Optimal Multiple Importance Sampling

Ivo Kondapaneni
Charles University, Prague

Petr Vévoda
Charles University, Prague
Chaos Czech a.s.

Pascal Grittmann
Saarland University

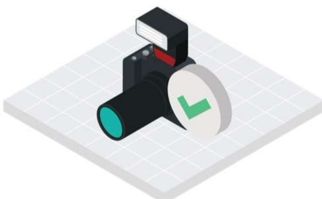
Tomáš Skřivan
IST Austria

Philipp Slusallek
Saarland University
DFKI

Jaroslav Křivánek
Charles University, Prague
Chaos Czech a.s.

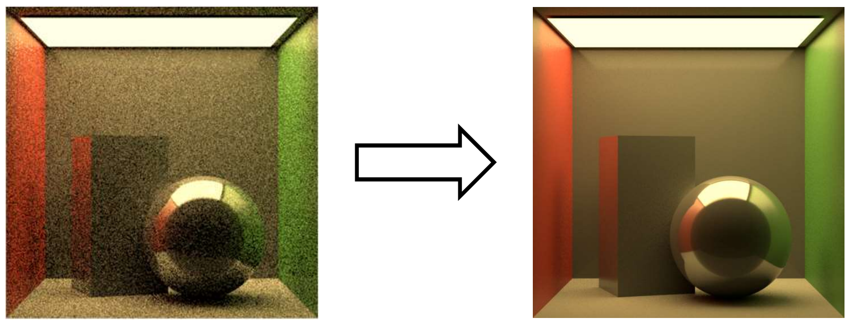
Computer Graphics Charles University
UNIVERSITÄT DES SAARLANDES
DFKI
thrive SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

Welcome to our presentation about optimal multiple importance sampling



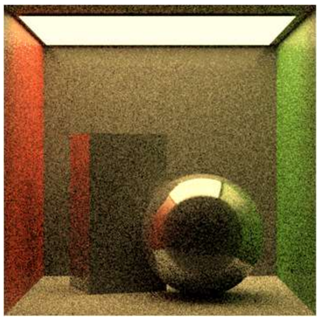
PHOTOGRAPHY & RECORDING ENCOURAGED

Goal: Low noise in short time



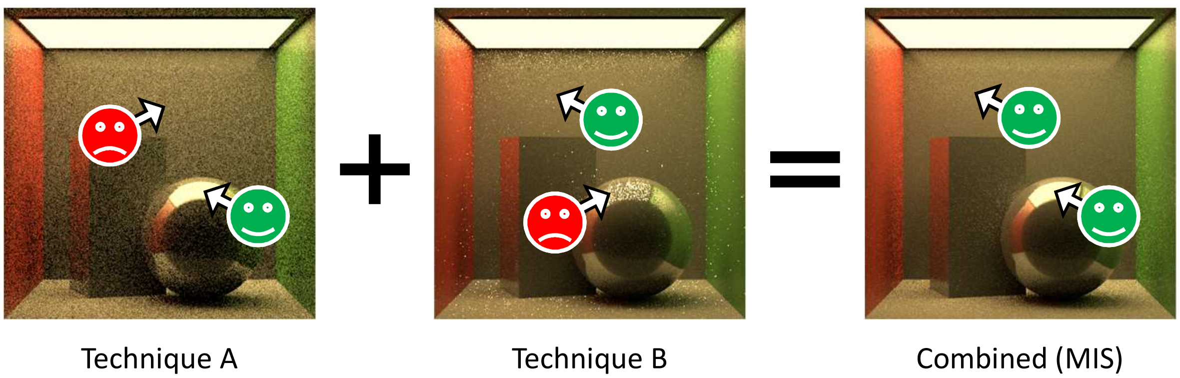
- In Monte Carlo rendering, images contain noise.
- Eventually, given enough time, that noise will go away.
- Our goal is to minimize the time it takes to arrive at a converged image.

Goal: Low noise in short time



- In Monte Carlo rendering, images contain noise.
- Eventually, given enough time, that noise will go away.
- Our goal is to minimize the time it takes to arrive at a converged image.

MIS: Two wrongs can make a right



- There are different algorithms for rendering images.
- Each has benefits and disadvantages.

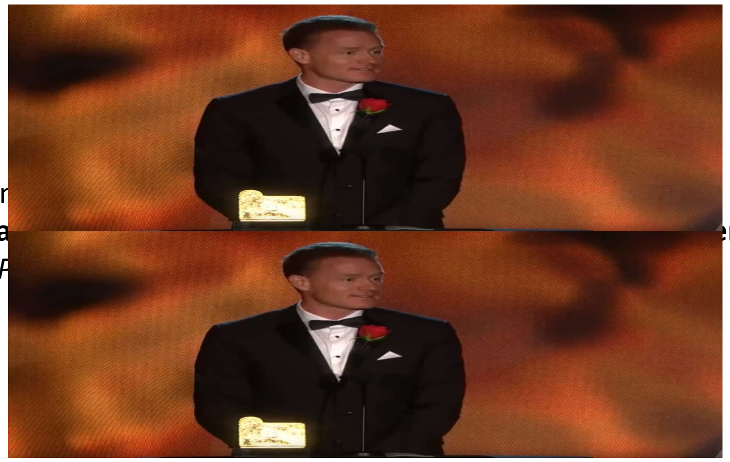
- <click> In the example on the left, the direct illumination on the wall is handled poorly <click> while in the example on the right it is captured nicely

- The glossy reflection on the ball is the exact opposite : <click> it is captured nicely on the left example, <click> but it is poor on the right.

- <click> Fortunately, MIS can be used to combine both, achieving a nice image overall.

MIS: Essential for the path tracing revolution

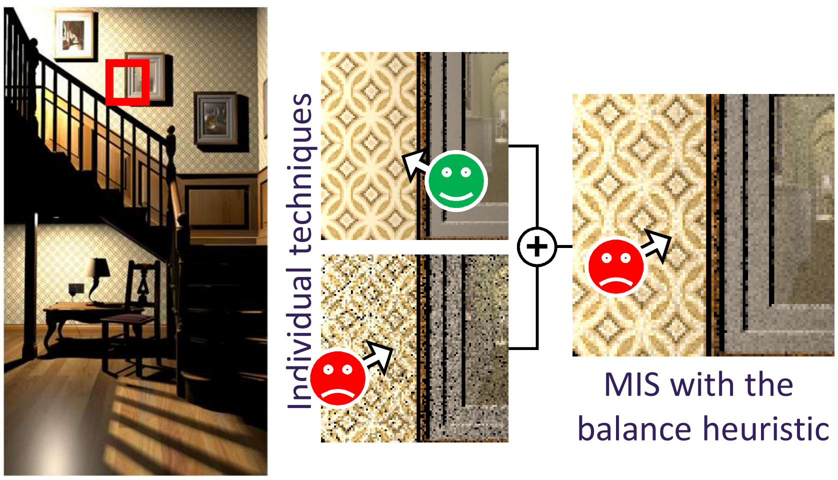
Veach ar
“**Optima** **endering**.”
SIGGRAF



6 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- MIS was invented in 1995 by Veach and Guibas
- It had such a tremendous impact that Eric Veach was awarded the Scientific and Engineering Award in 2014.

MIS: Increases robustness, may reduce efficiency



Can we do better?

- But MIS is not perfect.
- In this example, we combine two techniques to render the scene on the left.
- <click> One is almost perfect on the wall above the stairs, the other performs poorly there.
- <click> Unfortunately, MIS with the balance heuristic keeps some of the error of the worse technique.
- <click> So the question is: can we somehow do better than that?

- And indeed, we can.
- We derived the optimal MIS weights.
- In this case, they keep the lower error of the almost perfect technique.
- and the image in overall has a ten times lower level of noise.

Contributions

- A: Balance heuristic’s variance bounds reconsideration
- B: Optimal weights derivation
- C: Control variate interpretation
- D: Practical implementation

- (Entry) The complete list of our contributions starts by
- showing that the balance heuristic is further from the optimum than believed so far.
- Deriving the new optimal weights that provably minimize an MIS estimator variance.
- Proving that the optimal weights are equivalent to a control variate.
- And last, we show that the optimal weights are not a mere theoretical construct, they lend themselves to a practical implementation in light transport.

Background and Related Work

Computer Graphics Charles University

UNIVERSITÄT DES SAARLANDES

thrive SIGGRAPH2019 LOS ANGELES • 28 JULY - 1 AUGUST

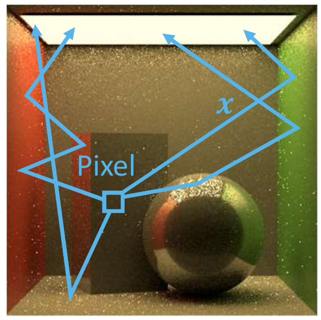
- Before explaining you our approach, let me give you some high-level background on MIS.

Light transport simulation

Path $\boldsymbol{x}$

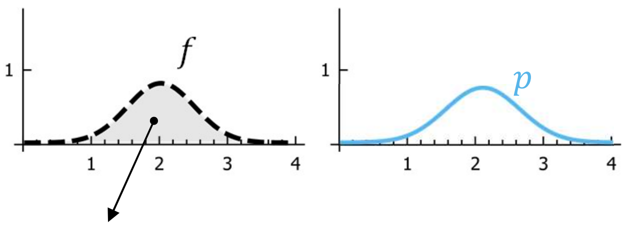
$$I_{\text{pixel}} = \int f(\boldsymbol{x})\, d\boldsymbol{x}$$

Contribution



- When we perform light transport simulation,
- <click> we have to find a light path incident with a pixel,
- <click> take its contribution
- <click> and integrate over all such paths. To estimate that integral efficiently by Monte carlo, we use the importance sampling approach.

