Title: Mathematics, Physics, and Machine Learning

Speakers: Sergei Gukov

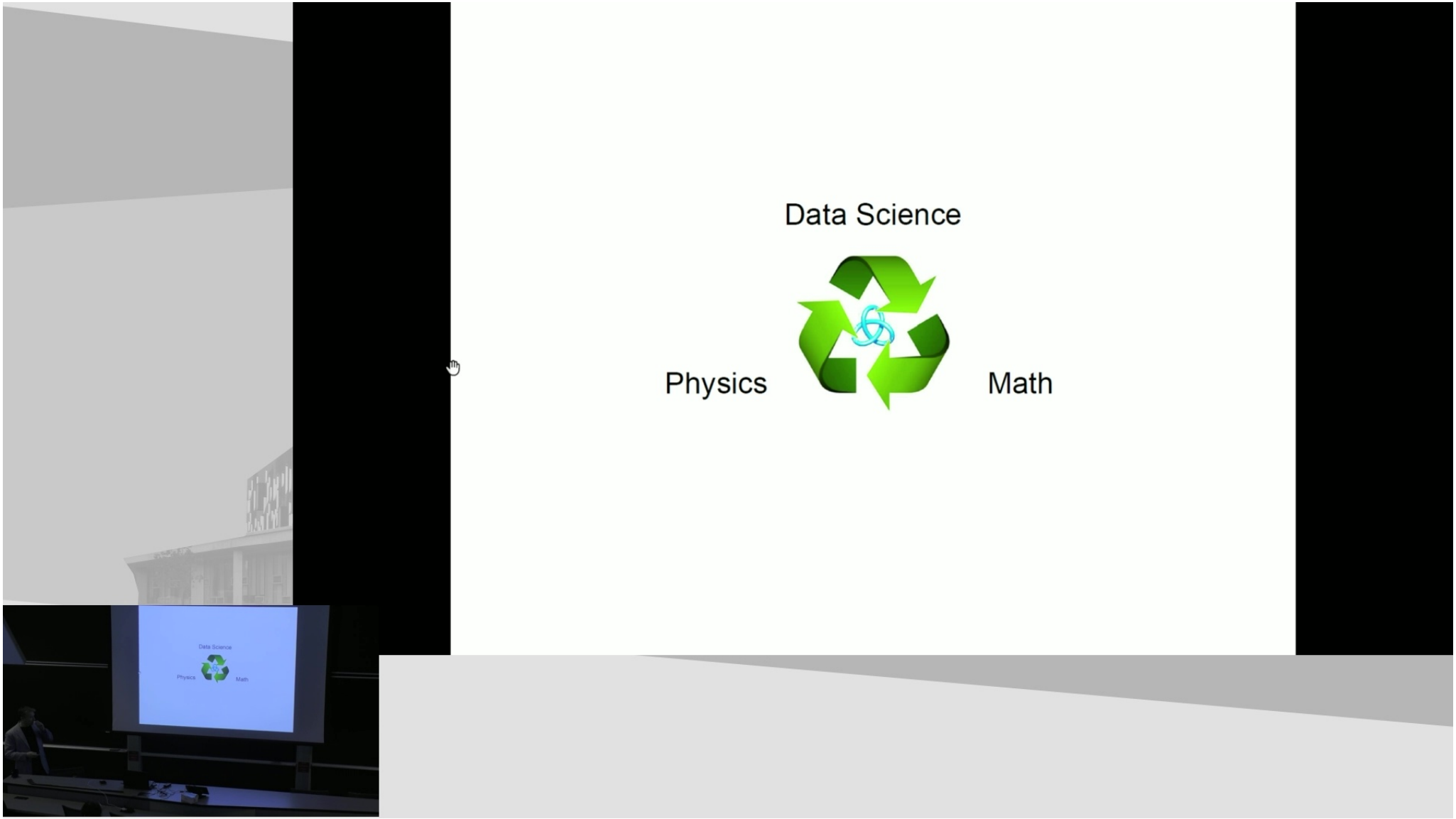
Series: Quantum Fields and Strings

Date: November 27, 2023 - 11:00 AM

URL: https://pirsa.org/23110087

Abstract: Abstract TBA

Zoom link https://pitp.zoom.us/j/93671537911?pwd=UjFJK3B5Y1ZQbU9ucm9yK3dNVmNSZz09


Data Science
Physics
Math

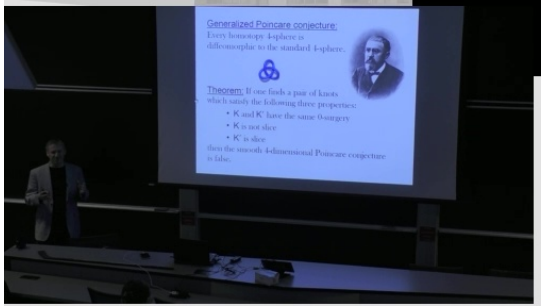
Generalized Poincare conjecture:

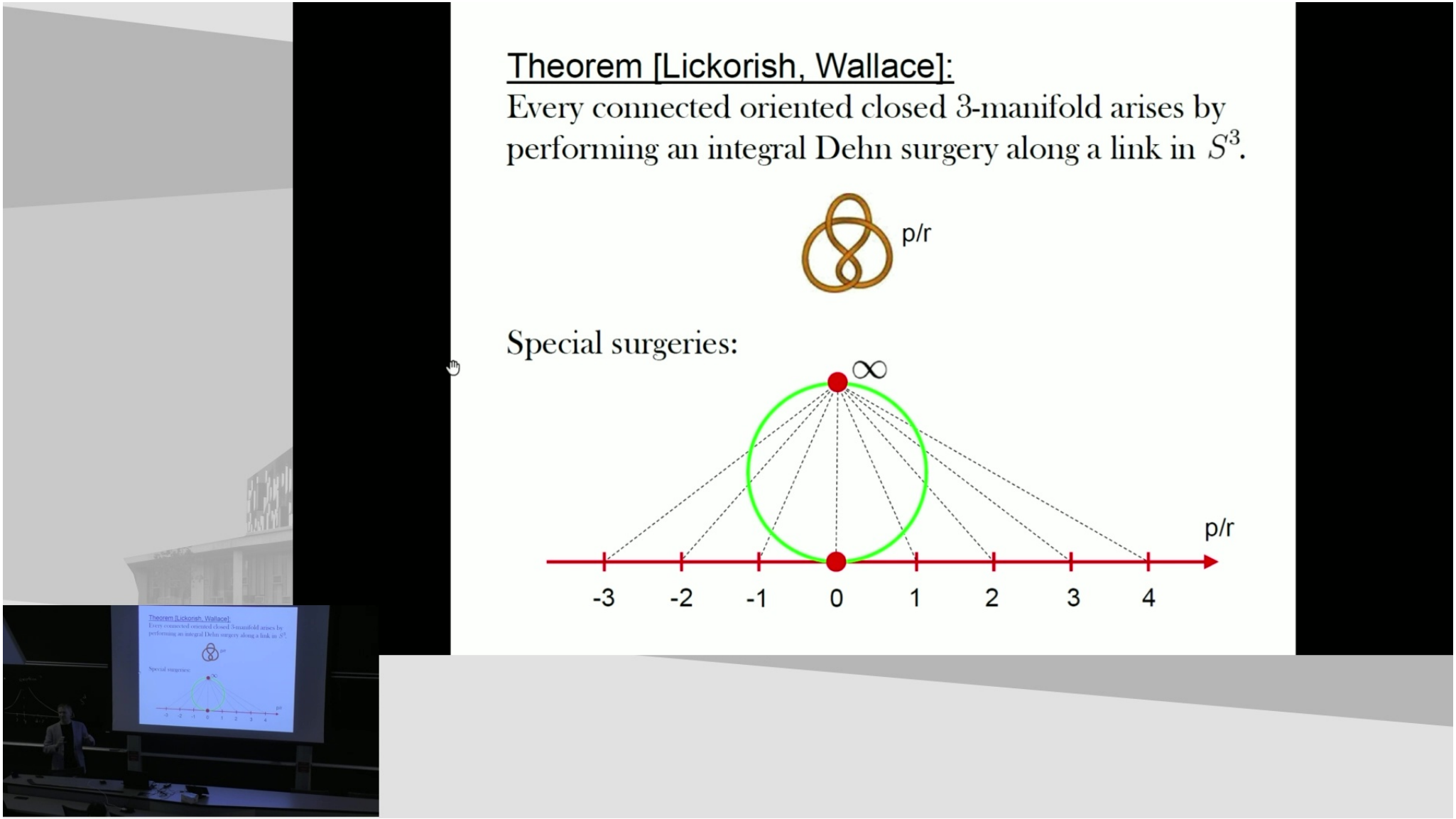
Every homotopy 4-sphere is diffeomorphic to the standard 4-sphere.

Theorem: If one finds a pair of knots which satisfy the following three properties:

- K and K′ have the same 0-surgery
- K is not slice
- K′ is slice

then the smooth 4-dimensional Poincare conjecture is false.

