Exploration of Parallel Optimization Techniques for Accelerator Design

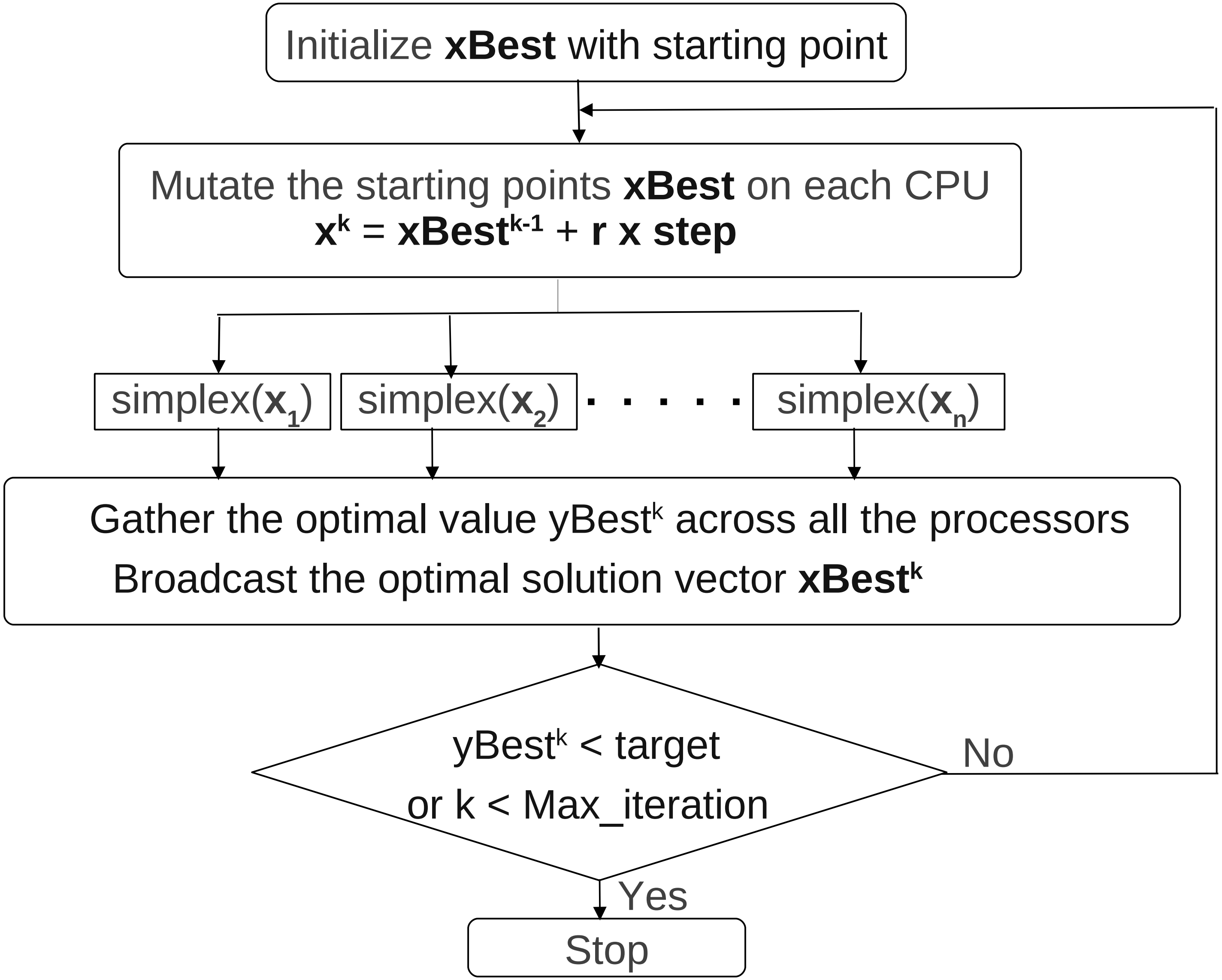
3/29/2011

Yusong Wang   Michael Borland   Vadim V. Sajaev

Motivation

- Optimization through simulations is very time consuming for real-world accelerator designs
  - High dimensions
  - Long simulation time
  - Noisy searching space
- Traditional method, such as simplex, could be easily trapped to a local optimum
- Optimization time is limited by on-demand accelerator adjustment requirements
- Computer performance is improved by adding more cores instead of increasing processor speed at present