Importance sampling

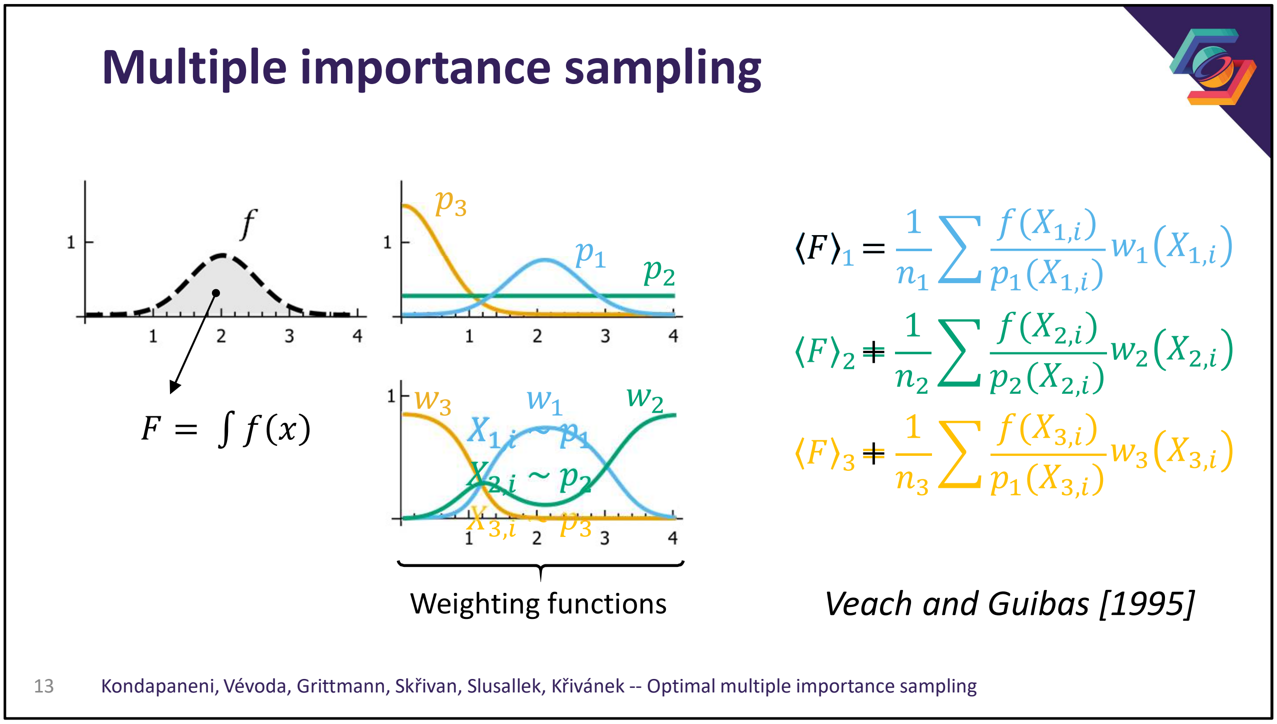


$$F = \int f(\boldsymbol{x})\, d\boldsymbol{x}$$

$$X_i \sim p$$

$$\langle F \rangle = \frac{1}{n} \sum \frac{f(X_i)}{p(X_i)}$$

- (Entry) Consider a simple integration problem, where we integrate a function f
- <click>To estimate its integral we use an importance sampling technique p and we draw samples according to it
- <click>Finally, we obtain an estimate of the integral by combining all samples. The better we sample important parts of the integrand the lower the variance of the estimator is. But what if we cannot find a single technique that would sample f well?



- (Entry) Then Multiple importance sampling can help us.
- <click> There might be another technique suited for sampling a different part of f.
- <click> It generates another set of samples and <click> defines another estimator.
- <click> And a third technique, <click> with its samples and <click> estimator.
- <click> Now we introduce a set of weighting functions, one weighting function for each sampling technique.
- <click> Finally, we form an MIS estimator by re-weighting the contributions of individual estimators by means of weighting functions. That will yield a more robust MIS estimator.

Previous attempts to improve MIS

- Sample allocation
  - *Pajot et al. [2011]*
  - *Lu et al [2013]*
  - *Havran and Sbert [2014], Sbert et al. [2016], Sbert and Havran [2017], …*
- Weighting functions
  - *Georgiev et al. [2012]*
  - *Elvira et al. [2015; 2016]*

$$\langle F \rangle = \frac{1}{n_1}\sum \frac{f(X_{1,i})}{p_1(X_{1,i})} w_1(X_{1,i}) + \frac{1}{n_2}\sum \frac{f(X_{2,i})}{p_2(X_{2,i})} w_2(X_{2,i}) + \frac{1}{n_3}\sum \frac{f(X_{3,i})}{p_1(X_{3,i})} w_3(X_{3,i})$$

- The most of research effort in improving MIS falls into two categories:
- <click> One category represents optimal sample allocations among techniques, because fixed sample allocations have its shortcomings.
- <click> In the other category, people were investigating weighting functions which would improve upon the existing ones.
- But one set of weights was always considered as almost optimal – the balance heuristic.

Balance heuristic

- Simple weighting functions $$w_i(x) = \frac{n_i p_i(x)}{\sum n_k p_k(x)}$$
- Close to optimal
  - ~~tight variance bounds by *Veach and Guibas [1995]*~~
  - ~~no other strategy can do much better~~

We show that it does not hold!

- ⇒ A de facto universal solution

But what about our 10x speedup?

- The balance heuristic weights are easy to compute.
- <click> They are proportional to the sampling technique times the number of samples taken from it.
- <click> So after normalization we get this simple formula.
- At the same time, these weights are close to optimal. According to ~~the tight variance bounds derived by~~ Veach and Guibas no other weights can achieve much lower variance.
- Therefore it has been used as a de facto universal solution.
- <click> But how could we get the 10-times speedup we showed you in the beginning?
- <click> Because the variance bounds do not hold!

Exhibit A
Variance bounds do not hold

Computer Graphics Charles University

UNIVERSITÄT DES SAARLANDES

thrive SIGGRAPH2019 LOS ANGELES • 28 JULY – 1 AUGUST

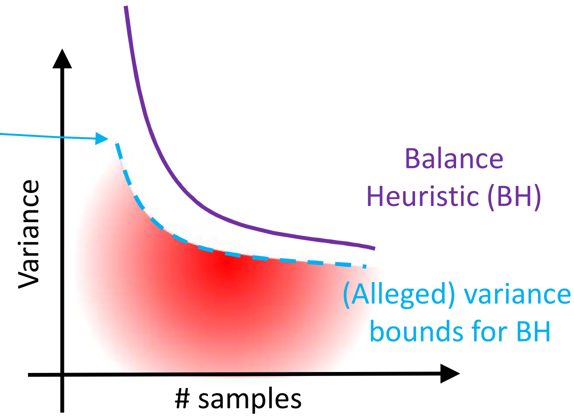
Realizing this is our first contribution. Let’s take a look on it.

Variance bounds derivation

Step 1: Find w_i that minimize $M2$

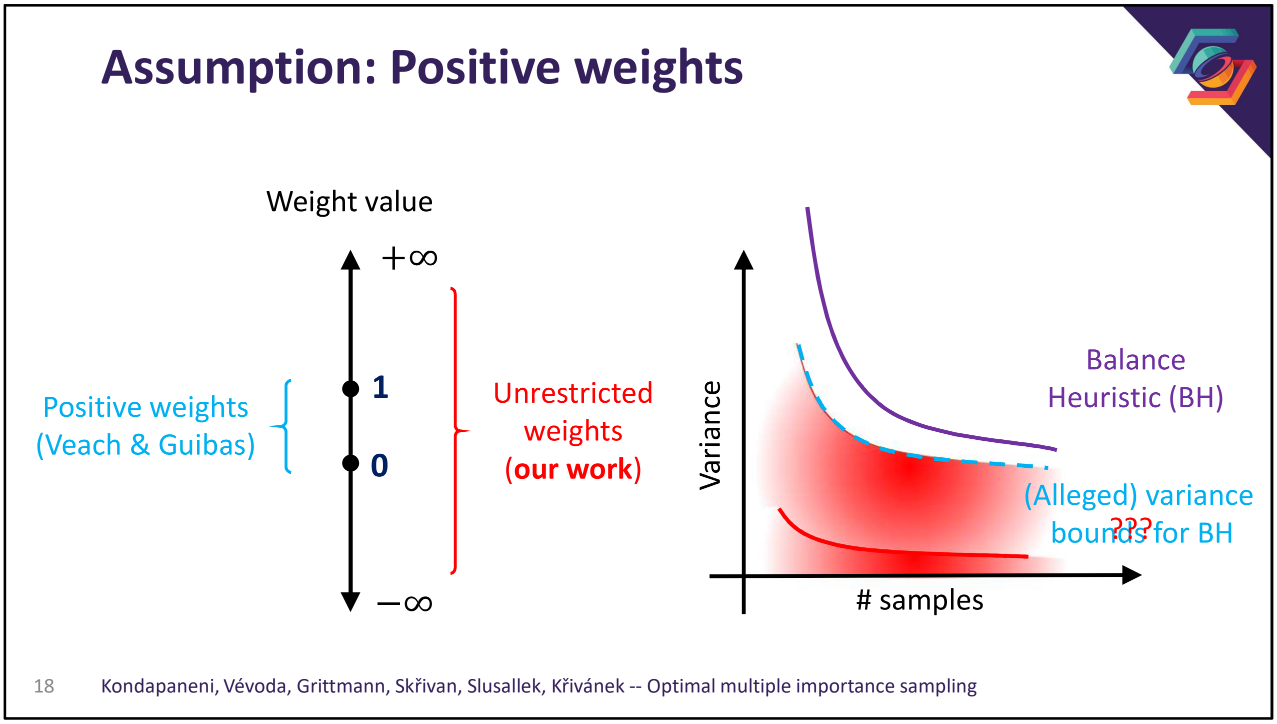
- $\mathrm{V}[\langle F \rangle] = \mathrm{M2} - \mathrm{M1}^2$

Step 2: Find upper bound



Veach and Guibas [1995]

- When deriving the bounds Veach and Guibas considered the variance of an MIS estimator split into two terms.
- <click> By minimizing the first one the balance heuristic is obtained. We plot the variance of the resulting estimator vs. total number of samples on the right.
- <click> Then the authors bounded the second term from above and got the assumed variance bounds.
- <click> No alternative weighting functions can have variance in the red area.



- However, we investigated their derivation further and realized that
- <click> they assumed only positive weights restricted to the interval 0 and 1.
- <click> But the MIS framework allows for weights which are not restricted! This simple fact has not been recognized up until now,
- <click> Removing the restriction of positive weights invalidates Veach's bounds`.
- <click> and it opens up to possibility that the truly optimal MIS weights have much lower variance.

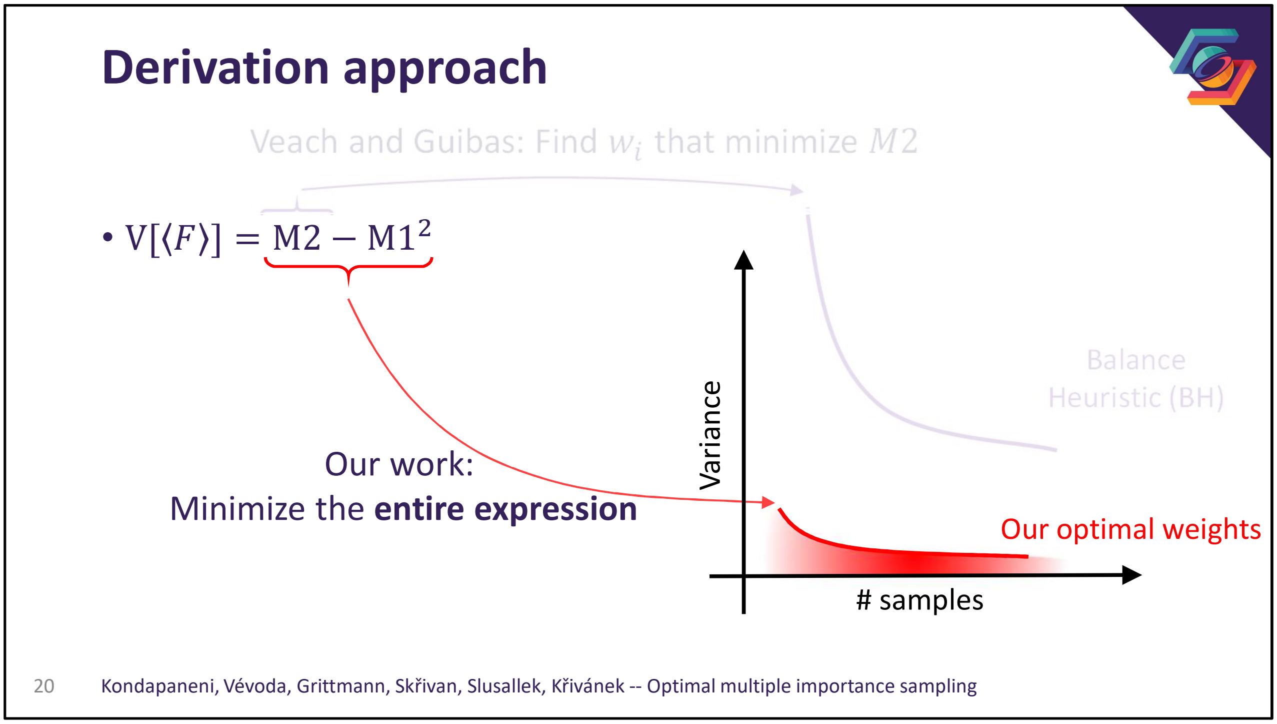
Exhibit B
Optimal MIS weights

Computer Graphics Charles University

UNIVERSITÄT DES SAARLANDES

thrive SIGGRAPH2019 LOS ANGELES • 28 JULY – 1 AUGUST

Now we know that the optimal solution can be much better than the bounds suggest. But how can we compute it?