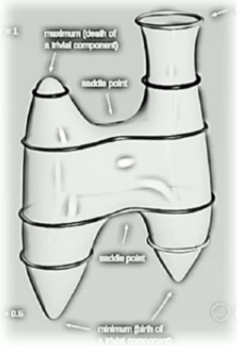



Theorem [Lickorish, Wallace]:
Every connected oriented closed 3-manifold arises by performing an integral Dehn surgery along a link in S^3.
p/r
Special surgeries:
∞
p/r
-3
-2
-1
0
1
2
3
4

FIBERED KNOTS AND POTENTIAL COUNTEREXAMPLES TO THE PROPERTY 2R AND SLICE-RIBBON CONJECTURES

ROBERT E. GOMPF, MARTIN SCHARLEMANN, AND ABIGAIL THOMPSON

48 crossings



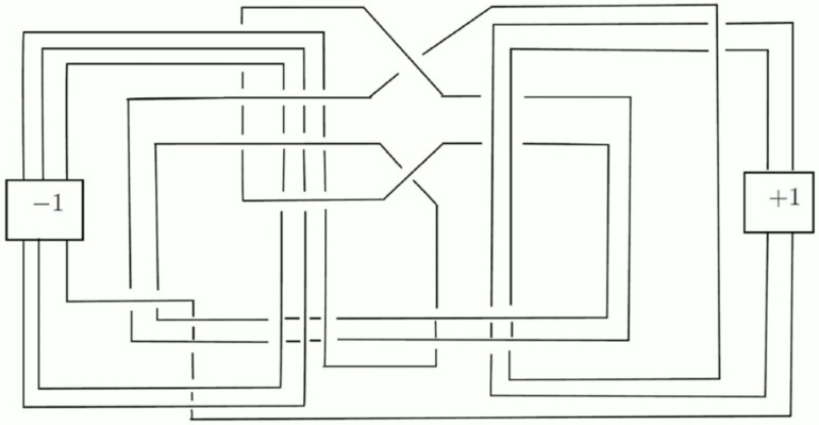
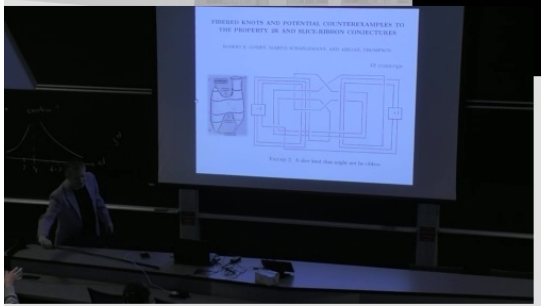
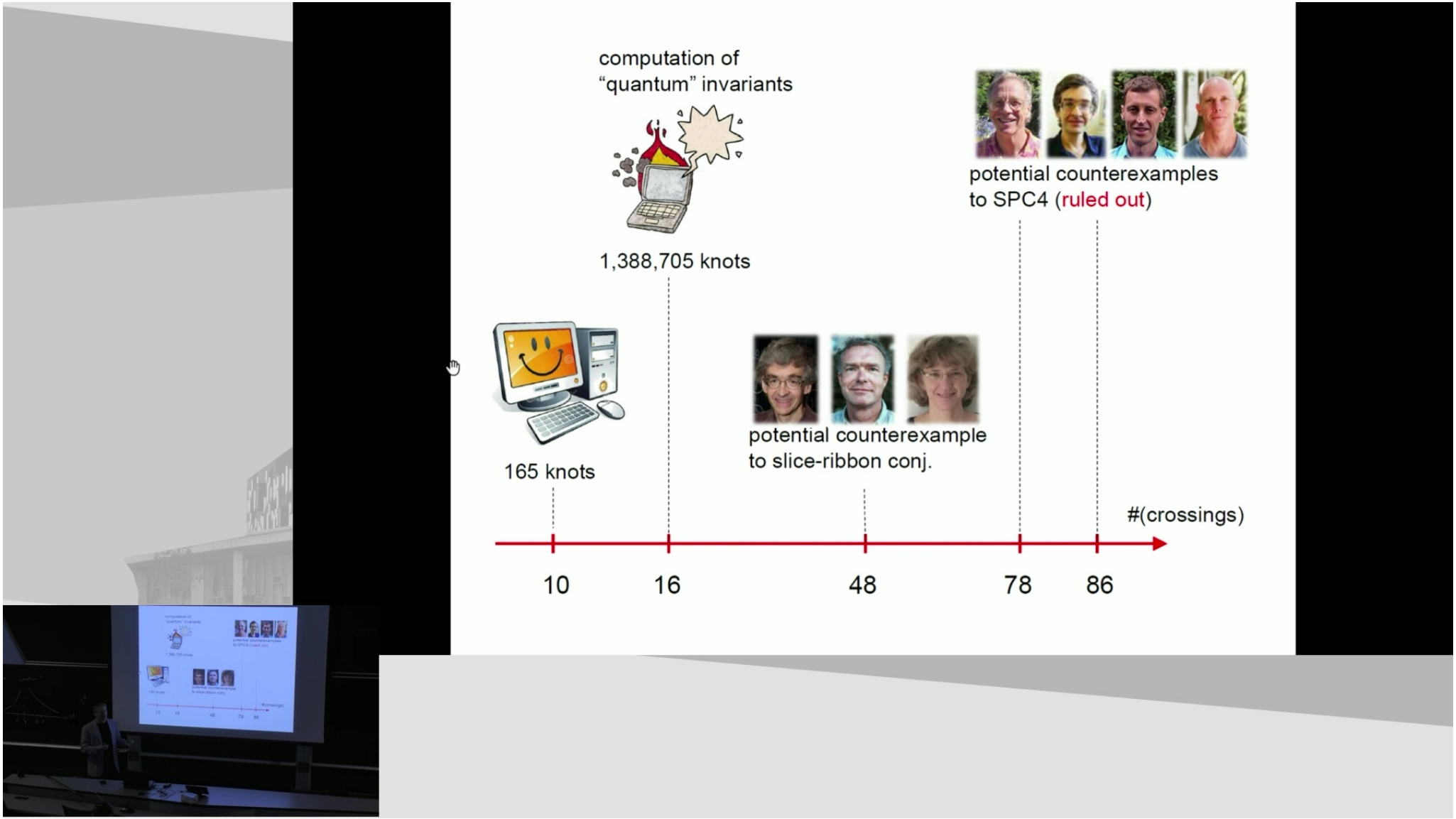
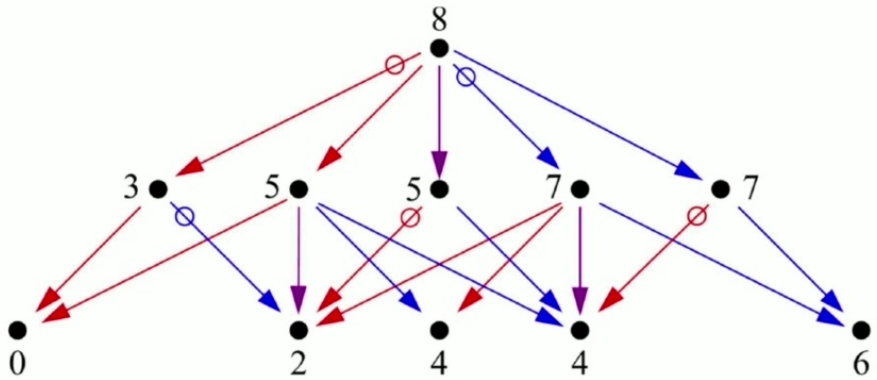


FIGURE 2. A slice knot that might not be ribbon



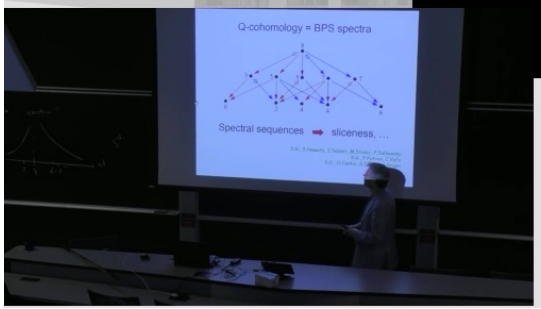

computation of
“quantum” invariants
1,388,705 knots
potential counterexamples
to SPC4 (ruled out)
165 knots
potential counterexample
to slice-ribbon conj.
#(crossings)
10
16
48
78
86

