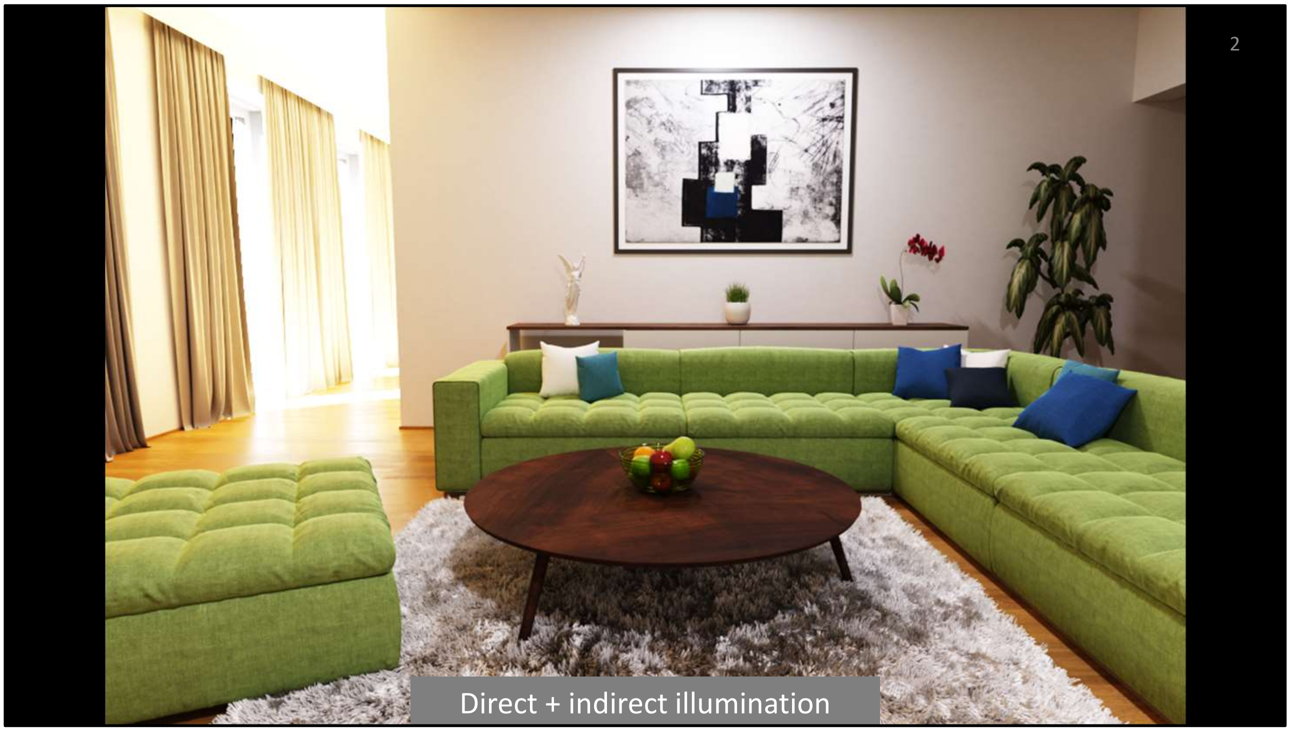
Bayesian online regression for adaptive direct illumination sampling

Petr Vévoda, Ivo Kondapaneni, and Jaroslav Křivánek

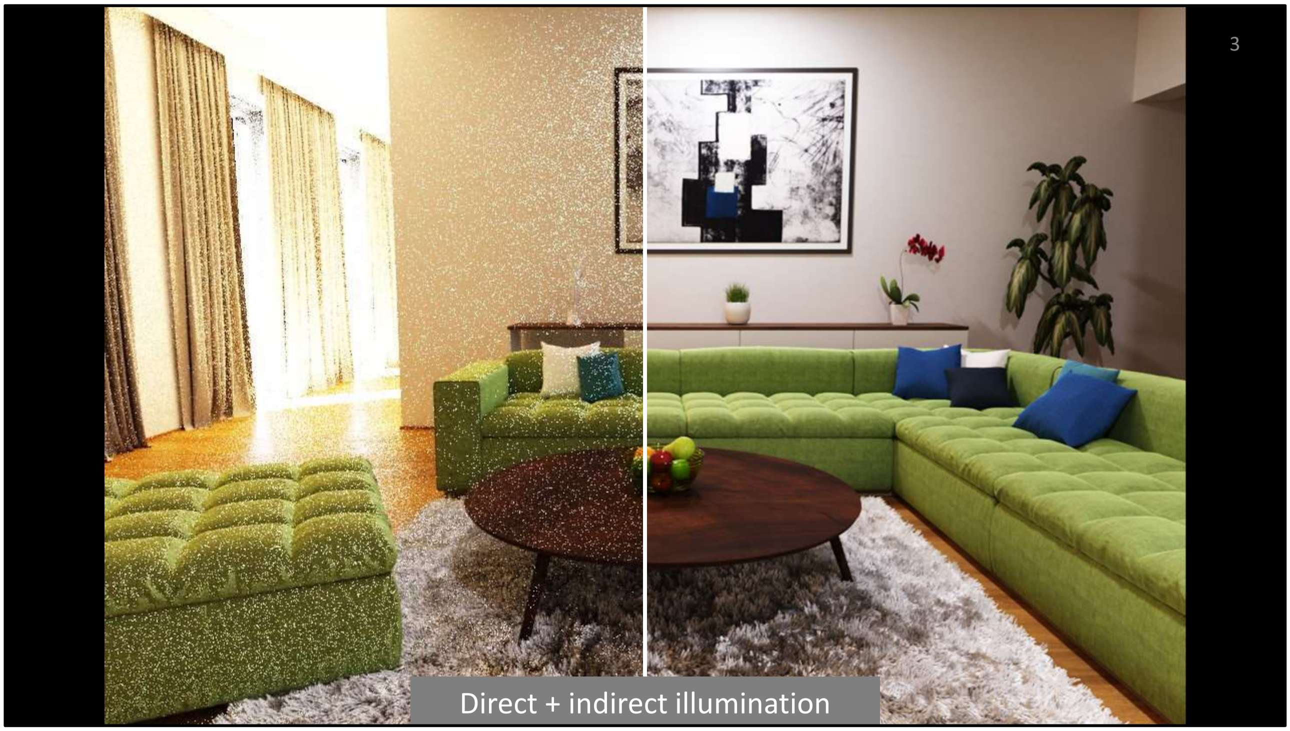
Render Legion, a.s.
Charles University, Prague

Computer Graphics
Charles University

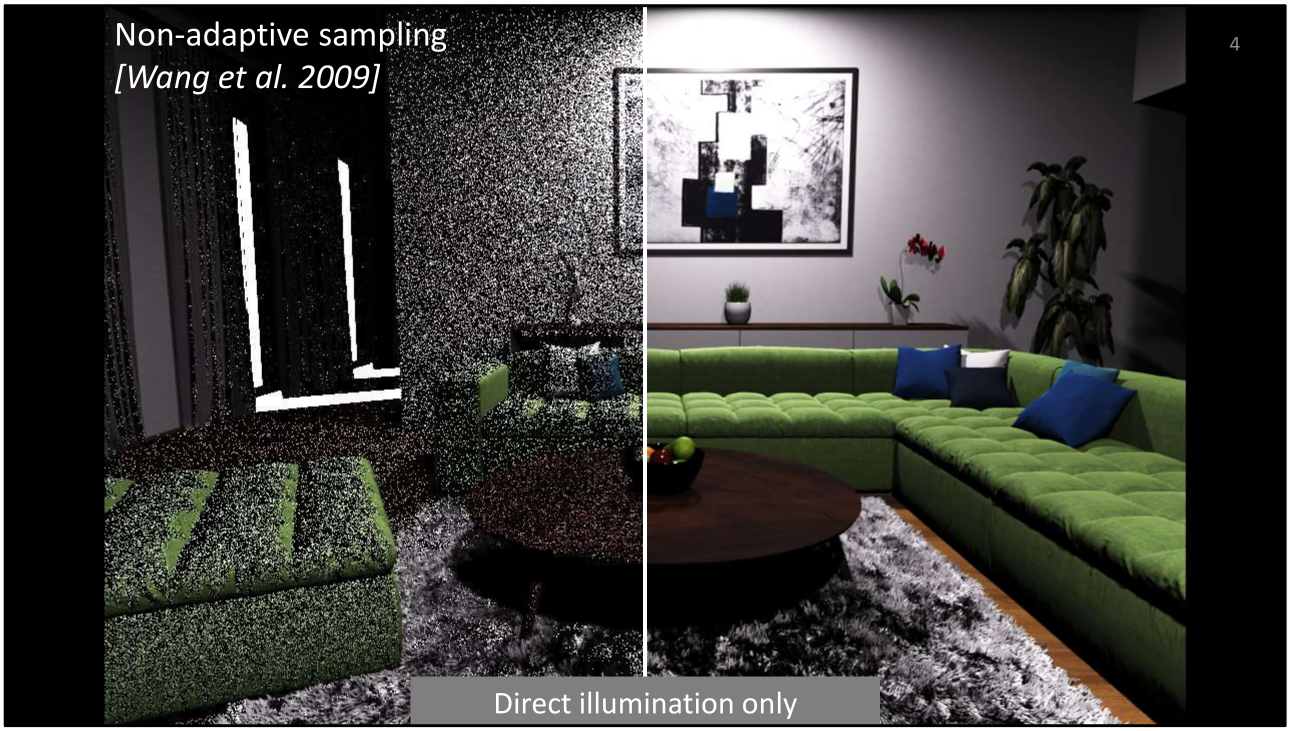
Hello, welcome to our presentation of our novel algorithm for robust adaptive direct illumination.



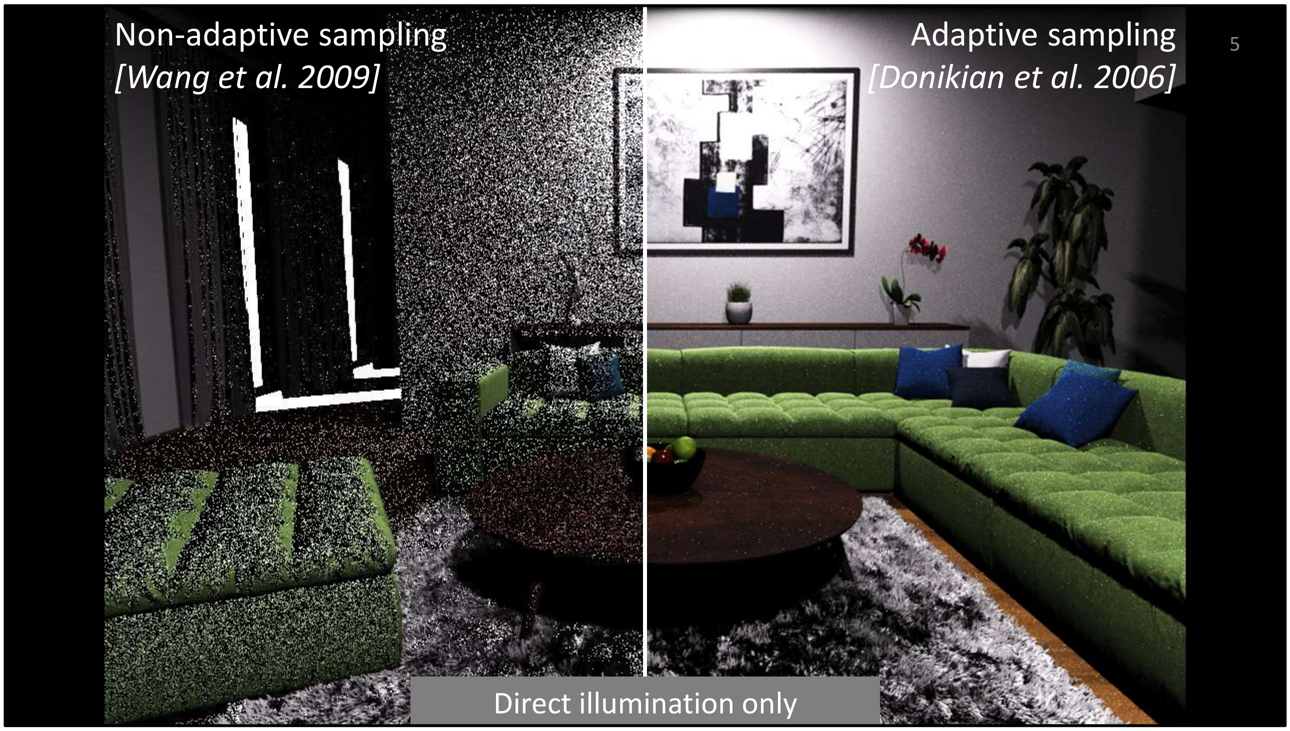
- MC rendering algorithms are currently getting more and more popular, but they suffer from noise.



