I thank the reviewers for their reading of this paper and their constructive comments. I have written in bold all changes in the manuscript.

As stated in the introduction the goal of this paper is to provides numbers for the resolution and a field of view of a given lensless holographic setup. Indeed several works show super-resolution [13-15,19-21,22-25] and/or field of view extrapolation [16-19] capabilities without providing theoretically grounded limits on the maximum super-resolution factor or the size of the extrapolated field of view. The evaluation of the derived resolution and field of view formula is only performed by assessing the modeling error in several conditions. I do not rely on reconstructed hologram to assess resolution and field of view as only few algorithms are able to perform both super-resolution and field of view extrapolation and in any case it is cumbersome to disentangle the effect of the reconstruction algorithm (and introduced priors) from the effect of the lensless setup alone. However, I provide in conclusion several examples from the literature where the bounds derived in this paper fits with the claimed resolution or field of view.

As a global answer to the comments about resolution and pixel size, this paper cannot provide prescription on camera pixel size or area. It gives only the maximum pixel size and the minimum field of view extent to properly model numerical propagation (that are used in numerical reconstruction methods) and to prevent any aliasing both in frequency and space domain. These number are not directly linked with the camera pixel size. Indeed, without any cutoff imposed by a lens, the spatial bandwidth of the complex amplitude in the detector plane is bounded only by $n \lambda^{-1}$ imposing a maximum pixel size of $\lambda/2$ for the propagation and $\frac{\lambda}{2(1+\sin(|\theta|)}$ for the object. This is independent of the camera sampling rate as (as stated page 2) the spatial spectrum of the intensity is the auto-correlation of the spatial spectrum of the complex amplitude. Because of the "folding" effect of the auto-correlation, any high spatial frequencies in complex amplitude may generate arbitrary low intensity spatial frequency.

As a thought experiment, if a sample contains two spatial frequency ν_a and ν_b lower than $n \lambda^{-1}$, then these frequencies will be also present in the complex amplitude in the detector plane and the intensity will contains the spatial frequencies $\nu_b - \nu_a$, 0 and $\nu_a - \nu_b$. Even if both frequencies ν_a and ν_b are much higher than the camera sampling rate, $\nu_b - \nu_a$ and $\nu_a - \nu_b$ can be sufficiently low to be detected by the camera. Hence, if the numerical propagation is done at the camera sampling rate, these frequencies can be cut off and the measured intensity not properly modeled.

This explains the size of the pixels $104 \text{ nm} < \lambda/4$ (accounting for $\sin(|\theta|) \leq 1$) used in the numerical experiments. However, as stated in the introduction, this paper does not say anything about how the propagation kernel should be numerically implemented and how the object should be reconstructed in further numerical processing steps or even if it is possible. It just gives bounds. As said in the concluding remarks, these bounds correspond to the resolution and/or to the field of view claimed in some already published experiments [17,19,21].

Reviewer 1:

1. What does the word 'bandwidth' exactly mean in the title and introduction, spectral bandwidth or the spatial bandwidth product? The author should make it clear.

It is now precised as the spatial-frequency bandwidth

2. Some previous works have realized about 200nm resolution of lensless imaging, which in my opinion is determined by the optical diffraction limit, and is quite different from TIRF (capturing the evanescent wave) or some computational super-resolution techniques (SIM, STED and so on). Is the above limitation equivalent to the analysis of Wigner distribution in the work, since common lensless set-ups can just detect signals in Fresnel distance and cannot record the evanescent wave.

The proposed propagation model does not take into account evanescent waves. The spatial frequency bandwidth is limited by the optical diffraction limit. However, for highly inclined illumination wave, the resolution can be up to $\lambda/2$ (pixel size of $\lambda/4$) thanks to the aperture synthesis.