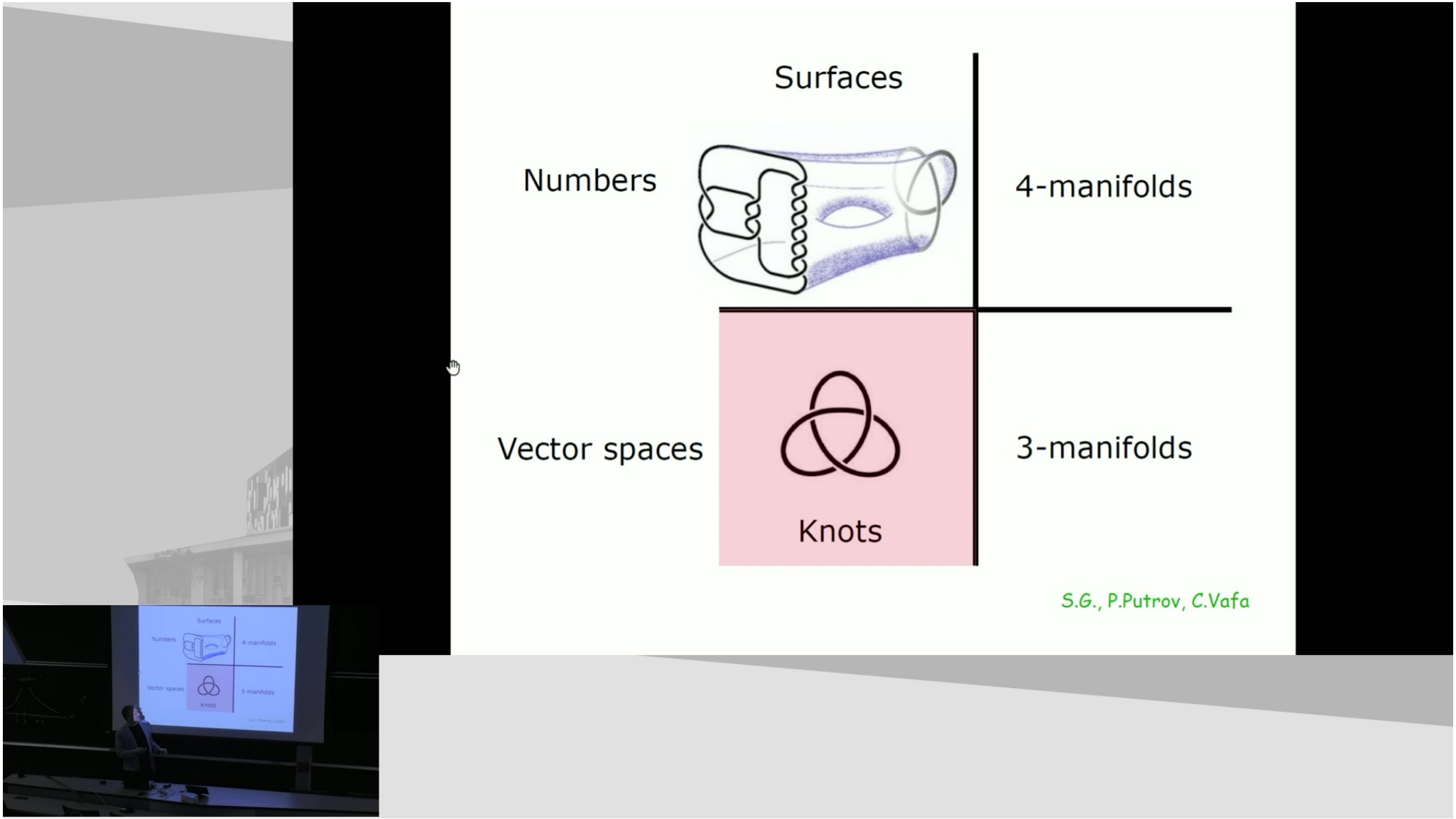
Q-cohomology = BPS spectra



Spectral sequences ➡ sliceness, …

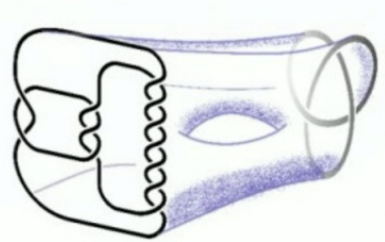
S.G., S.Nawata, I.Saberi, M.Stosic, P.Sulkowski
S.G., P.Putrov, C.Vafa
S.G., O.Costin, G.Dunne, A.Gruen
:




Surfaces
Numbers
4-manifolds
Vector spaces
3-manifolds
Knots
S.G., P.Putrov, C.Vafa

Surfaces

Numbers



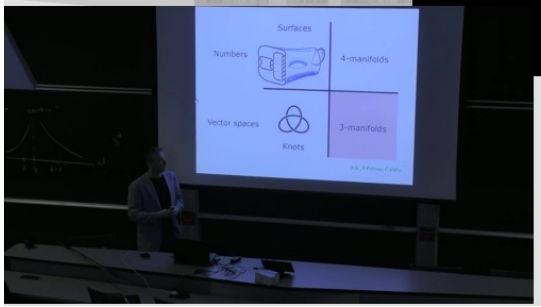
4-manifolds

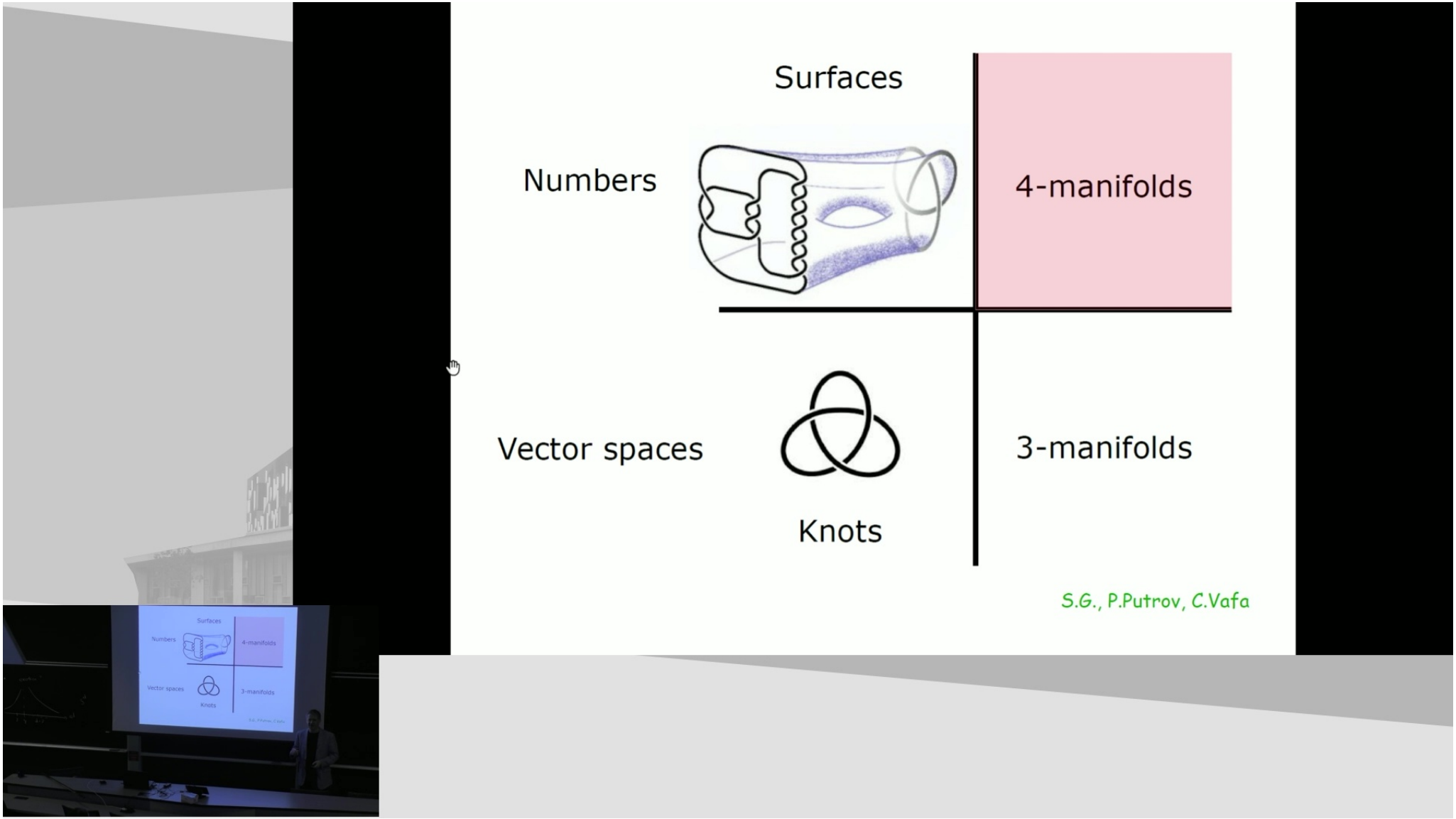
Vector spaces

Knots

3-manifolds

S.G., P.Putrov, C.Vafa




Surfaces
Numbers
4-manifolds
Vector spaces
3-manifolds
Knots
S.G., P.Putrov, C.Vafa