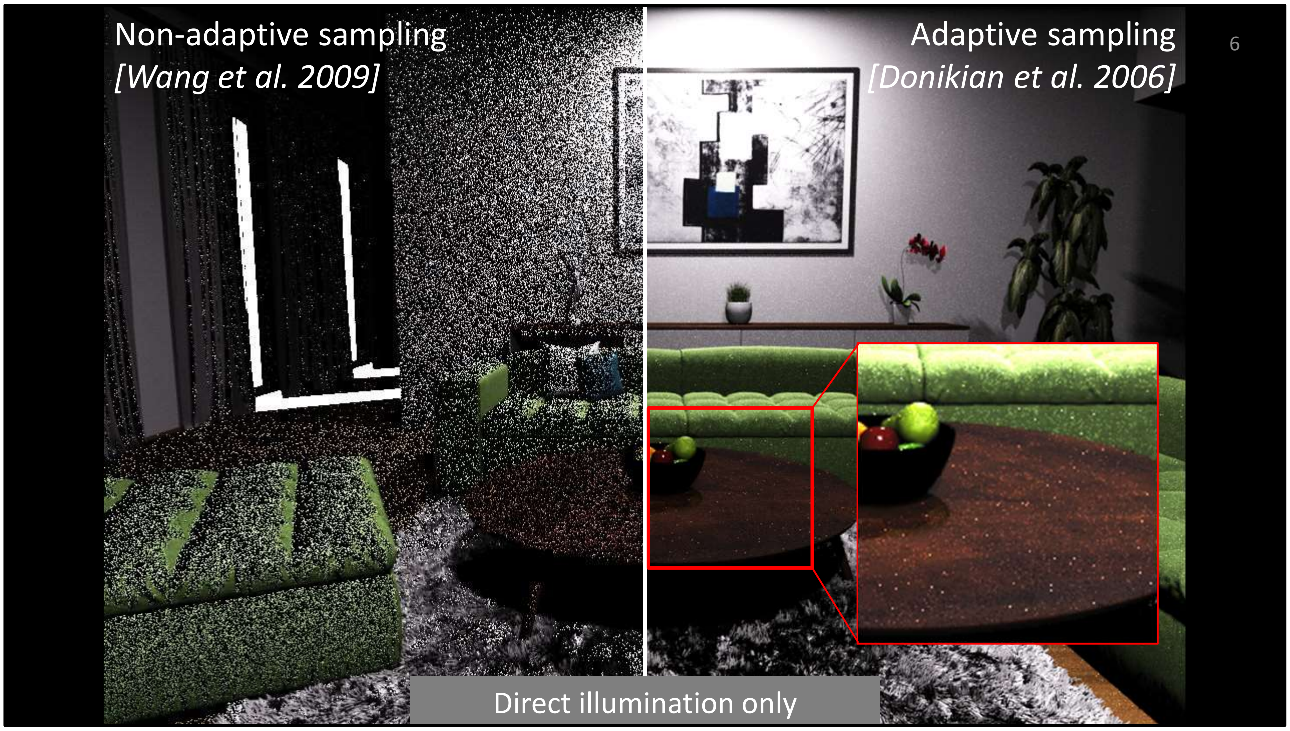
- Traditionally, the *indirect illumination* component has been considered as the main source of noise, and it's been subject to lot of research, … but in this scene it is actually the direct component which causes the trouble.



- In this image with direct illumination only we can see it clearly. The non-adaptive method shown on the left struggles to work efficiently, because it wastes lot of samples on strong but completely occluded sun.

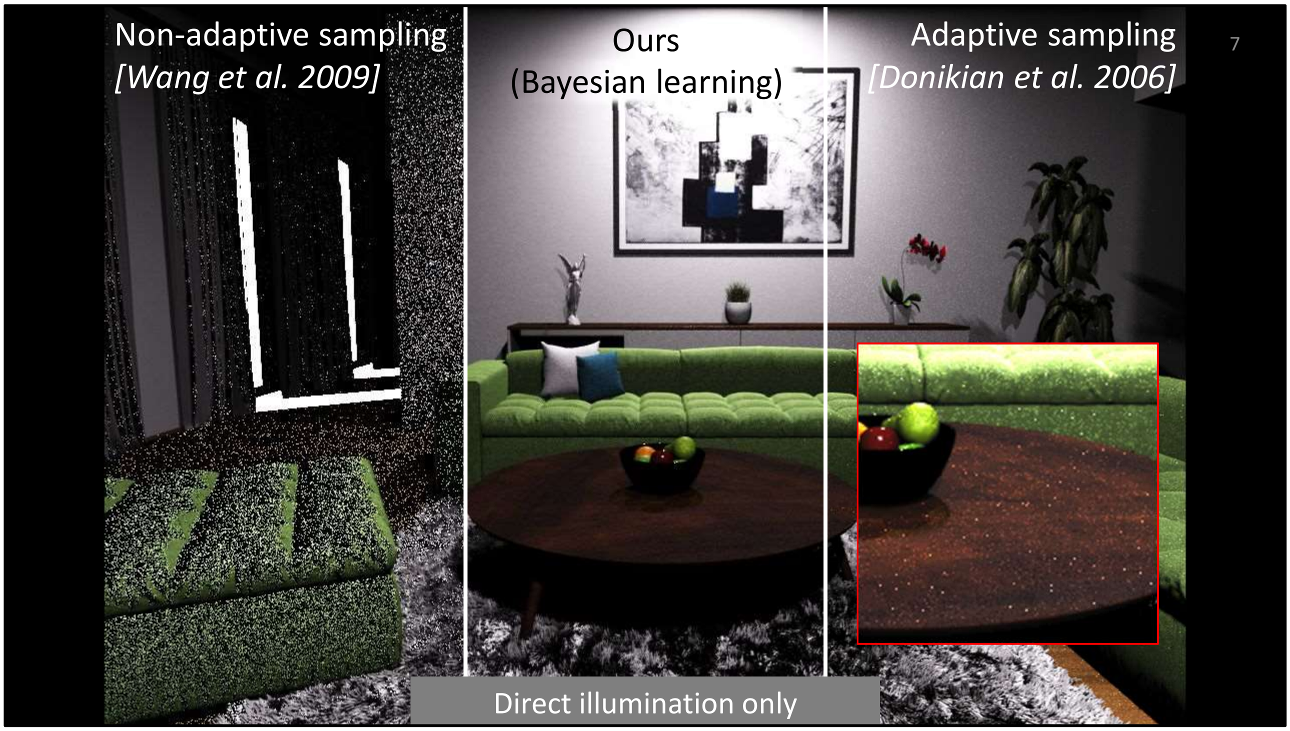


- Possible solution are *adaptive methods* which try to improve sampling based on past samples. But while they can decrease the amount of noise significantly, …

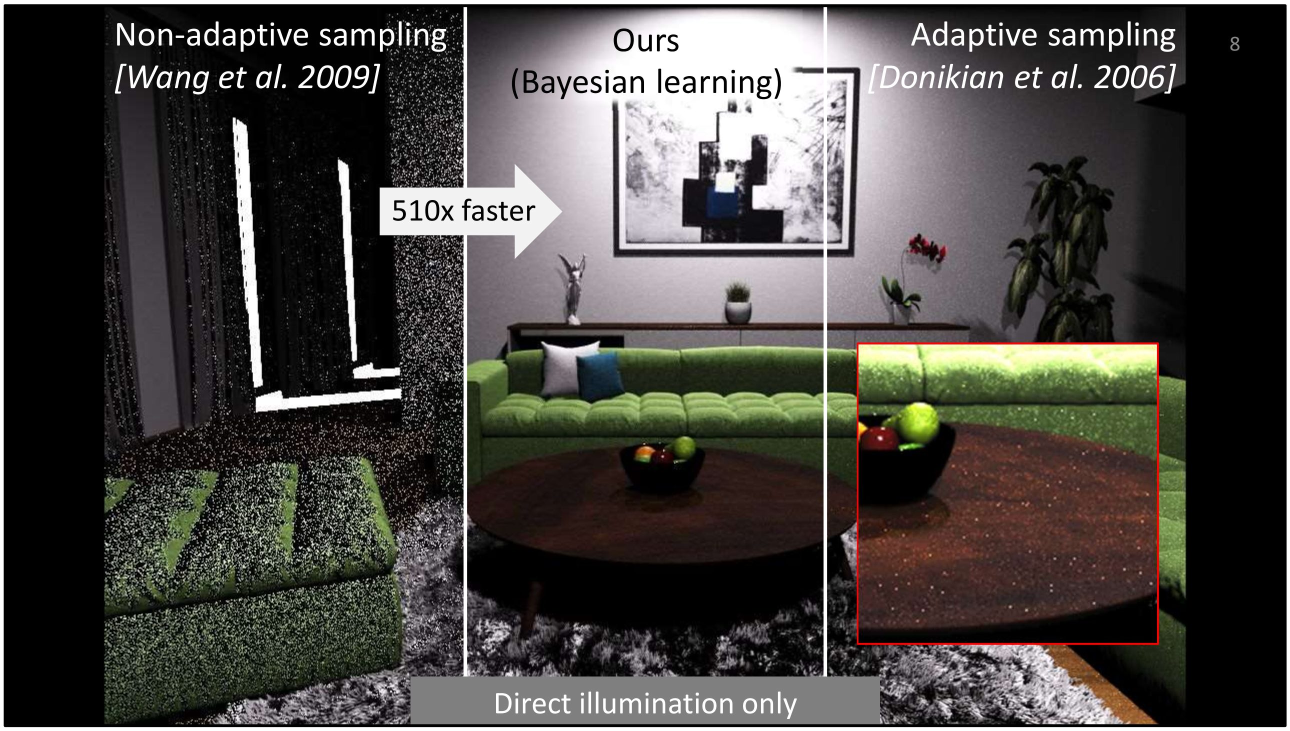


- they can also introduce artifacts or spiky noise because they are based on adhoc solutions and they tend to overfit.

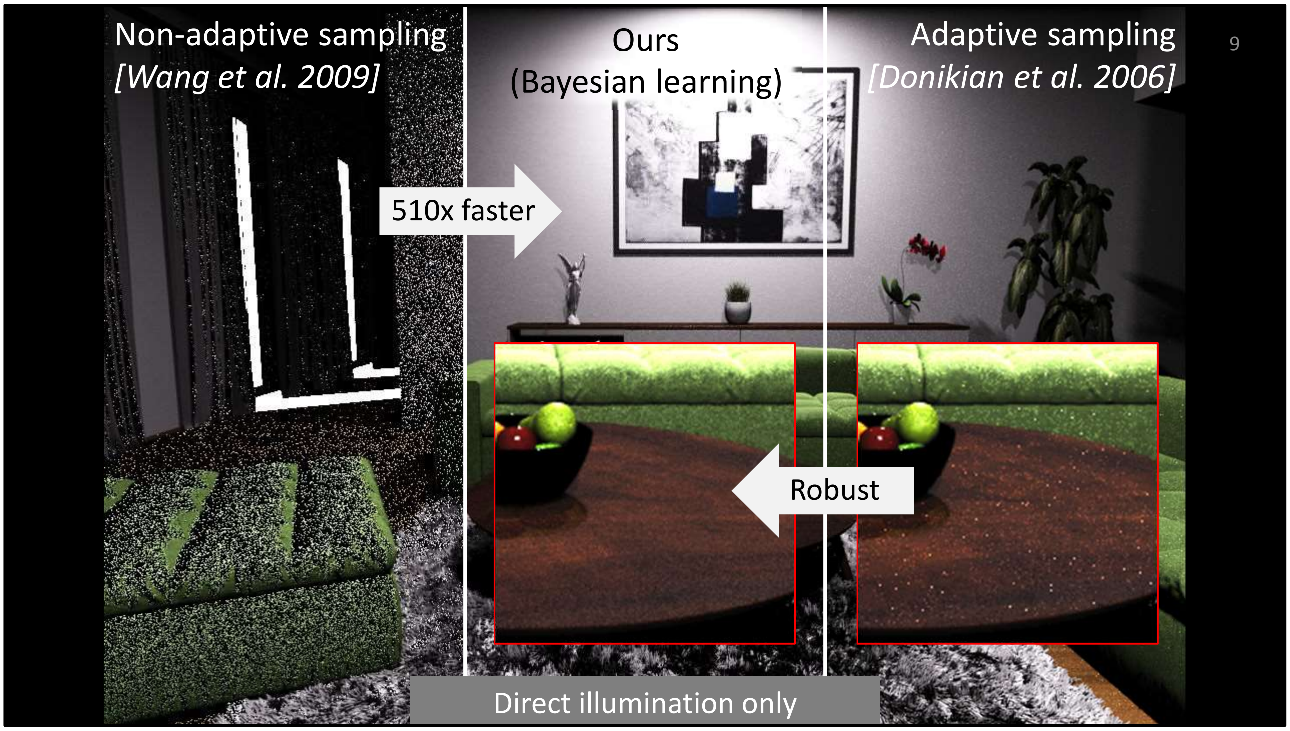
- This lack of *robustness* is a consequence of adhoc solutions to crucial questions in adaptive sampling: when is it safe to use the samples and how they should be combined with any previous knowledge?



- Therefore we propose a *first* solid theoretical framework for robust adaptive sampling in rendering.



- In this scene, our solution is more than 500 times faster than the non-adaptive solution…



- and we can achieve much better robustness.

- The new framework is not limited to the direct illumination. We are certain that other applications of adaptive sampling will benefit from it as well.

Previous work

In the context of Monte carlo simulation there is a lot of work related to ours.

Adaptive sampling

- General Monte Carlo
  - Vegas algorithm
    - [*Lepage 1980*]
  - Population MC
    - [*Cappé et al. 2004, ...*]
- Rendering
  - Image sampling
    - [*Mitchell 1987, ...*]
  - Indirect illumination (path guiding)
    - [*Dutre and Willems 1995*, *Jensen 1995*, *Lafortune et al. 1995, ...*]
    - [*Vorba et al. 2014, Muller et al. 2017*]
  - Direct illumination
    - [*Shirley et al. 1996*, *Donikian et al. 2006, Wang et al. 2009*]

- Adaptivity in Monte Carlo simulation is not a new concept.

- There is lot of work in the context of general Monte Carlo as well as in rendering, for example works dealing with image sampling, Indirect illumination and also in direct illumination.

- One of the oldest adaptive algorithms for Monte carlo estimation is Vegas by Lepage which works by histogramming integrand and using these histograms for sampling in next steps.
- Another example are population Monte carlo algorithms, which use population of particles and track how well they sample the integrand and based on that they keep the best individuals.
- In the context of rendering we can find a lot of work in image space sampling, we just mention the old work by Mitchell which deals with allocation of more samples into image parts with high-frequency content.
- In Indirect illumination computation first works were by Dutre and willems, where authors adaptively shot particles from lights (TODO: ziskat paper, nebo se zeptat Jardy), Jensen who used photon maps to construct sampling densities and Lafortune et al. who applied Vegas algorithm onto Monte carlo simulation of light transport.
- Along the years there were several other works dealing with this topic, but most recently there was work by Vorba et al. which uses gaussian mixtures for guiding

and Muller et al. which is a revamped version of algorithm by lafortune from 1995.

