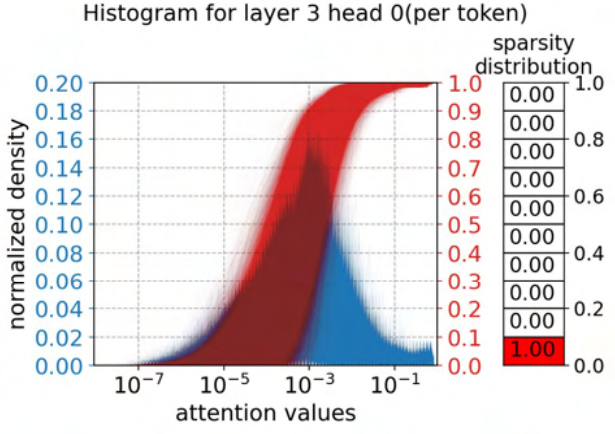

Histogram for layer 3 head 5(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.99
normalized density
attention values

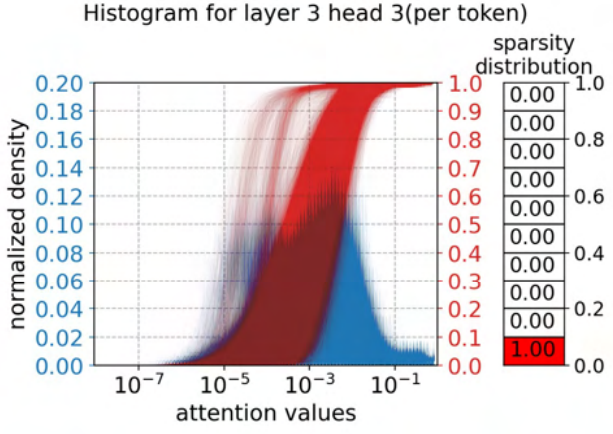

Histogram for layer 3 head 8(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

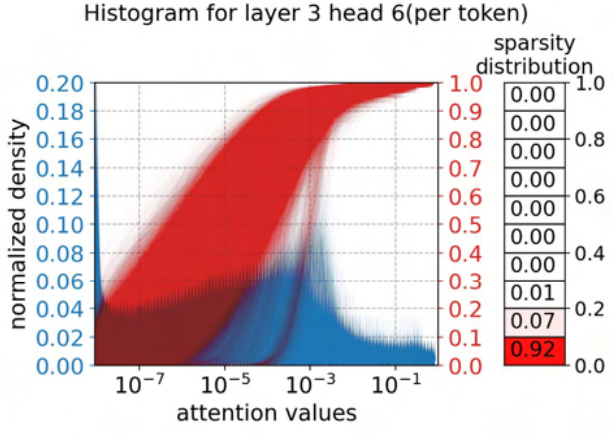

Histogram for layer 3 head 1(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.99
normalized density
attention values

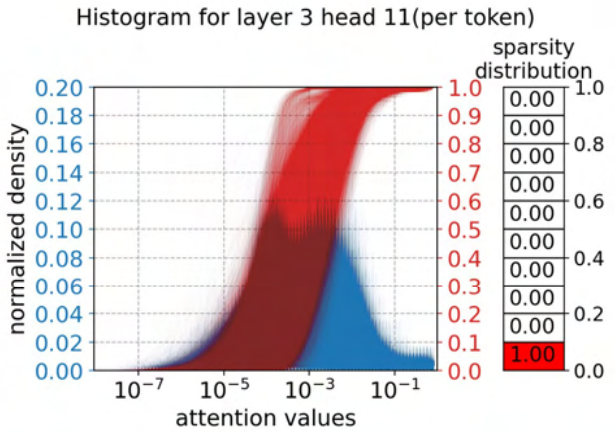

Histogram for layer 3 head 4(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.02
0.98
normalized density
attention values

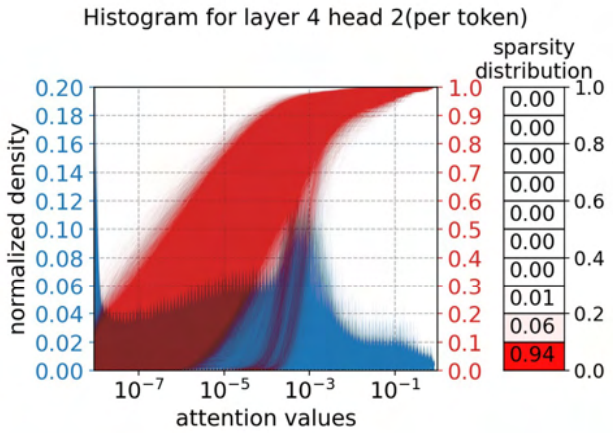

Histogram for layer 3 head 7(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.99
normalized density
attention values

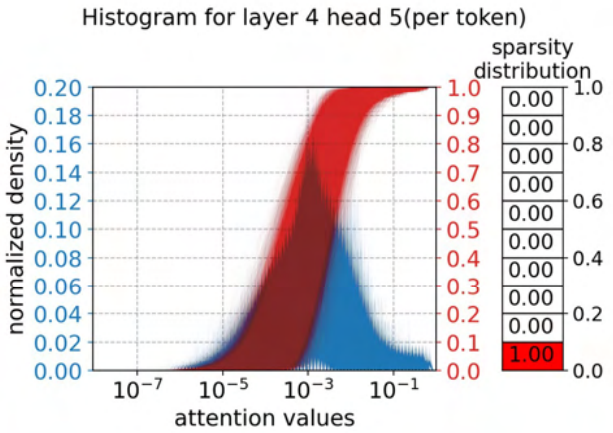

Histogram for layer 3 head 0(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

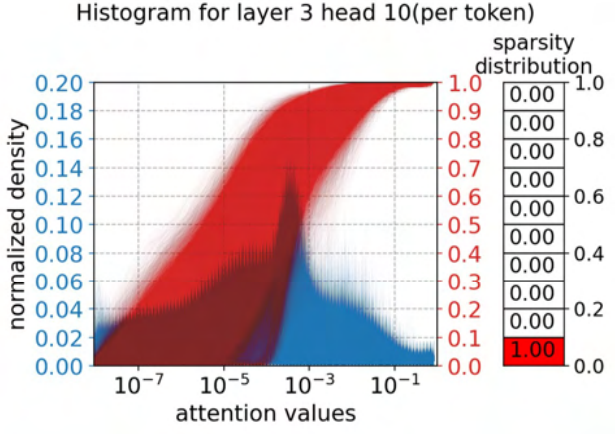

Histogram for layer 3 head 3(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

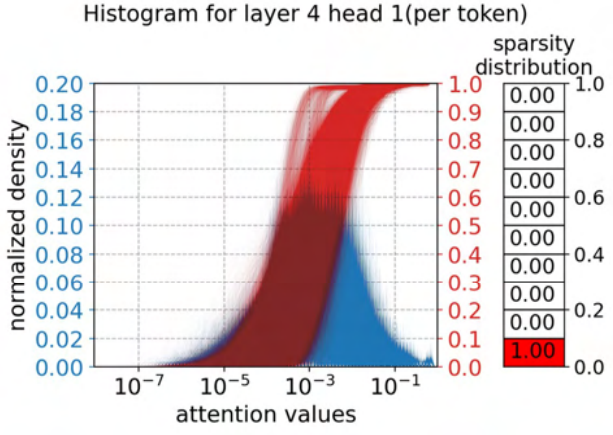

Histogram for layer 3 head 6(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.07
0.92
normalized density
attention values

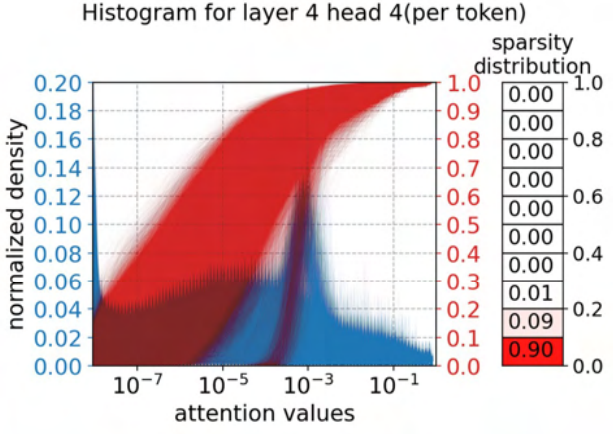

Histogram for layer 3 head 11(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

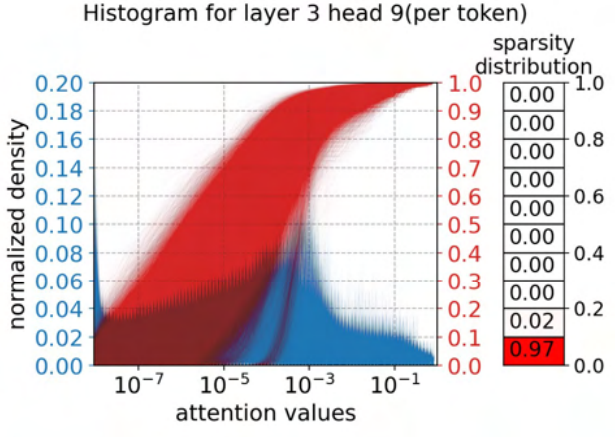

Histogram for layer 4 head 2(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.06 0.94

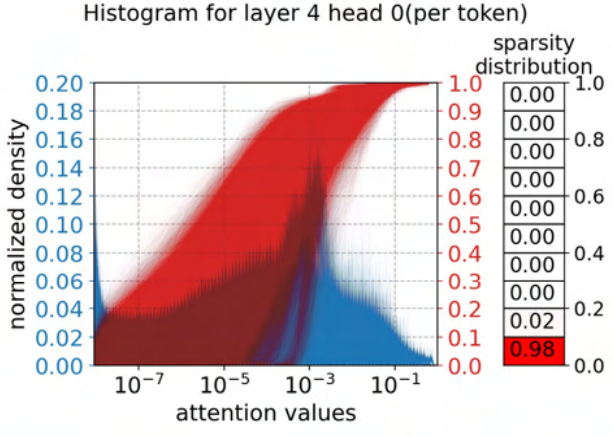

Histogram for layer 4 head 5(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

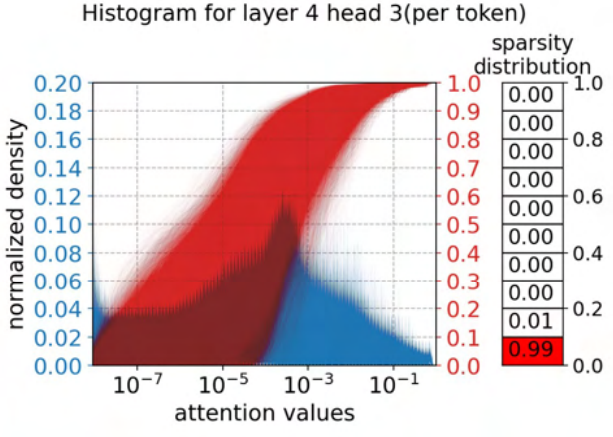

Histogram for layer 3 head 10(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

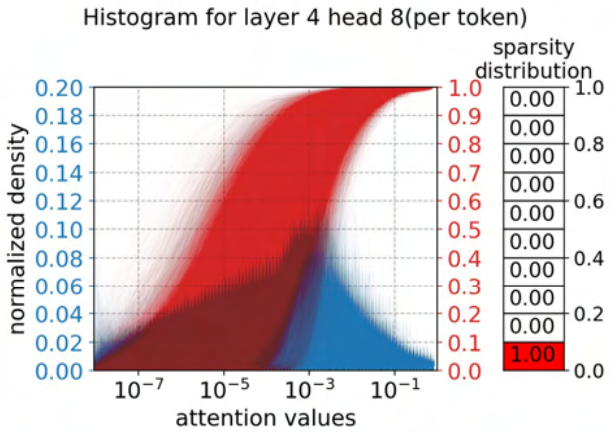

Histogram for layer 4 head 1(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