- 3. The explanations of simulations in this work are lack of some details. For example, why the samples in Fig. 9(b) are different from others, which seem have masked constrains in borders?
 The 5 images used in the experiments are of size 20160 × 20160. The image Fig.9(b) was a bit smaller (finding such large test images was quite difficult) and thus was expanded using mirror boundaries as said in Sec 6.1.
- 4. Why uses binary USAF-1951 image in Fig. 9(e)? Binary phase or absorption masks are very often used in the literature and we believe that it is representative of some experiments.
- 5. The size of pixels is 104nm in simulations, which means the realized resolution is 208nm? And why uses the 208nm resolution here, restricted by the optical diffraction limit?
 See the global comment on the pixel size

6. The FoV in simulations is 20160*104nm=2mm, which seems not 'much larger than the considered sensor area'.
I have amended the text to prevent misunderstanding. It only means that the propagation was perform over a 20160 × 20160 for much larger than the 2520 × 2520 area used to compute the

7. The author states that the resolution is worse in border and better in center, which cannot be found in simulations. I suggest the author provide more simulations with practical meanings.

modeling error.

In Sec 3.B and Sec. 5, I state that the resolution is worse at the center than in the border. This comes directly from the space

bandwidth analysis. As depicted on Fig 2, in the Fresnel regime, the bandwidth at the center (x = 0) is $\frac{\ell k}{z}$. It is half of the bandwidth at the edge of the sensor $(x = \ell/2)$: $\frac{2k\ell}{z}$. This also true with slightly different bandwidths using the angular spectrum. I however was not able to find any way to properly illustrate this on simulations without using reconstruction algorithms.

Reviewer 2:

1. Mathematical derivation are some times a bit "rude". See for instance Eqs. (39)-(42). Discussion should be a bit more progressive. Maybe addition of supplementary material for tedious derivation might help the reader follow the discussion.

Section 4.2 was rewritten to make the reasoning easier to follow.

- 2. Axes labels are missing in Figs. 10 to 13. I don't really understand as on my pdf version all axes are properly labeled: modeling error in dB as a function of the distance t in μ m.
- 3. From sec. 4.2.1 and 4.2.2 we would conclude that temporal and spatial coherency have the same effect on SBWP. What is the Author point of view on this aspect? Generally speaking, the submitted manuscript lacks for intermediate conclusion and in depth analysis of the results. Both temporal and spatial coherency have the same effect: it blurs the diffraction pattern reducing the visibility of high frequencies. As the propagation kernel is a frequency chirp, it reduce its size and its bandwidth. This is precised in section 4.2.
- 4. In Eq. (29), radial coordinate is defined by r. However, r is already used in the introduction to define the resolution.
 I have redefined the resolution with the upper case R
- 5. Eqs. (39) and (40) are derived considering an infinitely thin object. How do they scale if we consider a thicker object. This aspect should be discussed as the Author interests in z-sectioning capabilities in sec. 5. I have removed the hypothesis "neglecting the optical path length in the sample". In fact, the only hypothesis is that the optical path difference between the sample and the medium is much smaller than source coherence length. Otherwise, there will be no interference. In any case, the bound is an upper bound on the size, a non negligible optical path difference will only decrease the actual size of the diffraction pattern keeping it below the bound.
- 6. Cares have to be taken in the discussion about z-sectioning. As a matter of fact, even if presented as a 3D technique, digital holography is nonresolved on the light propagation axis (infinite depth of field). As far as I am concerned, Eq. (50) is associated with slicing, and gives the distance

in between two consecutive slices (some sort of two point resolution in z). Sectioning means to me that we can perform optical section throughout the object as it can be done in tomographic experiments.

I agree, Eq(50) describes the two points resolution rather than the optical sectioning power. Using born approximation, I only consider in this paragraph a 3D object as a stack of 2D slices. I have modified the text accordingly.

7. Parameters chosen for the simulation do not sound realistic. Especially pixel dimension scaling down to 100 nm. These might be associated with object plane pixel pitches working in microscopy configuration. As for as no extra details are given, it is hard to figure out: (i) if the illumination wave is divergent, (ii) where the simulated object is located compared to the point source, (iii) where is the sensor compared to the point source. I think that the simulation part lack for details for the reader to be convinced by the proposed results.

As stated in the first comment, the pixel size used in simulation it set to 104 nm to ensure a precise enough propagation modeling. Throughout the paper, the illuminating wave is plane (the source is assumed to be at infinity).