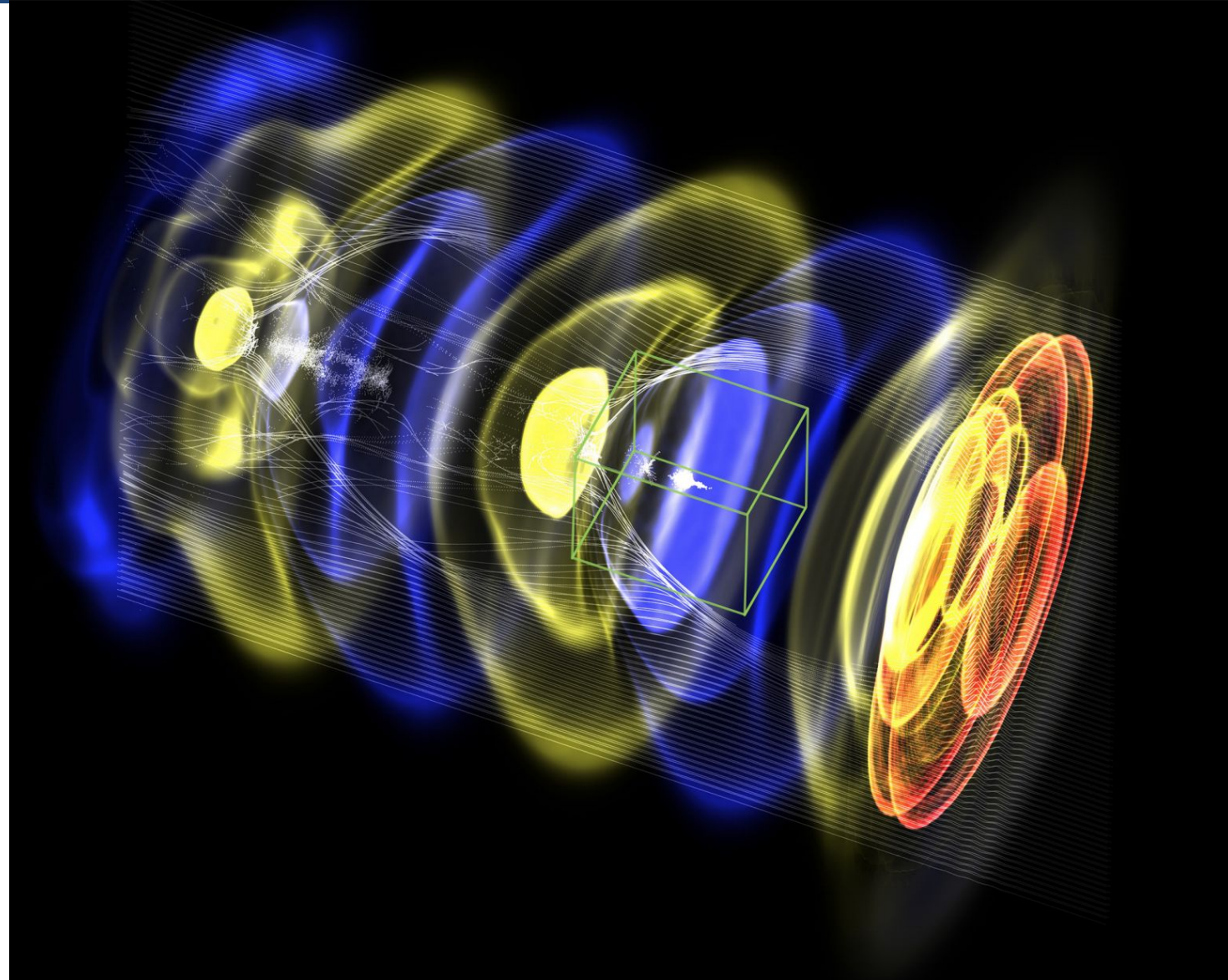


Real-time measurement of kernel execution time is needed for correct load balancing in GPU-accelerated codes

- On-the-fly measurement not possible with standard profiling tools (NVProf, Nsight)
- Developed a method (CUPTI Callback timing) for real-time measurement of kernel time
- Impact:
 - Provides accurate kernel timing **on-the-fly**
 - Enables correct **load balancing** in WarpX



Cannot measure execution time on GPU the same way as on CPU!

- What happens if we try to measure kernel execution time naively?

1. Start timer
2. Launch kernel
3. Stop timer