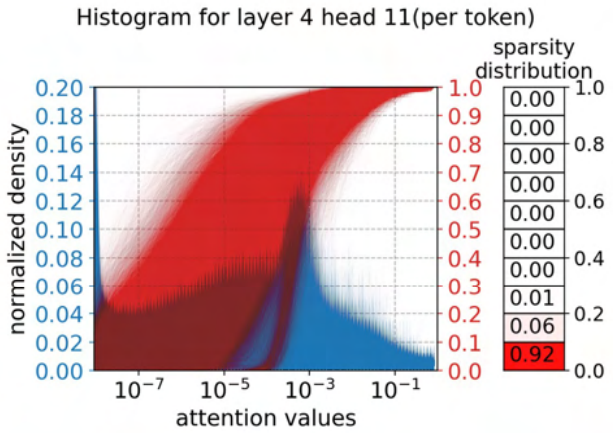

Histogram for layer 4 head 4(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.09 0.90

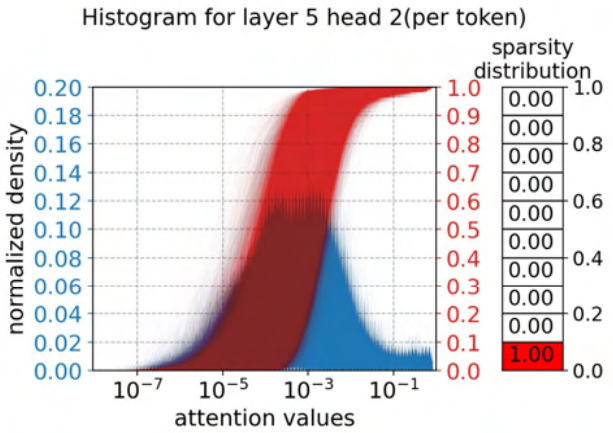

Histogram for layer 3 head 9(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.97

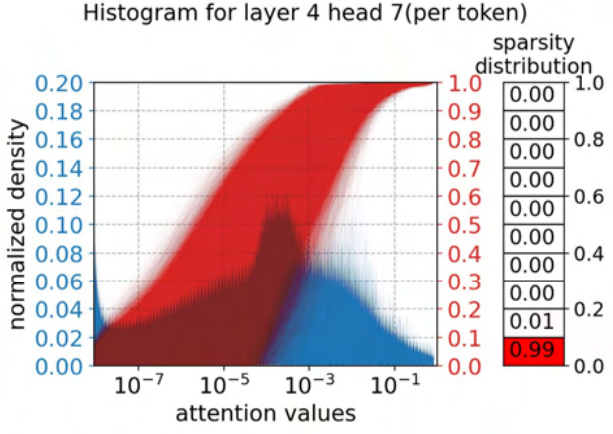

Histogram for layer 4 head 0(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.98

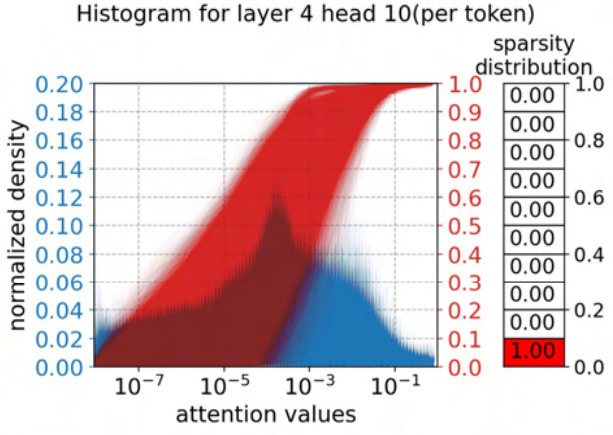

Histogram for layer 4 head 3(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.99

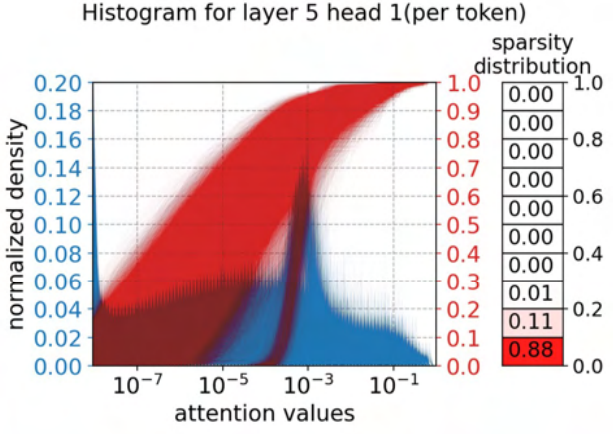

Histogram for layer 4 head 8(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

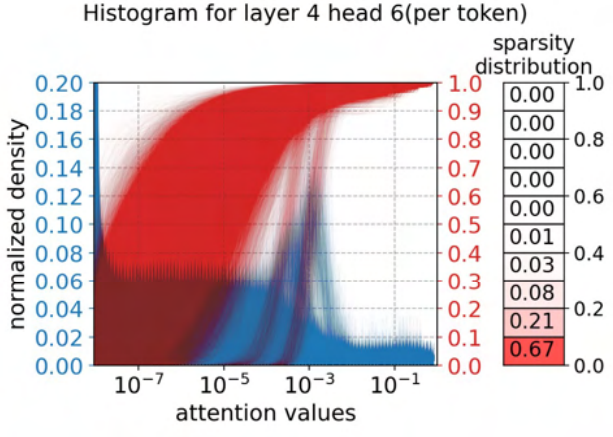

Histogram for layer 4 head 11(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.06
0.92

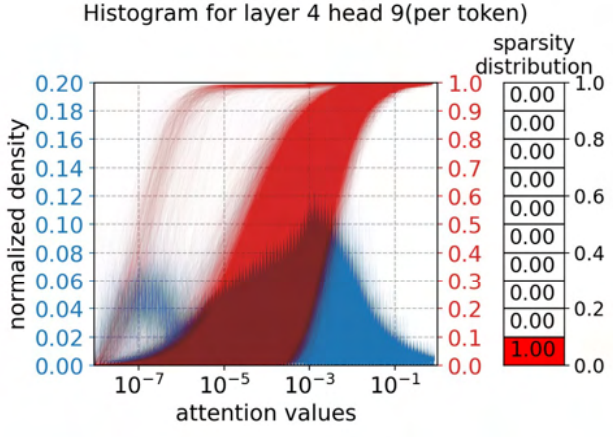

Histogram for layer 5 head 2(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

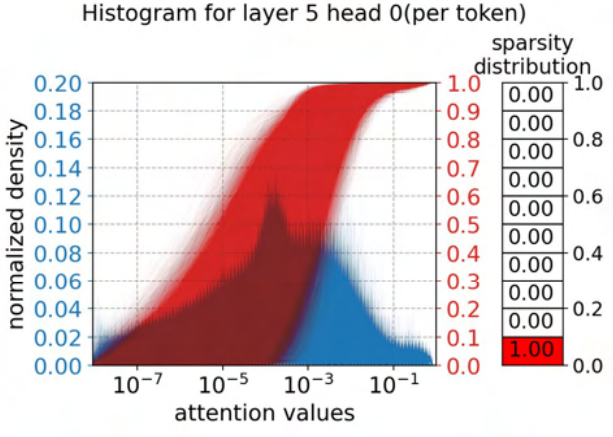

Histogram for layer 4 head 7(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.99

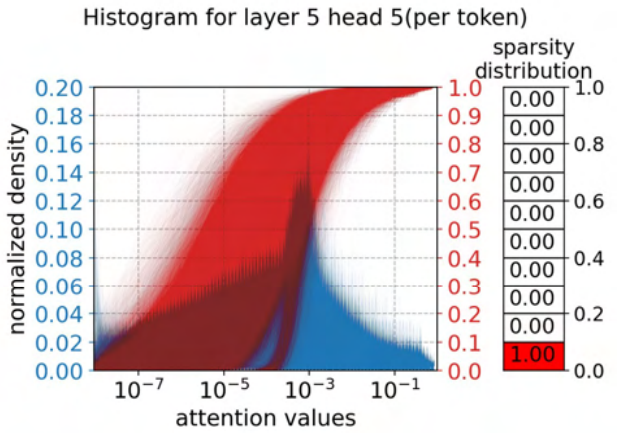

Histogram for layer 4 head 10(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

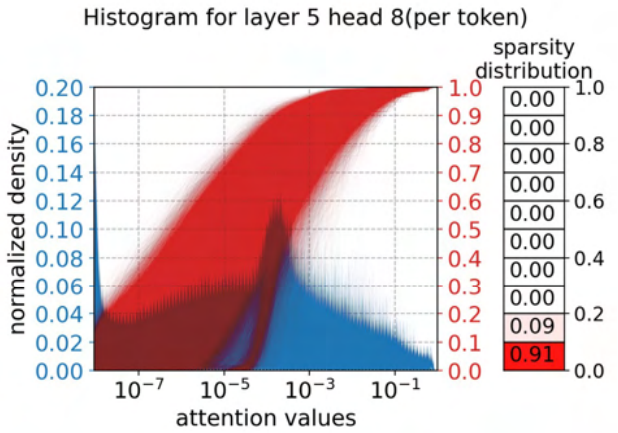

Histogram for layer 5 head 1(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.11
0.88

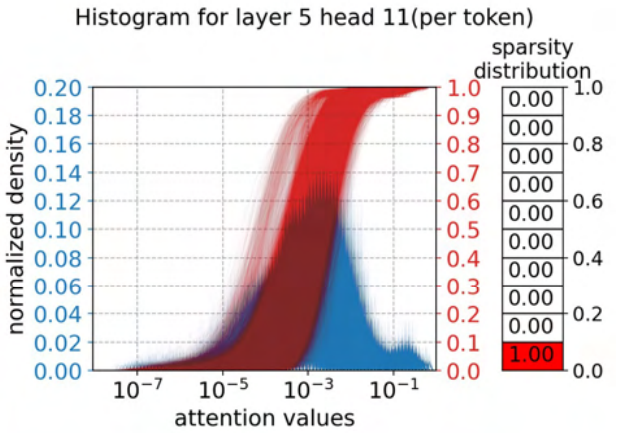

Histogram for layer 4 head 6(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.01
0.03
0.08
0.21
0.67

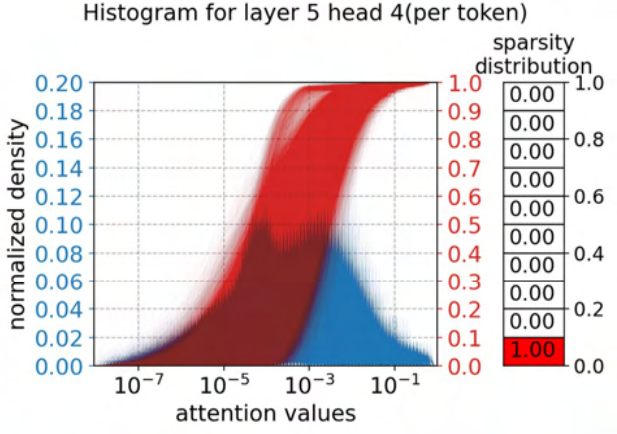

Histogram for layer 4 head 9(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

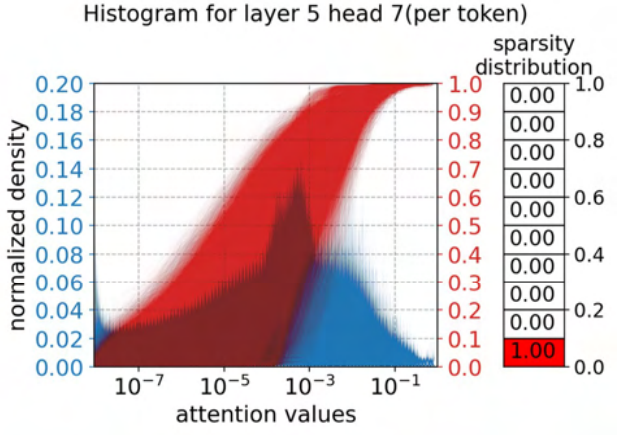

Histogram for layer 5 head 0(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

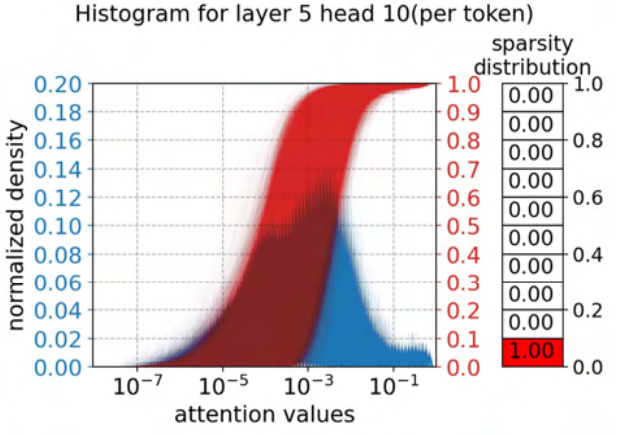

Histogram for layer 5 head 5(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