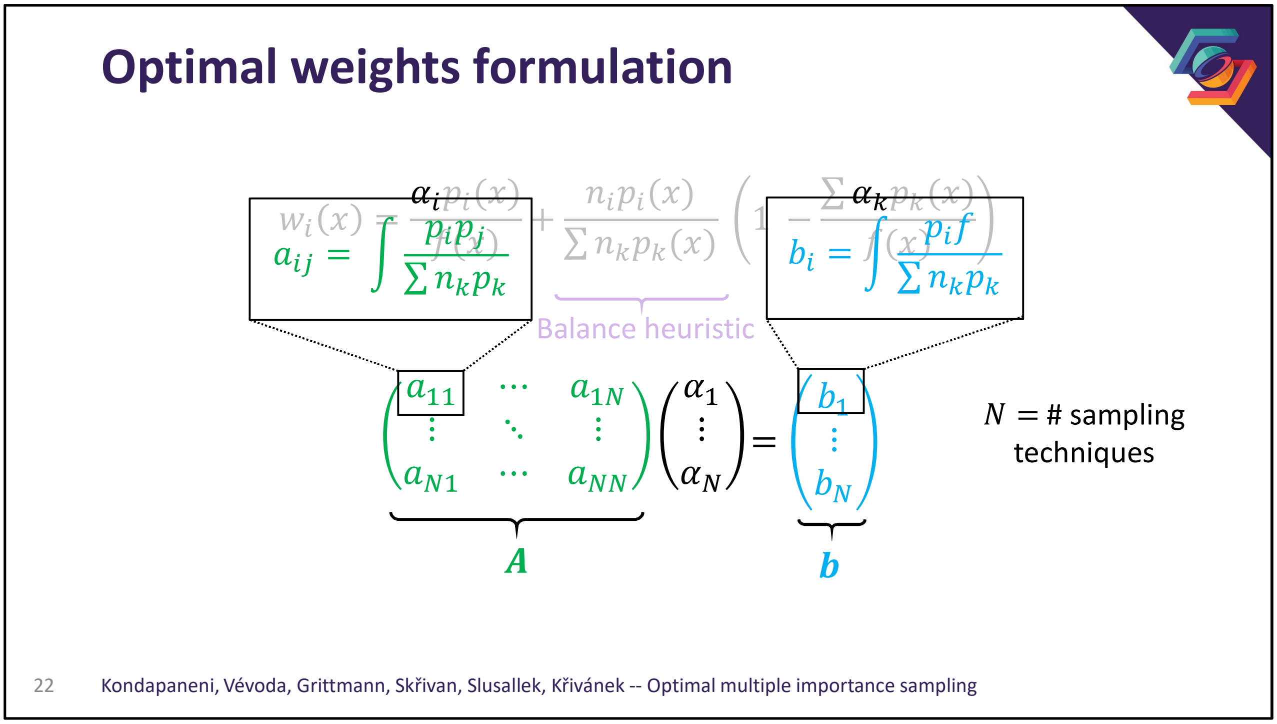
- <click>
- Our starting point is again the MIS variance formula.
- <click> But instead of minimizing just the first part, we apply calculus of variations to minimize everything in terms of weighting functions.
- <click> That gives us provably optimal weights.

Optimal weights formulation

$$w_i(x) = \frac{\alpha_i p_i(x)}{f(x)} + \underbrace{\frac{n_i p_i(x)}{\sum n_k p_k(x)}}_{\text{Balance heuristic}} \left(1 - \frac{\sum \alpha_k p_k(x)}{f(x)}\right)$$

21 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- (Entry) The optimal weights then have the following form
- <click> Note that they include the balance heuristic as part of their formulation. Also note that they include the integrand f itself in denominators, which is very uncommon among combination strategies widely used.

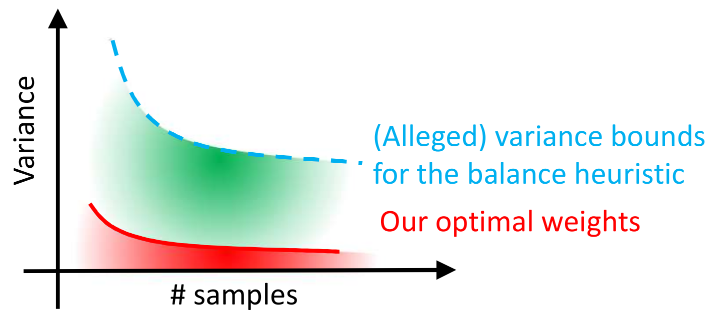


- (Entry) The formula also contains additional coefficients, which we denote alpha.
- <click> There is as many of these coefficients as there is sampling techniques. And how do we compute these alphas?
- <click> These are the solution to the following linear system represented by a matrix A and a vector b, where the matrix A's size is N-by-N and the vector b is a column vector of length N.
- <click> The individual elements forming the matrix A resemble projections of sampling techniques onto themselves, and elements of the vector b resemble projection of f into a system of sampling techniques.

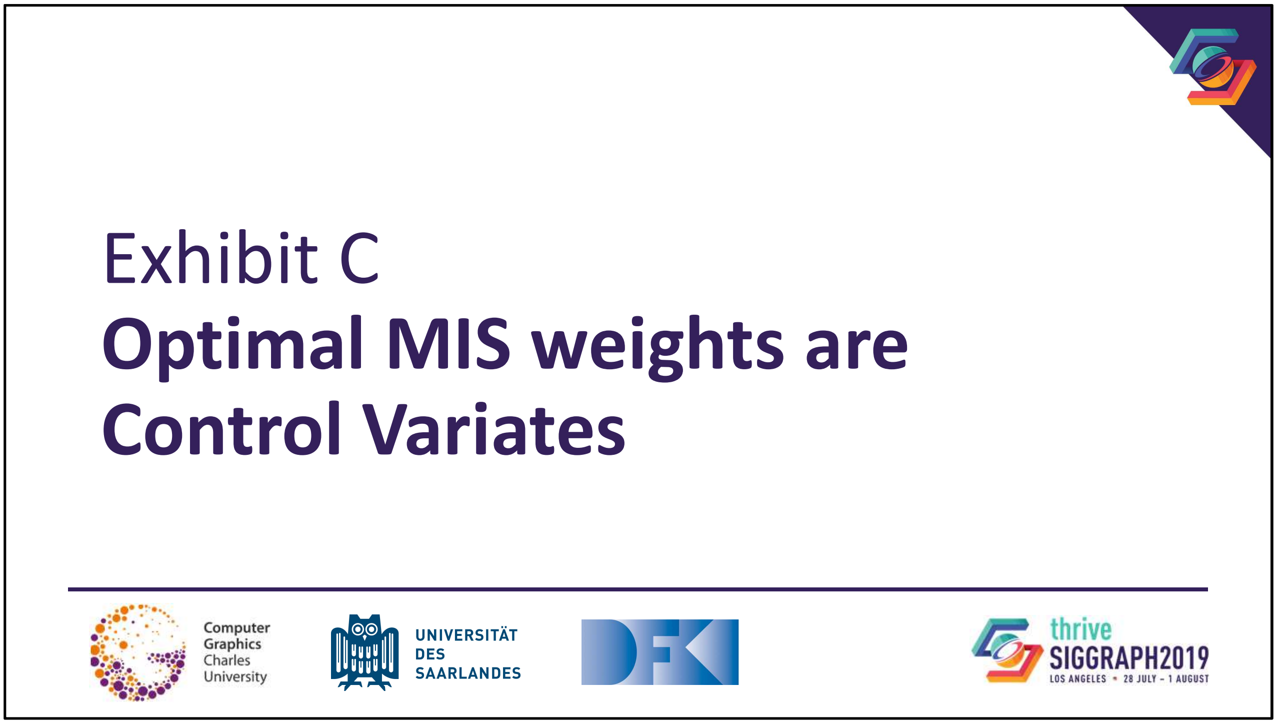
Summary

No weights can be better

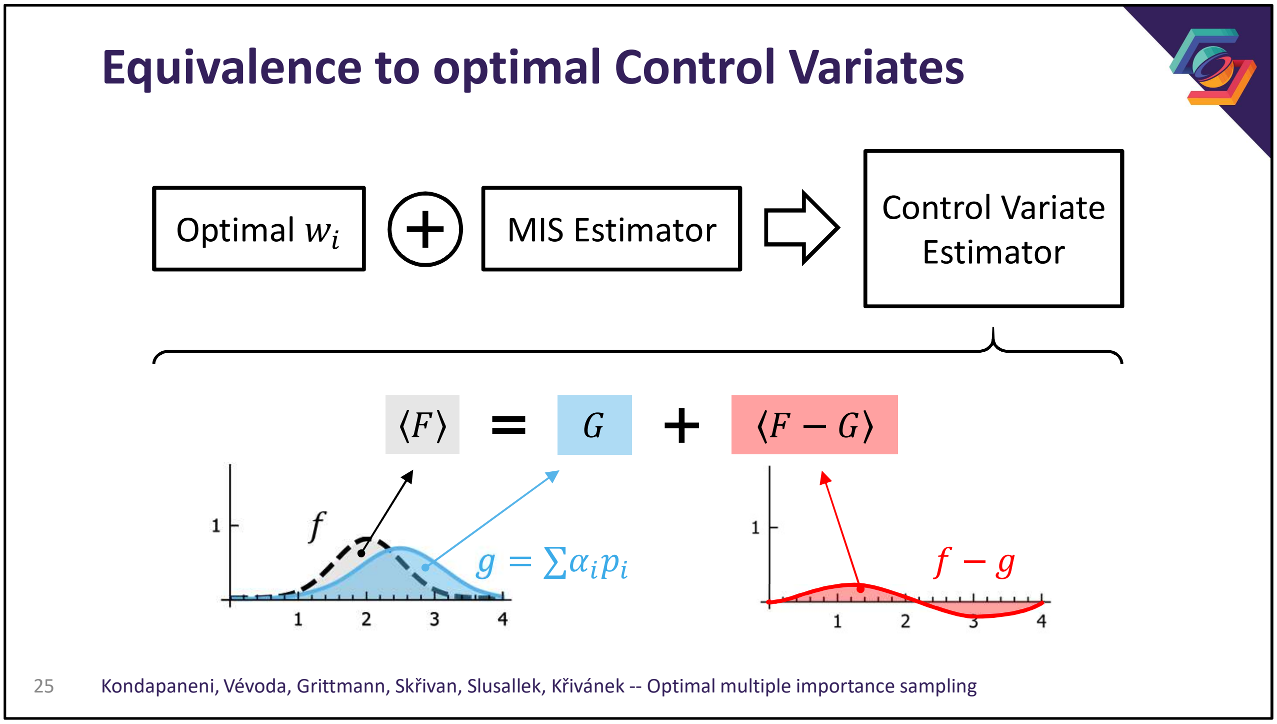
(for the same sample allocation)



- To summarize: we obtained a closed form formula for provably optimal MIS weights, and no weights can be better in terms of variance for the same sample allocation.



