AWAKE simulations:

Codes, computing resources, physics problems

A. Petrenko (Feb. 2, 2018, CERN)

Numerical codes to simulate plasma-based accelerators:

The basic problem is to find the trajectories of all beam and plasma particles under the influence of electro-magnetic fields created by the same particles – i.e. a self-consistent solution to equations of particle motion and Maxwell's equations.

Simulation of beam-plasma interaction

Fluid codes

Beam and/or plasma is treated as a liquid (magnetohydrodynamics). Very fast codes, but work only for linear or weakly nonlinear waves. Examples: LCODE, VORPAL, fusion codes.

Analytic theory is often available in this approximation.

Particle-in-cell (PIC) codes

Beam and plasma is treated as a large number of individual macro-particles each representing a large number of real particles. These codes are computationally expensive but can simulate highly nonlinear problems (e.g. the blow-out regime).

Full PIC codes.

Beam and plasma particles are treated equally. Examples: OSIRIS, WARP.

Quasi-static PIC codes.

Beam evolution is assumed slow compared to plasma oscillations. Examples: LCODE, Quick-PIC.

Codes can work in 1D, 2D, or 3D geometry. Full PIC 3D codes capture the widest range of physical effects but computationally expensive.

Numerical codes we use regularly (and people who can use them)

Full PIC codes:

VLPL 2D-3D (**A. Pukhov**), WARP 2D-3D (**A. Gorn**)

2D AWAKE runs will require ~1000 cores for many hours (yet to be tested for full-scale AWAKE). 3D runs only feasible for few micro-bunches.

Quasi-static PIC codes:

2D: LCODE (**K. Lotov** +) typical high-resolution full-scale AWAKE shot takes around 500-1000 cpu*hours, e.g. (100 cpu)*(1 night). 3D LCODE is under development (will require much more cpu though).

3D: QV3D **(A. Pukhov, A. Petrenko)** runs on a single core with ~100 GB of RAM. Full-scale AWAKE simulation (with 4x lower than LCODE resolution) takes 1-2 weeks (might become parallel soon => much faster, less memory per node, higher resolution can be achieved).

Learning a code takes time (~months at least) and we often need to modify the code in order to add the special conditions of the experiment. Close cooperation with code developers is very important! AWAKE is not a standard problem, few codes were tested in this regime. WARP 2D was used only to test the validity of quasi-static approximation so far.

Physics problems vs codes:

- **Proton beam defocusing or CTR**: full-scale AWAKE simulation. Routinely done with LCODE only (several shots can be simulated overnight), 3D only with QV3D (weeks).

- **Micro-bunch structure (streak-camera)**: LCODE, QV3D, probably WARP (calculation time depends on the location of the streak-window of course).

- **Electron injection**: LCODE can simulate e-injection without e-beam charge effects (because the wakefield is cylindrically-symmetric). As long as the injected e-beam density is low, the approximation is valid -- to be tested in 3D (QV3D, WARP maybe).

- A. Pukhov found that 3D and 2D wakefield might be significantly different after wave-breaking. This should be studied in detail via 3D vs 2D simulations (in progress already). This might limit the range of possible e-injection positions in xi.

- **Pre-pulse effect on the SMI**: only available in QV3D so far. Should be checked with respect to the e-beam injection (might be especially sensitive in the beginning).

- **Run 2 studies**: again, 2D wake is not yet fully reproduced in 3D (to be studied). **e-beam injection and acceleration** should be simulated in high-res 3D (QV3D).

Computing resources:

Clusters in Novosibirsk:
Typical runs: 100 cores overnight with LCODE. Or single-core QV3D over few days/weeks.

HTCondor at CERN: only tested in a single core mode with QV3D (64 GB). Have recently installed 1 TB RAM nodes – might allow to increase the QV3D resolution by a factor of 2-3 – interesting for 2D vs 3D and benchmarking of codes vs experiment.
MPI might be used with up to 10 cores, i.e. overnight LCODE simulation can be calculated within a week – interesting for large scale simulation vs experiment studies simulating many AWAKE shots (MPI is yet to be tested on this HTCondor).

Cloud computing option:

Multiple cloud-computing services (Amazon EC2, Google Cloud, MS Azure, etc.):

For example https://aws.amazon.com/ec2/pricing/on-demand/:

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage	
General Purpose - Current Generation						
t2.medium	2	Variable	4	EBS Only	$0.0464 per Hour	
m4.16xlarge	64	188	256	EBS Only	$3.2 per Hour	10h LCODE run: **$30**
Memory Optimized - Current Generation						
x1.32xlarge	128	349	1952	2 x 1920 SSD	$13.338 per Hour	10h QV3D: **$130** (probably will make sense only for parallel QV3D)

Spot instances are cheaper. Also MPI is available for low-cost vCPUs. This can bring the cost of quick (~hour or less) full-scale AWAKE LCODE simulation to below $10.

Possible future plans, needs & options (What do we want?)

Code vs experiment benchmarking (parameter scans) will require 100s (if not 1000s) of full-scale AWAKE runs, which is good to do at CERN. HTCondor is an interesting option for week-long parameters scans with LCODE (+maybe a limited res. QV3D), but it needs more work on MPI configuration (1-2 weeks probably).

Possibility of a quick (less than an hour) full-scale AWAKE simulation in LCODE: need large MPI-cluster with around a 1000 of cores. Cloud-computing service might be an interesting option especially if the number of such quick runs is not large --- the cost of a single run will be the same ($10 or less probably).

WARP and other full-PIC codes need similar or even bigger ~1000-10000 core clusters to produce the results with a reasonable resolution and within an acceptable time scale. WARP should be first configured for AWAKE (optimize resolution, Lorentz boost, with/without electrons). We don't really know the computing requirements for WARP yet (maybe similar to OSIRIS?).

QV3D is currently the only 3D code we have which can simulate the full-scale AWAKE. Now we are testing electron injection and acceleration in this code (this required some modifications). Parallelization would improve the performance a lot, detailed high-resolution benchmarking and parameter scans could become possible (on a suitable cluster). LCODE 3D might become available in the future as well.

Probably all full PIC/high-res quasi-static 3D/quick 2D are not feasible within the CERN infrastructure (not on HTCondor, which is the main CERN HPC cluster). Novosibirsk clusters are already full with 2D LCODE + QV3D, massive full-PIC and 3D quasi-static scans are not possible there as well.

LCODE (and limited res. QV3D) parameter scans are probably possible on HTCondor, but full-PIC and high-res QV3D will require an access to ~1000-10000 node MPI cluster or a cloud computing service.