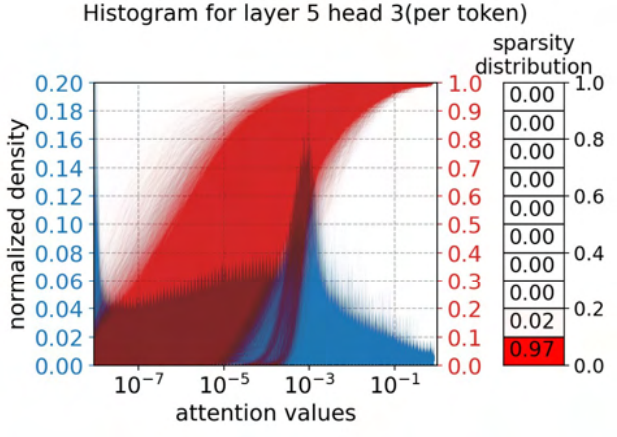

Histogram for layer 5 head 8(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.09
0.91

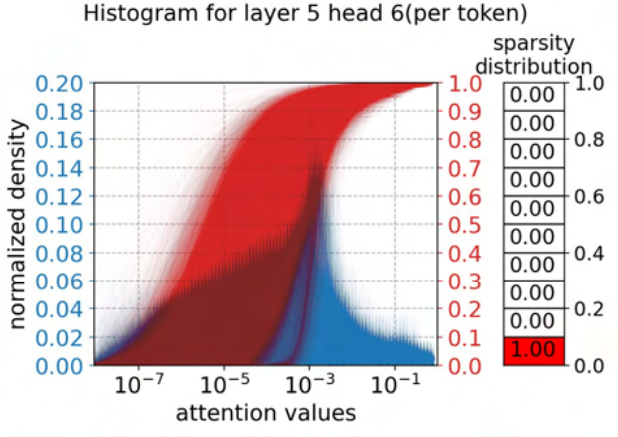

Histogram for layer 5 head 11(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

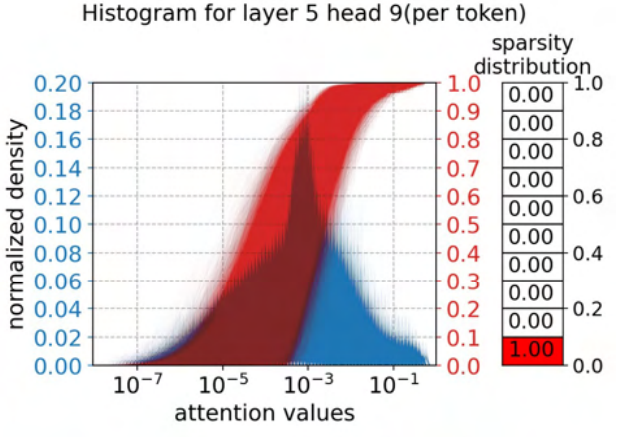

Histogram for layer 5 head 4(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

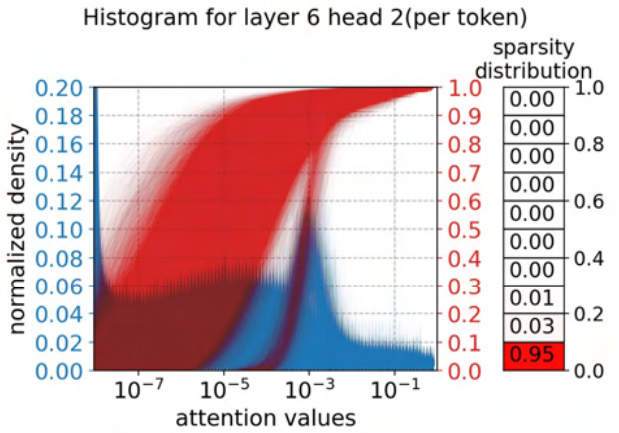

Histogram for layer 5 head 7(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

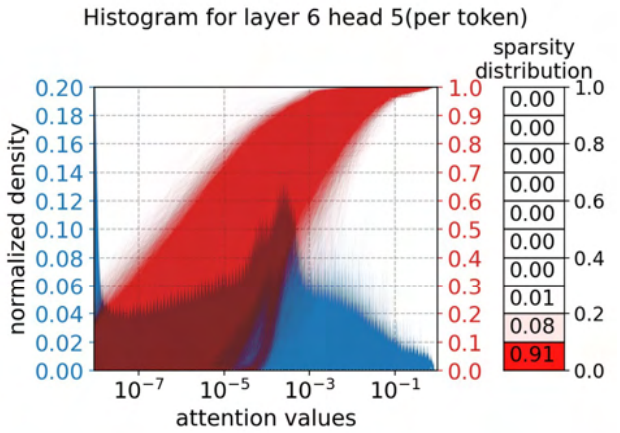

Histogram for layer 5 head 10(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

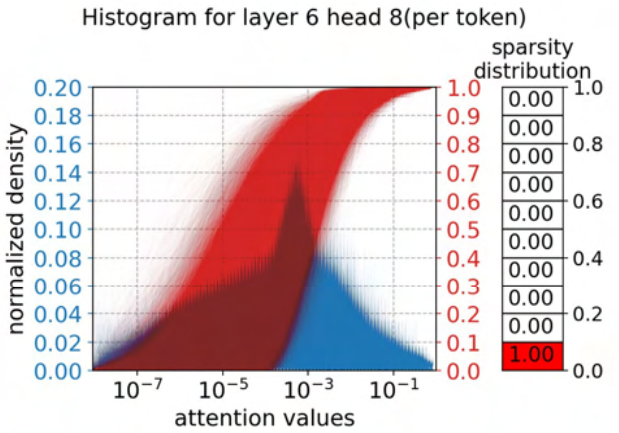

Histogram for layer 5 head 3(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.02
0.97

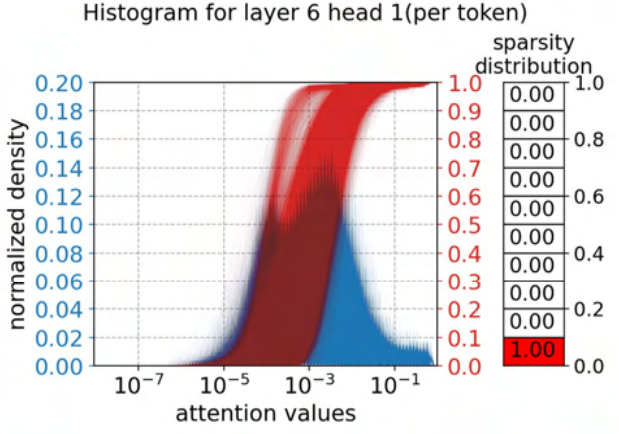

Histogram for layer 5 head 6(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

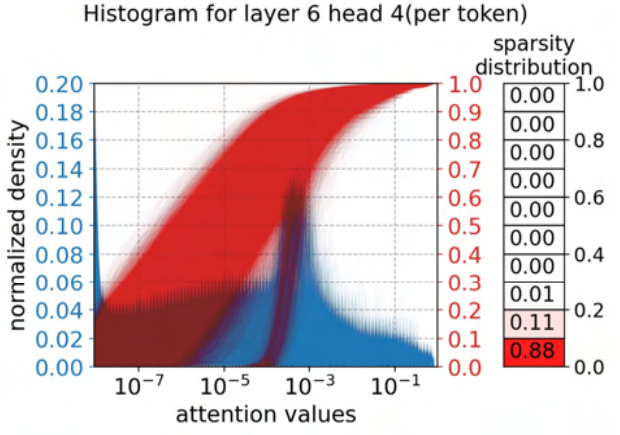

Histogram for layer 5 head 9(per token)
sparsity distribution
normalized density
attention values
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00

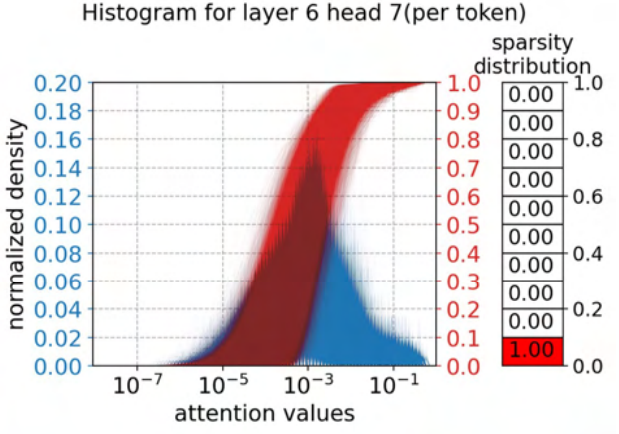

Histogram for layer 6 head 2(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.03 0.95

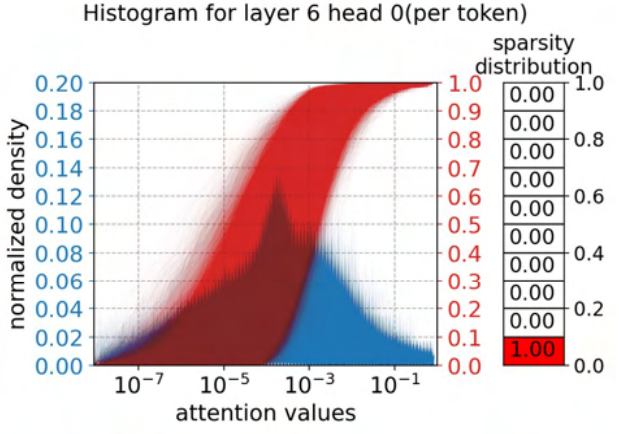

Histogram for layer 6 head 5(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.08 0.91

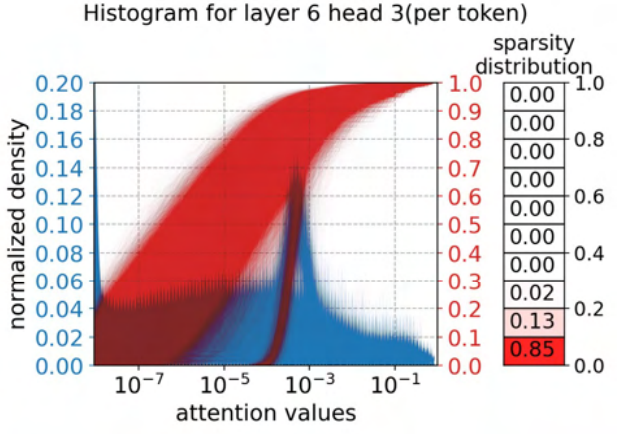

Histogram for layer 6 head 8(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

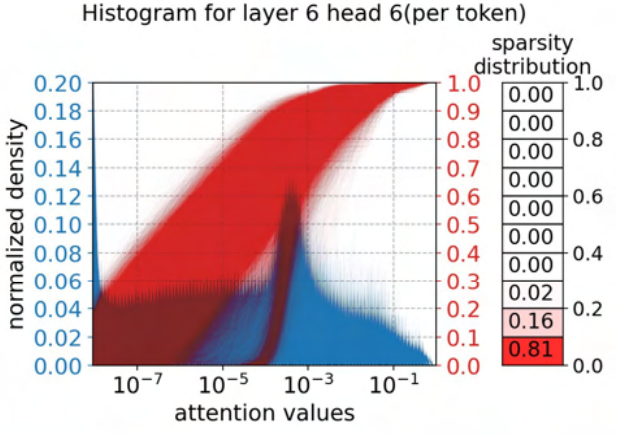

Histogram for layer 6 head 1(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

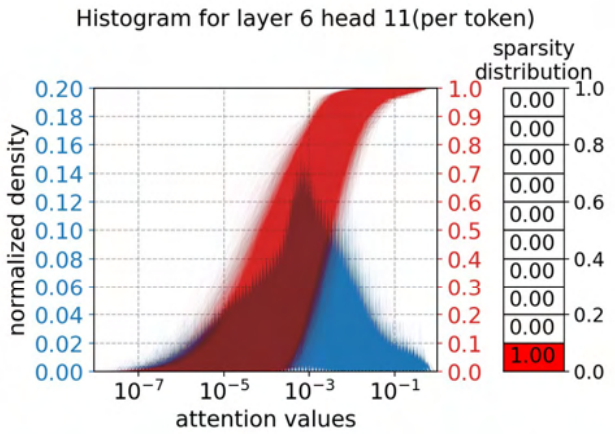

Histogram for layer 6 head 4(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.11 0.88