- Several widely-used global optimization methods, such as genetic algorithm, particle swarm optimization, are embarrassingly parallel
  - No communication needed during a simulation
  - Very little information shared between processors

Parallel Hybrid Simplex Algorithm



Particle Swarm Optimization

- First introduced by Kennedy and Eberhart in 1995
- Inspired by social behavior of birds, fish
- The penalty function needs not to be differentiable
  - Suitable for optimization through simulations
- Efficient global optimization algorithm

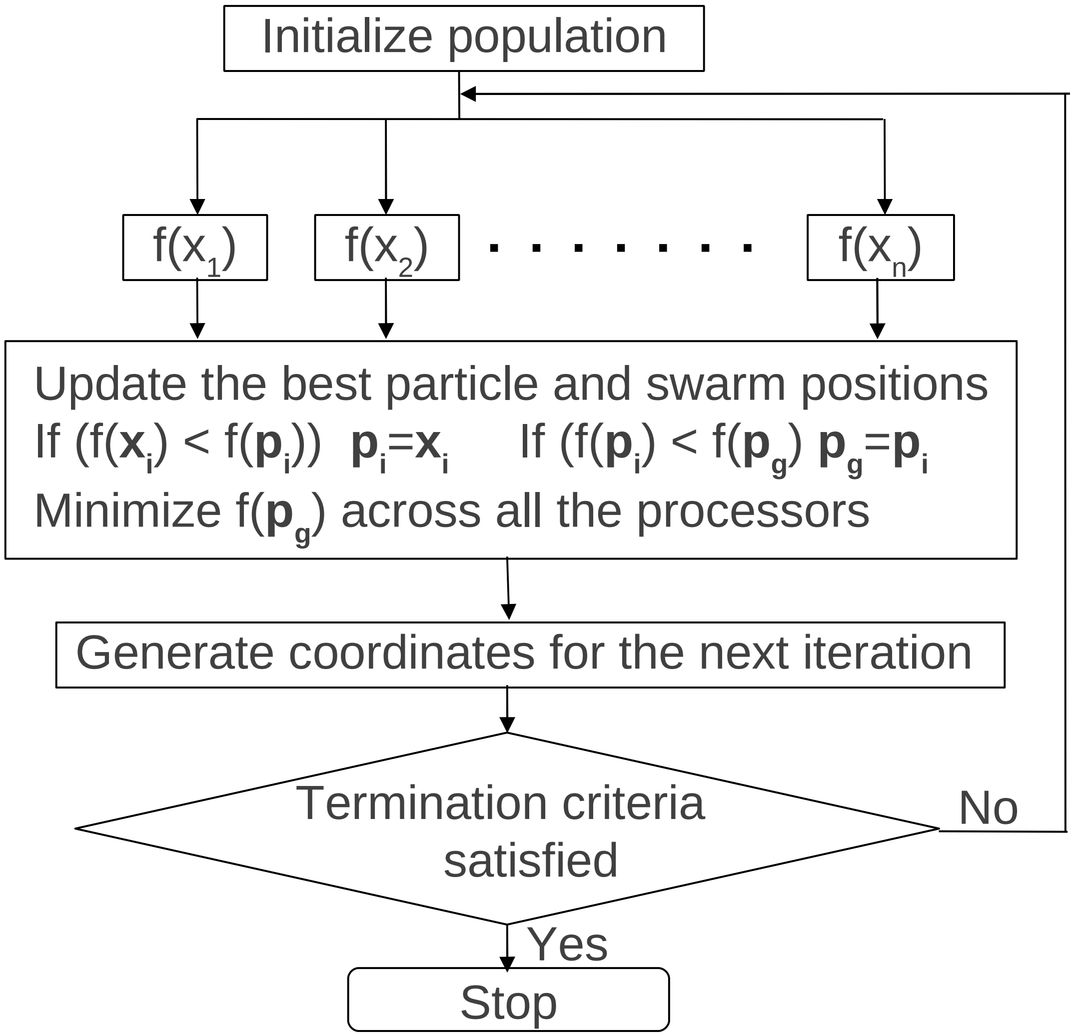
$$\mathbf{x}^{k+1}_i = \mathbf{x}^k_i + \mathbf{v}^{k+1}_i$$