- Regarding direct illumination we mention a pioneering work by Shirley et al. who adaptively classified lights into important and unimportant ones and more recent work by Donikian et. al. which is closely related. Wang et al. sampled lights adaptively based on surface reflectance and estimates of lights' contributions.
- *None of these works deal with a problem of determining when and how to incorporate new information into the current sampling model.*

Bayesian methods in rendering

- Filtering
  - NonLocal Bayes [*Boughida and Boubekeur 2017*]
- Global illumination
  - Bayesian Monte Carlo [*Brouilat et al. 2009, Marques et al. 2013*]
  - Path guiding [*Vorba et al. 2014*]

corona

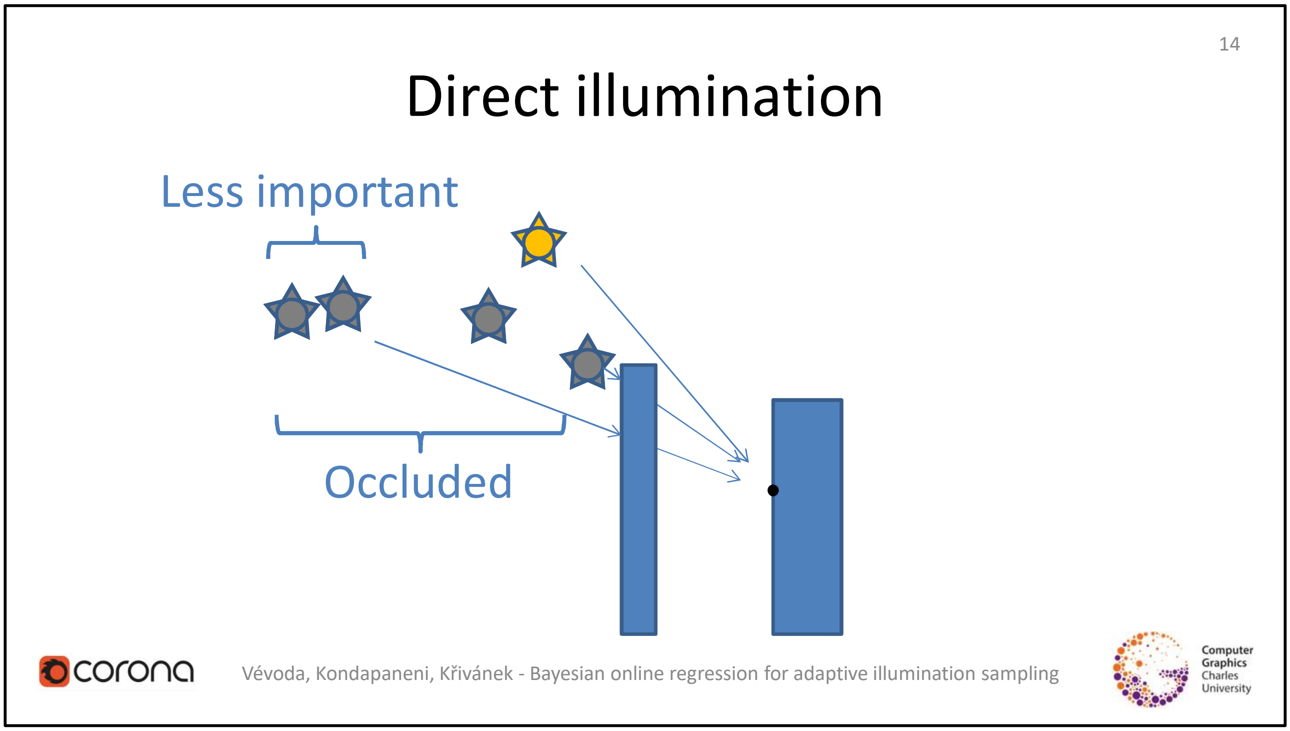
Vévoda, Kondapaneni, Křivánek - Bayesian online regression for adaptive illumination sampling

Computer Graphics Charles University

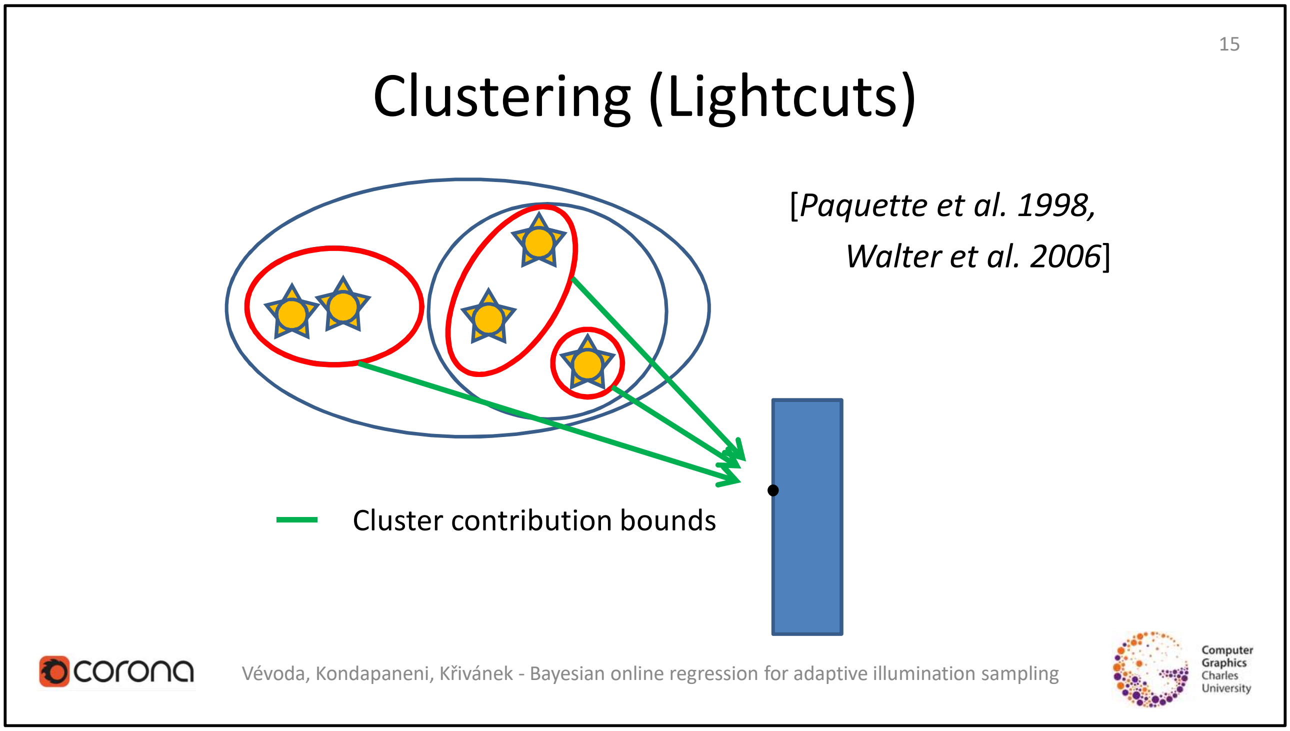
- Bayesian methodology on the other hand is not used very much in rendering.
- There are just few methods in filtering and Global illumination.

Background

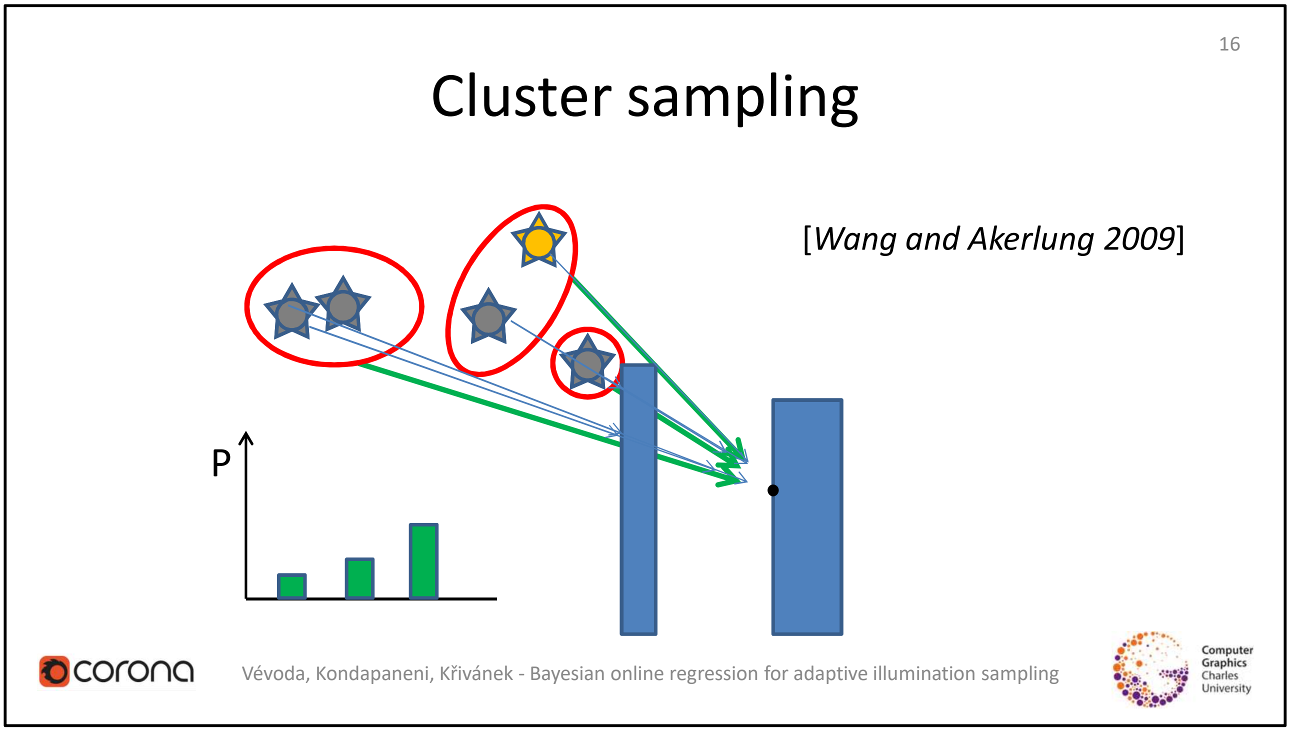
Now let me give you some background related to direct illumination problem



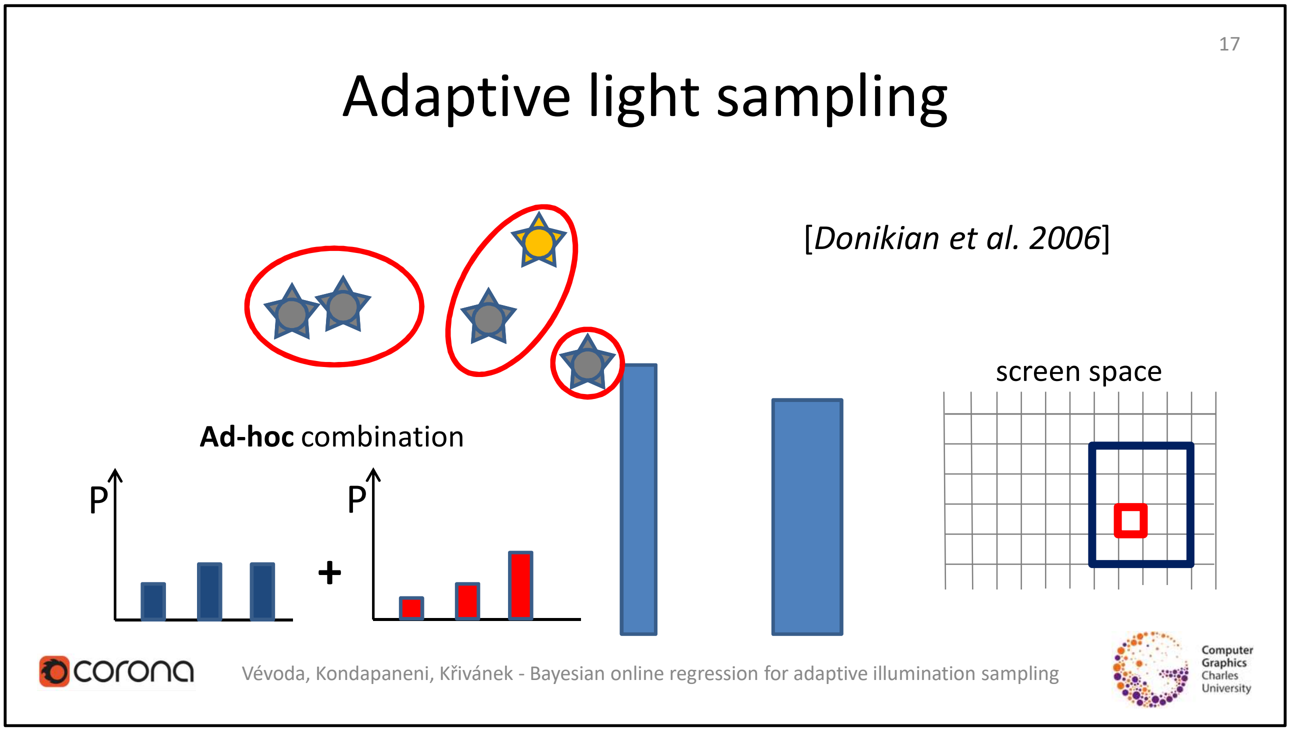
- What exactly is the direct illumination?
- We have a scene with several lights and some geometry
- And we estimate the ***direct*** contribution of each ligth onto each point in the scene. That is still a complex task ...
- due to uneven luminaire importances
- and due to occlusion



