

Branch divergence in GPGPUs has attracted a lot of interest in the computer architecture research community. This paper presents a new approach to reduce the overhead of branch divergence in GPGPUs. The approach is based on the observation that the overhead of branch divergence is not only determined by the number of branches, but also by the number of threads that are active in the same warp. The proposed approach uses a new scheduling algorithm to reduce the number of active threads in the same warp. The algorithm is based on the observation that the number of active threads in the same warp is not only determined by the number of threads, but also by the number of threads that are active in the same warp. The proposed approach uses a new scheduling algorithm to reduce the number of active threads in the same warp. The algorithm is based on the observation that the number of active threads in the same warp is not only determined by the number of threads, but also by the number of threads that are active in the same warp.

kernel.cu Codesnippet from an implementation of Breadth first search algorithm in CUDA
visit_node
??
??
??
Human Nature