Now we show a relationship between the optimal MIS weights and optimal control variates.

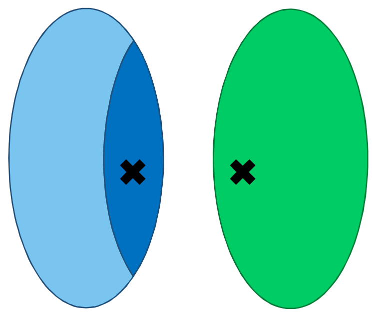


- (Entry) If we take the formula for our optimal weights ..
- <click> and plug it into the formula for an MIS estimator ..
- <click> the resulting estimator is equivalent to a control variate.
- <click> This means that we estimate the integral of f indirectly, based on a known integral of a function g <click> to which we add an estimated integral of the difference between f and g <click>
- <click> In our case the function *g* is a linear combination of the sampling techniques where the coefficients are the alphas.

Equivalence to optimal Control Variates

Owen and Zhou [2000]

- **Setup**
  - MIS + control variate
  - Form $g = \sum \beta_i p_i$
- **Result**
  - Optimal coeff. β_i



Our approach

- **Setup**
  - MIS framework
- **Result**
  - Optimal MIS weights

$$\beta_i = \alpha_i$$

- Control variates of this form have been studied before, the most related work is by Owen and Zhou.
- They assumed the Balance heuristic <click> used together with Control variate. Then they limited themselves to CV formed <click> as some linear combination of the sampling techniques.
- Then they found the optimal coefficients in this space <click>. ~~of the linear combination which are our alphas.~~
- On the other hand <click>, we took the MIS framework and without any further assumptions we found the provably optimal MIS weights <click>.
- Moreover, <click> we show that all CVs of the linear combination form are equivalent to some MIS weights and that the optimal solutions to both problems are the same.

- Knowing the optimal solution in theory is one thing. Now, we show that the weights can be actually used in practice.

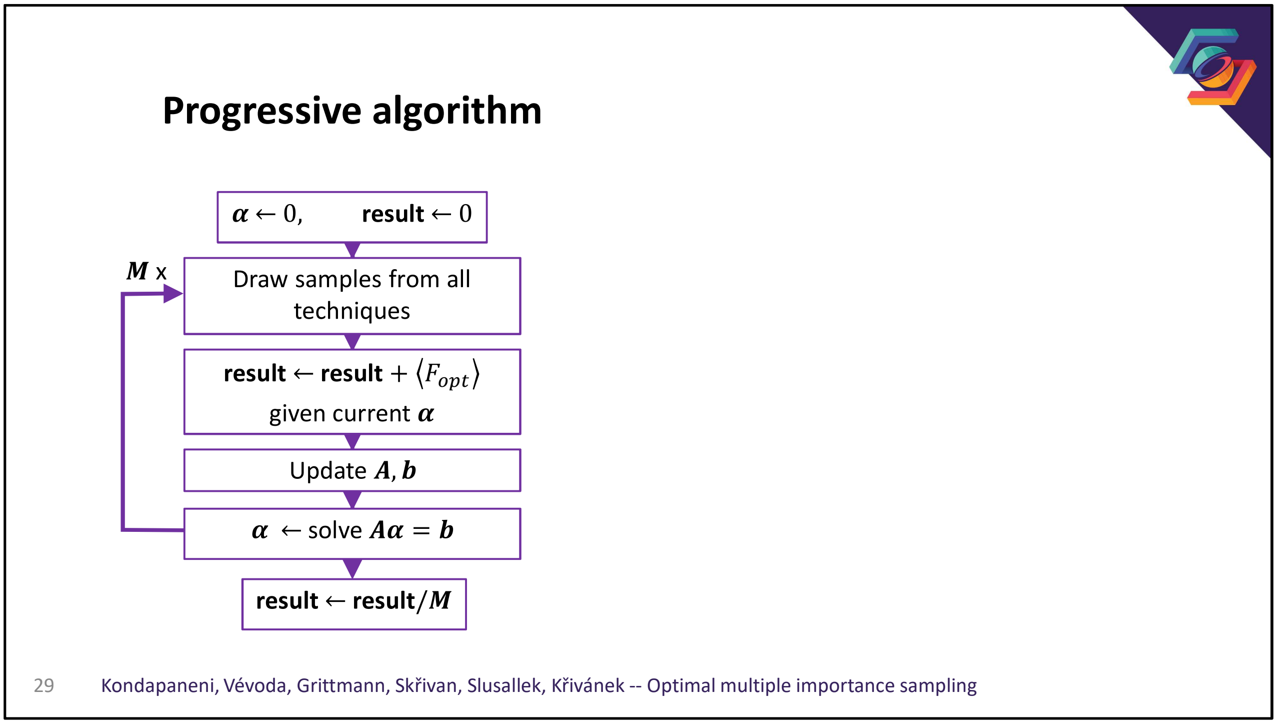
How to compute?

$$a_{ij} = \int \frac{p_i p_j}{\sum n_k p_k} \qquad b_i = \int \frac{p_i f}{\sum n_k p_k}$$

$$\underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}}_{\boldsymbol{A}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}}_{\boldsymbol{b}}$$

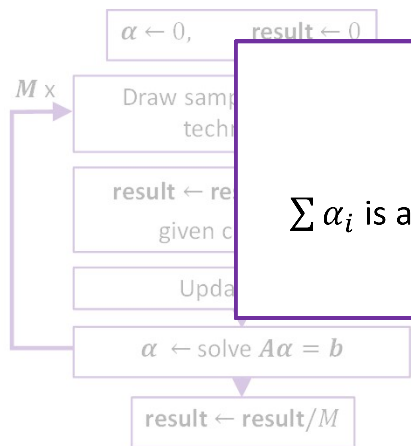
a_{ij}, b_i can be estimated from drawn samples

- (Entry) Recall the linear system we have to solve to obtain the alpha coefficients in the optimal weights.
- <click> The elements of A and b are defined as integrals
- <click> but they can be easily estimated from the samples we draw when computing the MIS estimator.
- We suggest two possible practical implementations.



- (Entry) The first one is called Progressive.
- After the initialization <click>,
- we <click> first draw samples from all techniques.
- Then <click> we accumulate the MIS estimate using the optimal weights but computed with alphas estimated from all previously seen samples.
- We <click> update the linear system
- and <click> re-compute the alphas.
- <click> This is repeated several times,
- and finally<click>, after leaving the loop, we return average of all the estimates.

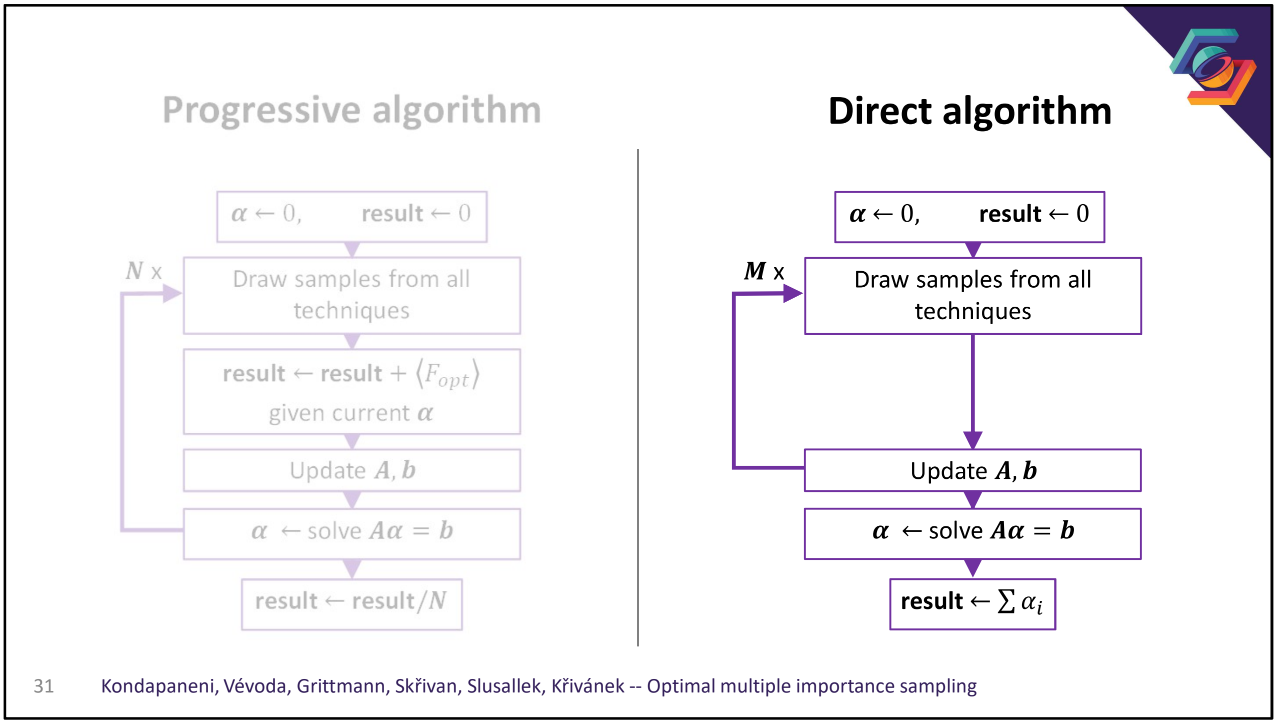
Progressive algorithm



Observation

$\sum \alpha_i$ is a (biased) estimator of F

- (Entry) The second approach how to implement the optimal weights is based on an observation that a sum of the estimated alphas forms also an estimator of the integral F.