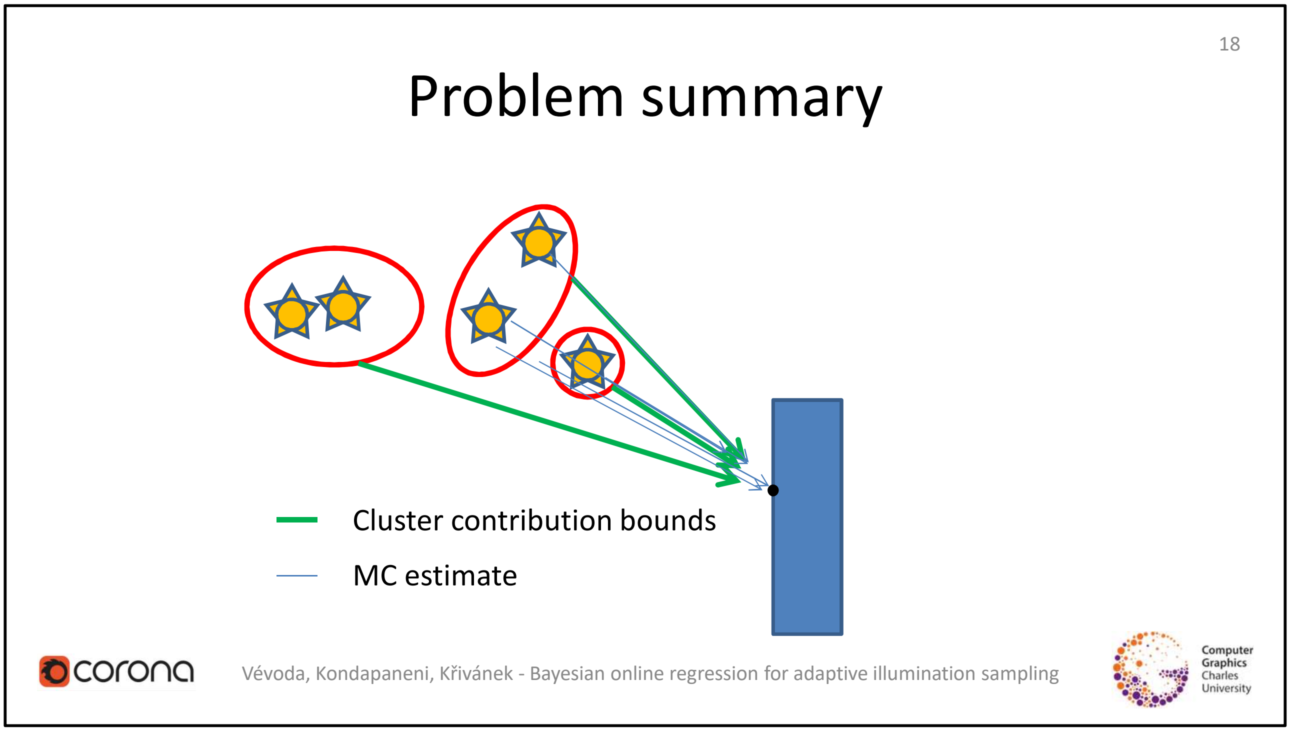
- One way to improve direct light estimation is to improve scalability when there is a lot of lights in the scene
- We can do so by hierachically clustering the lights
- and for each point in the scene we can choose clustering with the lowest approximation error.
- Each cluster gives us a conservative bound on its contribution to the point.



- Then we could use cluster contribution bounds to
- build a sampling distribution over clusters
- and then use it for getting MC estimates.
- At this point, we still wouldn't have adaptivity, and we wouldn't be able to capture complex visibility.



- We could achieve adaptivity as Donikian et al. [2006] did.
- They gather statistics about clusters in screen space and
- They do so for each pixel and from the statistics they build sampling over clusters. But these estimates are very noisy.
- Therefore they do the same also over the whole block of pixels and then
- they mix both information distributions, but in an ad-hoc way which may sometimes result into overfitting and artifacts in a picture.



- So our goal is to compute direct illumination by means of Monte Carlo, and for that we need to find **optimal** sampling distribution over clusters. We want to have **adaptive solution** because adaptivity has huge potential, but we strive for a **robust solution**.

- In order to achieve these goals, we have two kinds of information about clusters at hand:

    - We have cluster contribution bounds towards a point, which are conservative, noise-free. We have them right from the beginning so they can serve as our prior information.

    - And we have Monte-carlo estimates, which are noisy, and we get them over time.

- For a question how to combine these two sources of information together in a robust way we've found a good answer (as we explain shortly)

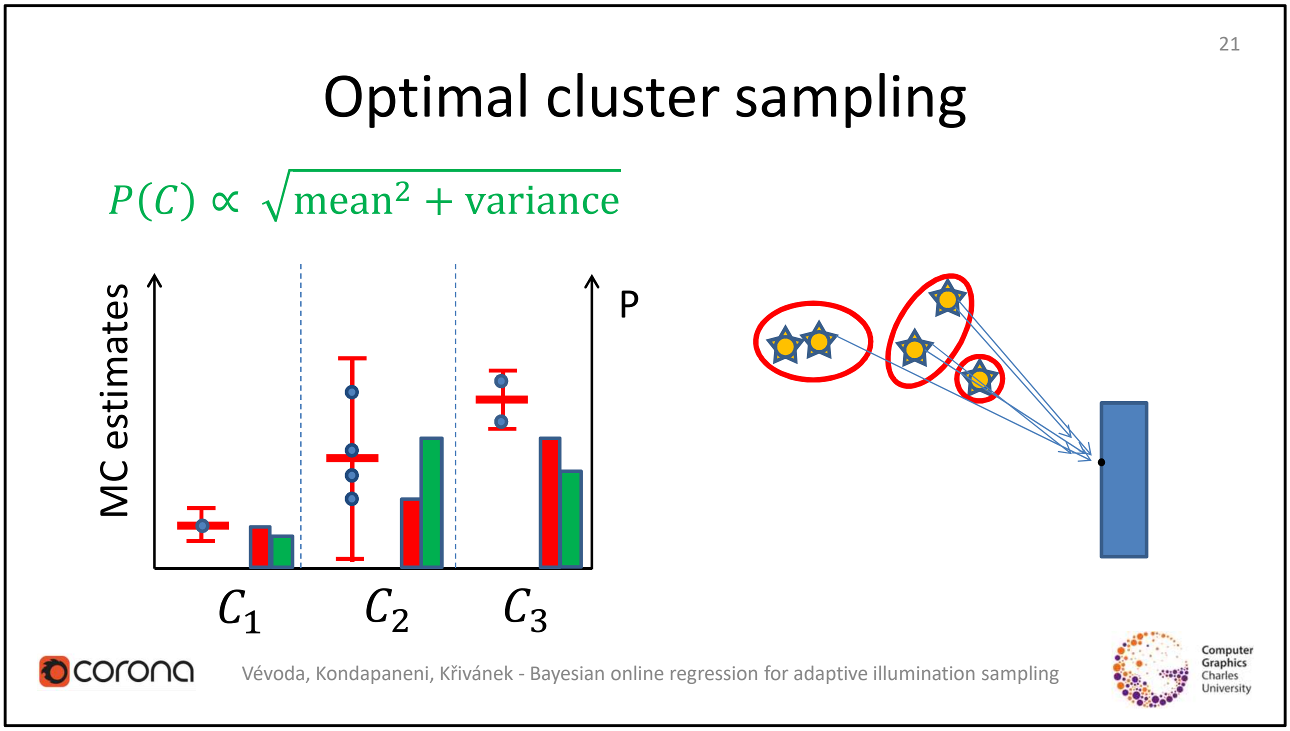
Our approach

Let us now introduce our approach ...

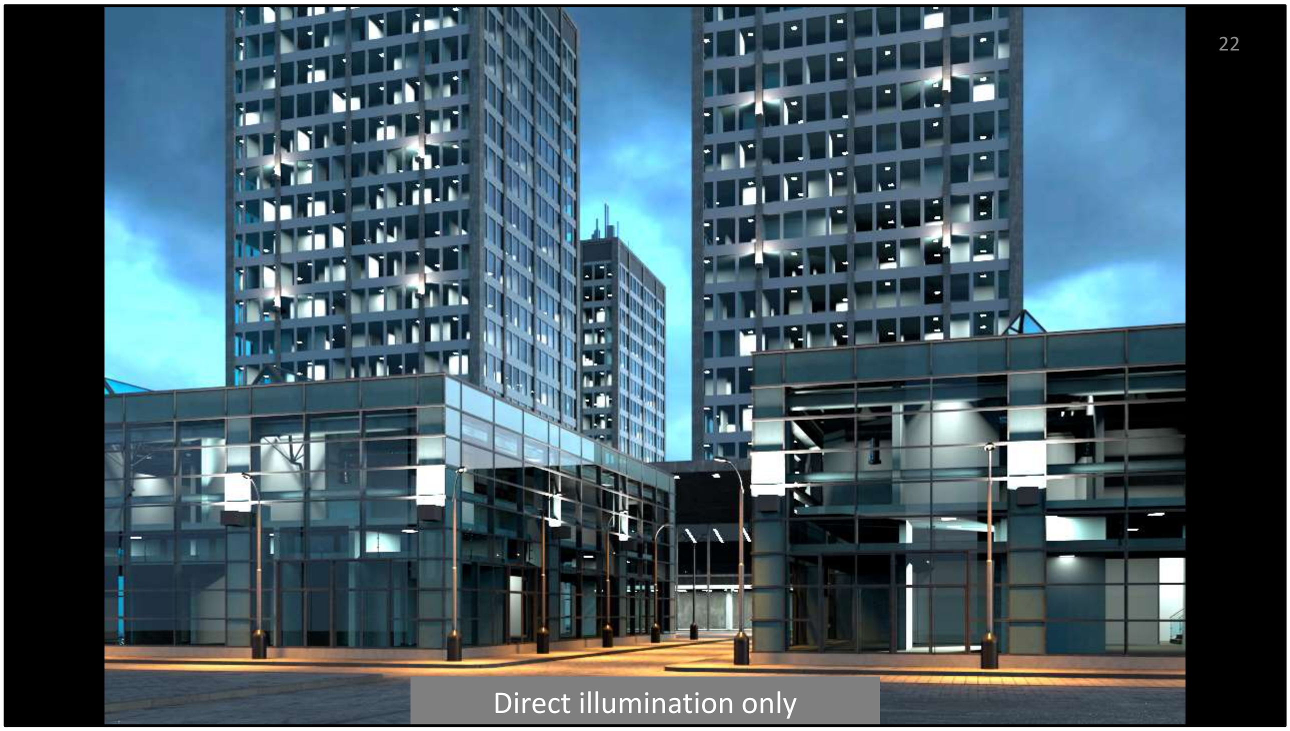
Contributions

- Optimal sampling of clusters
- Adaptive sampling by Bayesian inference

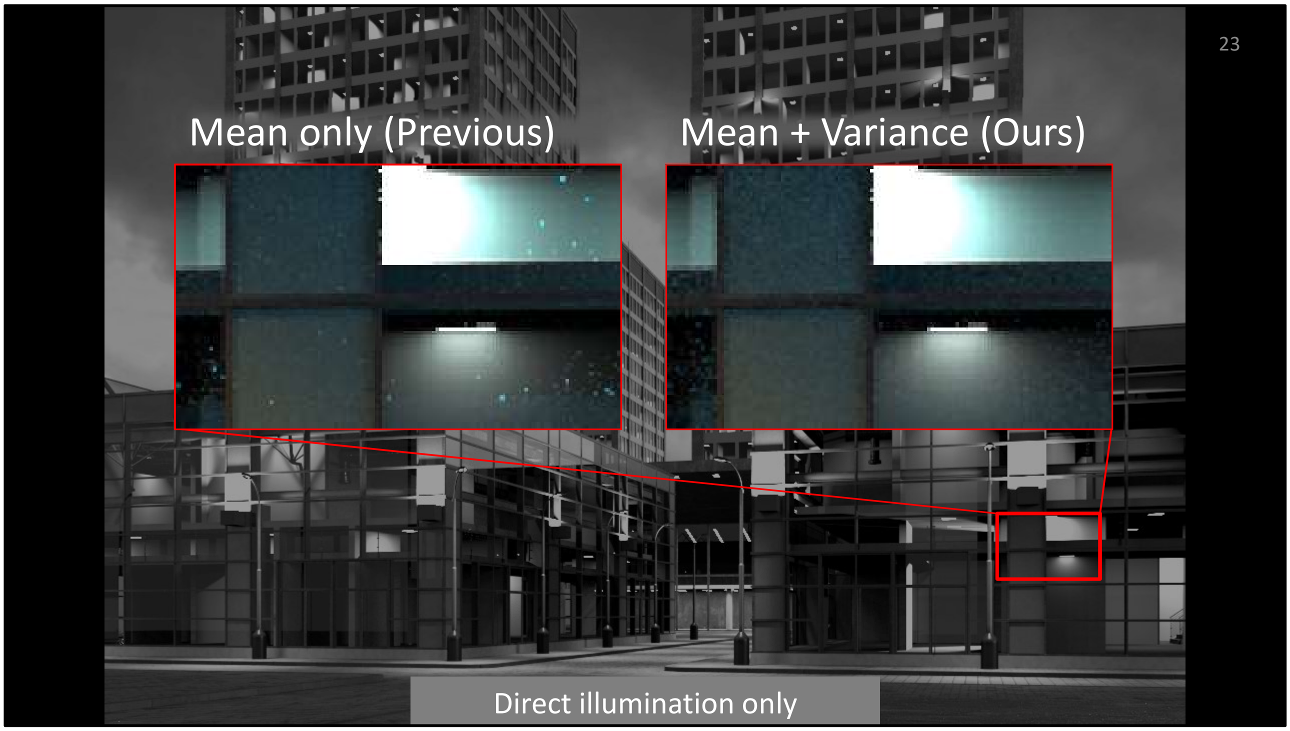
- Our main cotributions are two:
- We found the optimal sampling scheme of clusters
- And we do adaptivity with a help of Bayesian inference giving us a more robust solution and allowing us to combine Monte carlo samples with cluster contribution bounds in a principled way.
- Lets start with optimal sampling of clusters



- Given a scene, let’s have a look at how to derive sampling probabilities from MC samples
- We are sampling the clusters and getting our MC samples
- with some mean and variance.
- And normally, and what various approaches did in the past, the cluster's sampling probability would be proportional to the **mean only**.
- But we found out that variance of samples from each cluster has a big impact. The higher variance means that more samples should be allocated to it.
- Therefore optimal sampling should be proportional to the square root of the second moment of the samples.
- The sampling probability then can change drastically as is depicted by the green bars.



