As shown in the figure above, the first three bars represent the performance of CPU via MPI and the last three bars represent the performance of hybrid CPU/GPU on 256\*256 world with different number of nodes. Basically, the average runtime of CPU is higher than the runtime of GPU. The main differences come from the time spent on computing data and writing a file. The reason that CPU via MPI spends more time on writing a file than hybrid CPU/GPU may be that ranks spend more time on computing, and computing takes too much CPU memory. Therefore, when a rank would like to write a file after it completes its computing, it has to spend more time because of the limited memory. The reason why CPU via MPI spends more time on computing than hybrid CPU/GPU is that, each rank of hybrid CPU/GPU uses a parallel way to handle the data on different blocks in GPUs, while each rank of CPU via MPI handles data in a sequential way. Despite that, since the number of ranks is fixed in this figure, the larger the number of nodes, the more the number of states, and the larger the world. However, the number of nodes for CPU via MPI doesn’t actually affect the performance since no matter how large the world is, the program always has more ranks to handle the data sequentially. But for hybrid CPU/GPU, the runtime on file writing increases when the number of nodes increases. The reason is that, the number of ranks that is used to write the file increases, so, the communication among ranks also increases which increases the time for a rank to find a place to write the content. So, when the number of nodes increases, the overall runtime for hybrid CPU/GPU also increases.