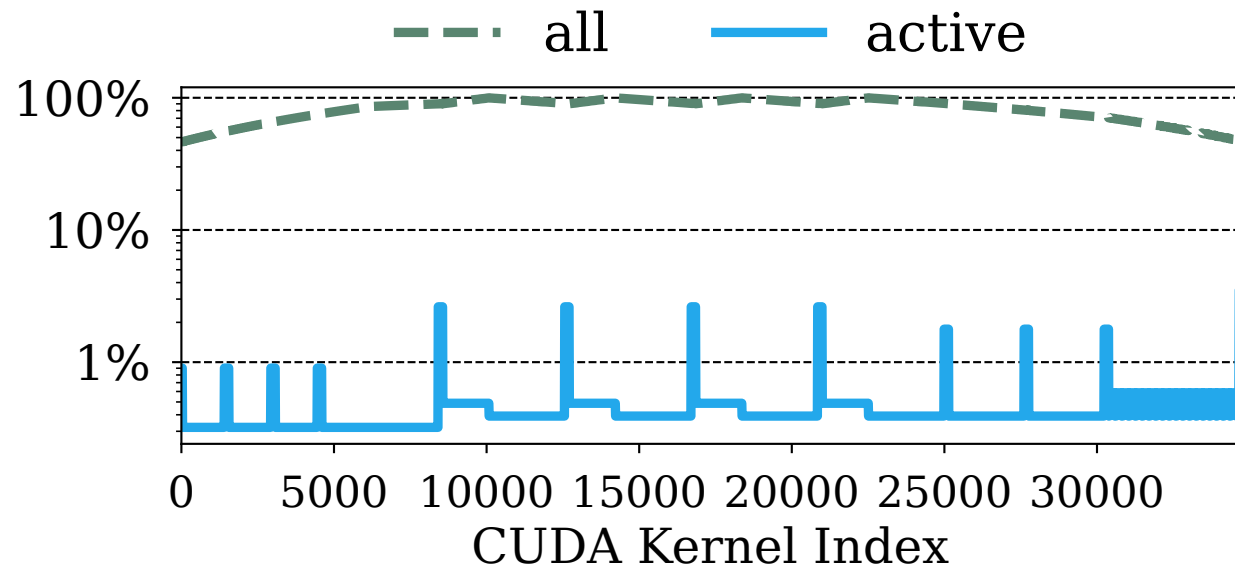
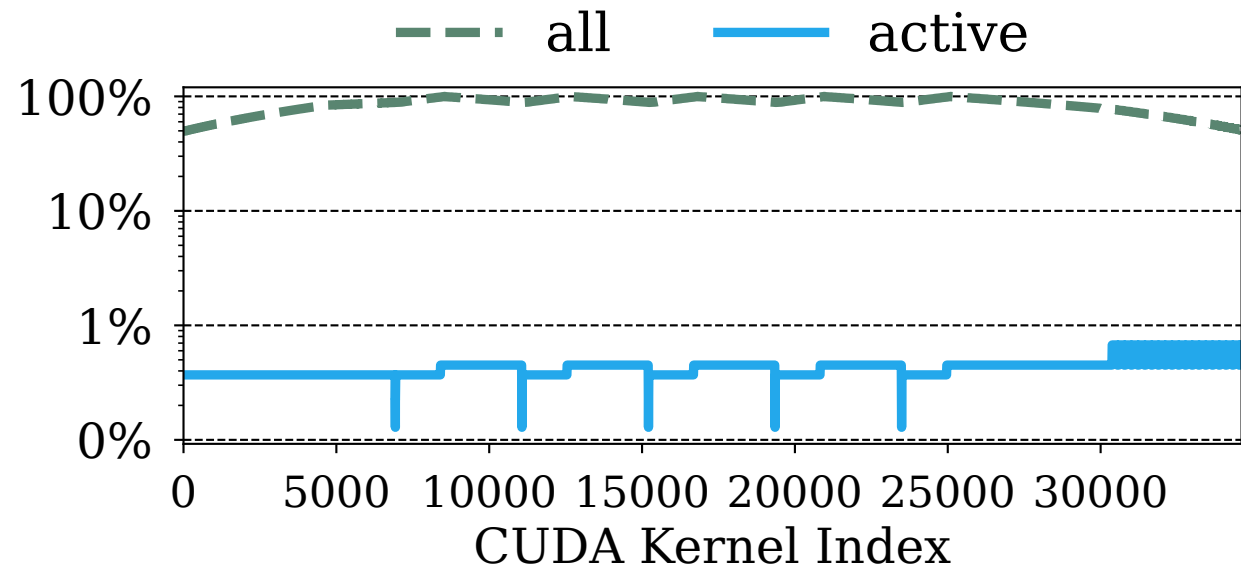