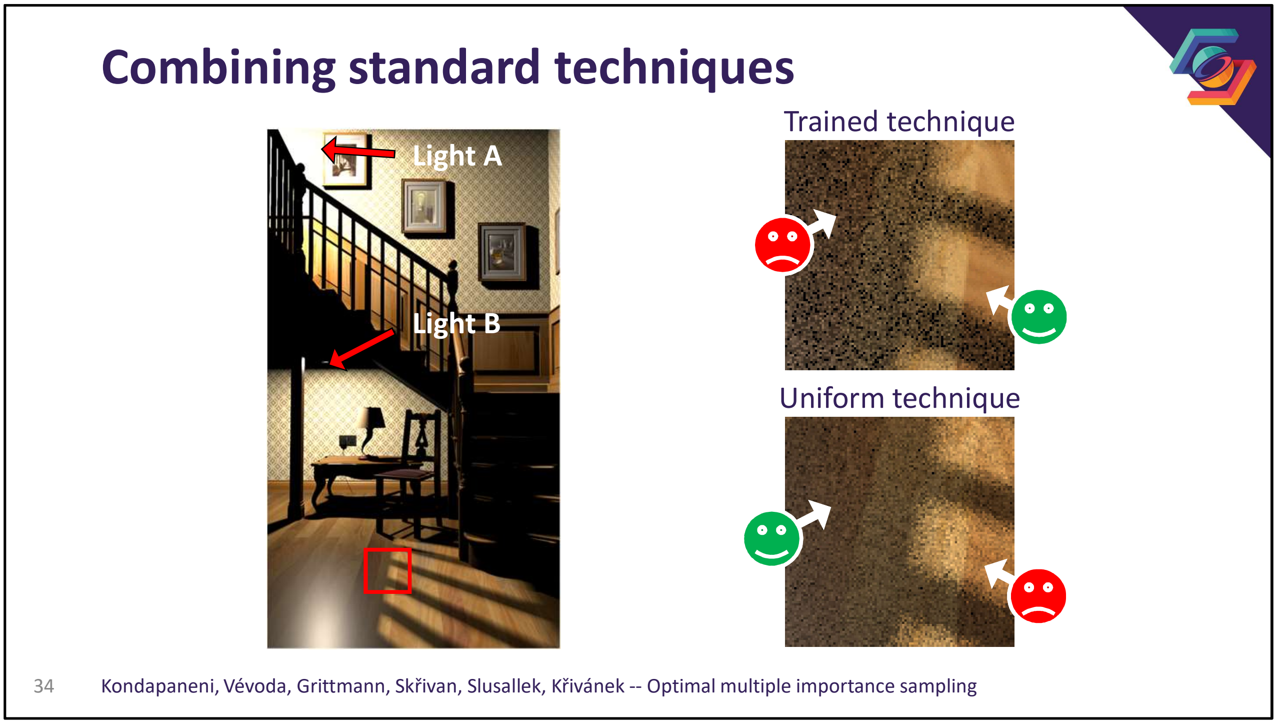
- (Entry) We call the resulting algorithm Direct.
- <click> Here, instead of using the optimal weights formula for mixing the individual contributions, we just keep updating the linear system.
- <click> After leaving the main loop we solve the system for alphas and <click> the result is then formed as their sum.
- While this algorithm is slightly biased, it is consistent and more efficient than the Progressive one.

- (Entry) Now we can show some of the results we were able to obtain with the optimal weights.

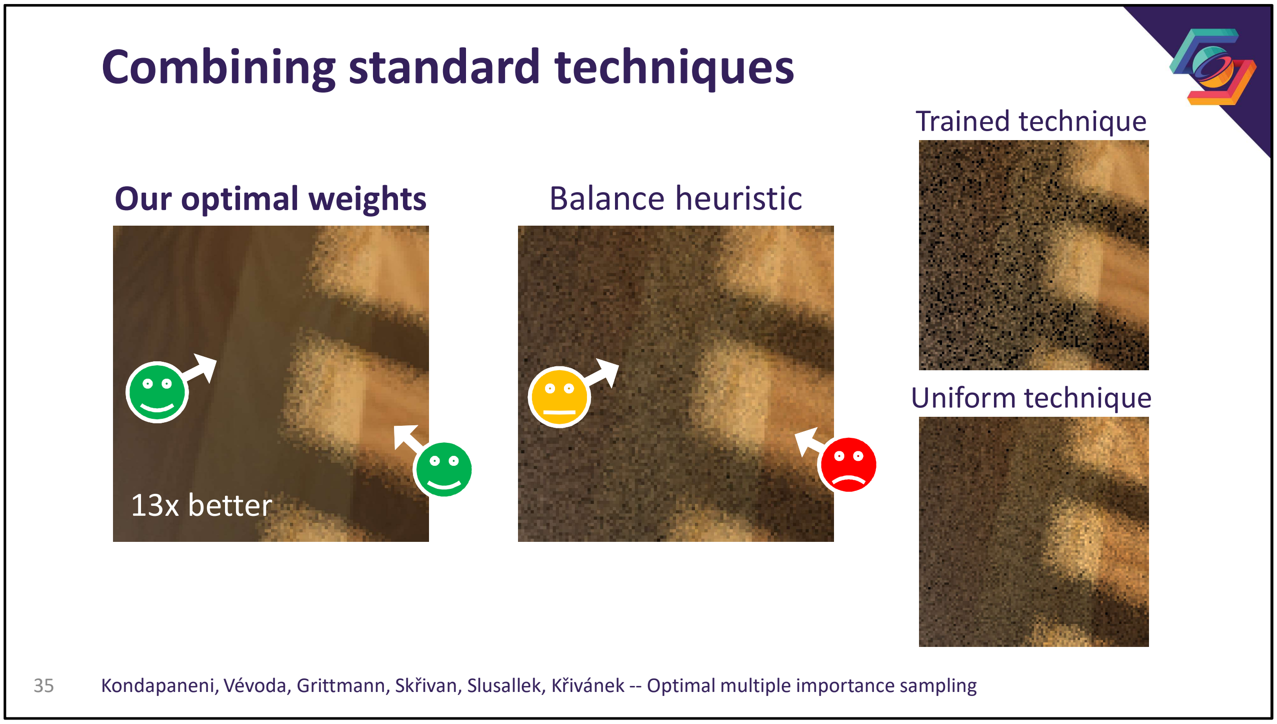
Direct illumination applications

- Applications I – combining standard techniques
- Applications II – designing novel techniques

- (Entry) We applied them to the problem of direct illumination and went in two directions: applying them to standard techniques and designing novel ones.
- <click> We start by showing the results for combining standard techniques.



- (Entry) For that we use our scene from the introduction.
- <click> The scene is illuminated by two lights.
- <click> When shading a point on the floor, we have to randomly select one of the lights.
- <click> Suppose we have a technique, which samples the lights according to their unoccluded contribution. We call this technique Trained. We can see it works nicely <click> in places illuminated by both lights and much worse <click> when light occlusion occurs.
- <click> Now, if we distribute samples across the lights uniformly, we see the opposite effect: it works much better in shadows.



- (Entry) Now we combine the two respective techniques using MIS.

- <click> By using the balance heuristic we obtain decent results, where we no longer see the excessive noise <click> from the Trained technique. But whenever the Trained technique performed good alone, <click> the result is now compromised by the uniform technique.

- <click> When using the optimal weights, we can see much better results. They even out-perform the individual sampling techniques where they performed good already. That is due to the optimal weights acting as the control variates.

Direct illumination applications

- Applications I – combining standard techniques
- **Applications II – designing novel techniques**

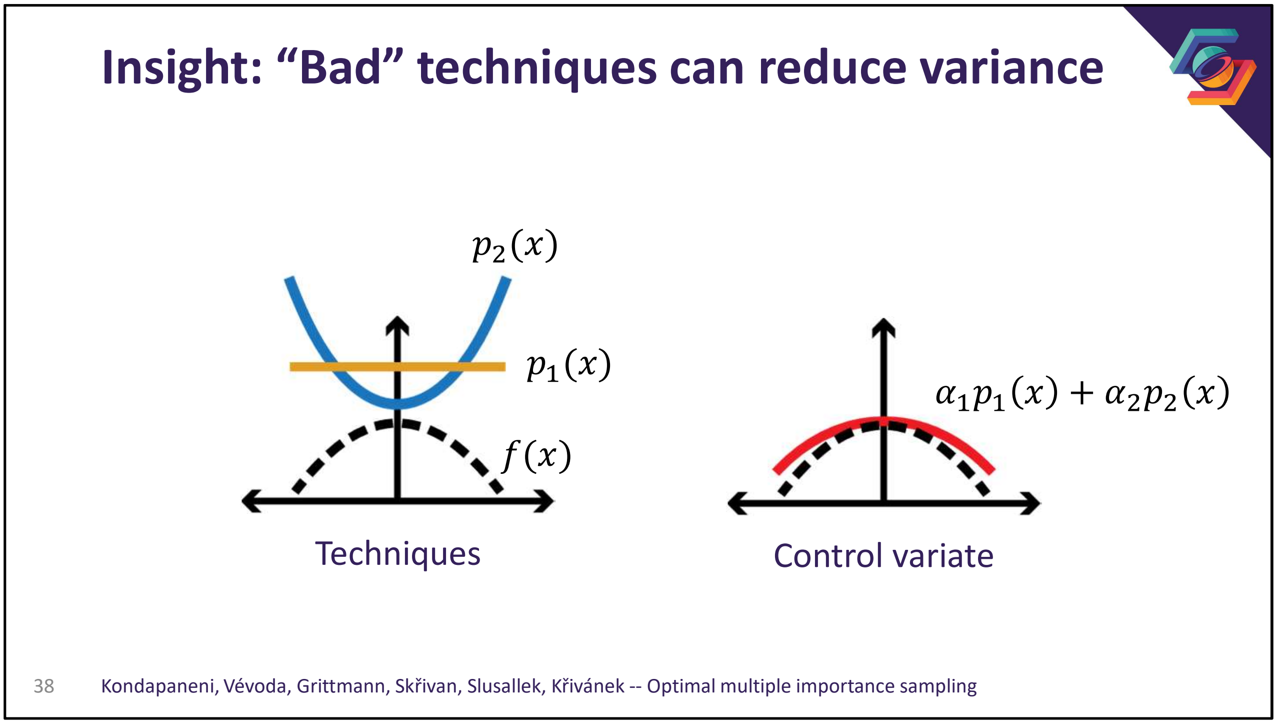
- (Entry) That gets us to why we felt motivated to design completely new techniques.