Update individual particle positions

$$\mathbf{v}^{k+1}_i = \overbrace{w^k \mathbf{v}^k_i}^{\text{inertia}} + \overbrace{c_1 r_1 (\mathbf{p}_i - \mathbf{x}_i)}^{\text{personal influence}} + \overbrace{c_2 r_2 (\mathbf{p}_g - \mathbf{x}_i)}^{\text{social influence}}$$

		w	Inertia weight
$\mathbf{x}^k_i$	Particle position	k	Iteration number
$\mathbf{v}^k_i$	Particle velocity	i	Particle ID
$\mathbf{p}_i$	Best particle position	c_1, c_2	Cognitive and social parameters
$\mathbf{p}_g$	Best swarm position	r_1, r_2	Random numbers

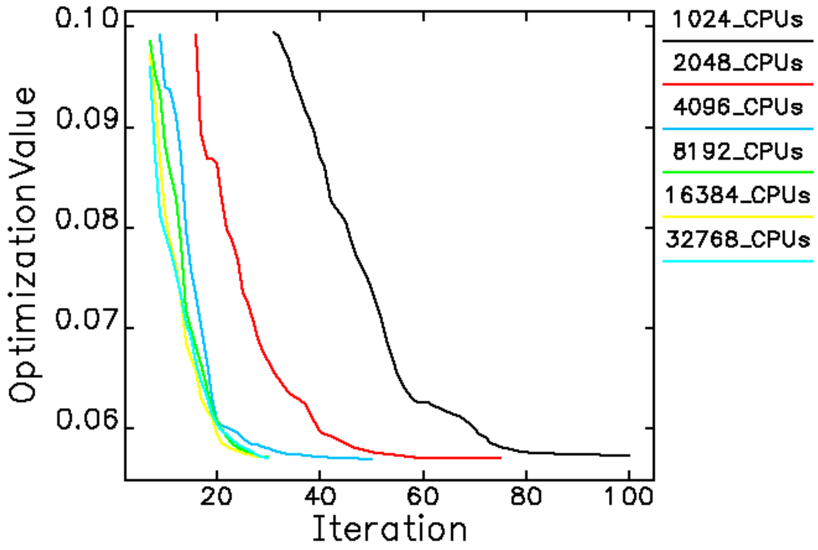
Parallel Particle Swarm Optimization



Coupling Minimization at APS

- Using a response matrix method and singular value decomposition performs poorly as a result of the small number of skew quadrupoles
- Minimize the sums of the squares of the vertical beam size at the source point by adjusting the strength of the 19 skew quadrupoles
- The optimization needs to be done within half an hour for online tuning
- It took more than 6 hours for the serial simplex method to reach an optimum of 0.057 on a computer with a 2.4 GHz CPU
- The hybrid simplex method requires several hundreds of function evaluations per iteration, which will take more than one hour
- As little as one function evaluation per processor is needed for the parallel particle swarm optimization

