This manuscript proposes a method to compute the optimal schedule of the update magnitude for adaptive flat-distribution-sampling (FDS) free energy simulation methods, including the Wang-Landau (WL) algorithm and metadynamics. This method characterizes each adaptive FDS simulation by its updating scheme encoded in an updating matrix, and then computes the optimal schedule by solving a numerical variational problem, which uses the eigenvalues and eigenvectors of the updating as input along with other quantities to the variational equations to be minimized. The authors show that the optimal schedule for the single-bin (WL) scheme is given by the well known inverse-time formula. Making an interesting analogy, they point out that for a general multiple-bin scheme, the optical schedule is implicitly given by an equation of motion of a free particle with a position-dependent mass. They suggest a "bandpass" updating scheme which they show has better convergence than the metadynamics Gaussian updating scheme.

There are several very interesting ideas introduced in this paper, and I recommend publication following revisions to address my questions and concerns.

The manuscript is highly mathematical, and it is difficult to follow all the details without devoting a very large amount of time to studying it. The authors can do a better job of explaining their mathematical analysis of FDS updating schemes, by reorganizing their manuscript, moving some sections of the manuscript to the appendix and expanding the explanations, while moving one section which is currently in an appendix to the main text. The procedure for computing the optimal schedule and error in Appendix D should be moved to the main text and the steps involved explained in more detail in a single section. An explanation of how eq. 55 is solved for q(T) is needed, this can be placed in an appendix, and an explanation of how eq. 49 is numerically integrated is also needed. A clearer discussion of what autocorrelation functions are being calculated, and how they are used in the variational optimization is also needed.

My biggest concern with the paper relates to the numerical example. It seems that the single test system studied in detail is a completely flat "potential". While it is interesting to note that different FDS schemes will exhibit different efficiencies on this "trivial" example, it doesn't give the reader any feeling of the corresponding results on the simplest "non-trivial" 1D or 2D problem with multiple wells and barriers, which at least begins to approach problems that motivate the need for advanced sampling schemes. The authors should report the results of at least one example of this kind in addition to the test system reported. While I am not requiring the authors to analyze how their (bandpass) FDS scheme and analysis methods perform on more realistic potentials than 1D or 2D potential with multiple barriers, I would point out that a key issue is whether metadynamics or any FDS scheme can really perform well on problems where there is a free energy balance between stable basins separated by free energy barriers, and the balance involves a competition between the enthalpic stabilization of one basin and the entropic stabilization of the other. The authors may want to comment on whether they think FDS sampling schemes are well suited for these kinds of landscape problems.

Some additional questions:

1. During the derivations of Eq.(22) and Eq.(49), the end point value $q(T)$ is fixed. Does this mean that the optimal schedule of the update magnitude suggested by Eq.(22) or (49)

is only for a subgroup simulations which have the same $q(T)$? Please clarify this issue or add the corresponding discussion to the manuscript.

2. How to determine the time of the preliminary adaptive FDS simulations with a constant updating magnitude?

3. The study of the updating matrix borrows techniques used in studying transition matrices. Does this require all the elements in the updating matrix w_{ij} to be nonnegative? If it is true, please mention this requirement after Eq.(28).

4. For Eq.(5): perhaps Change p_i to $p_{i(t)}$.

5. Eq. 11 expresses the total error in the cumulative bias potential as a weighted average over the errors in each bin, weighted by the density rho, which is the equilibrium eigenvector of the updating matrix w. Why is the equilibrium eigenvector of the updating matrix, the correct way to weight the errors in the cumulative bias in the bins? Can the authors provide a physical explanation of why this is the case, as well as a mathematical justification.