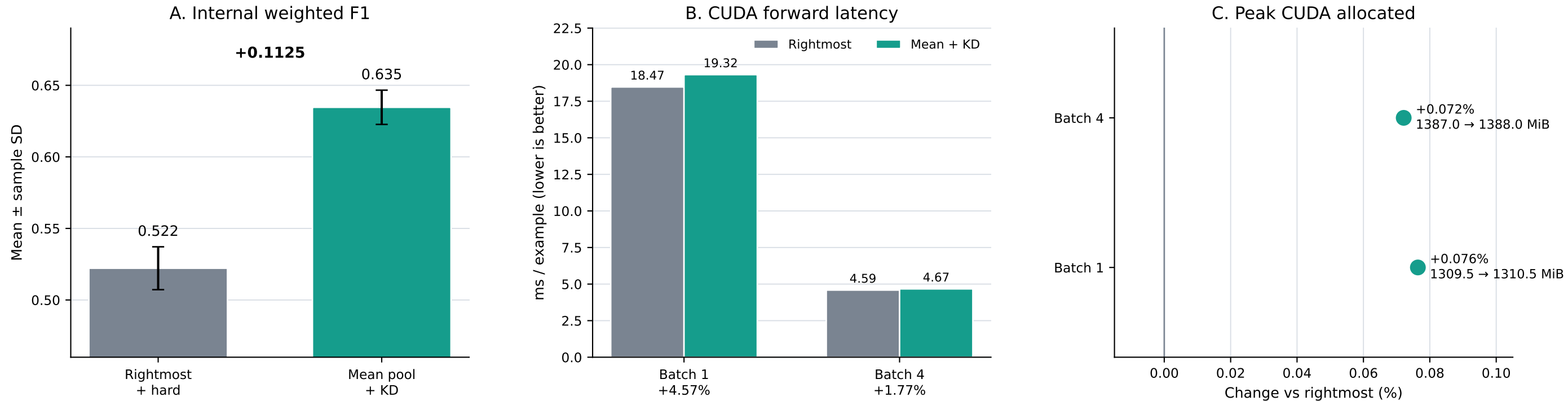


MALINT 4L: higher internal quality with a small measured forward cost



Quality and resource results come from separate frozen protocols. RTX 5070 Ti, BF16, length 256; timing is CUDA forward only. Mean+KD is not faster and GPU utilization is not claimed.