- Let me show you the practical example: this scene contains more than 5000 light sources so the clusters can be large and complicated….

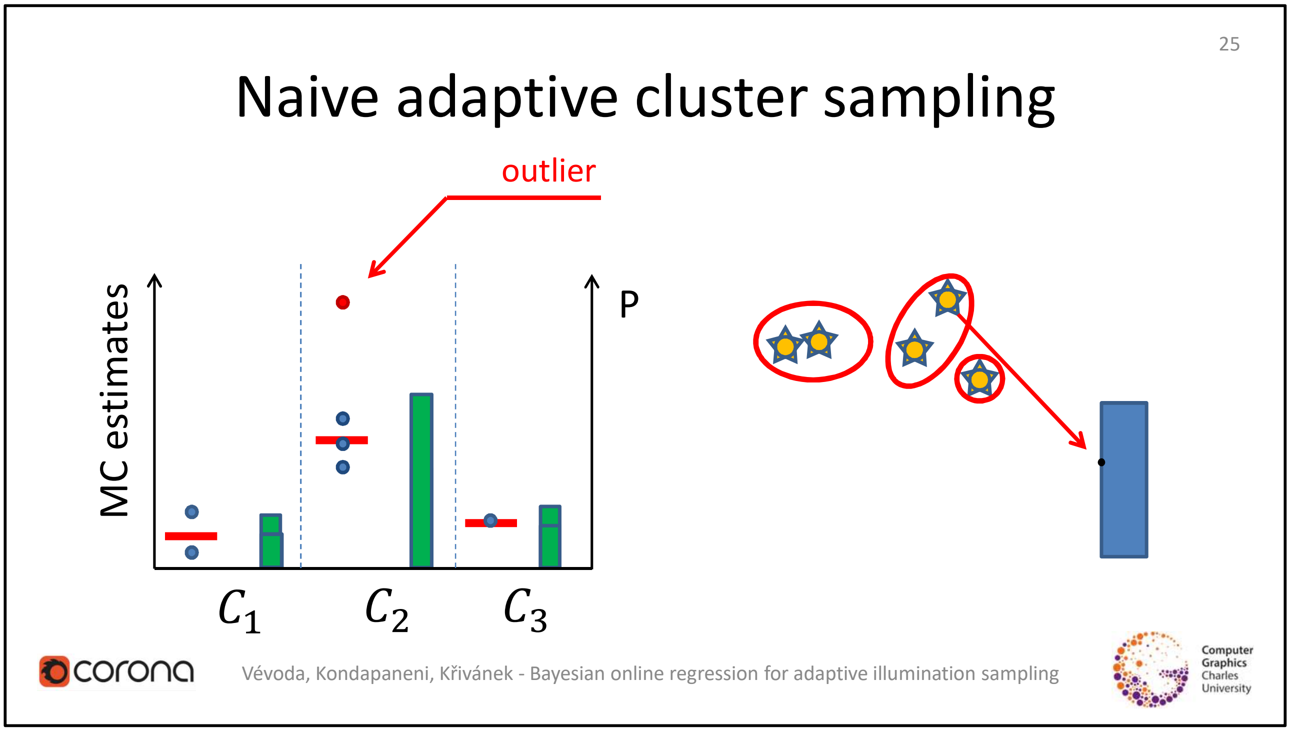


- On the left we see an inset showing how sampling according to a mean performs. It undersamples some tricky cluster which leads to spiky noise.
- And on the right we see that sampling according to both the mean and the variance eliminates this issue.

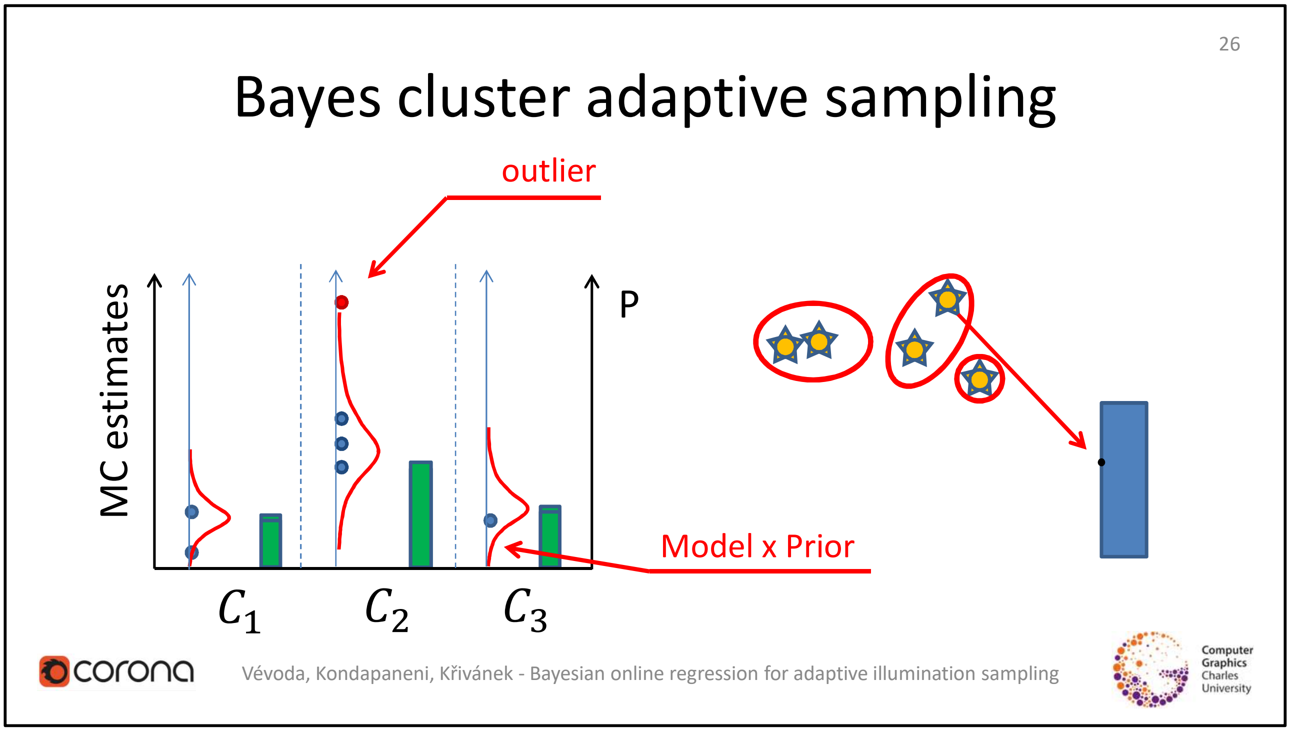
Contributions

- Optimal sampling of clusters
- Adaptive sampling by Bayesian inference

- Having explained the optimal sampling,
- we will show you now how to do the adaptivity the Bayesian way ...
- The issue is that the mean and the variance needed for the optimal sampling are not know upfront and need to be learned during rendering.

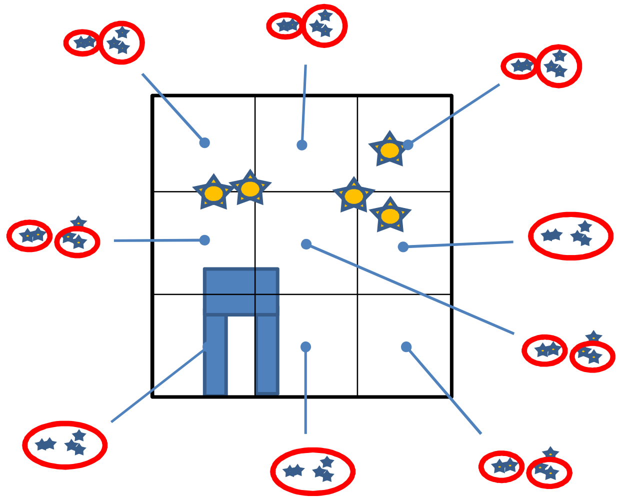


- First, lets have a look how naive adaptivity looks like. We have some samples from clusters and some cluster sampling probabilities
- And suppose we have gathered a new data point which happens to be an outlier
- If we estimate the sample means directly, our estimates can change abruptly and that will have a strong effect on further cluster sampling.
- (It might then take a long time to fix that decision. More probably it will cause an artifact in the final picture.)

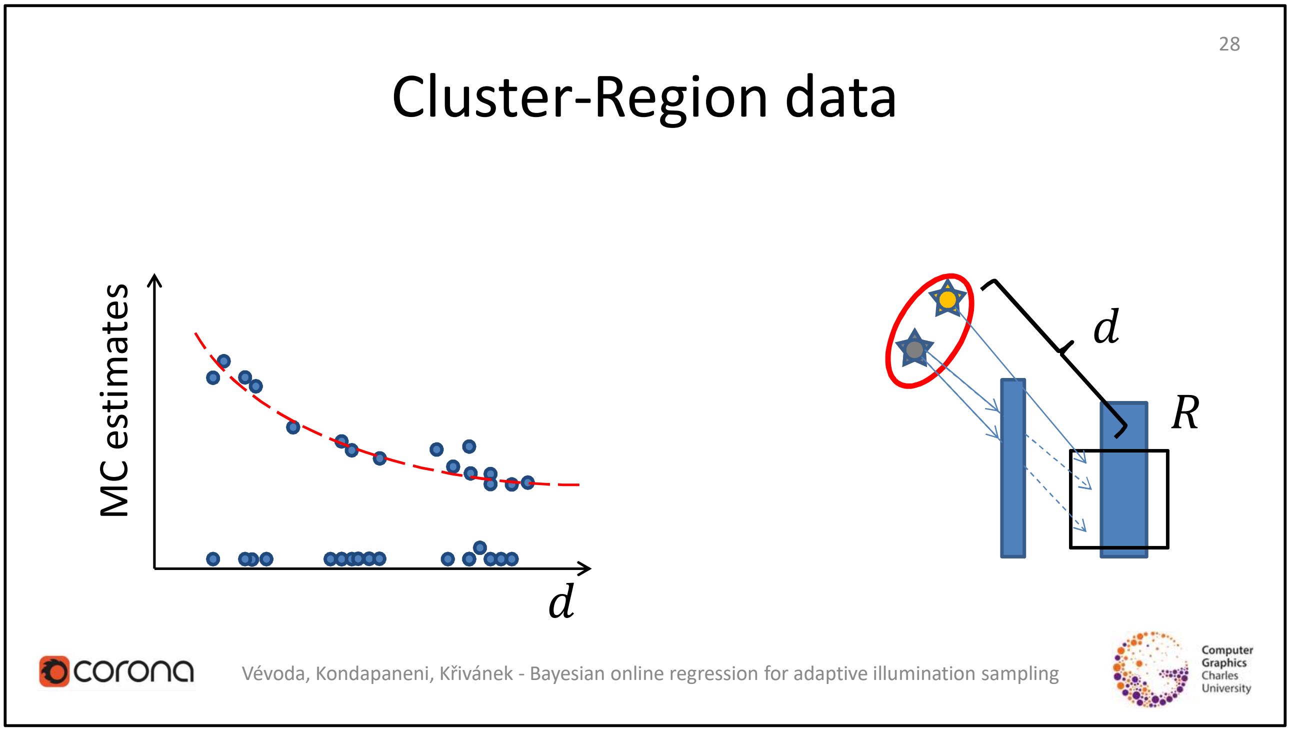


- If we estimate means in a Bayesian way, we model the **distributions** of MC estimates seen so far while we also have some prior information about parameters of that distribution,
- Therefore, when we get a new sample,
- Our distribution changes less abruptly as well as cluster sampling probabilities derived from it, which yields a more robust solution.

