## List of changes to the manuscript

1. Update Fig. 4 to include data for $r_s = 0.5$ and 0.25.
2. Report relevant benchmark values in the caption of Fig. 2 and S1.
3. Clarify the error analysis of effective mass from the original data; add data processing scripts to the public code repository for reproduction of the final results.
4. Make slight modifications to some phrases and sentences.

## Response to Report 1

1. The authors successfully benchmark their method against the ideal gas. Equally (if not more) important would be a benchmark against asymptotically exact analytic results for the (weakly interacting) small-$r_s$ regime, the green dashed line in Fig. 4. While the data for the largest studied system size $N = 57$ appears to be perfectly consistent with the small-$r_s$ analytic curve, this might be a mere illusion given the substantial drift of the data with $N$. Especially worrisome is the fact that the character of this drift dramatically changes with $r_s$ (cf. corresponding remark by Referee 1). Based on the above-mentioned circumstances, I suggest that the authors to produce more data:

   1. For $r_s = 0.5$ and 0.25.

   2. For $N$ substantially larger than $N = 57$ (if possible; if not, then explain why).

   **Response**: Thanks for the kind suggestion. We have extended the calculations to $r_s = 0.5$ and 0.25 for the electron numbers $N = 29, 49$ and 57 considered in the manuscript. The results are shown in Figure 1 below. One can see the data extrapolate smoothly to smaller values of $r_s$, where they also agree well with the low-density limit shown by the green dashed line as $N$ increases.

   The largest calculations of $N = 57$ electrons take around 5 days on 8 Nvidia V100 GPUs. The overall computational effort with respect to system size is $o(N^4)$. A substantially larger $N$ is possible but will go beyond the computational resources we have access to now.

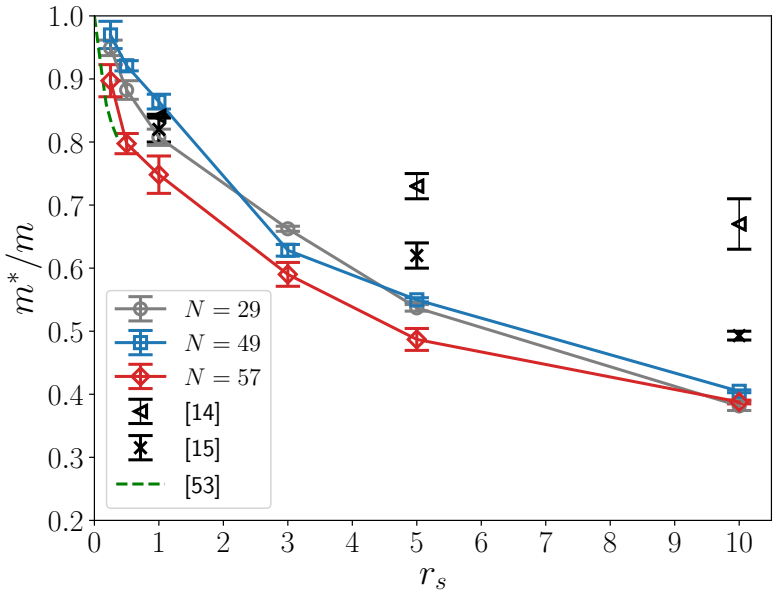


Figure 1: Updated version of Fig. 4 of the manuscipt, including additional data for $r_s = 0.5$ and 0.25 in the weak-interacting regime.

## Response to Report 2

1. The relation $m^*/m = s/s_0$ is based on the validity of Fermi liquid theory. Whereas it is very likely that FL applies to all parameters shown, I wonder if the machine learning data show any deviations from FL theory (ie, to make sure that the data is internally consistent). For instance, do the energies behave as $E \sim T^2$ for low enough temperature $T$? Or can we see a plot $S(T)$ for the interacting model?

   **Response**: Thanks for the kind suggestion. We have carried out a set of calculations for $N = 29$ electrons at $r_s = 10.0$ and temperatures $T/T_F$ ranging from 0.06 to 0.21, including the point $T/T_F = 0.15$ chosen for the extraction of effective mass. Figure 2 below shows the entropy $s$ per particle as the function of $T/T_F$. One can see that for electron densities as low as $r_s = 10$, the scaling behavior $s \sim T$ still holds nicely, as expected from the predictions of Fermi liquid theory.