Reminder: Optimal weights are control variates

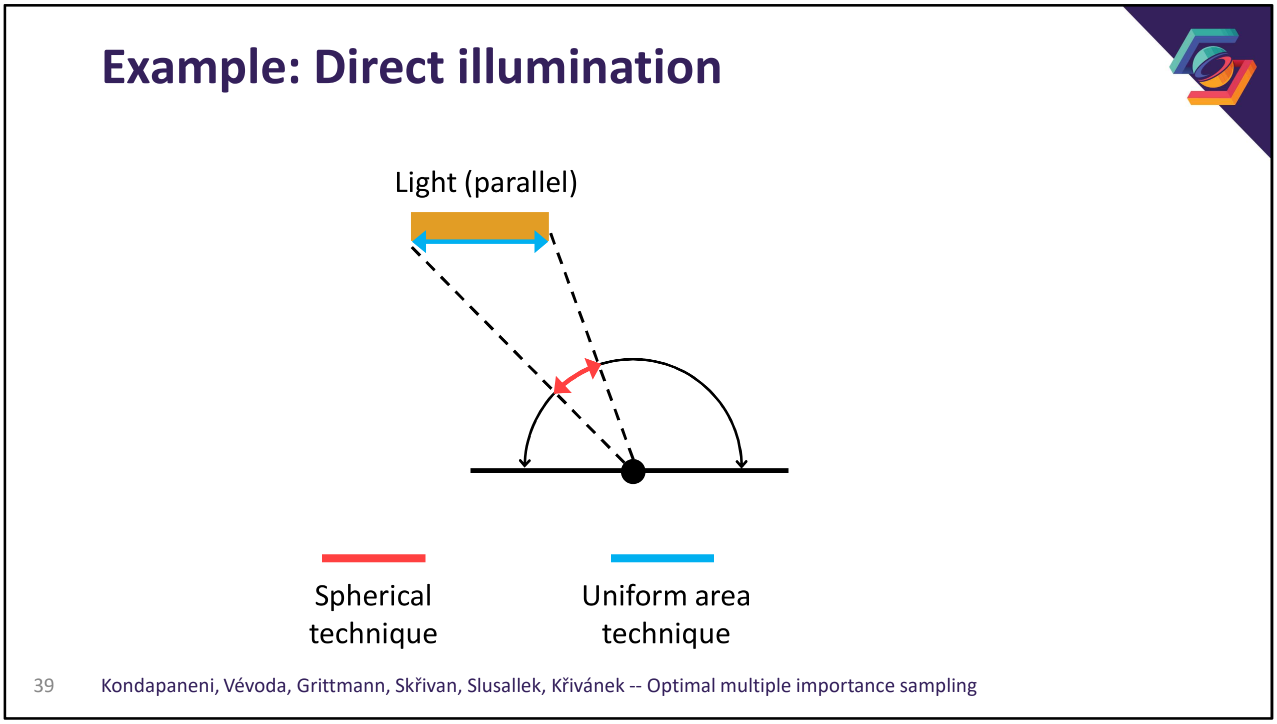
$$\langle F \rangle = G + \langle F - G \rangle$$

The closer $f - g$ to zero, the smaller the variance

- (Entry) To recap, the optimal weights represent a control variate formulation, where function g, formed by the linear combination of the sampling techniques, acts as a control variate for the integrand. And the closer the control variate g is to the integrand, the lower the variance is.



- Therefore, we design sampling techniques in a way that improves their linear combination.
- Consider this illustration where the dashed line is the function f that we want to integrate.
- <click> For that, we have a uniform sampling technique, shown in orange
- To improve the expressive power of the linear combination, we add this blue technique <click>. For importance sampling, this technique is a horrible choice.
- <click> For a control variate, however, we achieve a close to perfect linear combination, the red line on the right.
- Now we show how this can be applied in rendering.

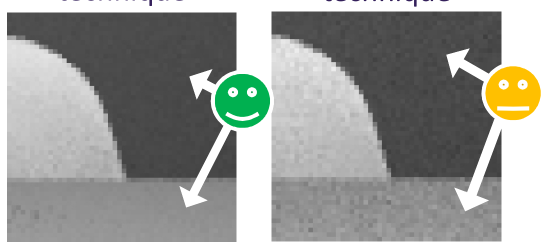


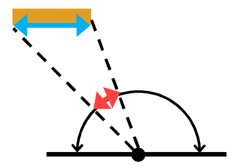
- (Entry) Consider we have a lambertian area light, and we want to apply our optimal weights to compute direct illumination on a diffuse surface.
- Also consider the following sampling techniques: the spherical technique <click> which samples uniformly the light's projection on the hemisphere, and the uniform area technique <click>, which samples uniformly the light's surface.

Comparison: Spherical projection, uniform area

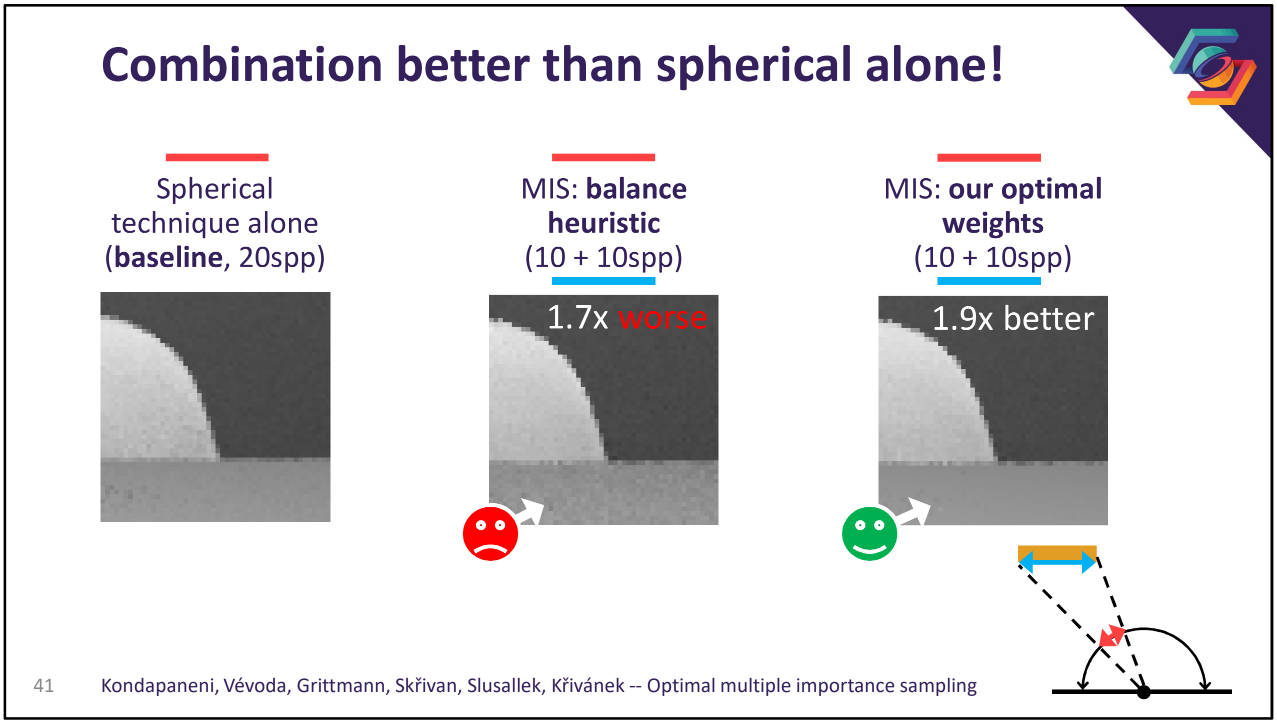
Spherical technique

Uniform area technique



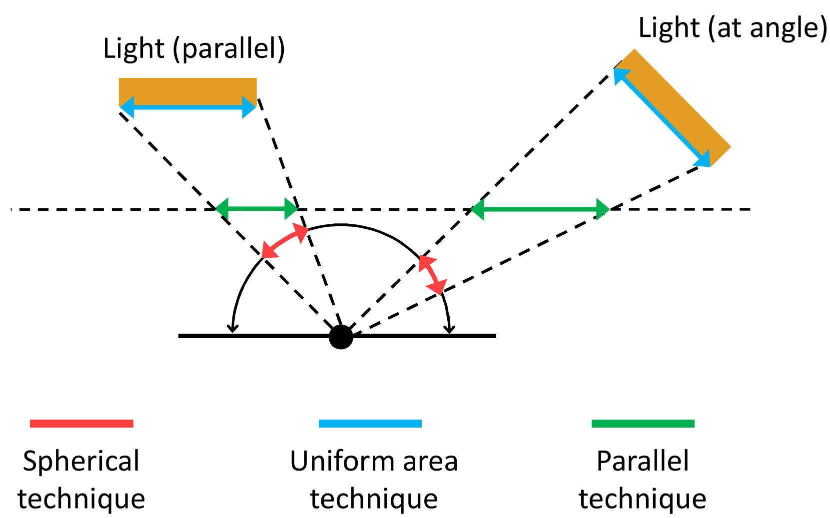


- Let us compare these two techniques in a simple scene. The dining room shown here is illuminated by a single rectangular light above and parallel to the table.
- <click> The spherical technique produces a nice image overall
- <click> while the uniform area sampling has a higher level of noise throughout



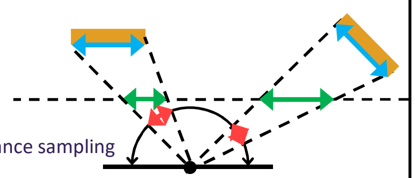
- And now, instead of taking 20 samples per pixel from the spherical technique alone, we replace half of them by samples from the worse uniform technique and combine them using the balance heuristic <click>. As expected the resulting image gets worse.
- <click> With our optimal weights, however, the opposite is the case. The combination actually almost doubles the performance.

Can we do even better?



- Now the question is, can we find a combination that works even better?
- If we look at this sketch, for a case where the light is not parallel to the surface<click>, the spherical and uniform area techniques will be similar. Their linear combination will have less expressive power.
- <click> So we introduce a completely new technique: sampling the parallel projection of the light source area.

Can we do even better?

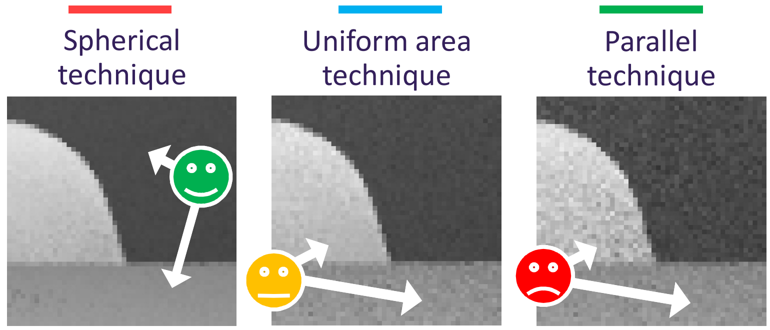


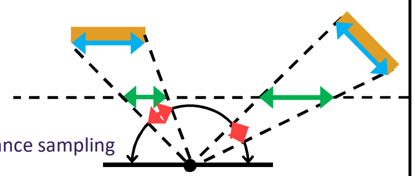
 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- If we look at the individual rendering results, it is clear that the parallel projection is not a sensible technique on its own.

- The level of noise is significantly higher than with either of the two other techniques.