Definition: (6-n)-dimensional theory

$$T[M_n, \ldots] := \text{6d (0,2) theory on } M_n$$

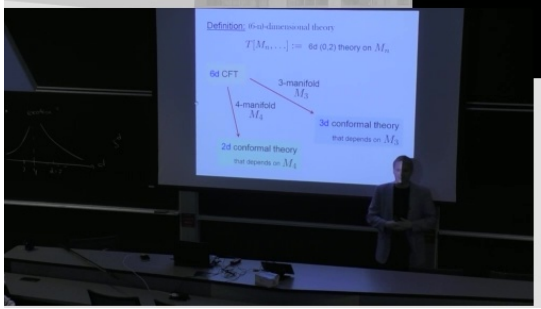
6d CFT

3-manifold M_3

4-manifold M_4

3d conformal theory that depends on M_3

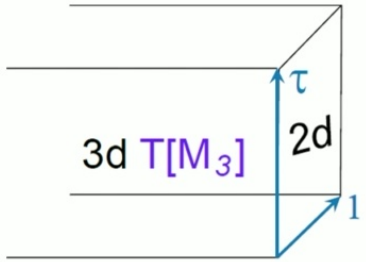
2d conformal theory that depends on M_4



Even though in most cases we don't know 3d $\mathcal{N}=2$ theory and can not determine its BPS spectrum, we can explicitly compute the graded trace

$$\sum_{i,j}(-1)^i q^j \dim H^{i,j}(M_3; b) = \widehat{Z}_b(M_3; q)$$

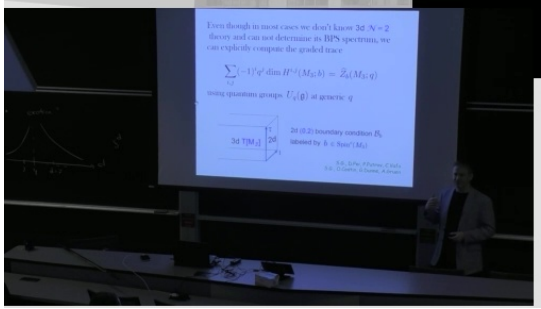
using quantum groups $U_q(\mathfrak{g})$ at generic q



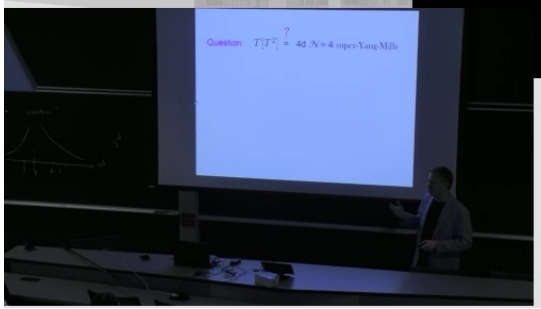
2d (0,2) boundary condition $\mathcal{B}_b$

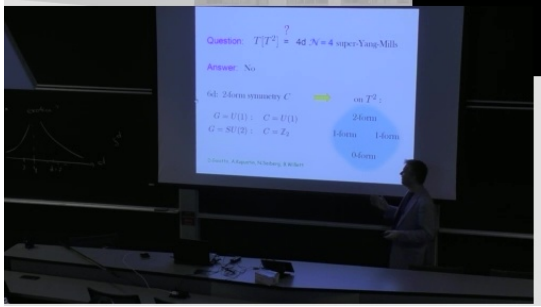
labeled by $b \in \mathrm{Spin}^c(M_3)$

S.G., D.Pei, P.Putrov, C.Vafa
S.G., O.Costin, G.Dunne, A.Gruen



Question: $T[T^2] \overset{?}{=}$ 4d $\mathcal{N}=4$ super-Yang-Mills





Question: $T[T^2] \overset{?}{=}$ 4d $\mathcal{N}=4$ super-Yang-Mills

Answer: No

6d: 2-form symmetry C ➡ on T^2 :

$G = U(1)$: $C = U(1)$

$G = SU(2)$: $C = \mathbb{Z}_2$

2-form

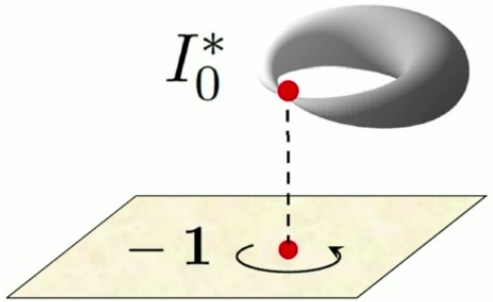
1-form 1-form

0-form

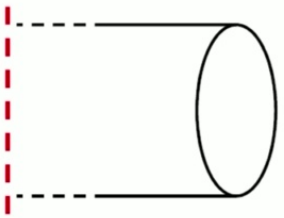
D.Gaiotto, A.Kapustin, N.Seiberg, B.Willett

Coulomb branches

4d $\mathcal{N}=4$ $G=SU(2)$



$T[T^2]$

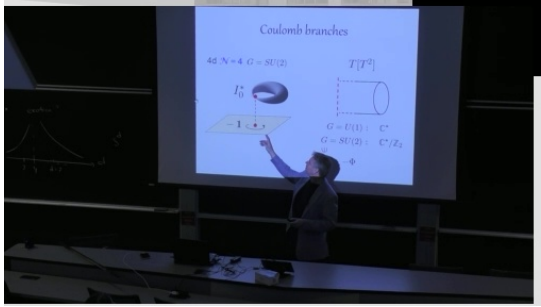


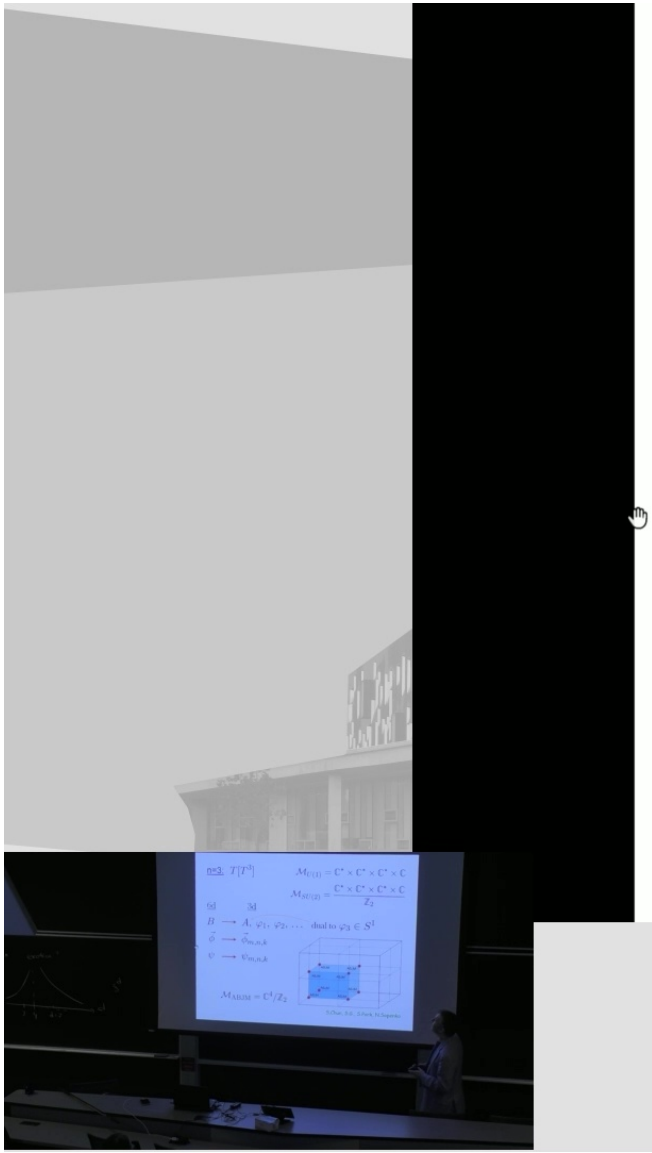
$G=U(1):\quad \mathbb{C}^*$

$G=SU(2):\quad \mathbb{C}^*/\mathbb{Z}_2$

$\cup$

$\Phi \to -\Phi$





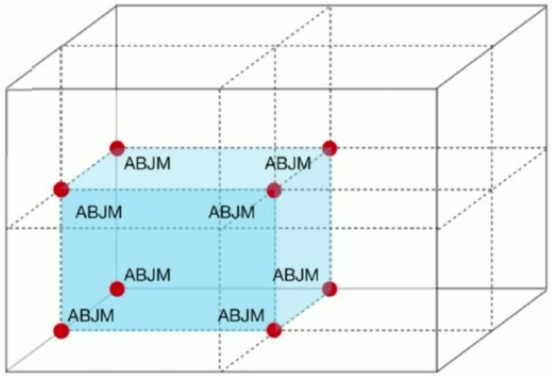
n=3: $T[T^3]$

$$\mathcal{M}_{U(1)} = \mathbb{C}^* \times \mathbb{C}^* \times \mathbb{C}^* \times \mathbb{C}$$