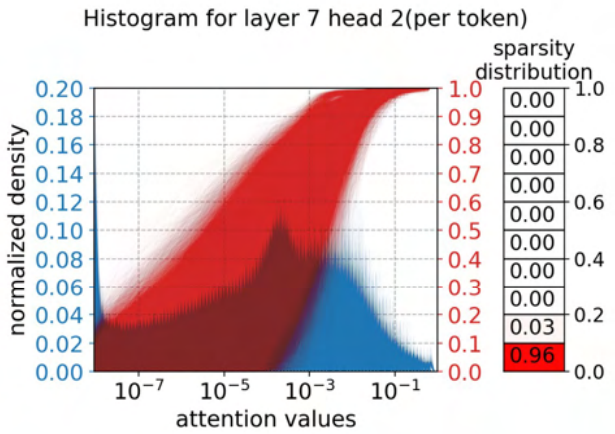

Histogram for layer 6 head 7(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

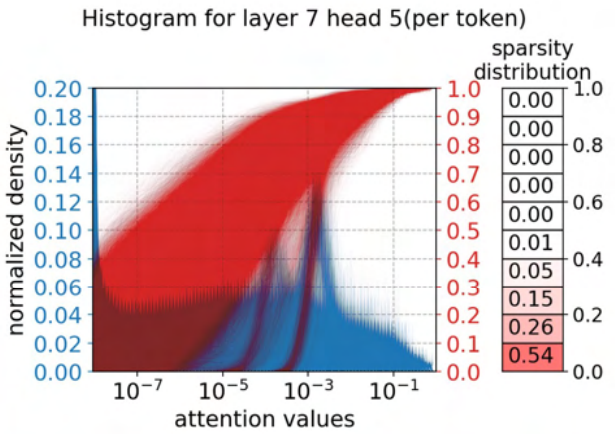

Histogram for layer 6 head 0(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

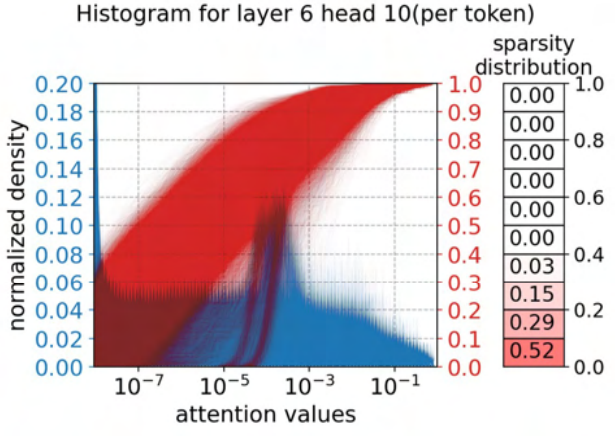

Histogram for layer 6 head 3(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.13 0.85

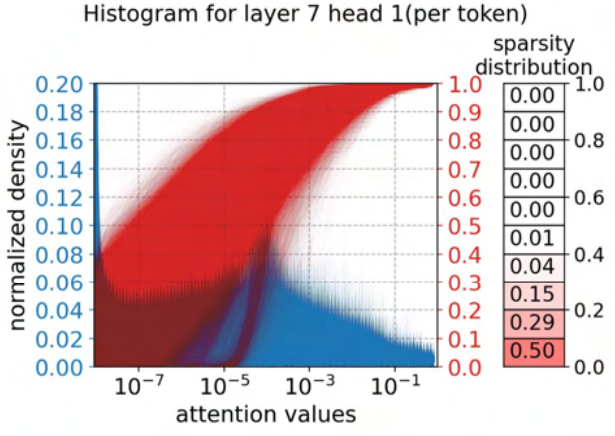

Histogram for layer 6 head 6(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.16 0.81

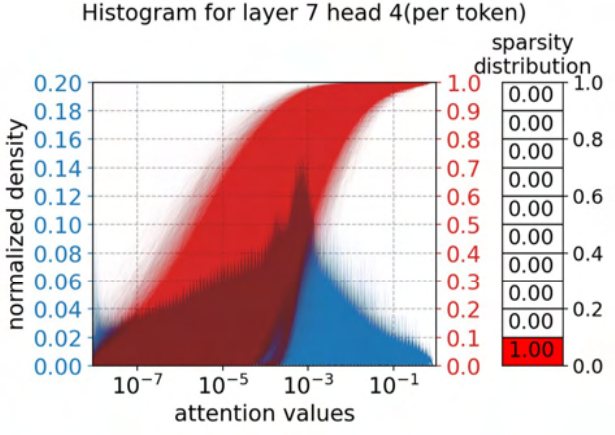

Histogram for layer 6 head 11(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

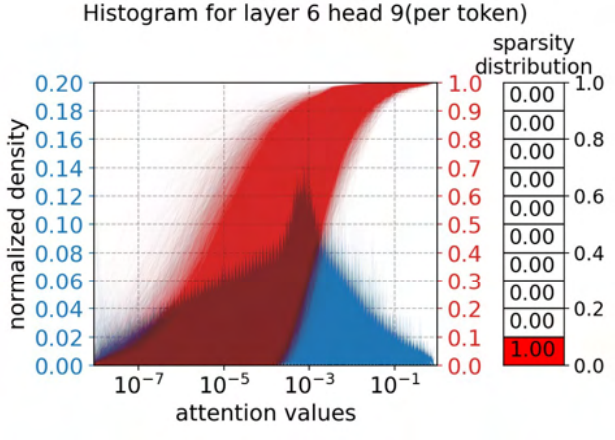

Histogram for layer 7 head 2(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.96

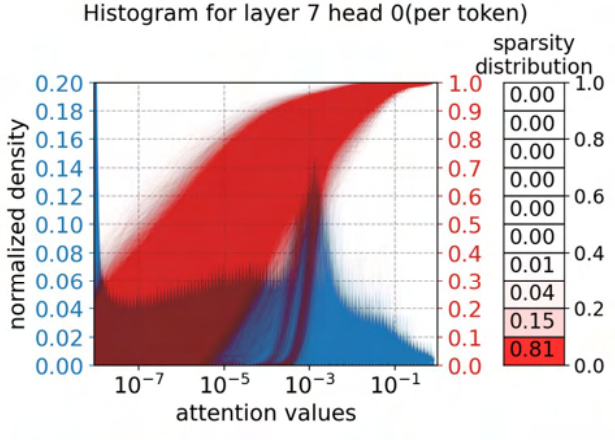

Histogram for layer 7 head 5(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.01 0.05 0.15 0.26 0.54

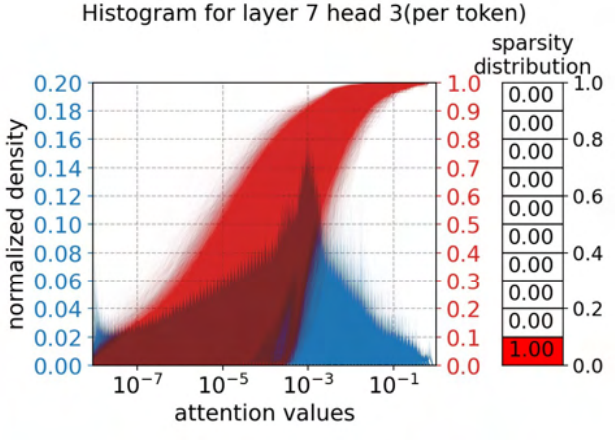

Histogram for layer 6 head 10(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.15 0.29 0.52

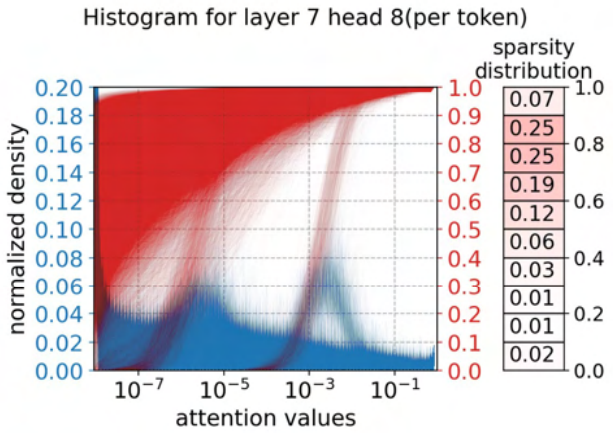

Histogram for layer 7 head 1(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.01 0.04 0.15 0.29 0.50

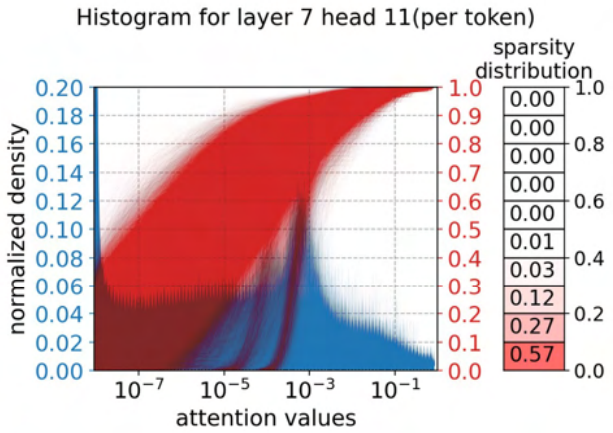

Histogram for layer 7 head 4(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

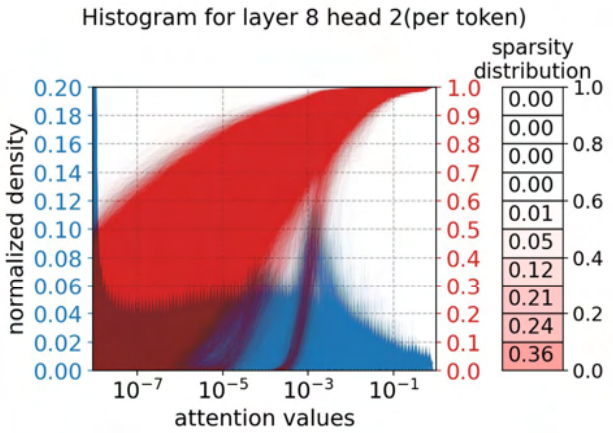

Histogram for layer 6 head 9(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

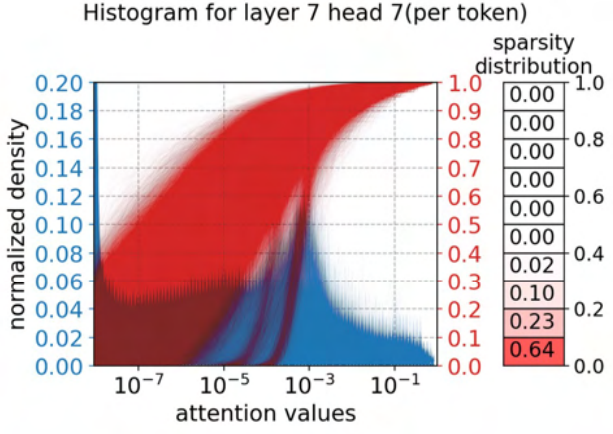

Histogram for layer 7 head 0(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.04 0.15 0.81

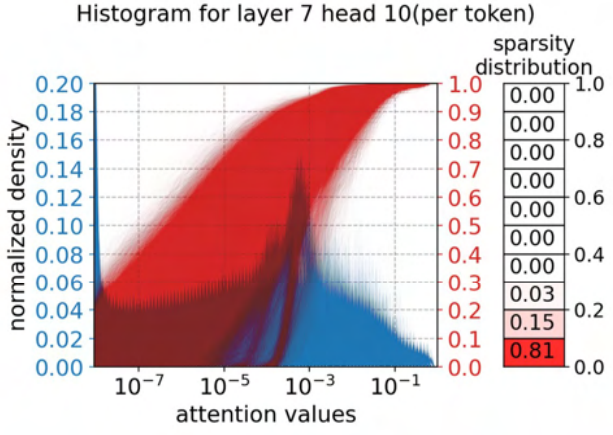

Histogram for layer 7 head 3(per token)
sparsity distribution
normalized density
attention values
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