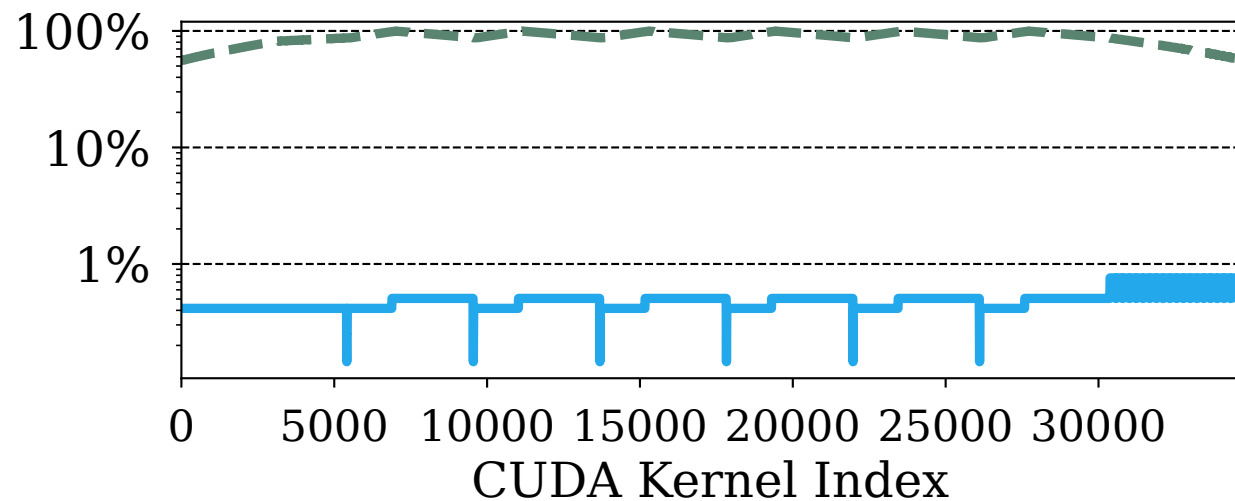
Memory Consumption



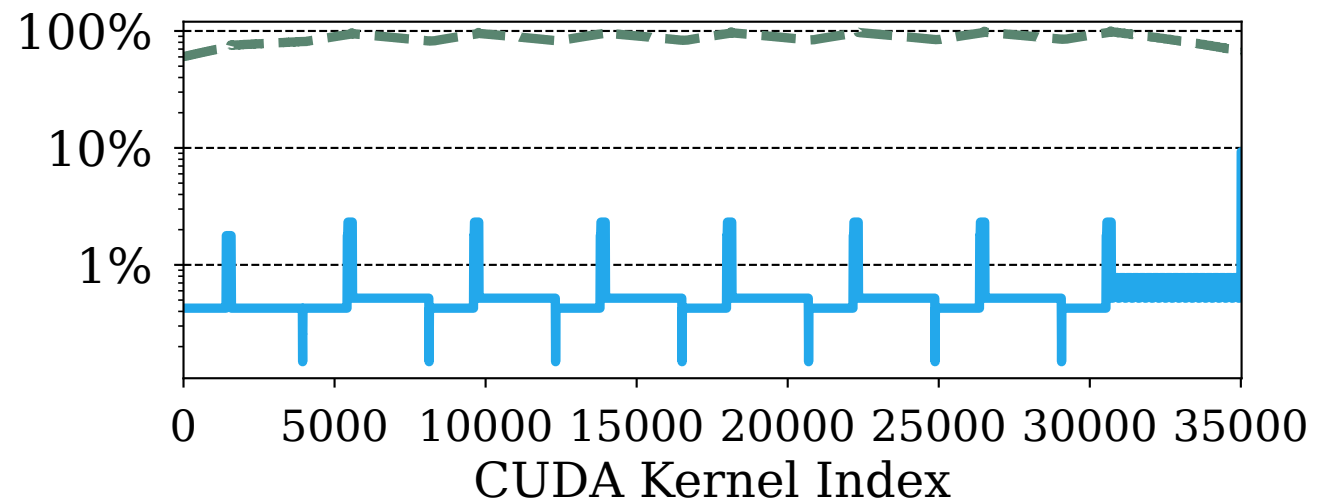
(a) llama-70B-GPU0(Stage-0)



(b) llama-70B-GPU2(Stage-1)



(c) llama-70B-GPU4(Stage-2)



(d) llama-70B-GPU6(Stage-3)