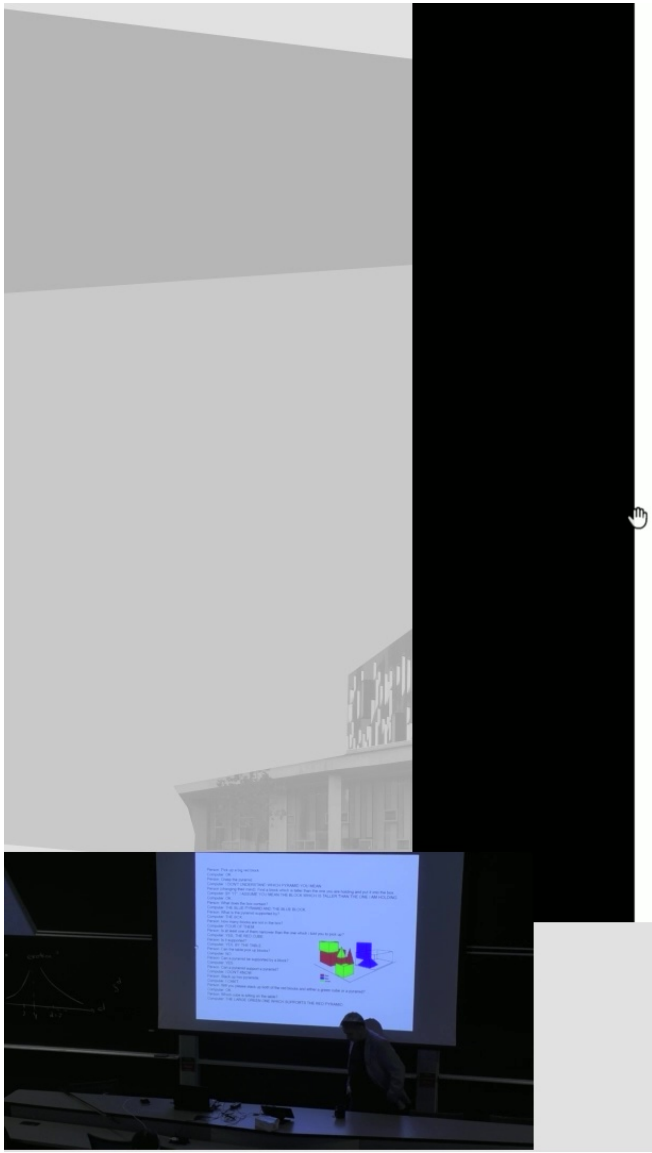
$$\mathcal{M}_{SU(2)} = \frac{\mathbb{C}^* \times \mathbb{C}^* \times \mathbb{C}^* \times \mathbb{C}}{\mathbb{Z}_2}$$

6d		3d
B	$\longrightarrow$	$A,\ \varphi_1,\ \varphi_2,\ \ldots$ dual to $\varphi_3 \in S^1$
$\vec{\phi}$	$\longrightarrow$	$\vec{\phi}_{m,n,k}$
ψ	$\longrightarrow$	$\psi_{m,n,k}$

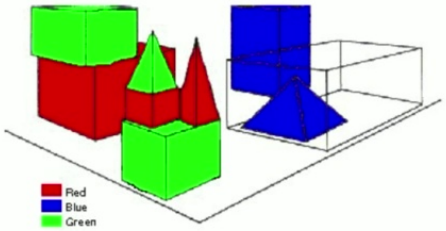
$$\mathcal{M}_{\text{ABJM}} = \mathbb{C}^4/\mathbb{Z}_2$$



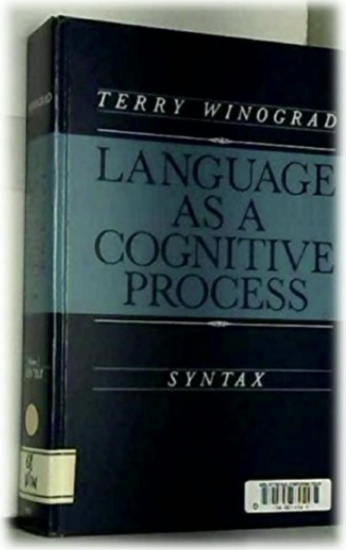
S.Chun, S.G., S.Park, N.Sopenko



Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK.
Person: What does the box contain?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: What is the pyramid supported by?
Computer: THE BOX.
Person: How many blocks are not in the box?
Computer: FOUR OF THEM.
Person: Is at least one of them narrower than the one which i told you to pick up?
Computer: YES, THE RED CUBE.
Person: Is it supported?
Computer: YES, BY THE TABLE.
Person: Can the table pick up blocks?
Computer: NO.
Person: Can a pyramid be supported by a block?
Computer: YES.
Person: Can a pyramid support a pyramid?
Computer: I DON'T KNOW.
Person: Stack up two pyramids.
Computer: I CAN'T.
Person: Will you please stack up both of the red blocks and either a green cube or a pyramid?
Computer: OK.
Person: Which cube is sitting on the table?
Computer: THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.
:



SHRDLU (1971)



1997


Google

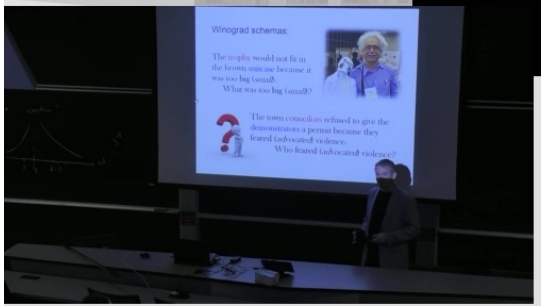
Winograd schemas:

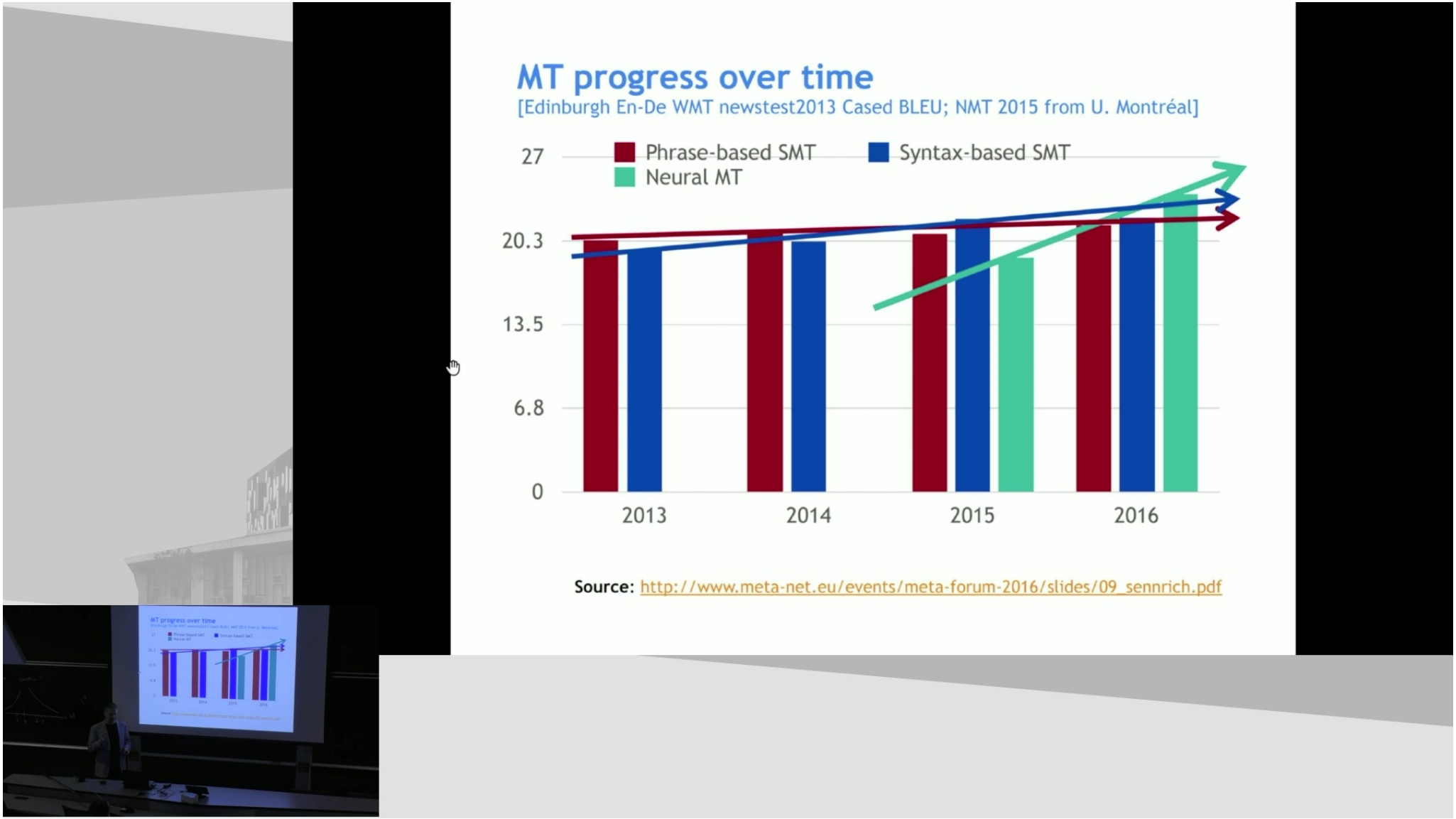
The trophy would not fit in the brown suitcase because it was too big (*small*).

