**We thank the reviewers for their comments and analysis.   
  
R1.  
1> Intro is short/broader coverage of the related work required:  
  
Symbolic pre-computation of Gibbs is completely novel and we have not found any related work for this part.  
We have already covered a brief literature related to piecewise distributions (In section 7 that can be moved to intro).   
In light of the reviewers comments we are also happy to expand the literature review with the existing works that attempt to address the observed determinism as follows:   
  
Observed determinism, can appear in applications such as failure detection (Huseby et al. 2004) where random variables are not directly observable but a function of them may be observed (indirect measurement).   
The following families of densities are closed under linear transformations and consequently allow linear determinism: piecewise polynomials with hyper-rhombus/linear partitioning constraints and conditional linear Gaussians (CLGs) (Lauritzen & Jensen 2001).   
In (Li et al. 2013) the restriction is imposed on the network topology rather than the densities. They can only handle the observed summation of variables on a very simple network. The generalization of this technique does not seem straight-forward (if possible).  
  
Handling nonlinear determinism is much trickier. Methods based on Hamiltonian MC are used for sampling under particular constraints (Hartmann 2008) but they cannot be used in piecewise models.   
Cobb & Shenoy (2005) approximate non-linear determinism with piecewise linear constraints via dividing the space into hypercubes. This method can hardly be used in dimensions more than 2 or 3 since the number of partitions required to preserve approximation accuracy is exponential in dimensionality.  
The solution often suggested by probabilistic programming languages is to approximate the observed determinism via adding noise to the observation (hard to soft constraint conversion via measurement error). But in practice if the added noise is large, the approximation bounds may become arbitrarily large and if the noise is small, the mixing rate may become extremely slow (near-deterministic problem) (Chin & Cooper, 1987).   
In short, inference conditioned on nonlinear determinism is still an open problem (Li et al., 2013) and even for linear constraints, the existing works are too restrictive.  
  
NEW REFERENCES:  
\* Chin & Cooper. Bayesian belief network inference using simulation. UAI 87  
\* Cobb & Shenoy. Nonlinear deterministic relationships in BNs. Springer 2005  
\* Hartmann "An ergodic sampling scheme for constrained Hamiltonian systems with applications to molecular dynamics." J. Stat. Physics 2008   
\* Huseby et al. System reliability evaluation using conditional MC. Stat. Res. 2004  
\* Lauritzen & Jensen. Stable local computation with conditional Gaussian distributions. Stat. & Comp. 2001  
\* Li et al. Dynamic scaled sampling for deterministic constraints. AIStat 2013  
  
  
R2.  
1> Existing attempts to address nonlinear constraints & piecewise distributions:  
Please see R1.1.  
  
2> Is the performance of SymGibbs significantly better than BaseGibbs?  
  
In Figures 5(a), 6(a) and 6(b), for a given time e.g. 0.5, the absolute errors of BaseGibbs is about 9, 2 and 3 times more than SymGibbs, respectively.   
  
R3.  
1> Lack of large model sizes makes one suspect that the method is perhaps not scalable for a large number of variables:   
  
Manual computations show that the number of PPF partitions do not grow significantly in the model size. So, model size should not considerably affect the results (except that all Gibbs samplers grow linearly in dimension).   
Since we substantially outperform existing state-of-the-art MCMC methods for the current range of model sizes, it is not clear to us that additional experiments would alter the conclusions from our empirical comparison.  
  
2> Short sampling time:  
  
In our experiments, sampling for 2 seconds is sufficient to show that SymGibbs consistently performs the best.  
Sampling for a longer period would not affect this conclusion (since SymGibbs has already been converged); however, it would make the plots hardly distinguishable. For instance, if Figure 6b was plotted for an hour then it would be quite hard to see the difference between SymGibbs and the second best algorithm (i.e. BaseGibbs) by looking at the plots.  
  
3> The experiments are simple collisions and trivial wiring models:   
  
The collisions & wiring models are not trivial in terms of the posterior densities that arise in these models with nonlinear determinism. They are already difficult enough to cause the state-of-the-art inference methods to perform poorly as shown in the experiments. In the case of the wiring model, SymGibbs and to a lesser degree, BaseGibbs (which is also based on the collapsing determinism and therefore, ours) outperform the state-of-the-art methods (HMC and SMC) so distinctly that we can claim that for the first time, we have introduced an effective & automated solution for handling ‘reduced mass’ deterministic relationships.   
  
4> Growth in the size of PPFs as variables are eliminated:  
  
The growth can be exponential (due to operations such as absolute value) though in practice many pieces may be infeasible (empty) and therefore removed.   
  
5> Most Bayesian models incorporate observation error:  
  
Consider a case with small observation error (i.e. a near-deterministic constraint, as mentioned in your review). The performance of MCMC methods on such models is typically very poor due to the "near-deterministic problem" (see R1.1).  
Interestingly, provided with our framework, for the first time we can suggest exactly the opposite of what probabilistic programing languages often do:   
To avoid the near-deterministic problem in models with small incorporated noise, approximate the noisy observation with a deterministic observation and reduce the dimensionality by Alg 1.   
  
6> RE introduction:  
Please see R1.1.**