

**IMPUTE2** uses a Markov chain Monte Carlo (MCMC) algorithm that alternates between phasing typed SNPs and imputing untyped SNPs. Each MCMC iteration includes two steps:

1. Sample a new phase configuration for each study individual, using information from other study individuals and reference panel haplotypes at SNPs typed in the study. (This step is excluded with “-use\_prephased\_g” option (?))
2. Given the newly sampled haplotypes for all study individuals, treat each haplotype as independent (conditional on the reference panel) and analytically impute the alleles at untyped SNPs. (with HMM algorithm)

This MCMC algorithm is run for a number of iterations (typically 30, including 10 burn-in iterations), then the probabilities from Step 2 are averaged across iterations to produce marginal posterior genotype probabilities at each untyped SNP.

The phasing and imputation calculations are driven by the hidden Markov model (HMM) of Li and Stephens (2003). This model can be used to update an individual's haplotypes by constructing them as “imperfect mosaic” copies of a set of template haplotypes. In the simplest case, the templates would include all reference+study haplotypes (minus the pair being updated) in Step 1 and all reference haplotypes in Step 2.