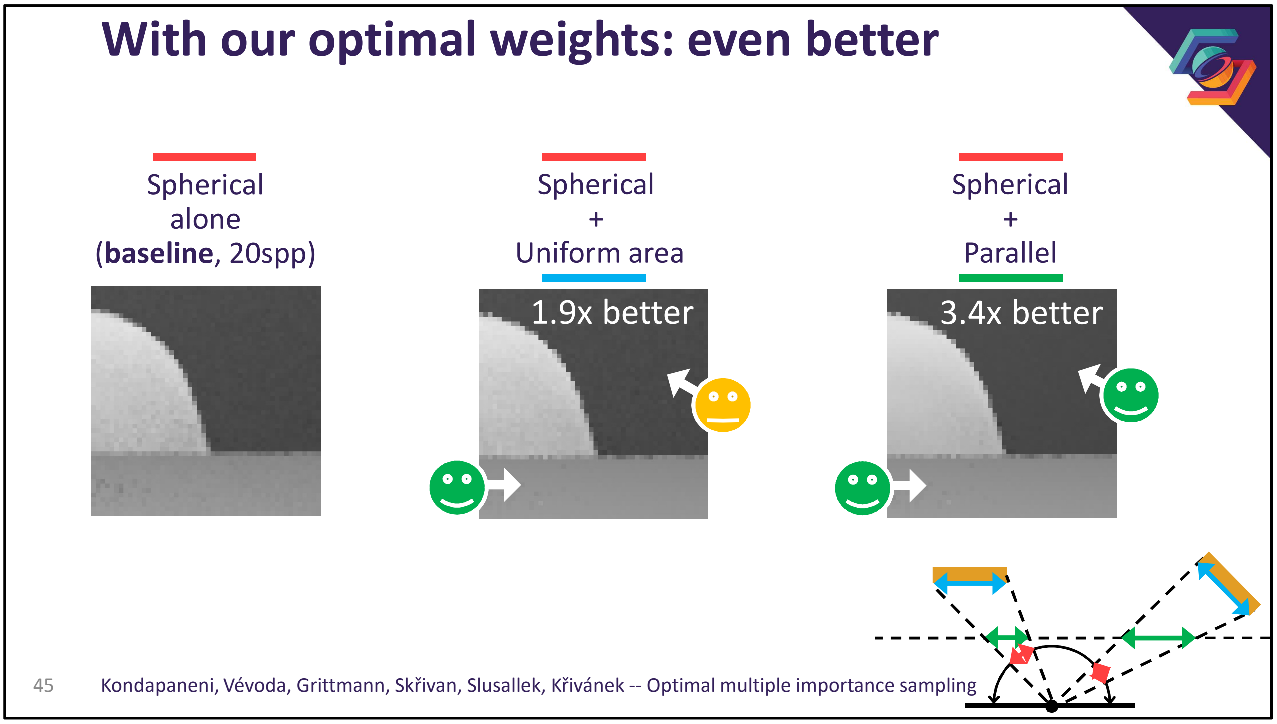
Alone, the parallel technique is even worse





44 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- If we look at the individual rendering results, it is clear that the parallel projection is not a sensible technique on its own.
- The level of noise is significantly higher than with either of the two other techniques.



- But if we look at the result when combining the parallel and the spherical technique via our optimal weights <click>,
- the image is not only better than the spherical technique alone but on the surfaces not parallel to the light <click> it is even better than the optimal weights combination with the uniform technique.

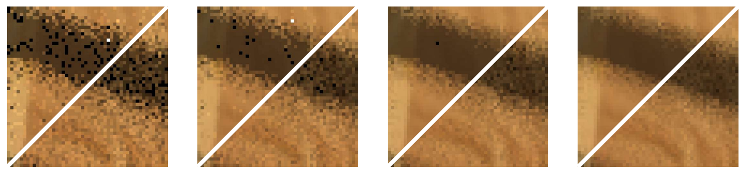
Limitations

Computer Graphics Charles University
UNIVERSITÄT DES SAARLANDES
thrive SIGGRAPH2019 LOS ANGELES • 28 JULY – 1 AUGUST

- Our approach also has its limitations.

Limitations: Salt & pepper for few samples

Optimal weights / Balance heuristic



2 spp | **4** spp | **8** spp | **16** spp

 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- When using either the progressive or direct version of our algorithm we can observe salt and pepper noise for very low sample counts. That is caused by instability of the linear system we need to solve for alphas. But this type of noise can be easily denoised if such low sample counts are really needed.

Limitations: Overhead

- Overhead for a large number N of techniques

$$\underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}}_{N \times N \text{ matrix}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$$

- Another issue is big overhead when using the optimal weights, as the linear system complexity is quadratic with the number of sampling techniques used.
- This is relevant for example in a bi-directional path tracer, where for each path length we have corresponding number of techniques which need to be combined.

Summary & Future work

Computer Graphics Charles University

UNIVERSITÄT DES SAARLANDES

thrive SIGGRAPH2019 LOS ANGELES • 28 JULY – 1 AUGUST

- Let us wrap up our presentation with list of our contributions and possible future work.

Contributions

- Balance heuristic bounds revisited
- Optimal MIS weights
- Connection to Control Variates
- Practical algorithms

- <click> We revisited the bounds on variance of the balance heuristic,
- <click> which lead us to investigation what is the optimum in space of all MIS weights
- <click> and we found this optimum to be equivalent to a particular control variates formulation.
- <click> Finally, we showed the practical application of the optimal weights.

Future work

- Path guiding, VCM
- Correlated samples
- Optimal sample allocation
- Simpler alternative heuristics: $w_{positive} \leq w_{??} \leq w_{opt}$

- (Entry) We admit that our applications are more a proof-of-concept, but we hope that our work will motivate further research.
- <click> For example application of the optimal weights in path guiding and bidirectional methods.
- <click> We see future work also in application to estimators with correlated samples,
- <click> and completing our approach by optimal sample allocation.
- <click> Also simpler alternative heuristics might exist, which would have lower overhead than the optimal weights but still outperform the ordinary heuristics.

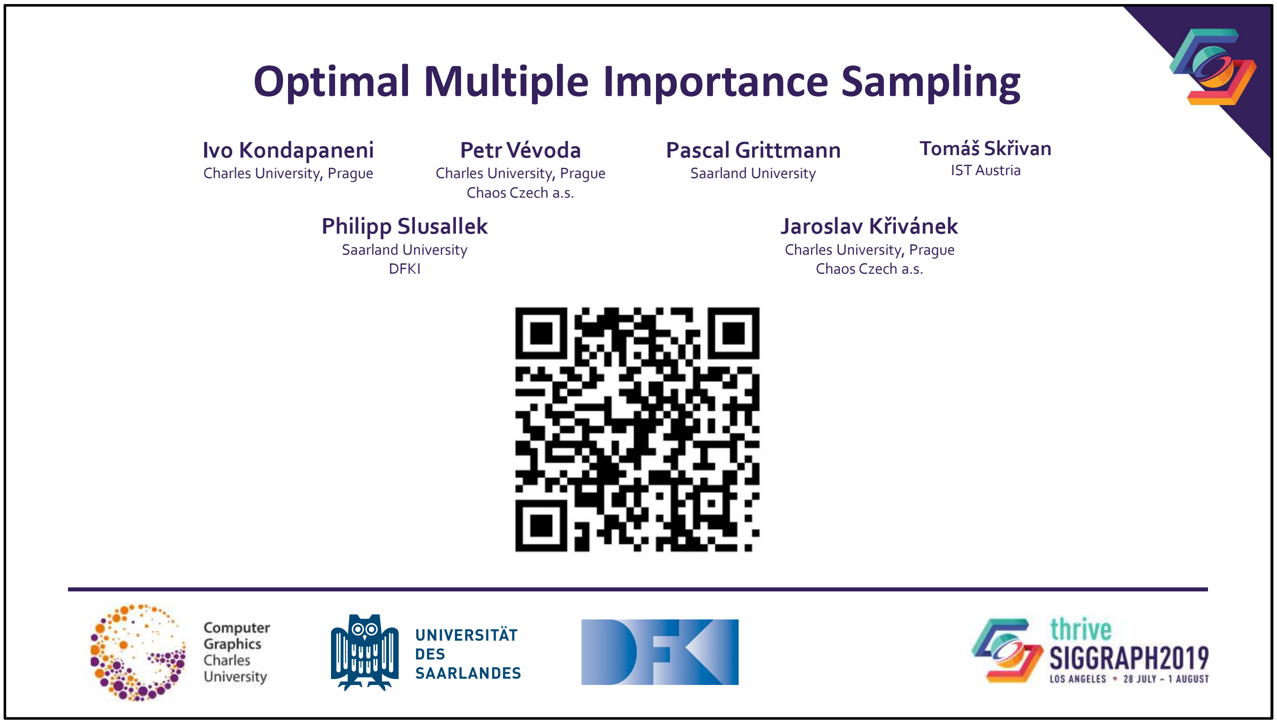
Acknowledgments

Funding:
- Charles University Grant Agency (GAUK 996218),
- Charles University (SVV-2017-260452)
- Czech Science Foundation (16-18964S and 19-07626S).

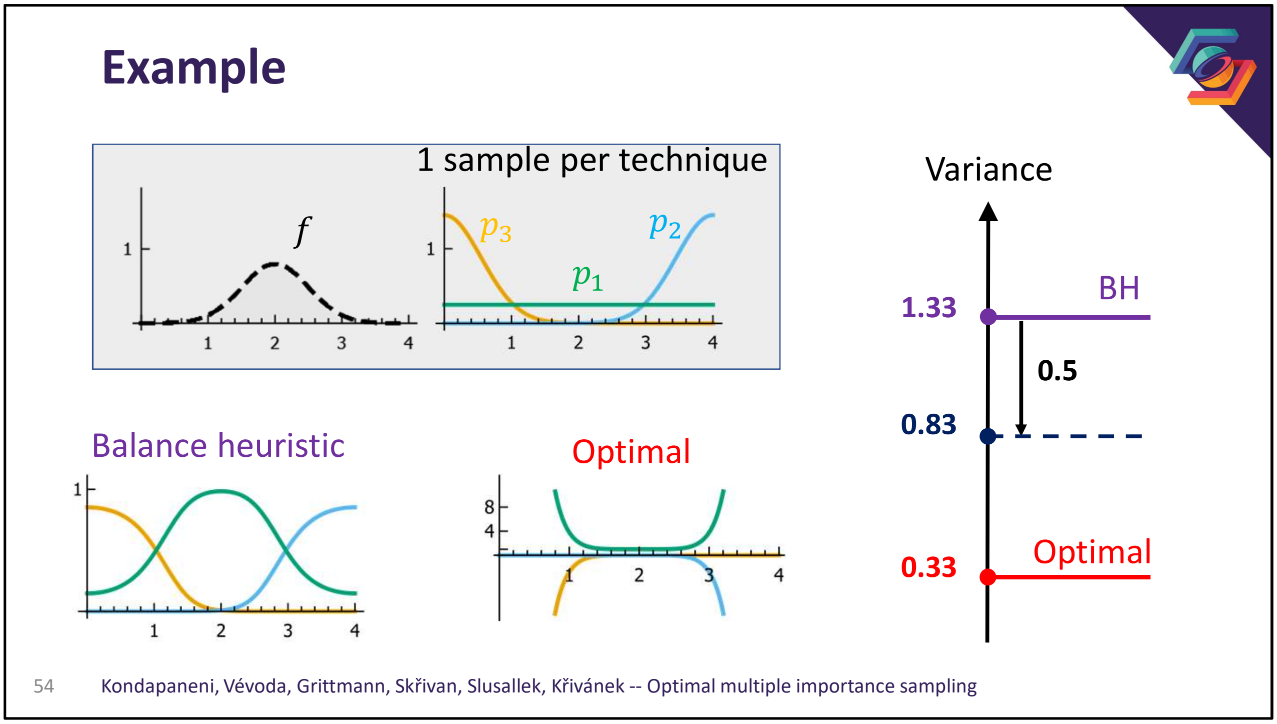
DiSTRO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642841.