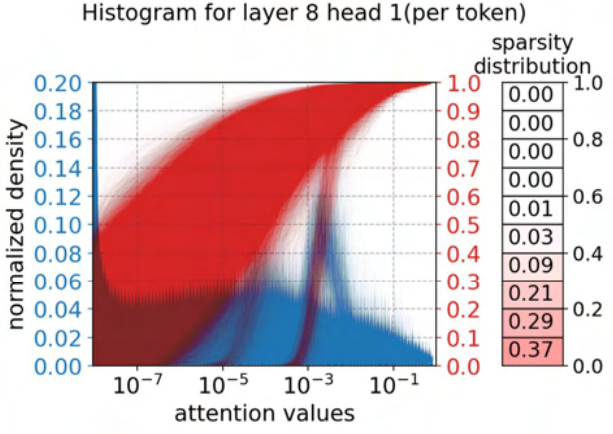

Histogram for layer 7 head 8(per token)
sparsity distribution
0.07
0.25
0.25
0.19
0.12
0.06
0.03
0.01
0.01
0.02
normalized density
attention values

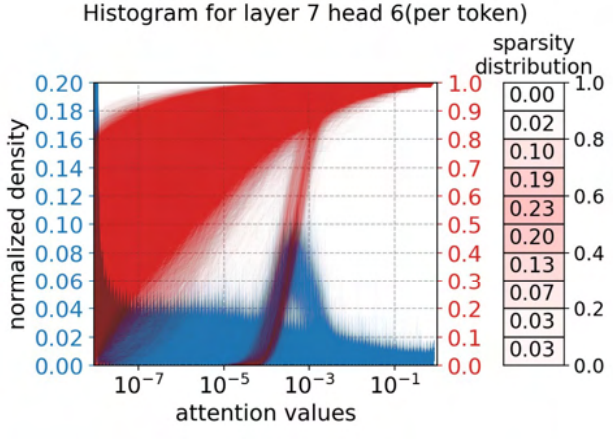

Histogram for layer 7 head 11(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.01
0.03
0.12
0.27
0.57
normalized density
attention values

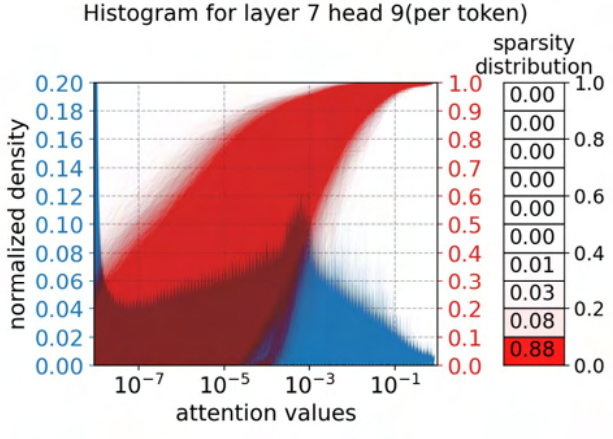

Histogram for layer 8 head 2(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.05
0.12
0.21
0.24
0.36
normalized density
attention values

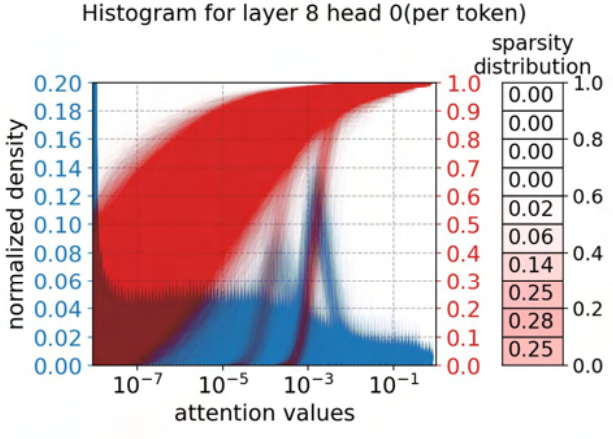

Histogram for layer 7 head 7(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.02
0.10
0.23
0.64
normalized density
attention values

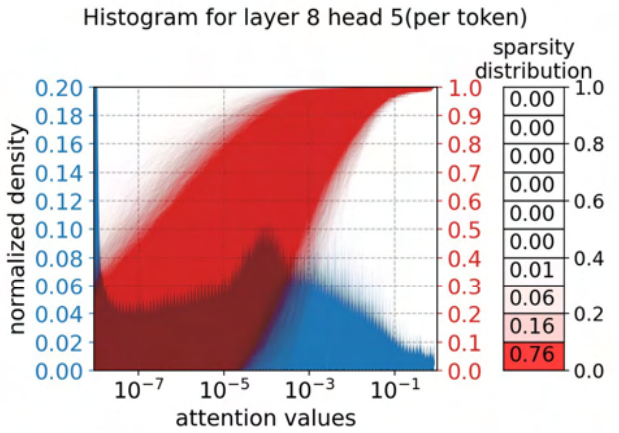

Histogram for layer 7 head 10(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.03
0.15
0.81
normalized density
attention values

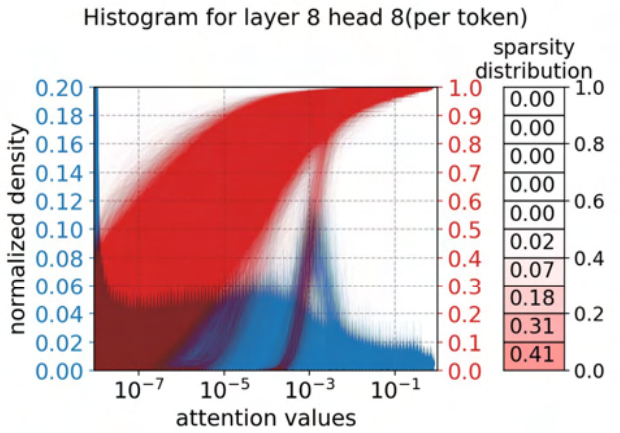

Histogram for layer 8 head 1(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.03
0.09
0.21
0.29
0.37
normalized density
attention values

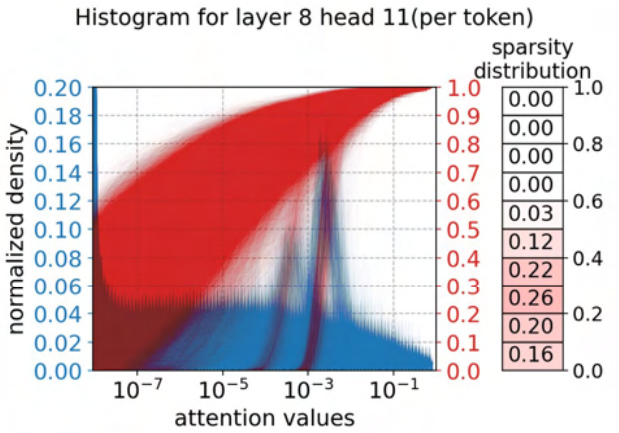

Histogram for layer 7 head 6(per token)
sparsity distribution
0.00
0.02
0.10
0.19
0.23
0.20
0.13
0.07
0.03
0.03
normalized density
attention values

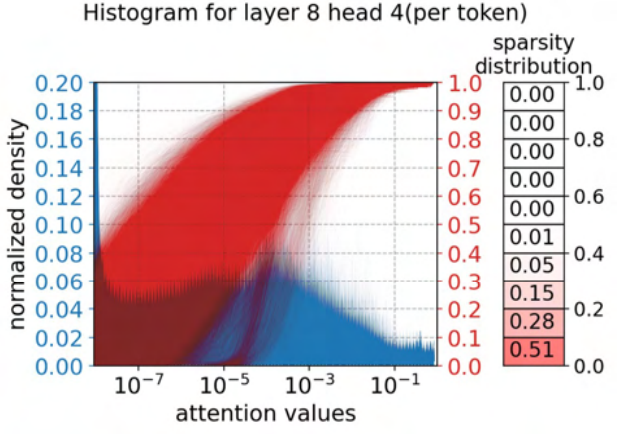

Histogram for layer 7 head 9(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.03
0.08
0.88
normalized density
attention values

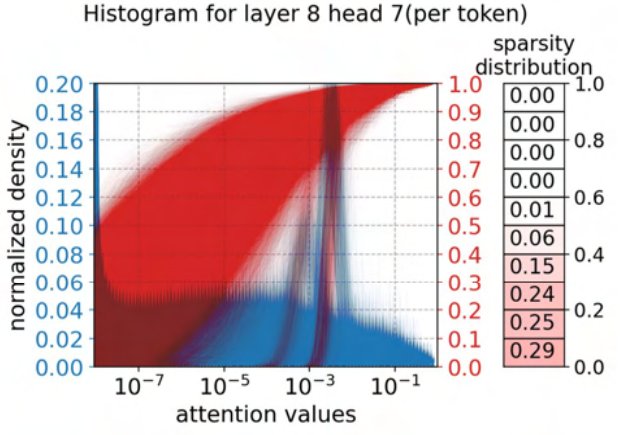

Histogram for layer 8 head 0(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.02
0.06
0.14
0.25
0.28
0.25
normalized density
attention values

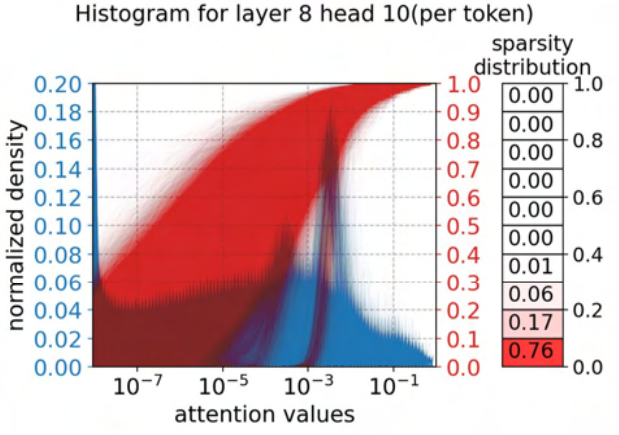

Histogram for layer 8 head 5(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.06
0.16
0.76
normalized density
attention values

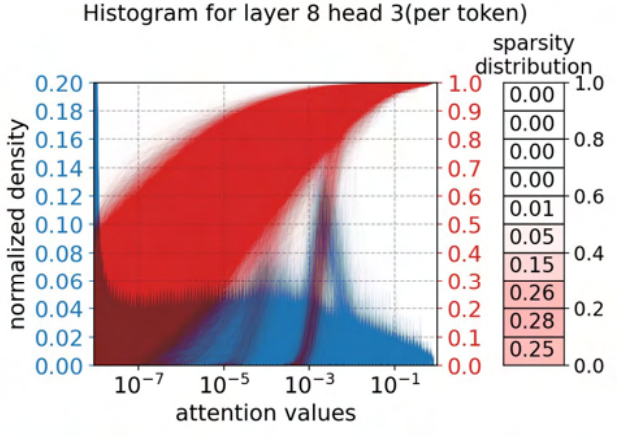

Histogram for layer 8 head 8(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.02
0.07
0.18
0.31
0.41
normalized density
attention values

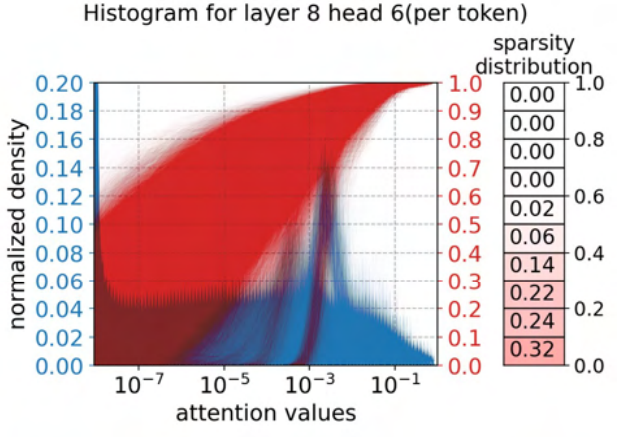

Histogram for layer 8 head 11(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.03
0.12
0.22
0.26
0.20
0.16
normalized density
attention values