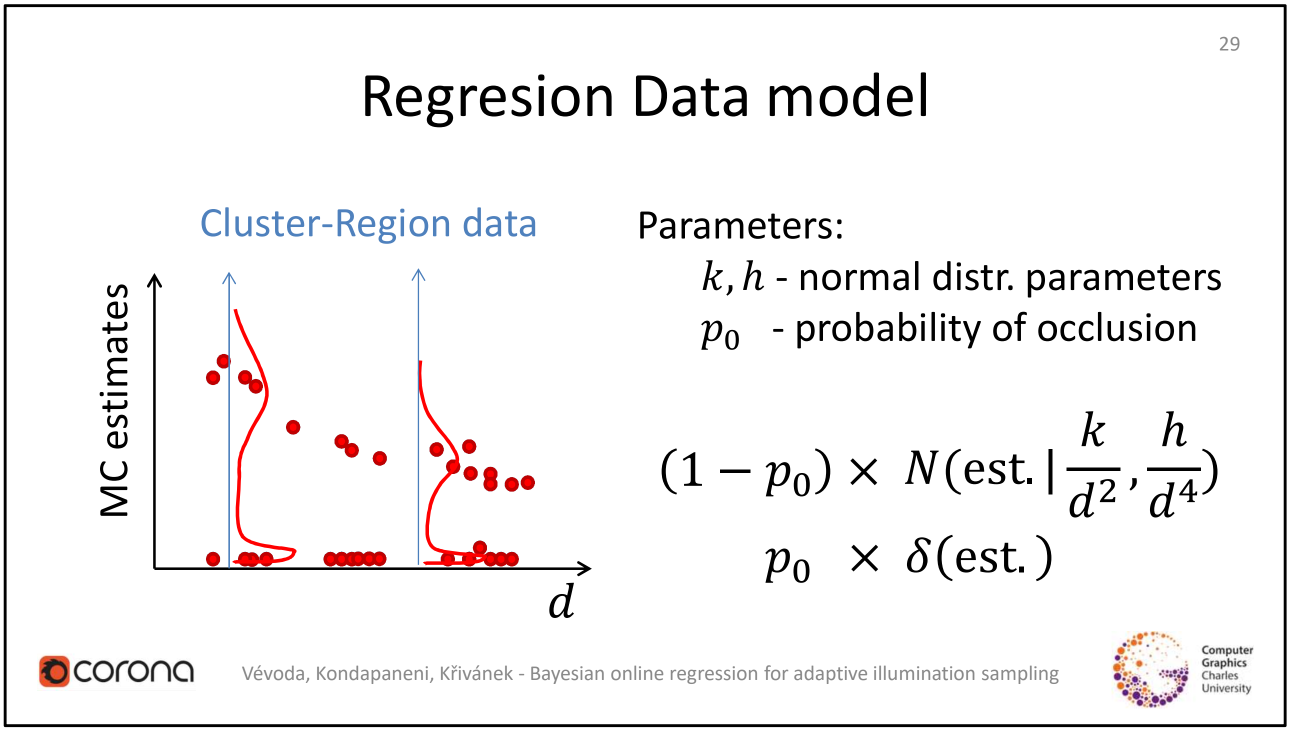
Cluster-region pairs



- Before we explain how we model the data we need to explain some basic context regarding clusters and scene subdivision.
- Contrary to the previous approaches, we split the scene into fixed **Regions**
- for each region we compute exactly one light clustering and keep it cached for that region. That speeds up clustering retrieval.



- Now we focus on samples collected from one cluster in one region in the scene.
- And we keep track of the distance *d* in the geometry factor in each estimate.
- We can then plot the data (i.e. MC estimates) for a cluster in a region which reveals relation of estimate values to distance. You may notice the inverse squared falloff with the distance and a number of zero-valued samples.



- Our model is therefore a parametric regression model, which for a distance yields a distribution of MC estimates. We design it as follows:

- Non-zero samples are modeled by a normal distribution with mean and variance being a function of a distance associated with samples. (This part has two parameters *k* and *h*.)

- The zero valued samples are incorporated by mixing the inverse-square distance falloff model with a delta function (and it is controled by p_0, which has a meaning of occlusion probability.)

- Now having designed our data model, we need to define prior for the model parameters

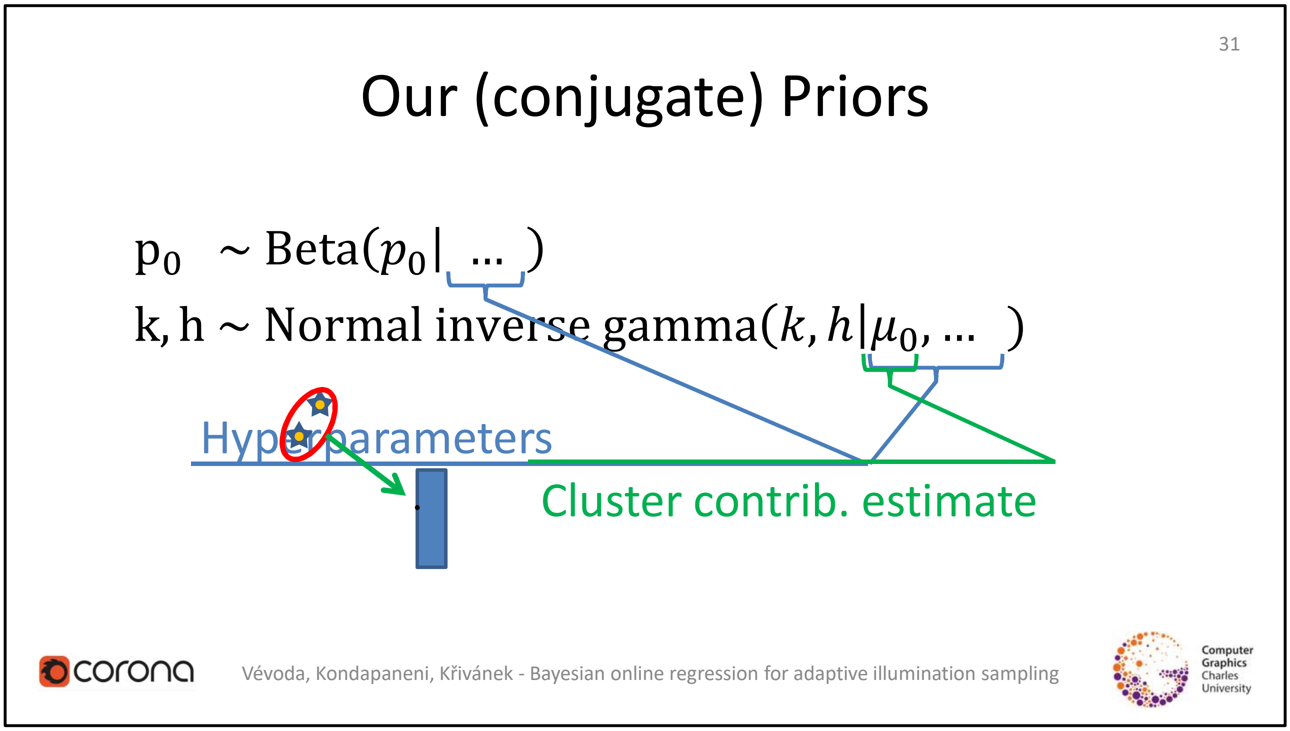
Conjugate prior

$$\underbrace{\color{red}{\mathbf{posterior}}} \propto \text{likelihood} \times \underbrace{\color{red}{\mathbf{prior}}}$$

Same functional form

- The model we have just defined has parameters k, h and p_0 and for them we use so called conjugate priors

- conjugate prior is such that when combined with likelihood it has the same functional form as the posterior.



- We proved that conjugate prior in our case is Beta distribution for p_0 and
- normal-inverse-gamma distribution for the parameters *k* and *h*.
- There are various hyperparameters in the equation, but one parameter which stands out is μ_0, for which we use the conservative cluster contribution bound

Summary

- Light preprocess (clustering)
- During each Next event estimation:
  – Obtain clustering (Cut) cached in a region
  – Compute distributions of estimates for each cluster in Cut
    -> mean, variance
  – Build distribution over clusters
  – Sample direct illumination
  – Record new data for sampled cluster

- To wrap it up, our algorithm is following:

Results

- Let us now demonstrate our solution in practice.

Tests

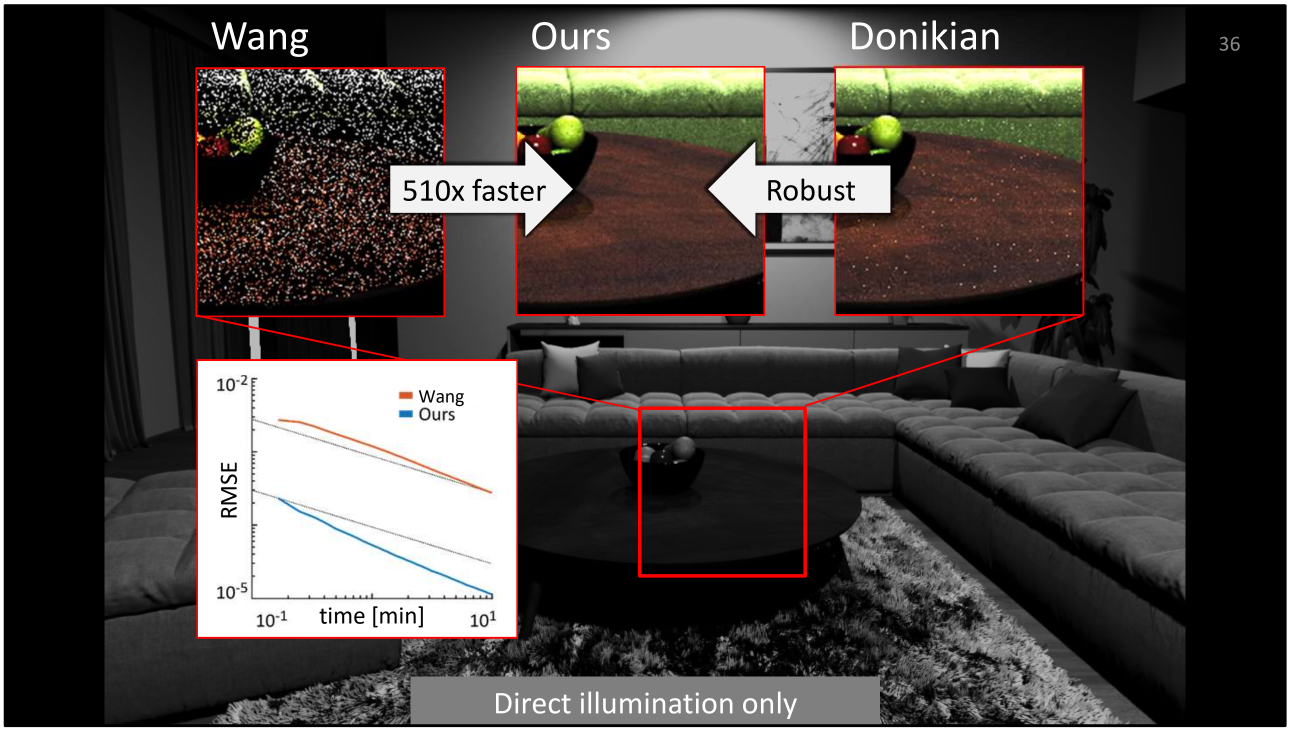
- Performance

	Direct only	Direct + indirect
Simple occlusion		
Complex occlusion		

- Grid resolution
- Temporal coherence

- This is a list of all comparisons you will see.
- <click>
- And I will start with performance testing in a scene with simple occlusion in direct illumination only setting.

- It is the living room scene from the beginning of our presentation.
- It is lit mostly by a few small area lights on the ceiling, only in the left part sunlight is coming through the windows.



- As you could already see, non-adaptive sampling of Wang at al. does not perform well in this scene.
- The sun is much stronger than the ceiling lights and is therefore sampled much more often even if it is actually occluded and so most of the samples are wasted.
- <click>
- Donikan’s algorithm improves the result significantly, as it quickly learns the sun occlusion.
- On the other hand, it struggles with the ceiling lights. They are covered by shades which block some of the samples. The method gives such samples too big weights, undersamples these lights and introduces spiky noise.
- <click>
- Our method also quickly learns the sun occlusion…
- <click>
- … and converges more than 500× faster than Wang.
- <click>
- We can even observe higher empirical convergence rate.
- <click>
- In the same time, thanks to the Bayesian treatment, our method is robust, does not get confused by the occluded samples and avoids the spiky noise.