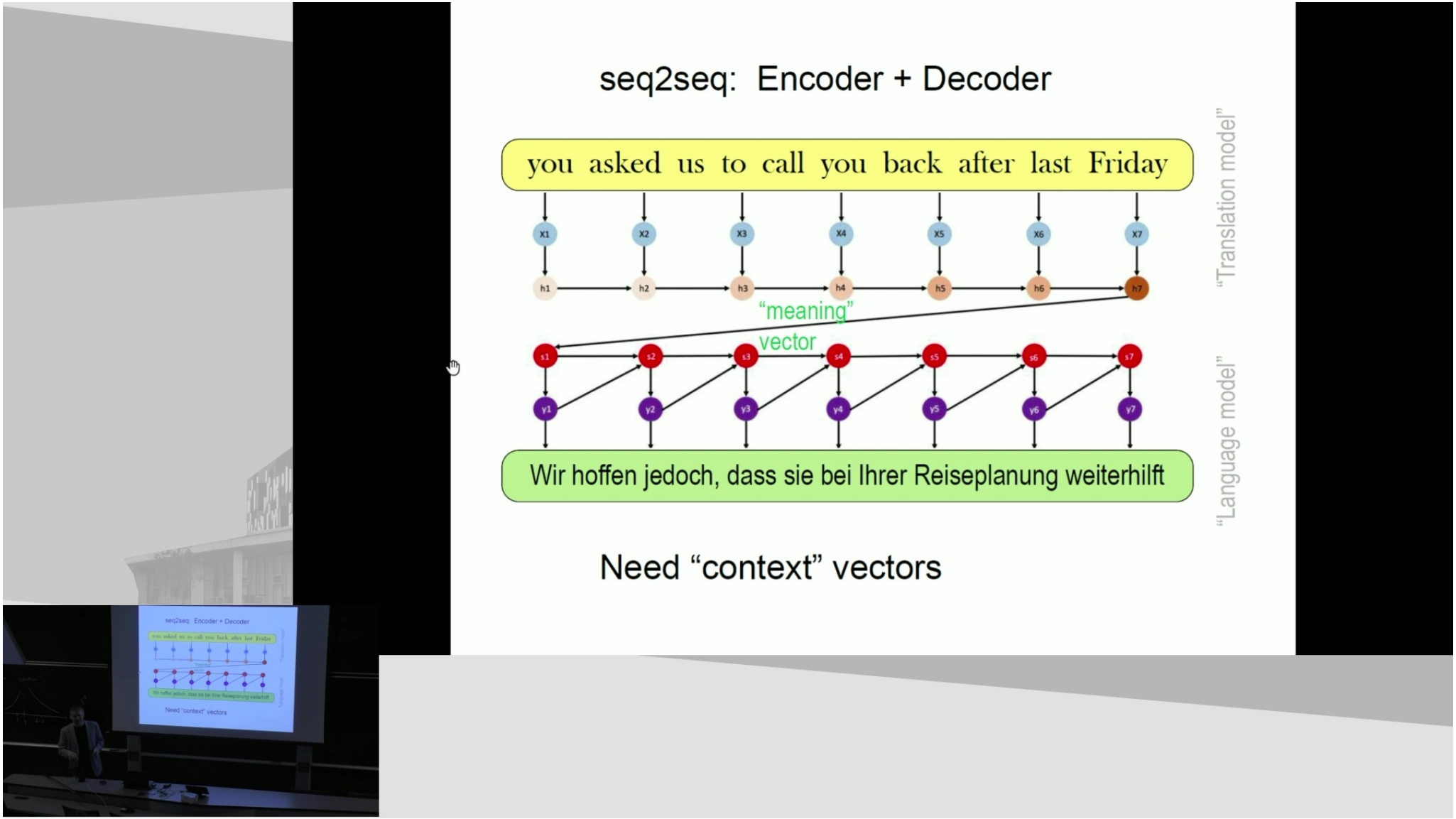
What was too big (*small*)?

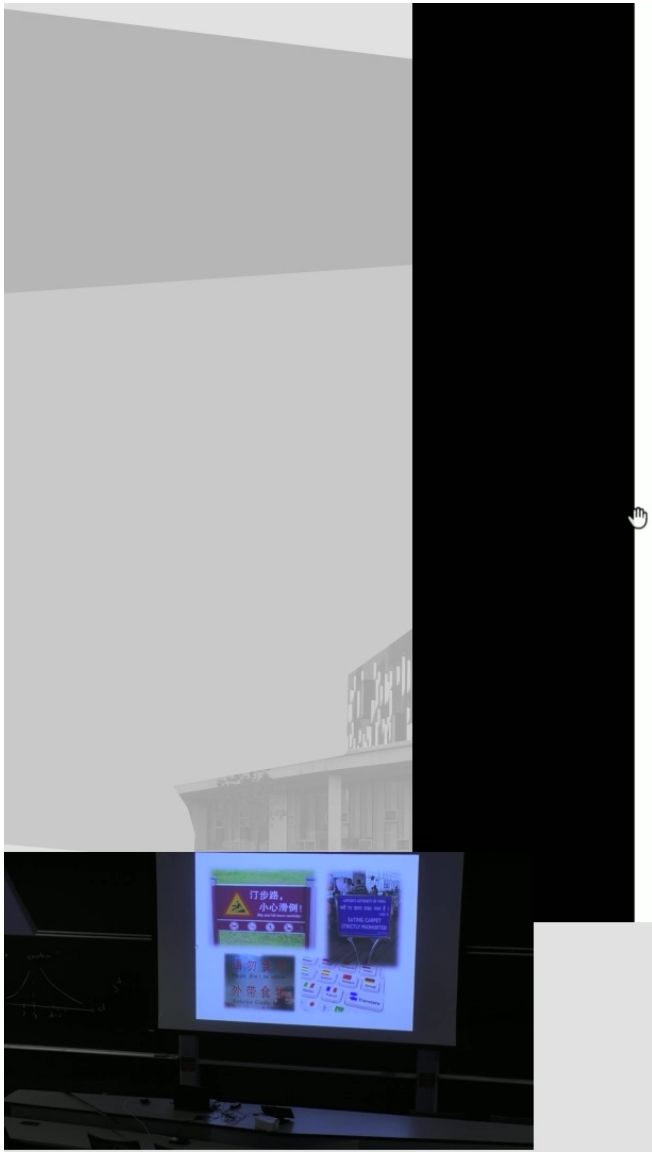
The town councilors refused to give the demonstrators a permit because they feared (*advocated*) violence.

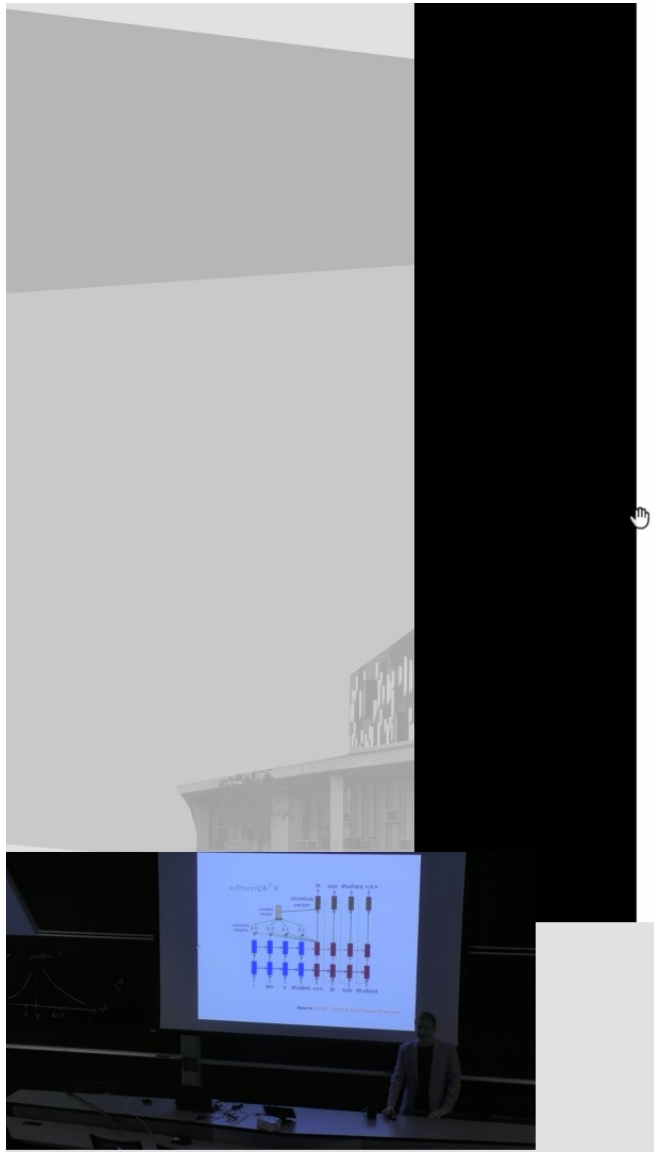
Who feared (*advocated*) violence?