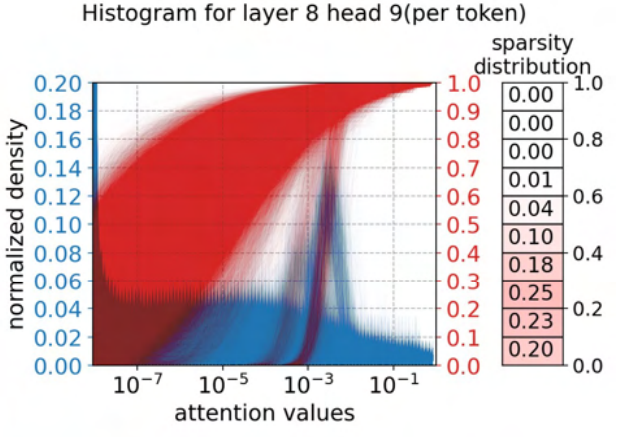

Histogram for layer 8 head 4(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.01
0.05
0.15
0.28
0.51
normalized density
attention values

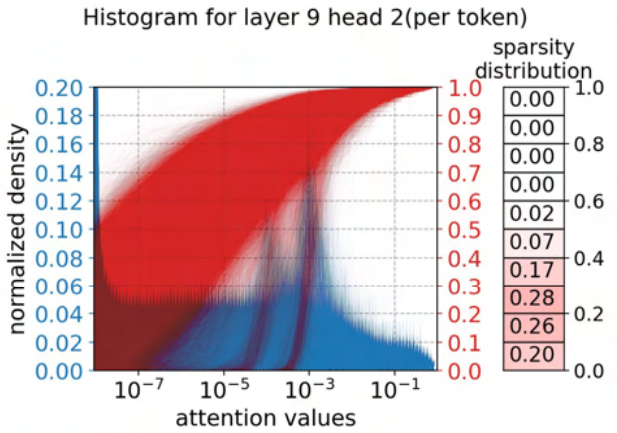

Histogram for layer 8 head 7(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.06
0.15
0.24
0.25
0.29
normalized density
attention values

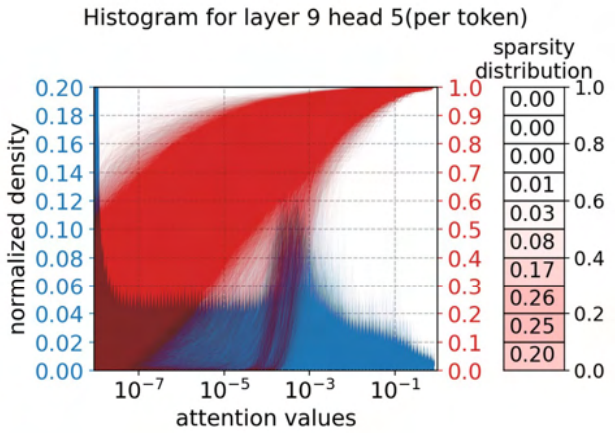

Histogram for layer 8 head 10(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.06
0.17
0.76
normalized density
attention values

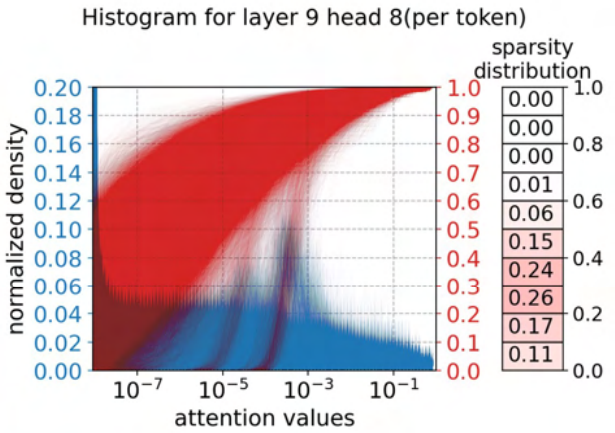

Histogram for layer 8 head 3(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.05
0.15
0.26
0.28
0.25
normalized density
attention values

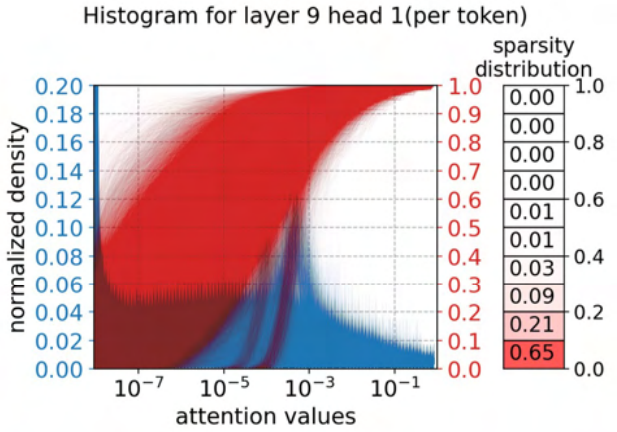

Histogram for layer 8 head 6(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.02
0.06
0.14
0.22
0.24
0.32
normalized density
attention values

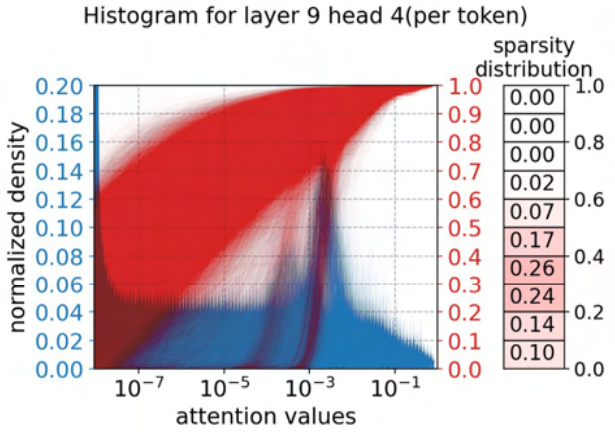

Histogram for layer 8 head 9(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.04
0.10
0.18
0.25
0.23
0.20
normalized density
attention values

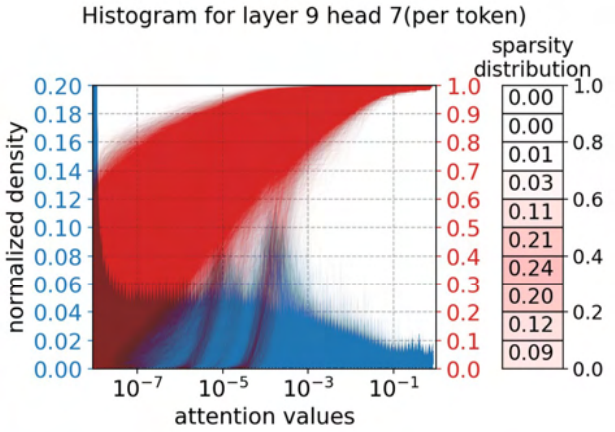

Histogram for layer 9 head 2(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.02
0.07
0.17
0.28
0.26
0.20
normalized density
attention values

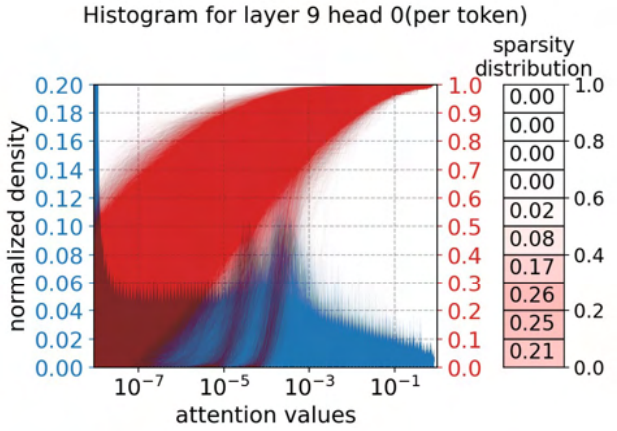

Histogram for layer 9 head 5(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.03
0.08
0.17
0.26
0.25
0.20
normalized density
attention values

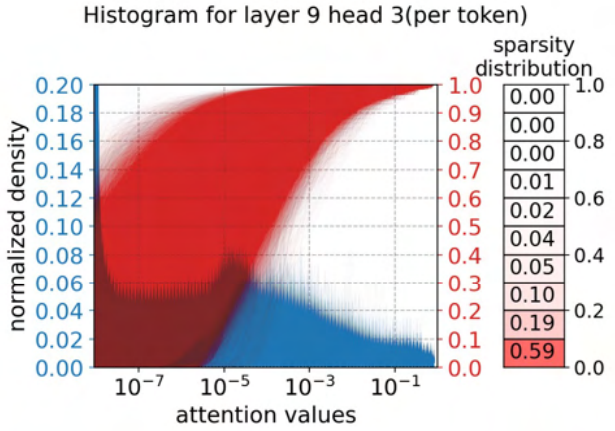

Histogram for layer 9 head 8(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.06
0.15
0.24
0.26
0.17
0.11
normalized density
attention values

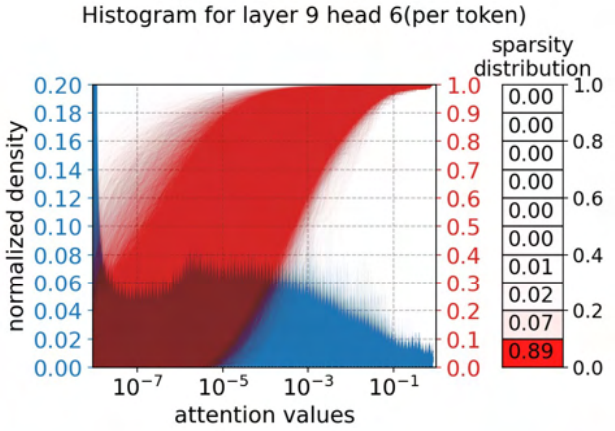

Histogram for layer 9 head 1(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.01
0.03
0.09
0.21
0.65
normalized density
attention values

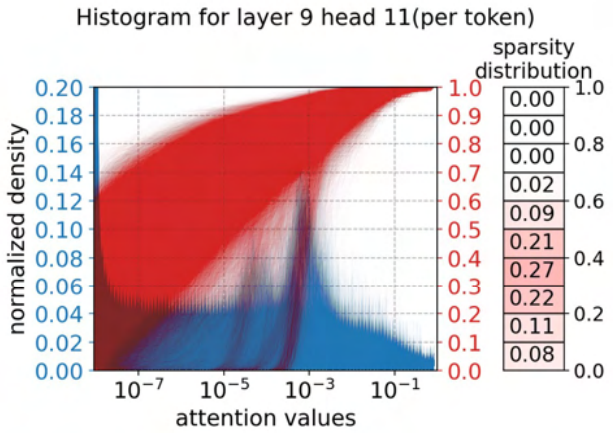

Histogram for layer 9 head 4(per token)
sparsity distribution
0.00
0.00
0.00
0.02
0.07
0.17
0.26
0.24
0.14
0.10
normalized density
attention values

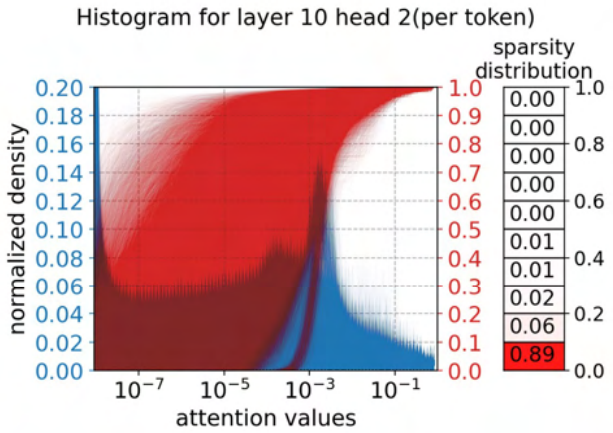

Histogram for layer 9 head 7(per token)
sparsity distribution
0.00
0.00
0.01
0.03
0.11
0.21
0.24
0.20
0.12
0.09
normalized density
attention values

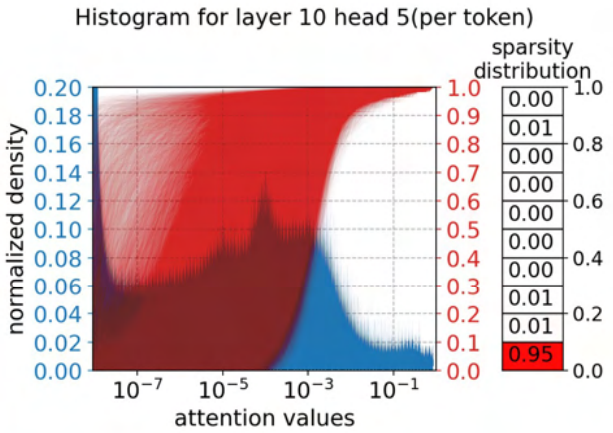

Histogram for layer 9 head 0(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.02
0.08
0.17
0.26
0.25
0.21
normalized density
attention values