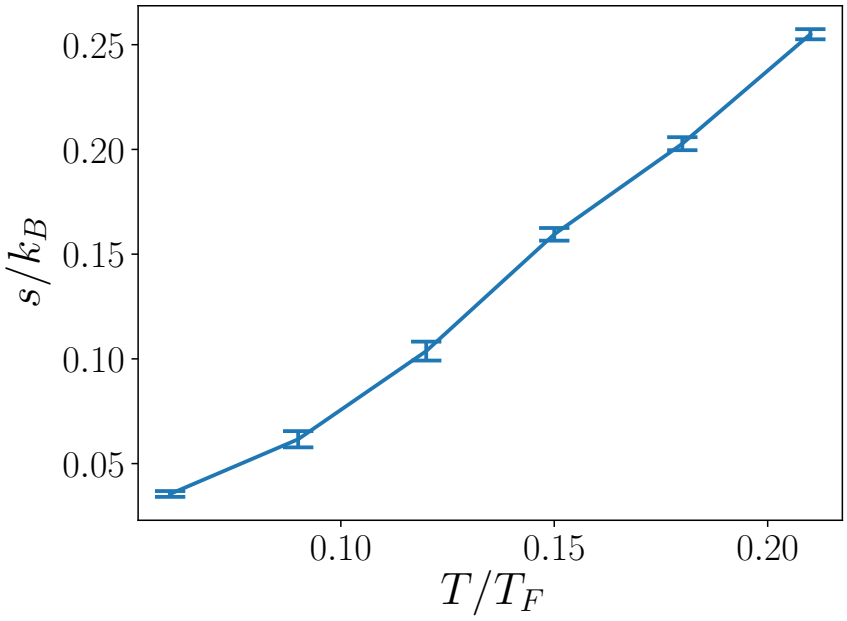


Figure 2: The entropy per particle of 29 interacting electrons at $r_s = 10$ as the function of temperature.

2. In many figures (Fig. 2, 3b, S1) the marker of the data points is big. Could the authors please write the value of the converged answer, with error bars, in the figure caption?

   **Response**: Below are the relevant data in Fig. 2 and S1 of the manuscript:

   **Fig. 2 left panel** $s_0 : 0.4227(4)$; benchmark : 0.4232

   **right panel** $s_0 : -0.28746(8)$; benchmark : $-0.2863(1)$

   **Fig. S1 left panel** $k : 0.04250(7)$; benchmark : 0.0426(1)

   **right panel** $v : -0.14360(7)$; benchmark : $-0.14358(1)$

   We have also added these numbers to the manuscript in the caption of corresponding figures. Thanks for the suggestion.

   Note we have not written the value of the original training data in Fig. 3b, as they need a little more processing to give final estimates of the effective mass. See the following response for details. Nevertheless the training data are available publicly.

3. The non-monotonicity of the data as a function of $N$ seen in Fig. 4 for $r_s = 3$ and 5 seems remarkable to me. What is the explanation? Or is it a consequence of effects seen in Fig. S2 (which would imply large systematic error bars)?

   **Response**: The $N$-scaling behavior of effective mass is jointly determined by the entropies of both interacting and non-interacting electron gas. The non-interacting entropy is monotonically increasing with $N$, as shown by the green line of Fig. S2. (Note we have used the twist-averaged boundary conditions in all calculations.) On the other hand, the scaling behavior of interacting entropies is more complicated and may be rather different for various values of $r_s$. In particular, it is hard to tell whether the interacting entropy is monotonically varying with $N$ for any given $r_s$. Answering such questions requires a rigorous finite-size scaling theory for the entropy of interacting electron gas, which to our knowledge is still lacking. Consequently, the effective mass may converge in a more regular way for certain values of $r_s$ due to "better" error cancellation between the interacting and non-interacting entropies.

4. The data shown in Fig. 3 seems to fluctuate a lot. Can the authors indicate how they extract the final entropies and error bars from these curves. Naively, the data fluctuate more than 15% with strong autocorrelations extending over many epochs, perhaps even drifting, and this is hard to reconcile with the rather tiny error bars in Fig. 4. The authors should provide a more detailed error analysis than the few sentences that are currently written in the text.

   **Response**: As you noticed, during training the entropies (like those shown in Fig. 3b) exhibit higher sensitivity than other quantities like the free energy. To take into account such fluctuations in the estimate of error bars, we have carried out an exponentially-weighted moving average over the original entropy data. This procedure is pretty standard and implemented in many data processing libraries like `pandas`.

   Below we show a simple code snippet that takes the original entropy data `s` and outputs the exponentially-weighted mean `s_ewm` and variance `s_ewm_var`, which are then used to plot the error bars shown in Fig. 1 above (including the additional data for $r_s = 0.5$ and 0.25 as suggested by the first referee). Note we have set the smoothing parameter $\alpha$ to be 0.01; see `pandas` doc for the definition. However, it's worth emphasizing that changing the smoothing parameter to reasonably different values will not essentially affect the conclusions of this work. As an example, Figure 3 shows the processed data in the setting of $\alpha = 0.05$, which reflects essentially the same behavior of the effective mass as Fig. 1.

```
import pandas as pd
import numpy as np

ewm = pd.Series(s).ewm(adjust=False, alpha=0.01)
s_ewm = np.array(ewm.mean(bias=True))
s_ewm_var = np.array(ewm.var(bias=True))
```

   We have also included a notebook in the public code repository for reproducing Fig. 1 above from the original data, and revised the manuscript accordingly. Hopefully these will clarify your (and others') concerns about the error bars of our predictions.