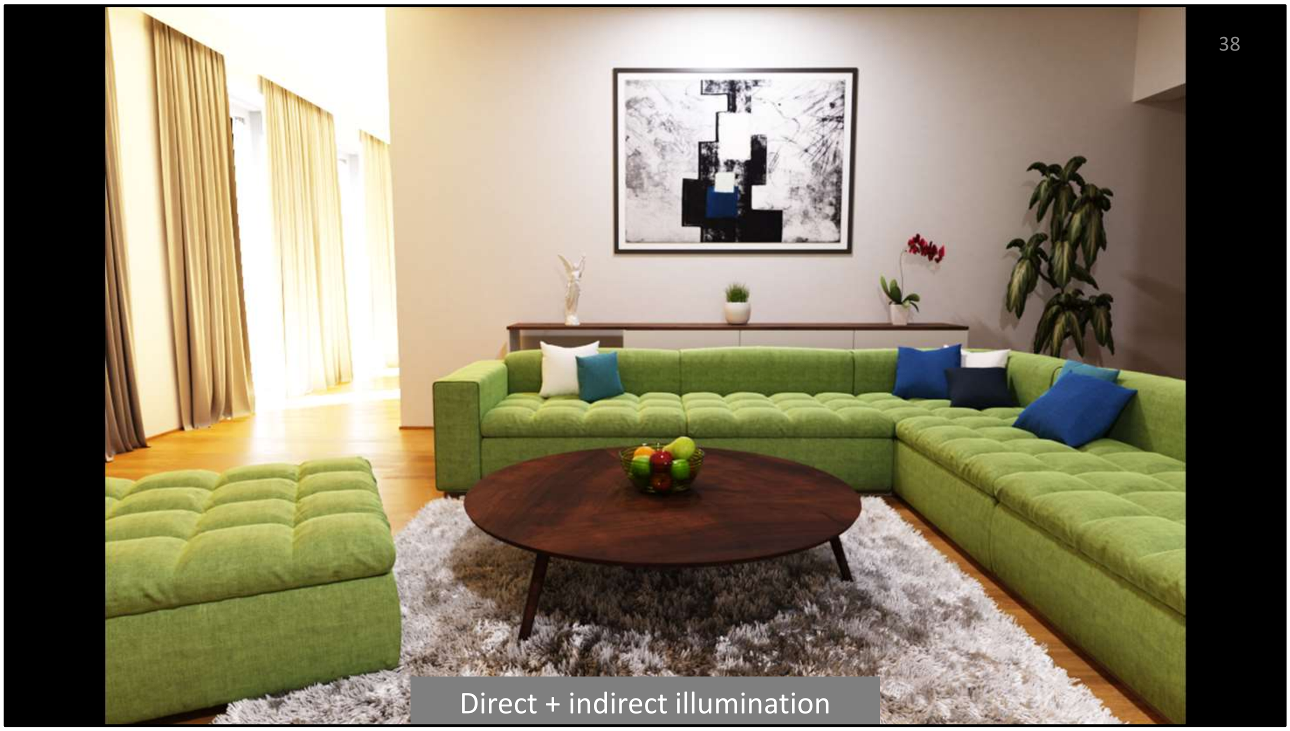
Tests

- Performance

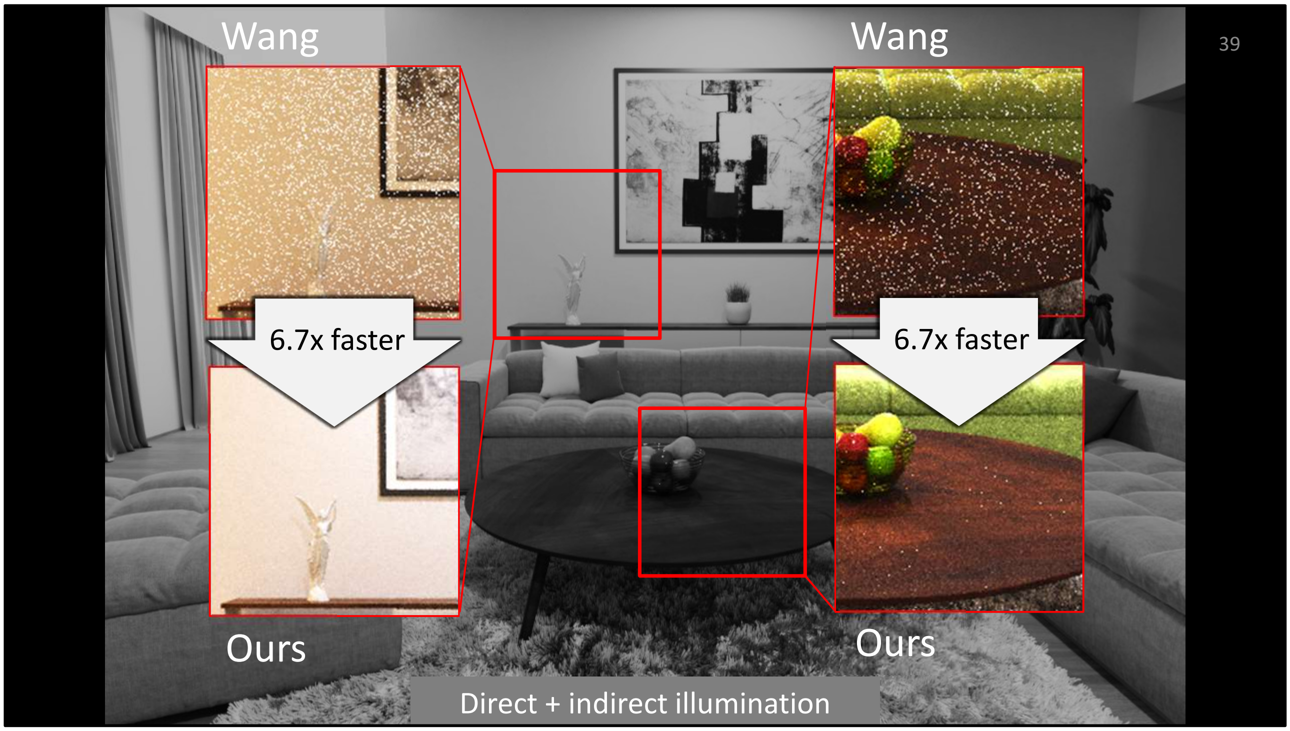
	Direct only	Direct + indirect
Simple occlusion	✓	
Complex occlusion		

- Grid resolution
- Temporal coherence

- So, that was the direct illumination.
- <click>
- However, in practice one is usually more interested in images containing both direct and indirect components….



- … like this one.



- We can see that the strong direct illumination noise of Wang dominates also in the complete image.
- The direct component is definitely the main source of noise in this scene.
- <click>
- By using our method in the next event estimation in path tracing we are able to improve the light sampling on every path vertex…
- <click>
- …and get more than 6x times speedup.
- Note that the remaining noise in the bottom right image is caused solely by the indirect component and cannot be influenced by our method.

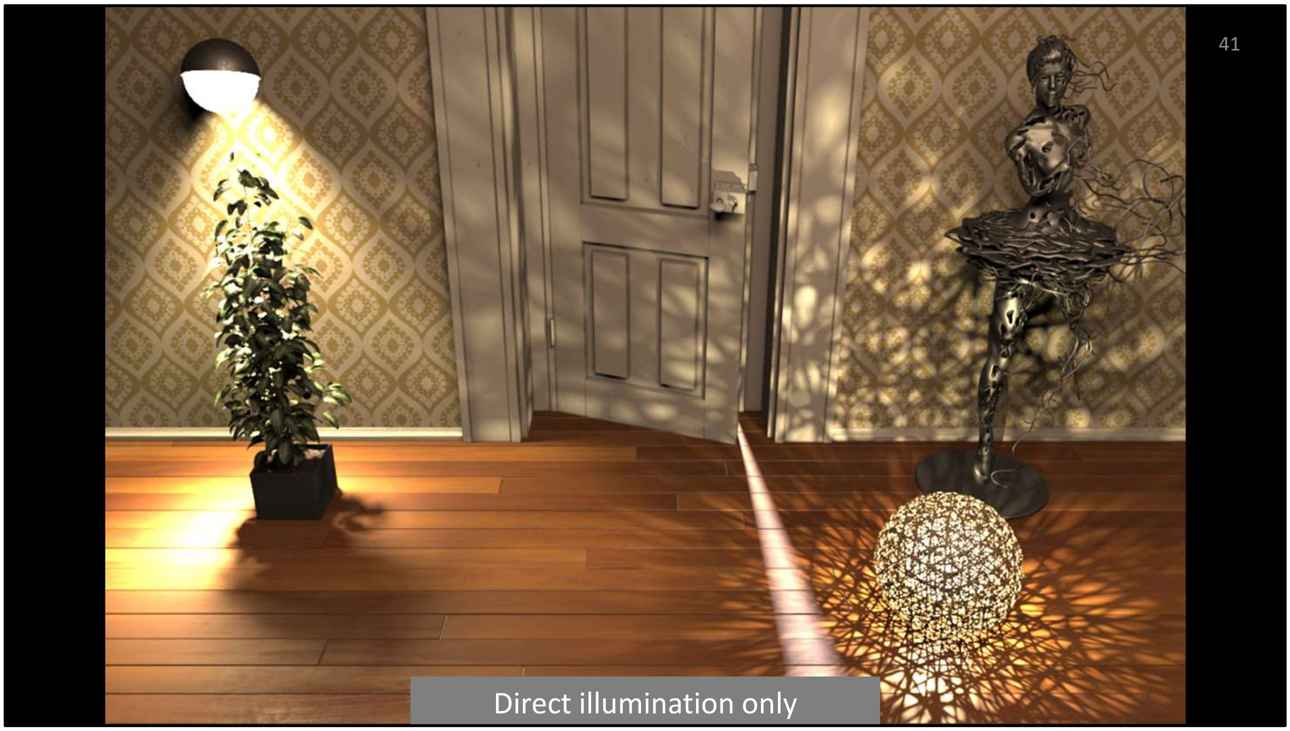
Tests

- Performance

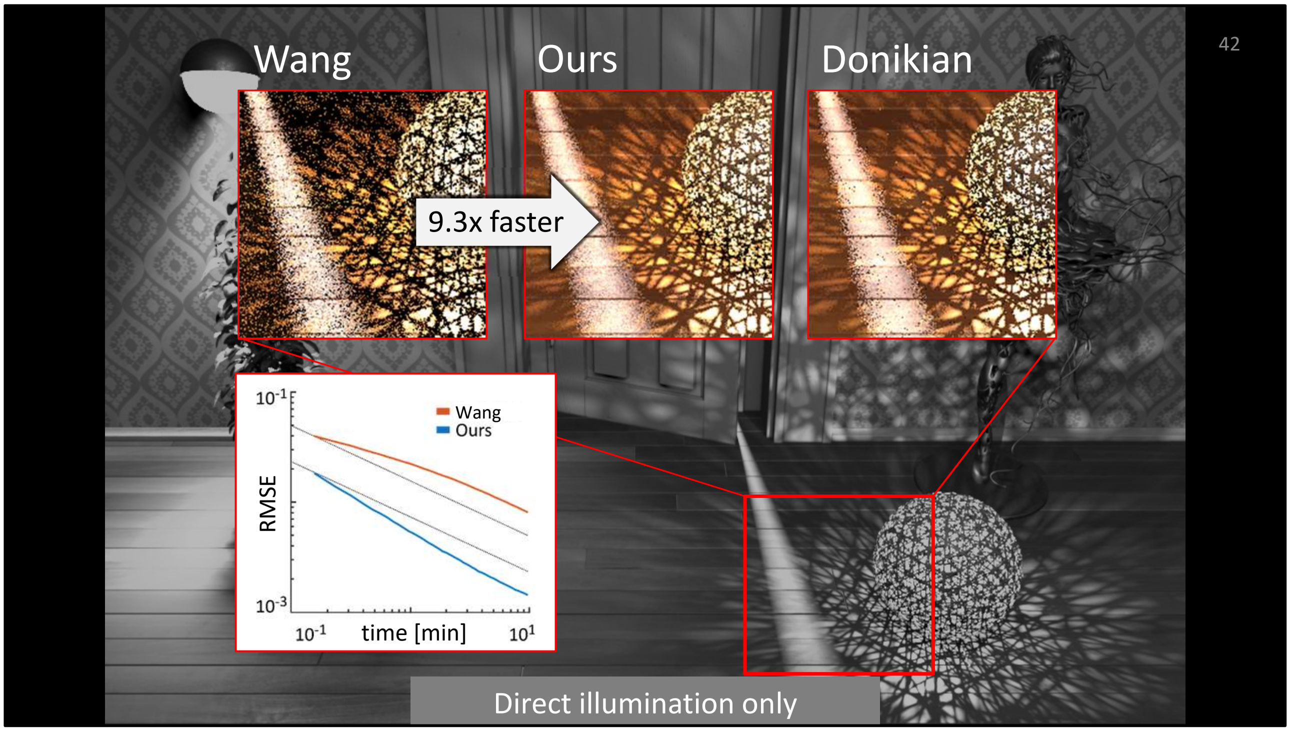
	Direct only	Direct + indirect
Simple occlusion	✓	✓
Complex occlusion		

- Grid resolution
- Temporal coherence

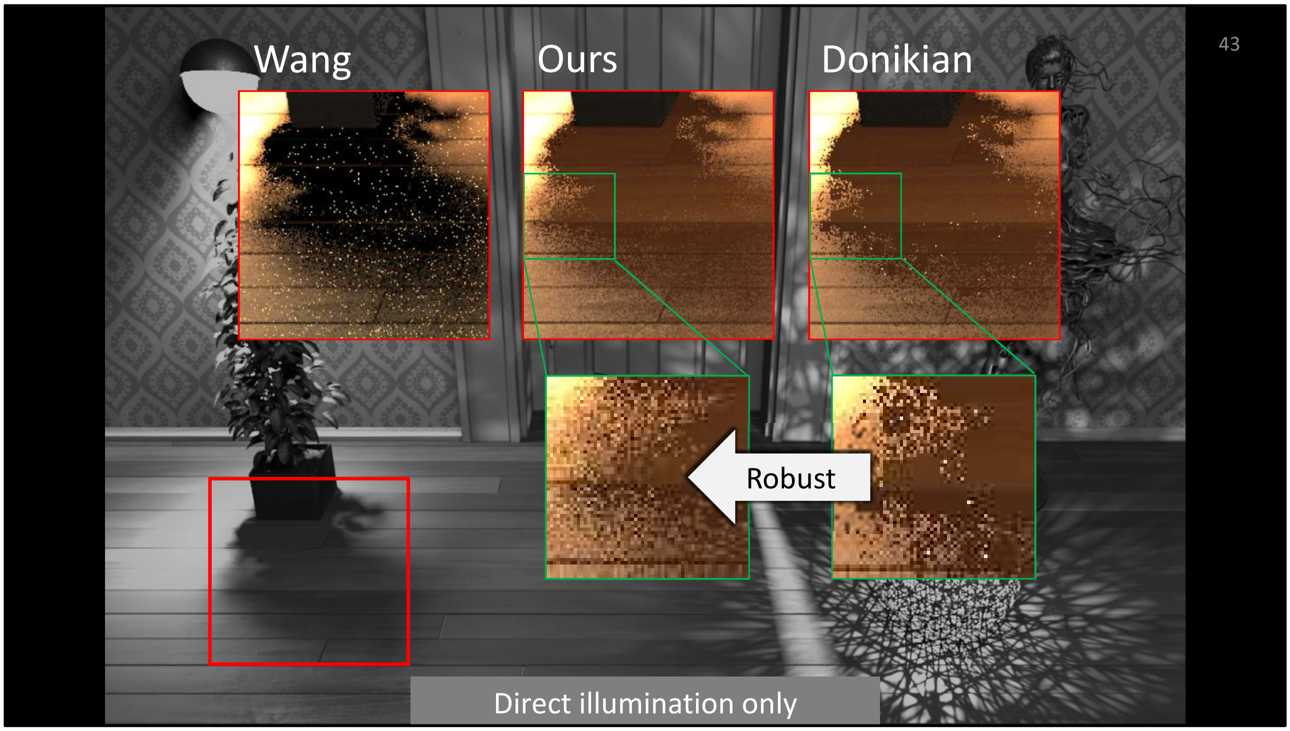
- <click>
- Now it is time to stress test our robustness in a scene with complex occlusion.



- This scene presents a real challenge due to its highly structured illumination plus there are lights in the other room behind the door.



- In this part…
- <click>
- … Wang's method produces a lot of noise again as it wastes samples on the lights behind the door.
- <click>
- On the other hand, our method performs great.
- <click>
- It is more than 9 times faster,
- <click>
- And again we can observe higher empirical convergence rate. And all that without introducing any artifacts in such a complicated illumination setting.
- <click>
- Donikian's method at first also seem to perform well but further inspection would discover small blocky artifacts in the shadows.