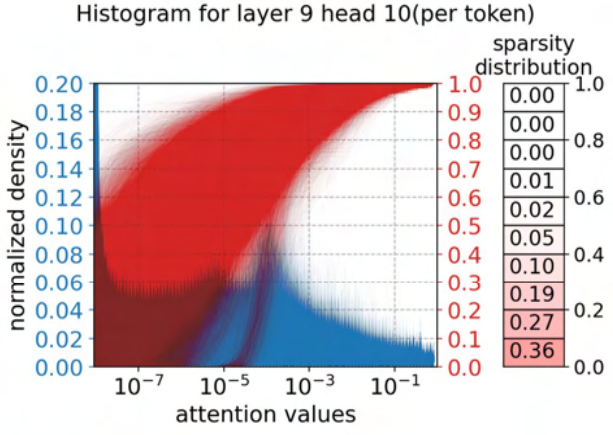

Histogram for layer 9 head 3(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.02
0.04
0.05
0.10
0.19
0.59
normalized density
attention values

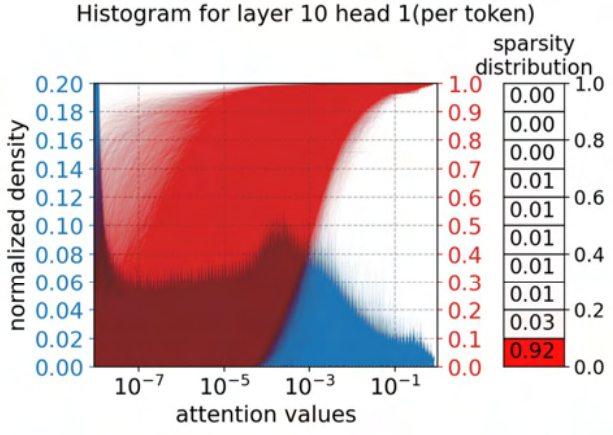

Histogram for layer 9 head 6(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.01
0.02
0.07
0.89
normalized density
attention values

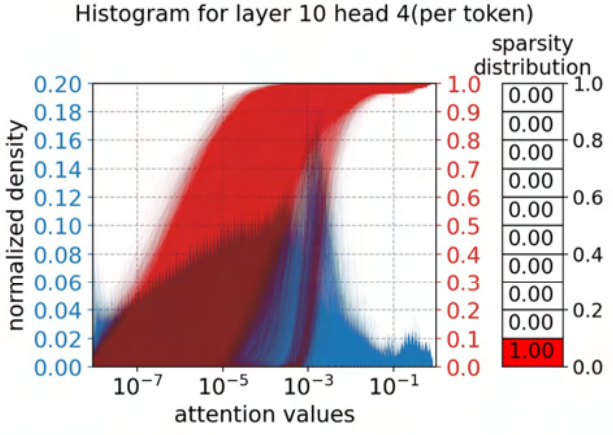

Histogram for layer 9 head 11(per token)
sparsity distribution
0.00
0.00
0.00
0.02
0.09
0.21
0.27
0.22
0.11
0.08
normalized density
attention values

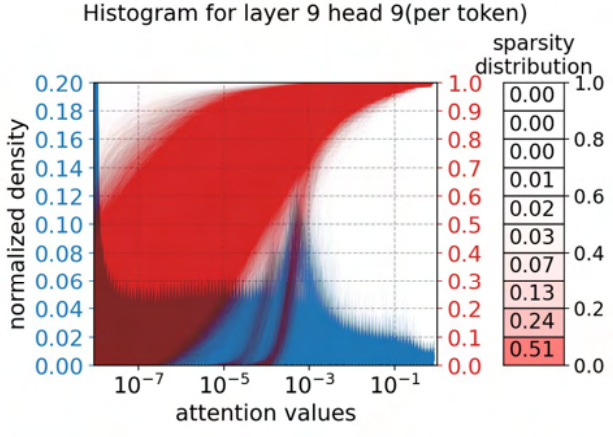

Histogram for layer 10 head 2(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.01
0.01
0.02
0.06
0.89
normalized density
attention values

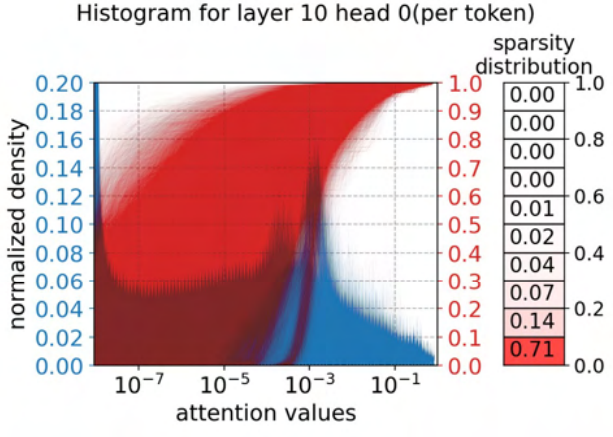

Histogram for layer 10 head 5(per token)
sparsity distribution
0.00
0.01
0.00
0.00
0.00
0.00
0.00
0.01
0.01
0.95
normalized density
attention values

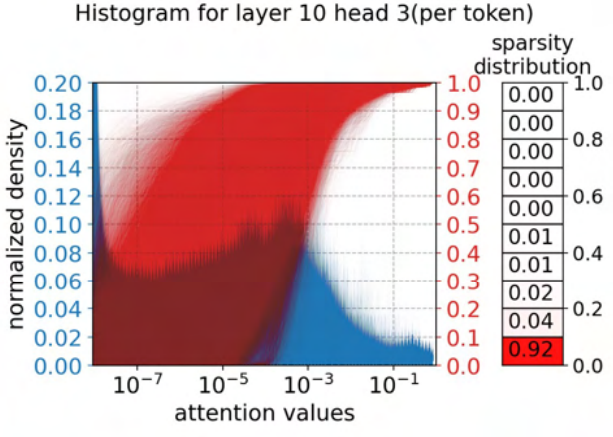

Histogram for layer 9 head 10(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.02
0.05
0.10
0.19
0.27
0.36
normalized density
attention values

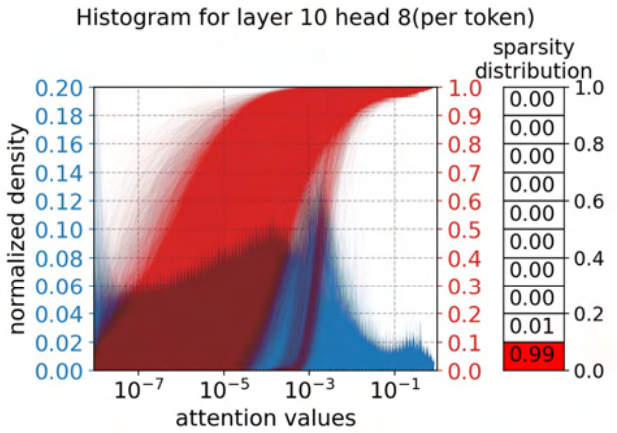

Histogram for layer 10 head 1(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.01
0.01
0.01
0.01
0.03
0.92
normalized density
attention values

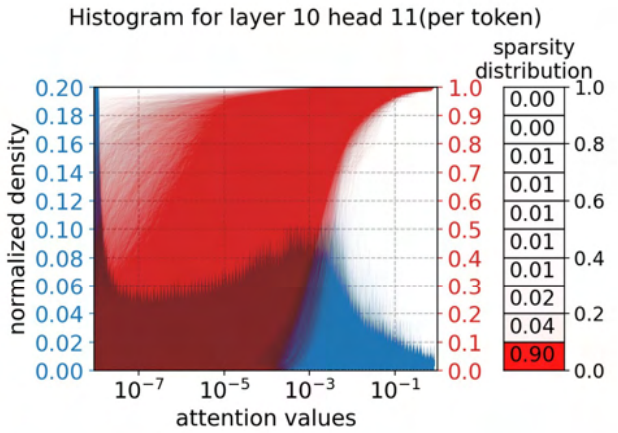

Histogram for layer 10 head 4(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
0.00
1.00
normalized density
attention values

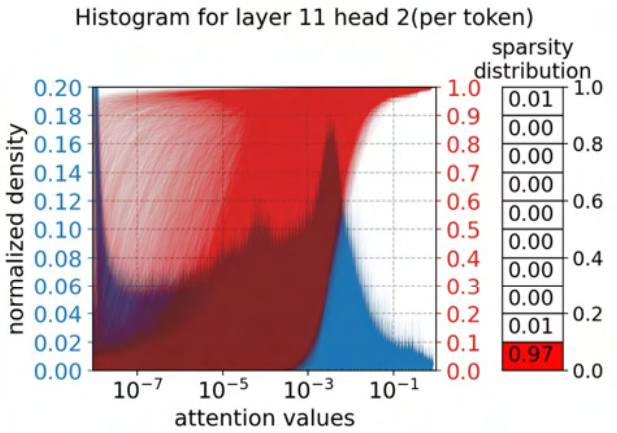

Histogram for layer 9 head 9(per token)
sparsity distribution
0.00
0.00
0.00
0.01
0.02
0.03
0.07
0.13
0.24
0.51
normalized density
attention values

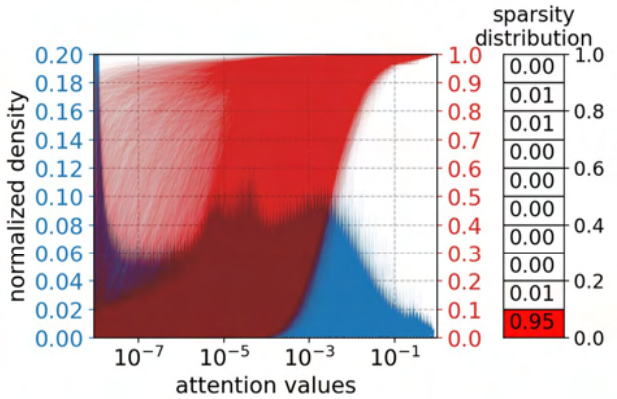

Histogram for layer 10 head 0(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.01
0.02
0.04
0.07
0.14
0.71
normalized density
attention values

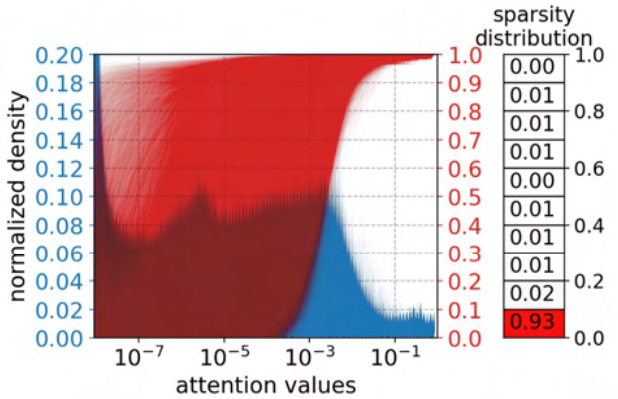

Histogram for layer 10 head 3(per token)
sparsity distribution
0.00
0.00
0.00
0.00
0.00
0.01
0.01
0.02
0.04
0.92
normalized density
attention values

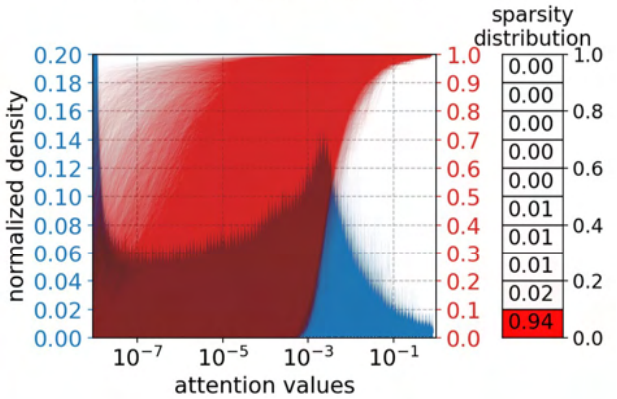

Histogram for layer 10 head 8(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.99
normalized density
attention values