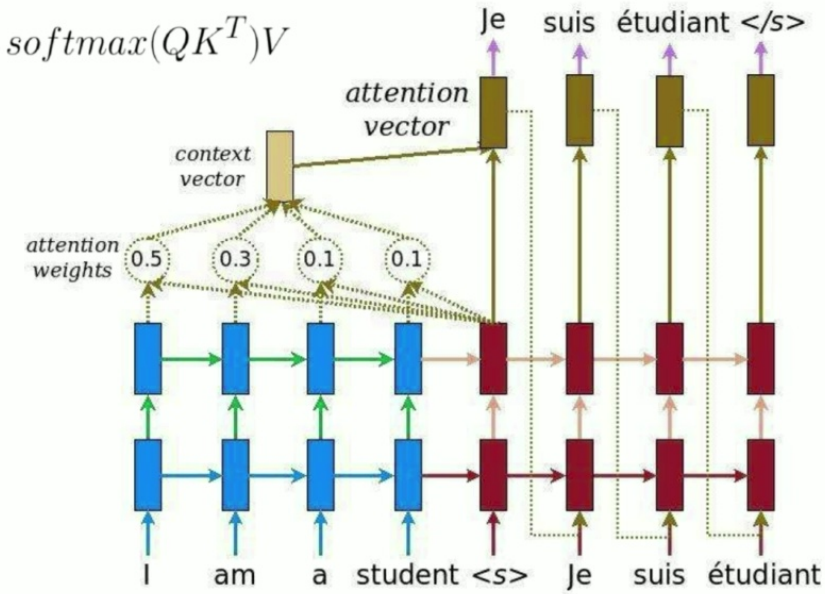

MT progress over time
[Edinburgh En-De WMT newstest2013 Cased BLEU; NMT 2015 from U. Montréal]
Phrase-based SMT
Syntax-based SMT
Neural MT
27
20.3
13.5
6.8
0
2013
2014
2015
2016
Source: http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf


seq2seq: Encoder + Decoder
you asked us to call you back after last Friday
x1 x2 x3 x4 x5 x6 x7
h1 h2 h3 h4 h5 h6 h7
"meaning" vector
s1 s2 s3 s4 s5 s6 s7
y1 y2 y3 y4 y5 y6 y7
Wir hoffen jedoch, dass sie bei Ihrer Reiseplanung weiterhilft
"Translation model"
"Language model"
Need "context" vectors

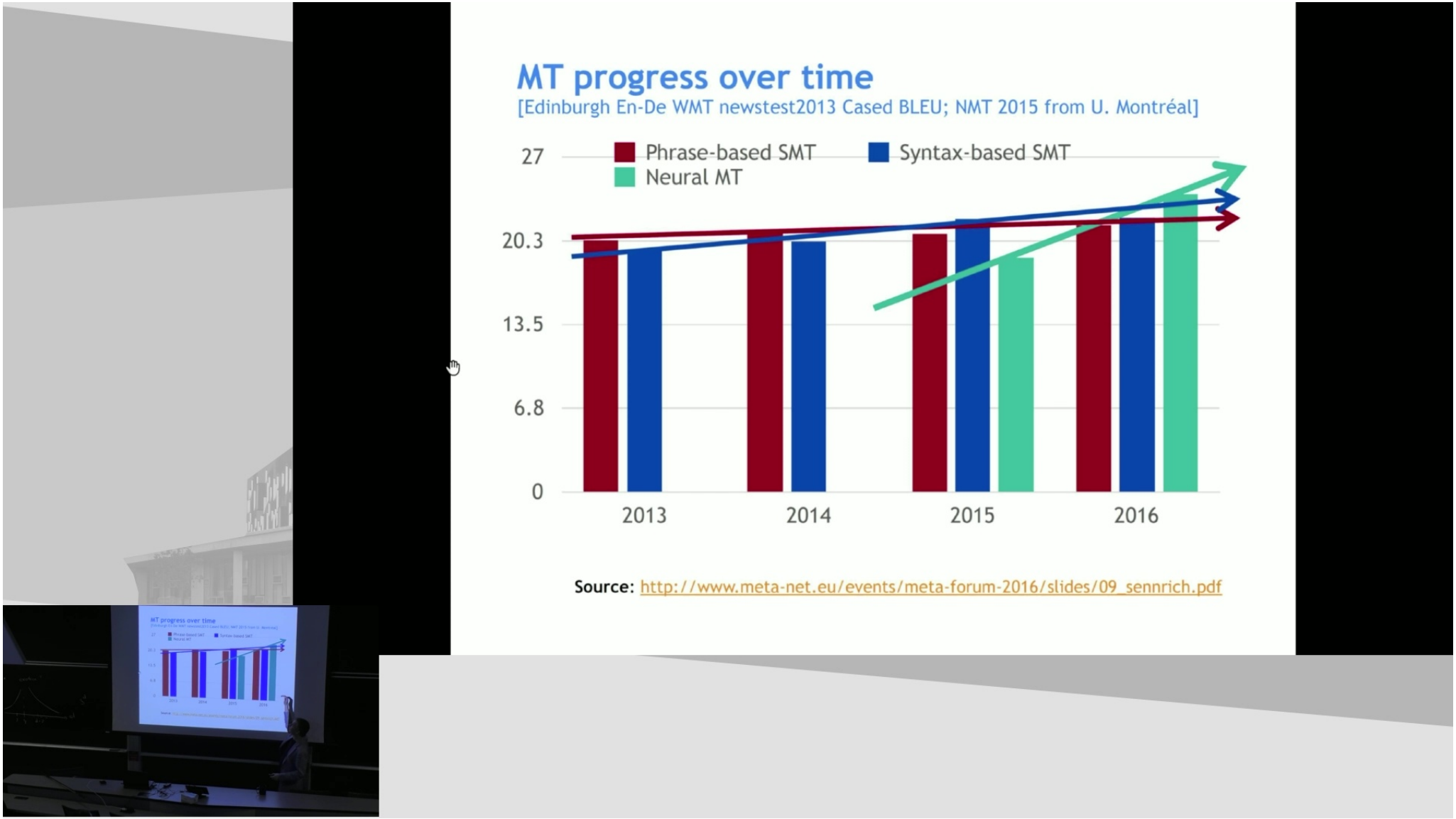

汀步路，
小心滑倒！
Slip and fall down carefully!
AIRPORTS AUTHORITY OF INDIA
फर्श पर खाना सख़्त मना है ।
—आदेश से
EATING CARPET
STRICTLY PROHIBITED
— By Order
请勿食用
Please don't be edible
外带食物
Exterior Girdle food
Bengali
Russian
Portuguese
Arabic
Hindi
Spanish
Mandarin
German
Italian
French
Translate