- (Entry) Finally, we would like to thank the funding agencies for supporting our work.

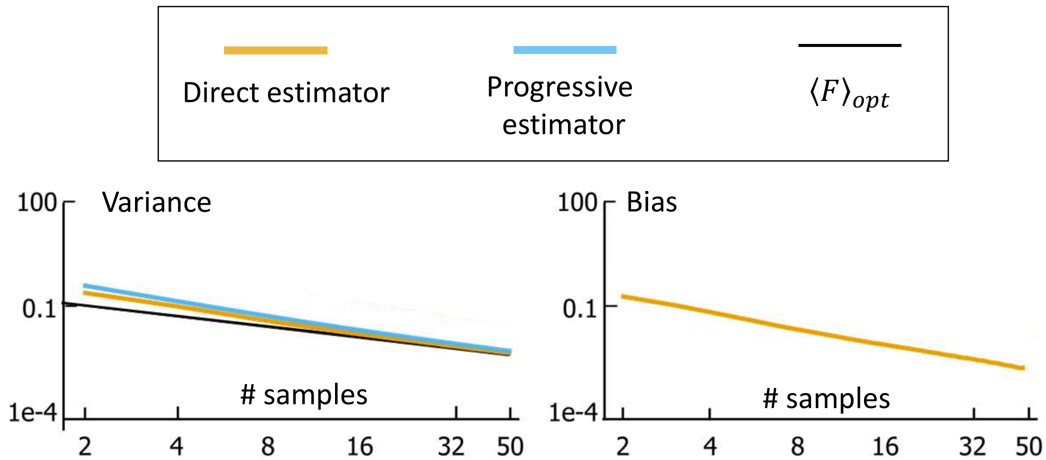


And that is all, thank you for your attention.



- (Entry) Let me illustrate the issue with bounds on a synthetic example
- Suppose we have our integrand and sampling densities as depicted, and suppose we draw one sample per each technique.
- Then the weights for the balance heuristic used for the above problem will yield variance 1.33, and the bounds for balance heuristic computed according to Veach and Guibass suggest that the best combination strategy shouldn't achieve lower variance than 0.83
- But the optimal weights for this case as shown in the picture attain also negative values in some parts of the domain, and they achieve variance little over 0.3, which is much lower than suggested by the usual bounds.

Comparison: Progressive and direct



55 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- (Entry) Now, using some of our results from synthetic tests, we compare the progressive and direct estimators in terms of MSE and Bias versus number of samples used in their computation.
- We can see that the direct estimator achieves better MSE for low sample counts. This is because the first iterations of the progressive algorithm are weighted with too coarse estimates of alpha. The noise from these iterations is only gradually averaged out.
- On the other hand, the direct estimator is biased, but consistent, and on the right we plot its bias for increasing sample counts. The bias is introduced by the inversion of the estimated linear system used for computing the alphas. Contrary to that the progressive estimator is unbiased, as it forms an unbiased estimator for any alphas used in its formula. By choosing between Direct and Progressive algorithm we tradeoff a little bias for lower noise.

Variance analysis

$$V[\langle F \rangle_{opt}] = V[\langle F \rangle_{bal}] - V[\langle G \rangle_{bal}]$$

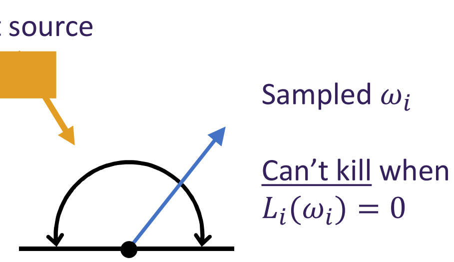
Zero when $\alpha \propto (n_1, \dots, n_N)^T$

- (Entry) Based on the control variate formulation, we performed a variance analysis of the optimal estimator ..
- and we reached to the following decomposition of variance into variance of balance heuristic and variance of control variate.
- The second term is zero when the elements of the vector α are proportional to the number of samples from the individual sampling techniques. In that case, the balance heuristic is optimal.

Zero samples can't be killed

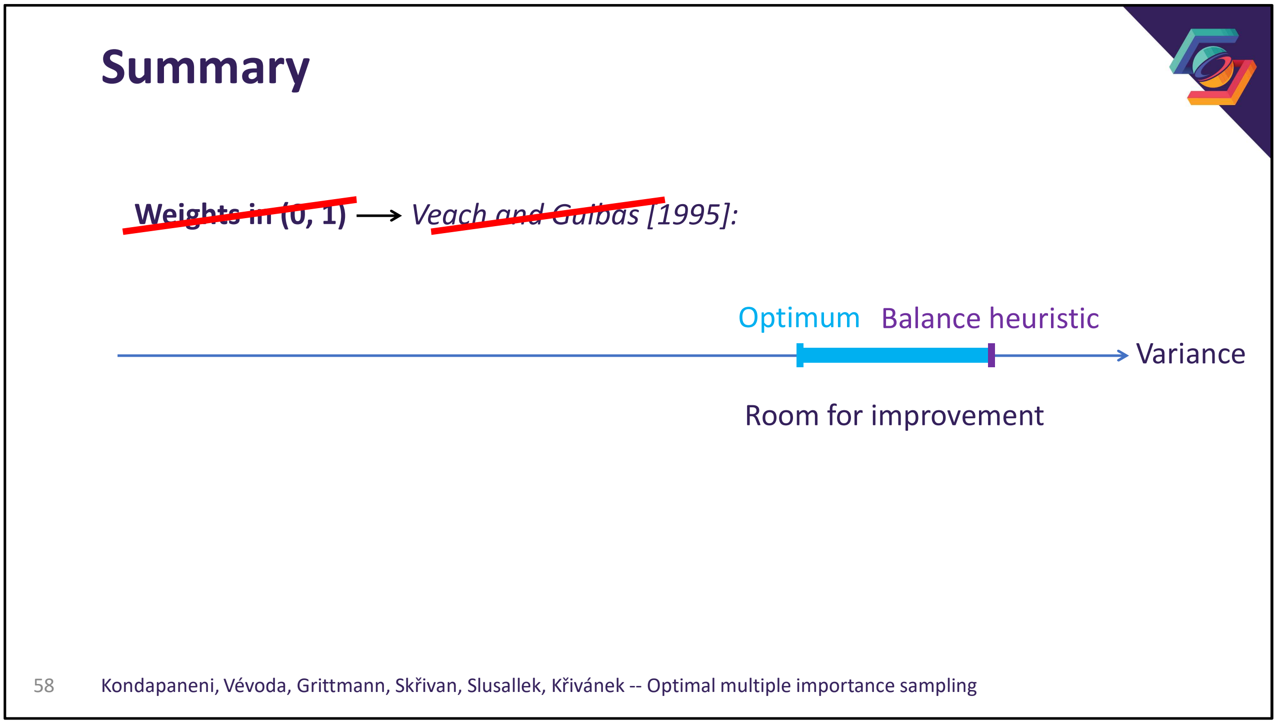
- Compute MIS <u>even</u> when $f(x) = 0$
- Because we use CV!

<u>Example</u>: Next-event estimation

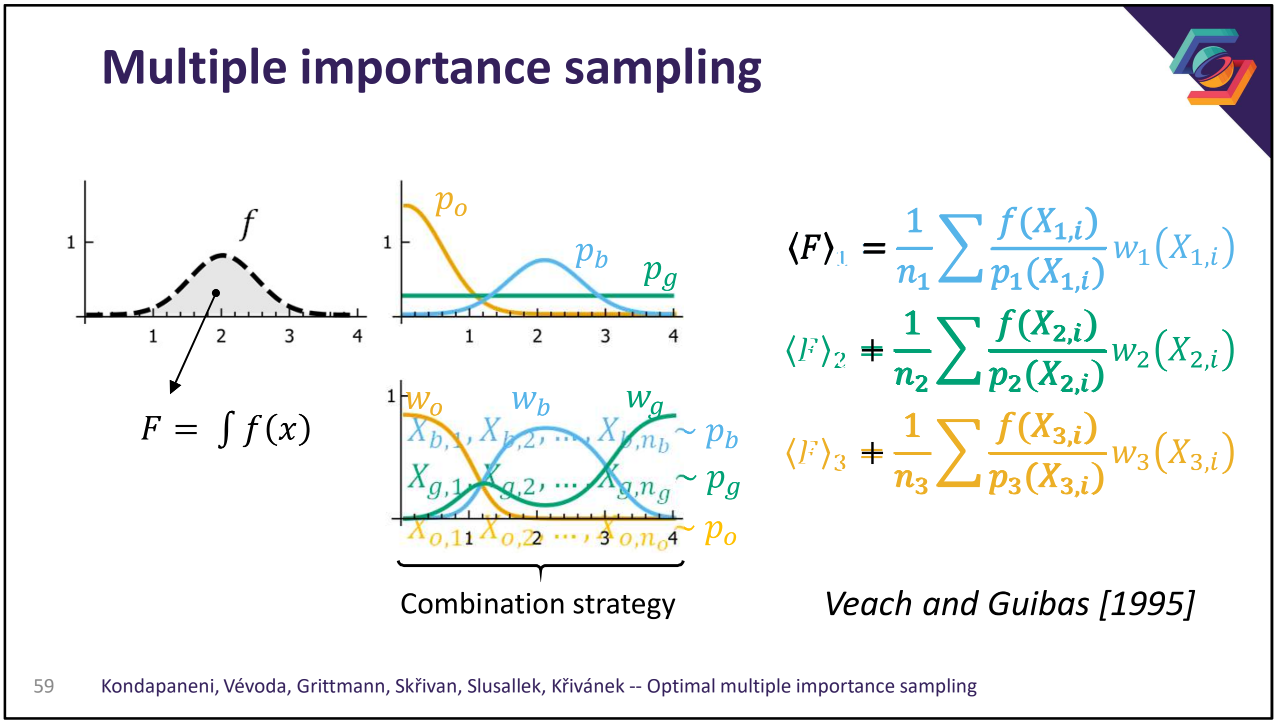


57 Kondapaneni, Vévoda, Grittmann, Skřivan, Slusallek, Křivánek -- Optimal multiple importance sampling

- (Entry) Also the usual optimization approach of ignoring samples for which an integrand is zero can not be used, because we still need to compute Control Variates part for each of the samples.
- This typically happens when we do next-event estimation, where the contribution is zero for a lot of BSDF samples!



- To sum up: variance bounds proven by Veach and Guibas make people believe that the balance heuristic is almost optimal and there is not much room for improvement.
- But the proof of the bounds assumes only weights between 0 and 1
- Removing this restriction invalidates Veach's bounds and opens up to MIS weighting with much lower variance that what's been believed so far to be possible.



- (Entry) Then Multiple importance sampling can help us.
- We have several sampling techniques, each possibly suited for sampling different parts of f,
- and they form several estimators.
- Now we introduce a combination strategy which is a set of weighting functions, one weighting function for each sampling technique.
- Finally, we form an MIS estimator by re-weighting the contributions of individual estimators by means of weighting functions. That will yield a more robust estimator, an MIS estimator.