Check 1D lattice gas model of 1D Protein Interaction Code

Fangzhou Xiao

20150827

We go from small (N, m) to large (N, m). Exact theory is

$$Z(N, m, \beta) = \sum_{k=1}^{k_{max}} Z_{cl}(N, m, k, \beta) \exp((m-k)\epsilon\beta) \tag{1}$$

$$Z_{cl}(N, m, k, \beta) = \delta_{N,m} \exp(\epsilon\beta) + \frac{N}{k} F(k, N-m) F(k, m) \tag{2}$$

$$F(k, n) = C_{k-1}^{n-1} I_{\{k \geqslant 1, n \geqslant k\}}, \qquad k_{max} = \min(N-m+1, m) \tag{3}$$

1 $N = m$

We first consider $N = m$ case. $k_{max} = 1$. $Z_{cl}(N, N, 1, \beta) = \exp(\epsilon\beta) + NF(1, 0)F(1, N) = \exp(\epsilon\beta)$ because $F(1, 0) = 0$. So

$$Z(N, N, \beta) = \exp(\epsilon\beta) \exp((N-1)\epsilon\beta) = \exp(N\epsilon\beta)$$

Therefore

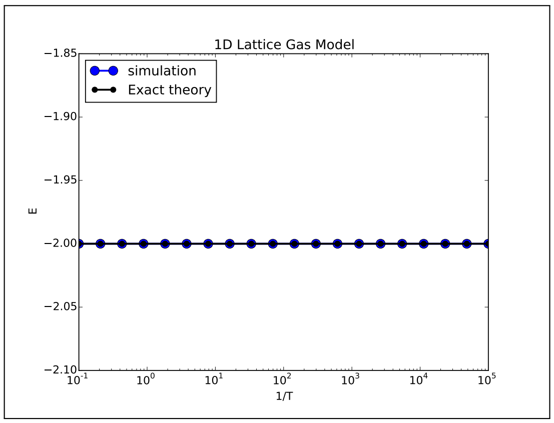
$$E = -N\epsilon$$

which is independent of temperature.

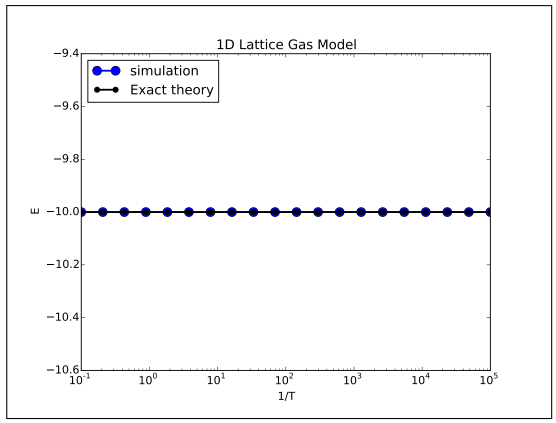
2 $m \neq N$

For this case. the δ_{Nm} term disappears in Z_{cl}.

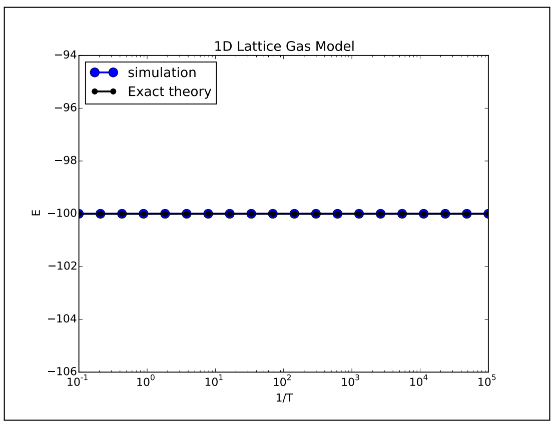
$$Z = \sum_k Z_{cl}(k) \exp((m-k)\epsilon\beta)$$



(a) $N = m = 2$.



(b) $N = m = 10$



(c) $N = m = 100$.

Figure 1: $N = m$ case. (obs,dur) = (150, 300) unless specified.

So energy

$$E = \frac{\sum_k (m-k)\epsilon Z_{cl}(k) e^{-k\epsilon\beta}}{\sum_k Z_{cl}(k) e^{-k\epsilon\beta}}$$

where we canceled $e^{m\epsilon\beta}$ term.

For numerical stability, we can time both the numerator and denumerator with $\exp(\frac{1}{2}k_{max}\epsilon\beta)$. But I did not implement this yet.

The major difference between theory and simulation is when T is large. The theory says when $\beta \to 0$,

$$E = \frac{\sum_k (m-k)\epsilon Z_{cl}}{\sum_k Z_{cl}}$$

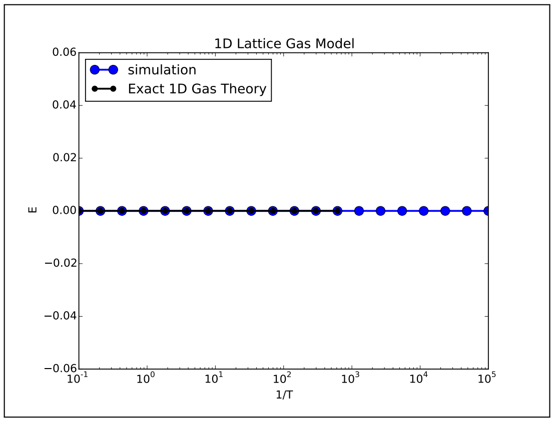
From computation we see for$(N, m) = (4, 1)$, $E = 0$, which agrees trivially with simulation. On the other hand, $(N, m) = (4, 3)$, $E = -2$, which also agrees completely with simulation. The latter is because no matter how we move there is always one and only one chain, of length 3. Only $(N, m) = (4, 2)$ gives non-trivial simulation results. When (obs, dur) = (150, 300), simulation and theory don't agree well at small β. Taking (obs, dur) = (1000, 1000), we have similar agreement.... So duration isn't problem here.