- However, in this part …
- <click>
- Wang does not perform well again, but this time also the Donikian’s method fails.
- <click>
- The illumination coming through the leaves of the plant is too complex for the ad-hoc learning to handle, the method overfits and produces square artifacts. This is exactly the problem of previous adaptive methods. They can provide substantial speedup but they don’t fail gracefully.
- <click>
- Bayesian learning makes our method much more robust and artifact-free.

Tests

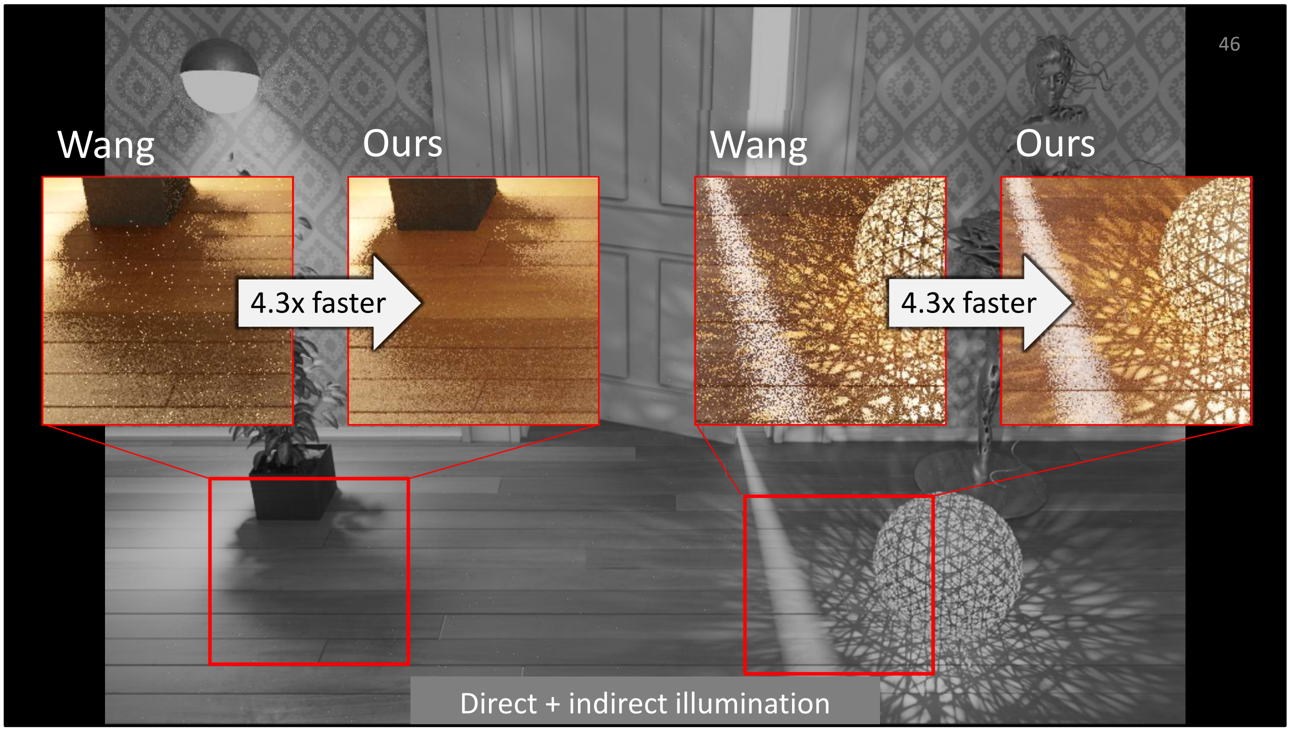
- Performance

	Direct only	Direct + indirect
Simple occlusion	✓	✓
Complex occlusion	✓	

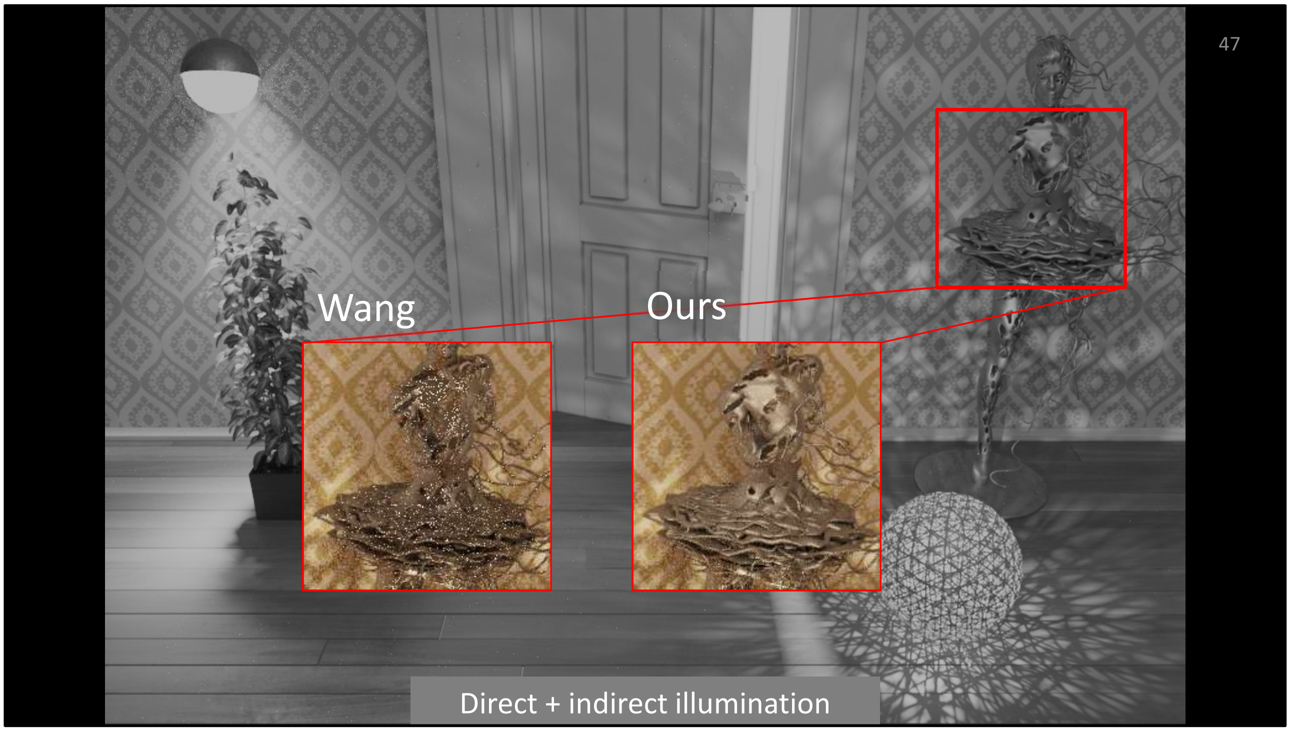
- Grid resolution
- Temporal coherence

- Finally, let's test the complex occlusion also with the indirect component.

- If we take a look at the same scene…



- …we will see the direct illumination noise dominates the complete image the same way as in the previous scene.
- <click>
- Our method eliminates it
- <click>
- and renders the complete image more than 4 times faster and without any artifacts.



- There is one more interesting place in this scene.
- <click>
- It is this statue.
- It is made of glossy metal and even though our method does not take the surface BRDF into account, it performs significantly better even there.

Tests

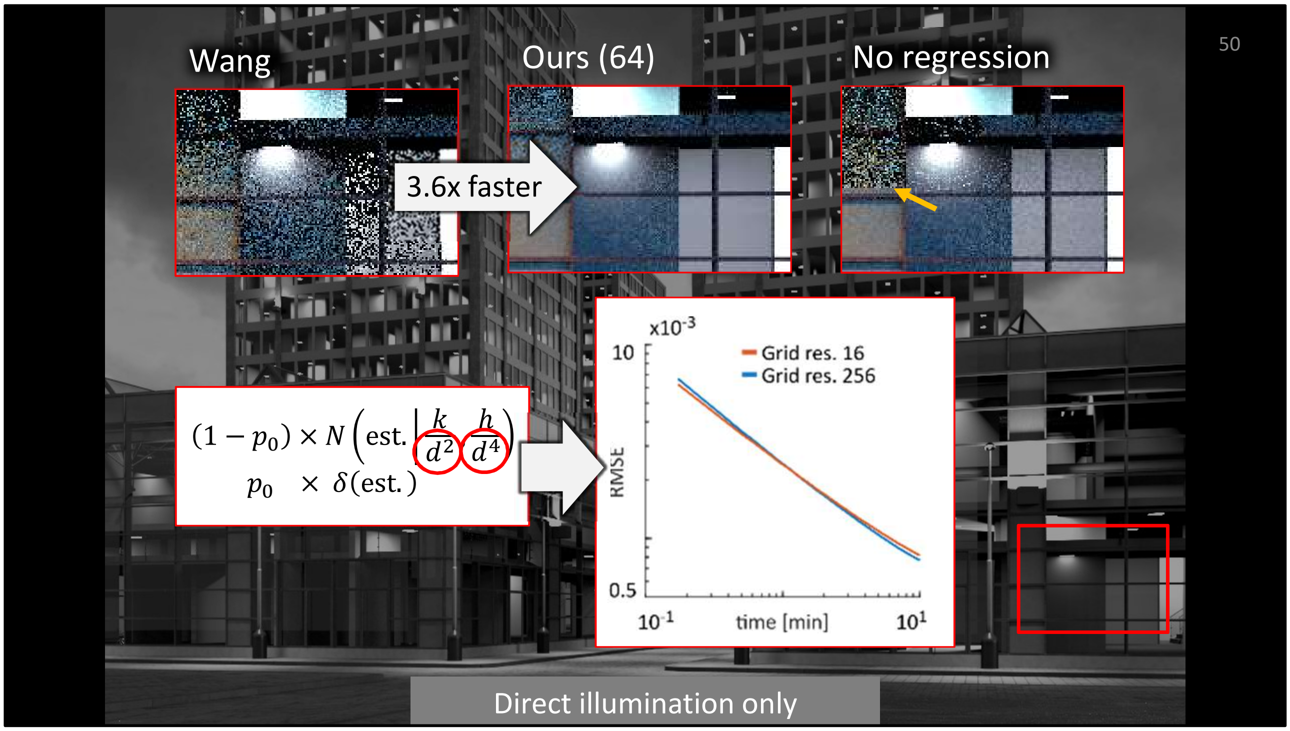
- Performance ✓

	Direct only	Direct + indirect
Simple occlusion	✓	✓
Complex occlusion	✓	✓

- Grid resolution
- Temporal coherence

- Since we divide a scene into regions by a uniform grid of a fixed resolution, we need to test how this resolution affects the performance.

- For that we will use this large scene containing a lot of lights.



- With our default choice of 64 regions per the shortest grid dimension
- <click>
- Our method performs more than 3 times faster than Wang.
- So what about other resolution?
- <click>
- The regression modeling of the distance falloff makes our method rather insensitive to the actual grid resolution.
- <click>
- And so even much smaller as well as much higher resolutions all perform roughly the same.
- <click>
- Without the regression we would have to use much higher resolution otherwise we would see sudden noise transitions between regions.

Tests

- Performance ✓

	Direct only	Direct + indirect
Simple occlusion	✓	✓
Complex occlusion	✓	✓

- Grid resolution ✓
- Temporal coherence

- Ok, so our method works well in all our static scenes.
- Klik
- Now to our final test. It will verify temporal coherence in animations.

- We now show a short animation of this scene.
- The only thing that will change is the position of the sun.
- It starts behind the scene and will go towards the camera.
- Notice that right now Wang works well on the back wall because the sun is behind it, so it is culled and no samples are wasted.
- <fwd animation>
- The sun is moving towards us
- <stop animation>
- And now the sun already starts to contribute to the back wall and so performance of Wang will decrease.
- On the other hand, our method works well all the time without any artifacts or blinking.
- <play the rest>

Conclusion

Future work

- BRDF incorporation
- Adaptive scene subdivision
- Rigorous hyperparameters derivation
- Combination with path guiding [*Vorba et al. 2014, Muller et al. 2017*]

- While our method performs well not only in our test scenes but also in a production renderer as there are several aspects we didn't address which could be the possible topics for a future work.
- We could incorporate the BRDF in learning, use adaptive scene subdivision instead of our grid and more rigorously derive values for our hyperparameters.
- Finally, our method improves mainly the direct illumination not the indirect component. Combination with path guiding which has exactly the complementary focus could be interesting.

55

Contribution

- **Bayesian framework for robust adaptivity**
- Optimal cluster sampling
- Algorithm for direct illumination
  - Unbiased, adaptive, robust
  - Easy to integrate into a path tracer

corona Vévoda, Kondapaneni, Křivánek - Bayesian online regression for adaptive illumination sampling Computer Graphics Charles University

- The main contribution of our work is creating a Baysian framework for adaptive Monte Carlo quadrature.
- It enables exploiting the big potential of the adaptive approach while avoiding the biggest weakness of previous attempts – the lack of robustness.

- We applied this framework on the problem of direct illumination sampling.

- In the process we derived the optimal sampling of clusters and developed an unbiased adaptive direct illumination algorithm with online learning of light sampling distributions. It is easily integrable into a path tracer and suitable for interactive rendering.

- Our new framework is not limited to the direct illumination though and we are certain that other applications of adaptive sampling will benefit from it as well and it opens the path for many other tools of statistical machine learning (such as full Bayes or variational Bayes).

Acknowledgments

- Ludvík Koutný (a.k.a. rawalanche)
- Charles University Grant Agency project GAUK 1172416, by the grant SVV-2017-260452
- Czech Science Foundation grant 16-18964S

corona

Computer Graphics Charles University

- Finally, We would like to thank Ludvík Koutný for modeling the test scenes and Charles University and Czech Science Foundation for supporting our work.

And that’s all. Thank you for your attention.