   By the way, here may also be a suitable place to give an explanation to your following concern:

   I find it counterintuitive that the error bars for $r_s = 10$ are the smallest. There are no indications in the data leading up to Fig. 4 that indicate so, and one may hence worry that the systematic errors are severely underestimated in this work.

   For small $r_s$, the entropy of 2D interacting electron gas is larger. This implies that there are more states that are actively populated and contribute significantly to the entropy. These state levels may change dramatically during optimization of the normalizing flow model, which then leads to larger

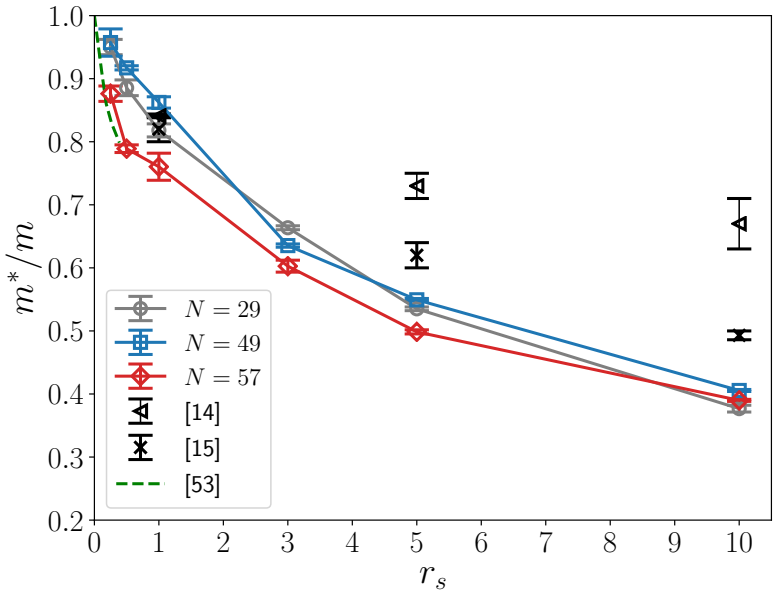


Figure 3: The same as Fig. 1, except that the smoothing parameter $\alpha$ for the exponentially-weighted moving average is set to be 0.05.

fluctuations of entropy for small $r_s$. Enhancing the expressive power of the normalizing flow model may help to further "stabilize" the involved energy levels and hence reduce the error bars of our predictions.

5. The energy shown in Fig. 2b for $r_s = 5$ goes below the value of the energy reported in the literature whereas the energy for $r_s = 10$ in Fig. S1 seems to agree. Could the authors elaborate more? Is there a systematic trend where the method introduced here performs better than other methods? Knowing (ground) state energies as a function of $r_s$ would certainly also be a plot of interest.

   **Response**: The benchmark data in Fig. 2b is from a **variational Monte Carlo** calculation for the ground-state electron gas published in 1989. It is therefore not surprising that our result is significantly lower than this rather early result. On the other hand, the benchmark data in Fig. S1 are from a **restricted path integral Monte Carlo** calculation for the electron gas at finite temperature. This result can be seen as nearly exact for such a low density as $r_s = 10$, where the fermion sign problem is not severe. The good agreement between our results and this benchmark, therefore, indicates great expressive power of the models used in our approach.

6. What possibilities exist to compute other, common Fermi liquid parameters?

   **Response**: One important example is the quasiparticle-quasiparticle interaction parameters $F(\boldsymbol{k}, \boldsymbol{k}')$, which can be written as the second-order functional derivatives of the total energy $E(\boldsymbol{K})$. (Note $\boldsymbol{K} = \{\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_N\}$ collectively denotes the momenta occupied by each electrons.) $E(\boldsymbol{K})$ is known as the Landau energy functional and is directly related (up to a constant) to the logarithm of the momentum occupation $\log p(\boldsymbol{K})$, which is represented as an autoregressive model in this work. One therefore can obtain $F(\boldsymbol{k}, \boldsymbol{k}')$ by naively taking energy differences between the ground state and various particle-hole excitations. Another way is to calculate other thermodynamic quantities such as the specific heat, which also incorporate certain information about $F(\boldsymbol{k}, \boldsymbol{k}')$.

7. I see no particular reason why a short-range potential (like a Yukawa potential) cannot be studied in the current approach. Is there a particular reason why the authors stayed away from such simpler

problems?

**Response**: You are right. It is straightforward to study problems with short-range interactions using the current approach. In fact, the implementation will be even simpler because we do not need to carry out Ewald summation as we have done for the calculation of Coulomb potential.

We chose to study the interacting electron gas with Coulomb interaction because it is a fundamentally more important problem with open issues, as also pointed out by you in the report.