To see whether dynamics is the problem, I then uses the alternative dynamics, i.e. not allowing hopping. (obs, dur) = (10000, 10000). Turns out this is not large enough, so I'm trying (obs, dur) = (100000, 100000).

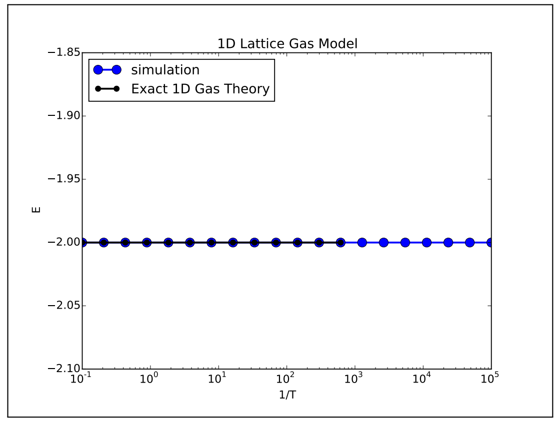
From here on, I changed the temperature space to be 20 points in log scale between 1e-3 to 1e3 for better numerical stability.

Then I look at large N, m results too see whether this discrepancy disappears. (obs, dur) = (10000, 10000). $(N, m) = (400, 200)$

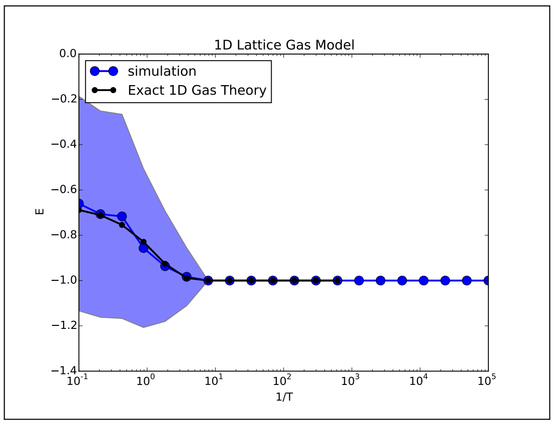
!! Eventually, as it turns out, the error in my "exact theory" implementation is forgetting "-1" in the combination function for $F(k, n)$.....



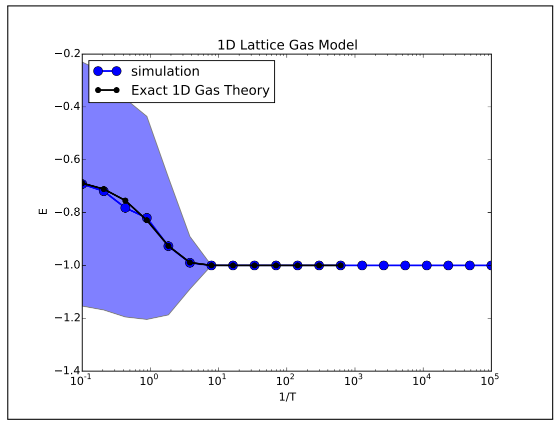
(a) $(N, m) = (2, 1)$.



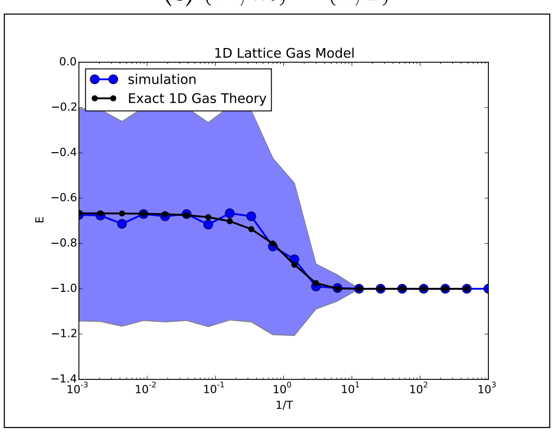
(b) $(N, m) = (4, 3)$



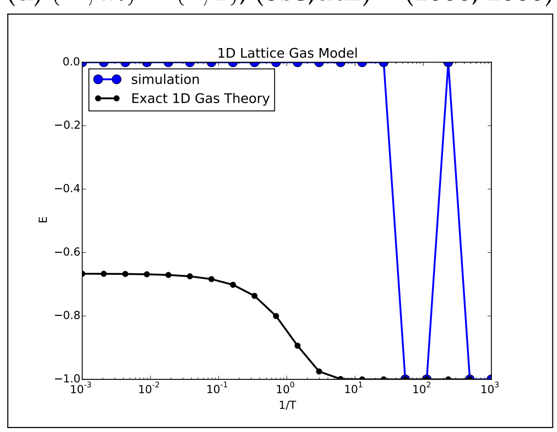
(c) $(N, m) = (4, 2)$



(d) $(N, m) = (4, 2)$, (obs,dur) = (1000, 1000)



(e) $(N, m) = (4, 2)$, same as 2c, but temperature range is different.



(f) $(N, m) = (4, 2)$, (obs,dur) = (10000, 10000), with no-hopping dynamics

Figure 2: $N \neq m$ case. (obs,dur) = (150, 300) unless specified.