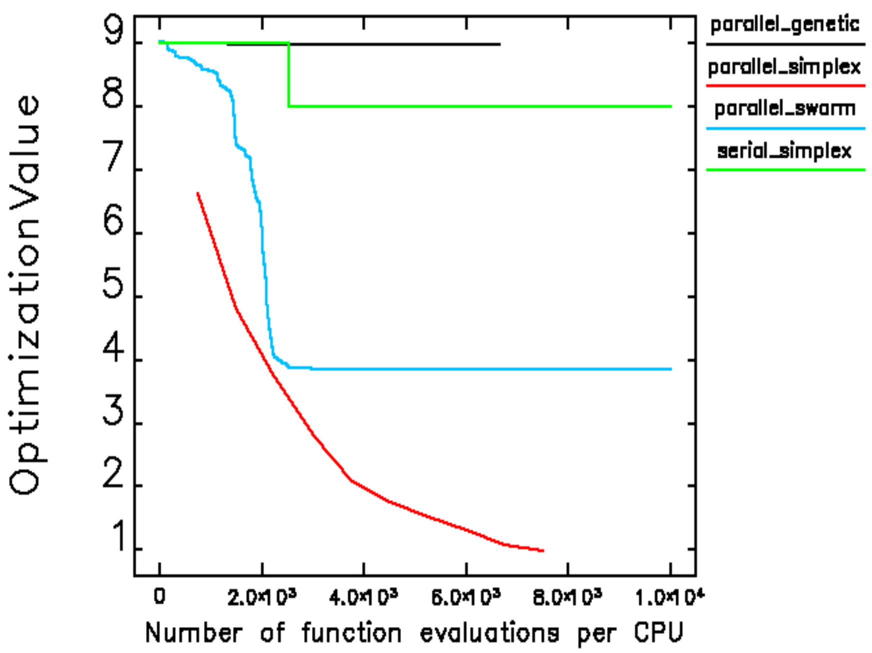
Coupling Minimization at APS (continued)



Parallel particle swarm optimization on the Intrepid supercomputer at Argonne Leadership Computing Facility (ALCF)

- The number of CPUs is the same as the number of individuals
  - One function evaluation per CPU
- The algorithm converges faster with a larger number of CPUs/individuals
- It took 28 minutes on 4k CPUs to finish 50 iterations to reach the target
- The optimization time can be reduced to 20 minutes with 30 iterations on 16k CPUs

Twiss Parameter Optimization at APS



Twiss parameter optimization results with different algorithms on the Intrepid supercomputer. The target value is 1.

- 38 independent quadrupoles
- Can not reach the target after 6 hours with the serial simplex (>260k evaluations)
- Using 1k compute nodes (4k CPU cores)
  - One individual per CPU core
- The parallel hybrid simplex converges to the target within 1.5 hours (7.5k function evaluations)
- The solution of the problem is very close to the given starting point
  - Very fine tuning in a local area

Conclusion and Future Improvement

- The parallel optimization algorithms will reduce the optimization time and achieve a better optimization result on parallel computers.

- Parallel optimization methods will allow taking advantage of hardware trend toward massively multi-core computers.

- A single algorithm can not fit different problem requirements
  - Particle swarm optimization and genetic optimization have their advantages in global optimization
  - Hybrid simplex is efficient for fine tuning in a local neighborhood

- The genetic optimization can be improved by applying adaptive step-size control for each of the dimensions

- Parallel function evaluations in the simplex algorithm can be used to reduce the optimization time

References

- M. Borland, Advanced Photon Source Light Source note LS-287, Sept. 2000.
- Y. Wang et al., Proc. ICAP09, p. 355 (2009).
- D. Levine, ANL Report ANL-95/18 (1996).
- J. Kennedy and R. C. Eberhart, Proc. of IEEE International Conference on Neural Networks, p. 1942 (1995).
- Y. Shi and R. C. Eberhart, Proc. of IEEE International Conference on Evolutionary Computation, p. 69 (1998).
- C. Wang et al., Proc. IPAC10, p. 4605 (2010).
- H. Shang et al., Proc. PAC05, p. 4230 (2005).