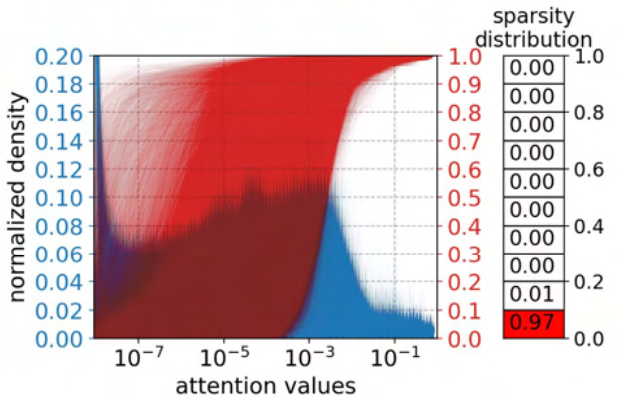

Histogram for layer 10 head 11(per token)
sparsity distribution
0.00 0.00 0.01 0.01 0.01 0.01 0.01 0.02 0.04 0.90
normalized density
attention values

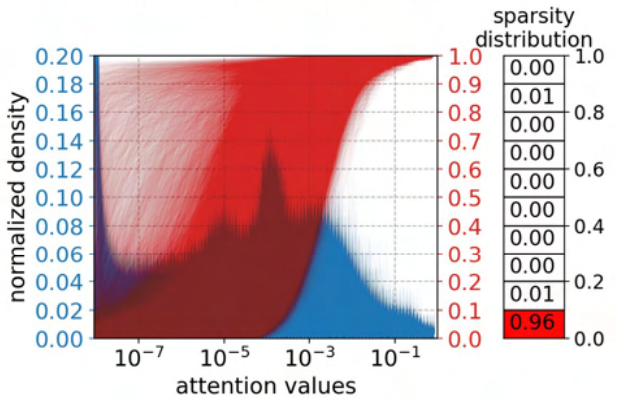

Histogram for layer 11 head 2(per token)
sparsity distribution
0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.97
normalized density
attention values

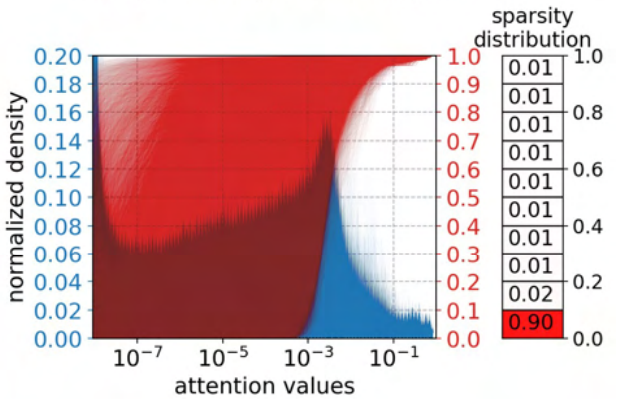

Histogram for layer 10 head 7(per token)
sparsity distribution
0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.01 0.95
normalized density
attention values

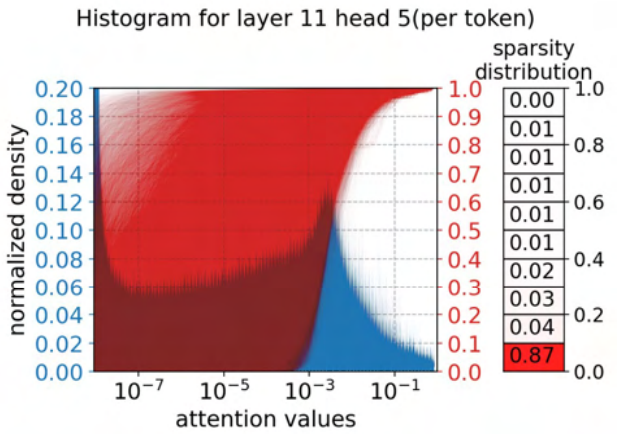

Histogram for layer 10 head 10(per token)
sparsity distribution
0.00 0.01 0.01 0.01 0.00 0.01 0.01 0.01 0.02 0.93
normalized density
attention values

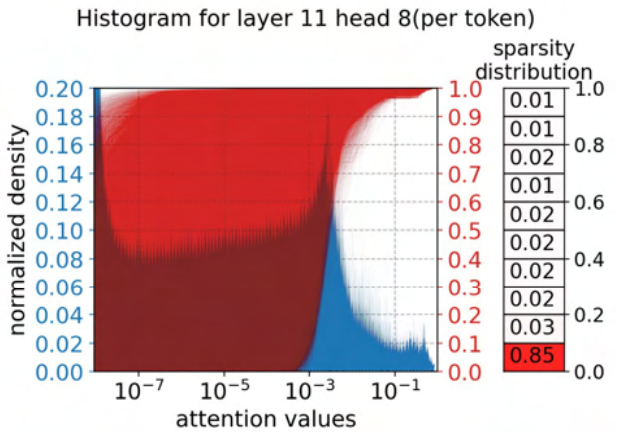

Histogram for layer 11 head 1(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.01 0.02 0.94
normalized density
attention values

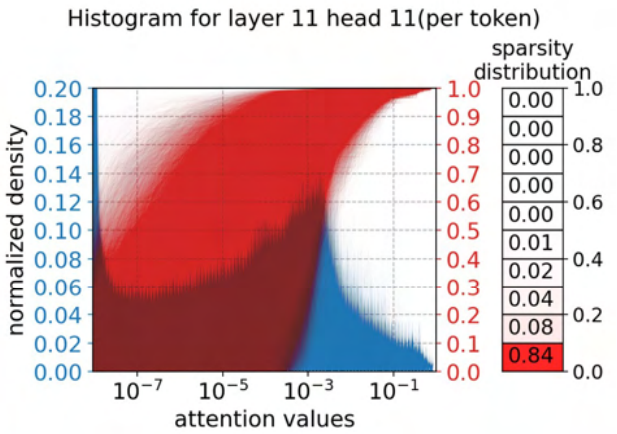

Histogram for layer 10 head 6(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.97
normalized density
attention values

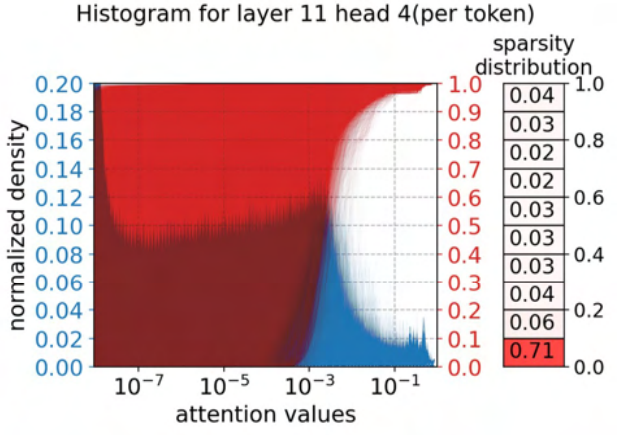

Histogram for layer 10 head 9(per token)
sparsity distribution
0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.96
normalized density
attention values

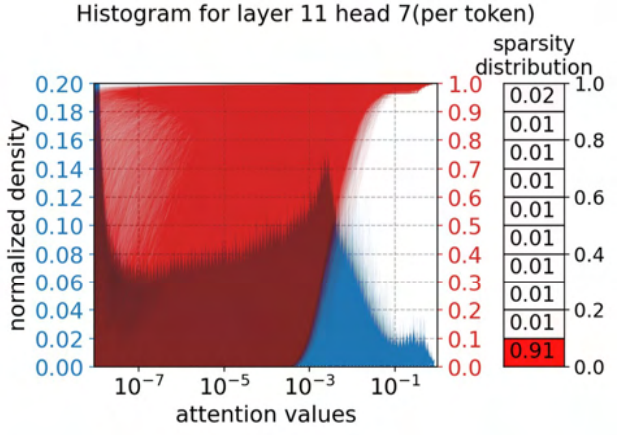

Histogram for layer 11 head 0(per token)
sparsity distribution
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.90
normalized density
attention values

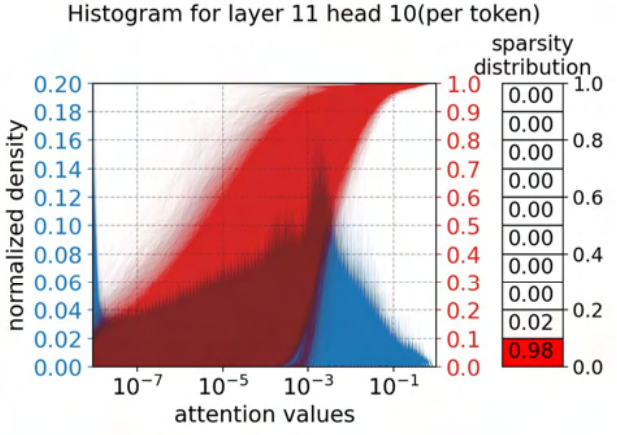

Histogram for layer 11 head 5(per token)
sparsity distribution
0.00 0.01 0.01 0.01 0.01 0.01 0.02 0.03 0.04 0.87
normalized density
attention values

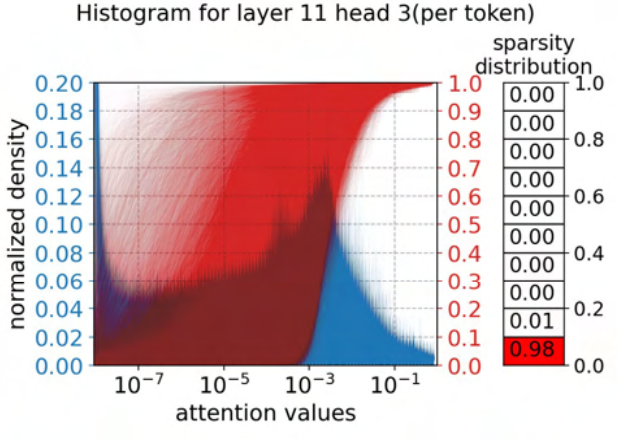

Histogram for layer 11 head 8(per token)
sparsity distribution
0.01 0.01 0.02 0.01 0.02 0.02 0.02 0.02 0.03 0.85
normalized density
attention values

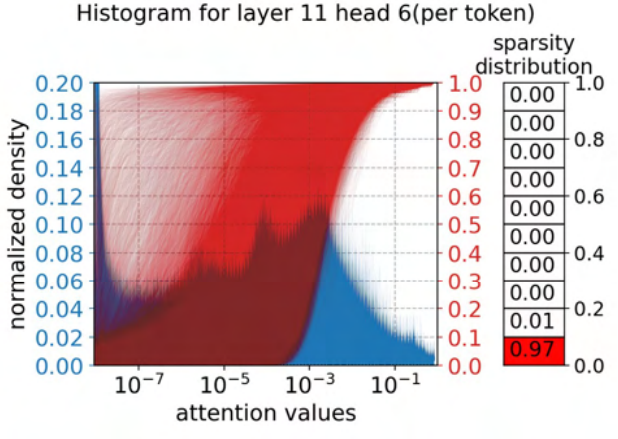

Histogram for layer 11 head 11(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.01 0.02 0.04 0.08 0.84
normalized density
attention values

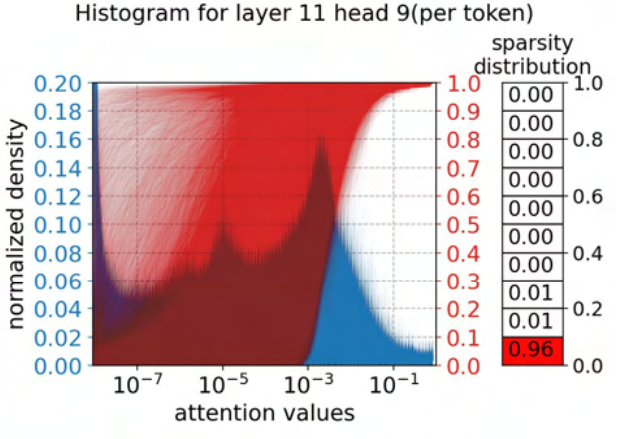

Histogram for layer 11 head 4(per token)
sparsity distribution
0.04 0.03 0.02 0.02 0.03 0.03 0.03 0.04 0.06 0.71
normalized density
attention values


Histogram for layer 11 head 7(per token)
sparsity distribution
0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.91
normalized density
attention values


Histogram for layer 11 head 10(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.98
normalized density
attention values


Histogram for layer 11 head 3(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.98
normalized density
attention values


Histogram for layer 11 head 6(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.97
normalized density
attention values


Histogram for layer 11 head 9(per token)
sparsity distribution
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.96
normalized density
attention values