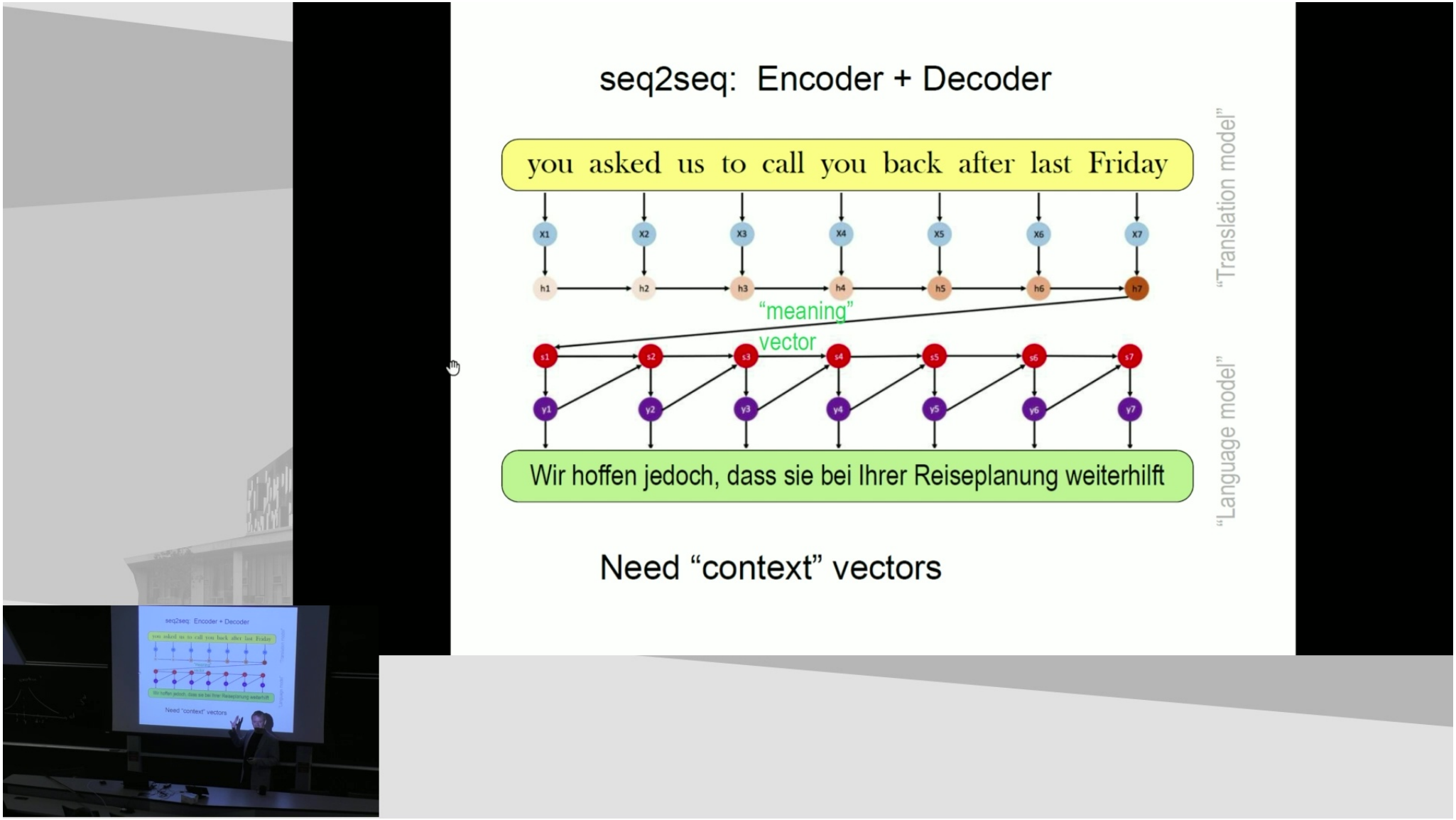


$softmax(QK^T)V$

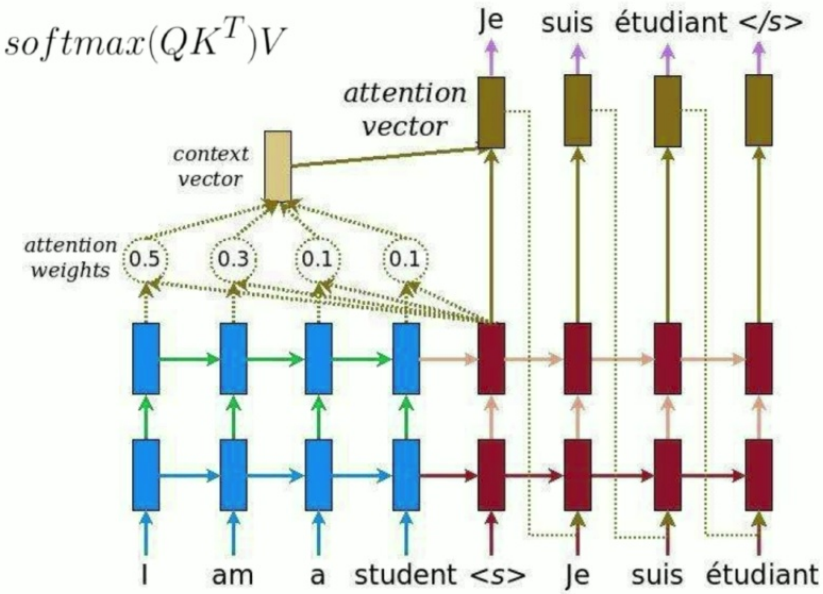


Source: https://github.com/tensorflow/nmt

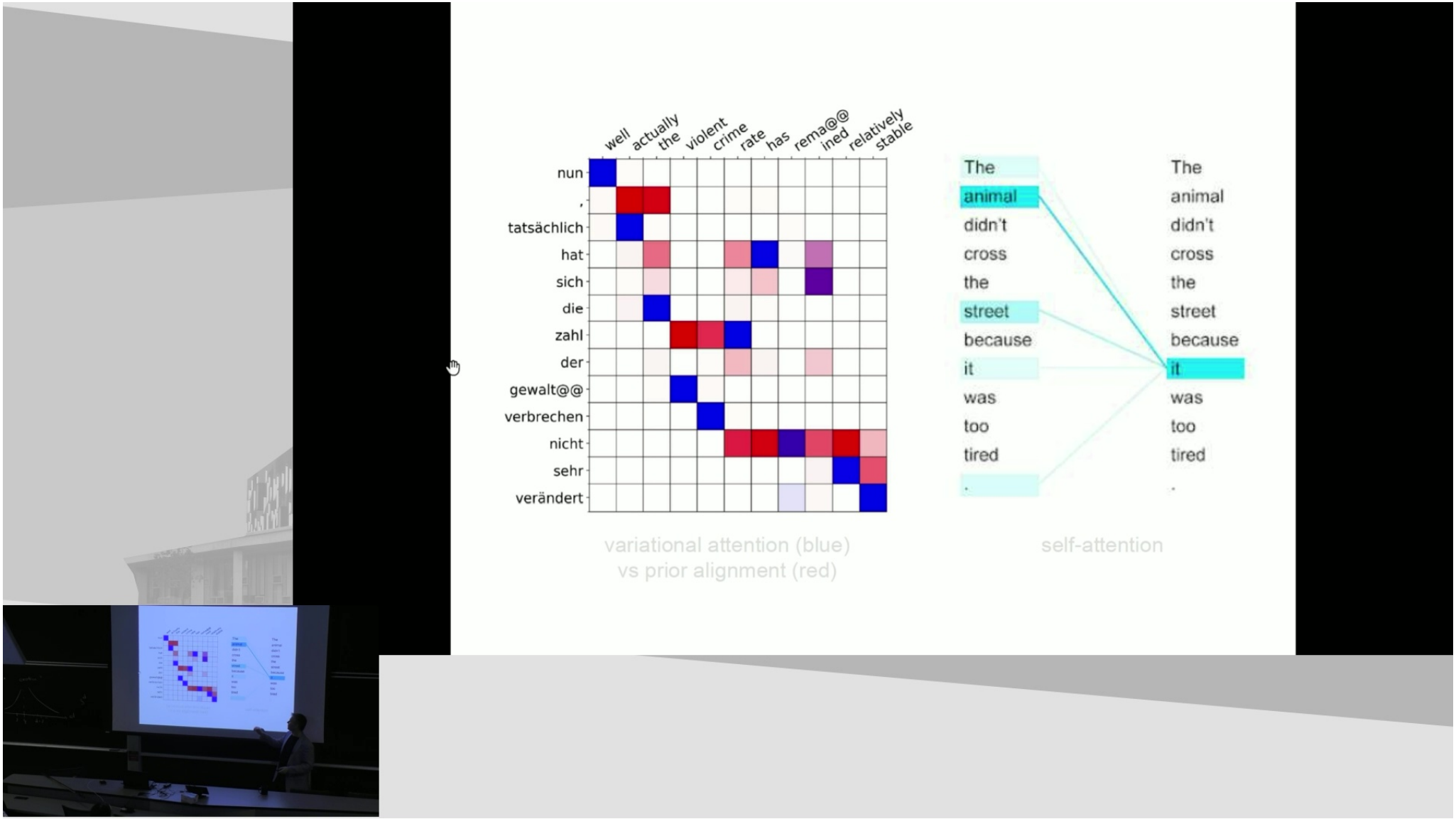

MT progress over time
[Edinburgh En-De WMT newstest2013 Cased BLEU; NMT 2015 from U. Montréal]
Phrase-based SMT
Syntax-based SMT
Neural MT
27
20.3
13.5
6.8
0
2013
2014
2015
2016
Source: http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf


seq2seq: Encoder + Decoder
you asked us to call you back after last Friday
x1 x2 x3 x4 x5 x6 x7
h1 h2 h3 h4 h5 h6 h7
"meaning" vector
s1 s2 s3 s4 s5 s6 s7
y1 y2 y3 y4 y5 y6 y7
Wir hoffen jedoch, dass sie bei Ihrer Reiseplanung weiterhilft
"Translation model"
"Language model"
Need "context" vectors

$softmax(QK^T)V$



Source: https://github.com/tensorflow/nmt


well
actually
the
violent
crime
rate
has
rema@@
ined
relatively
stable
nun
,
tatsächlich
hat
sich
die
zahl
der
gewalt@@
verbrechen
nicht
sehr
verändert
variational attention (blue)
vs prior alignment (red)
The
animal
didn't
cross
the
street
because
it
was
too
tired
.
The
animal
didn't
cross
the
street
because
it
was
too
tired
.
self-attention

Attention Is All You Need

Ashish Vaswani[*] Google Brain avaswani@google.com

Noam Shazeer[*] Google Brain noam@google.com

Niki Parmar[*] Google Research nikip@google.com

Jakob Uszkoreit[*] Google Research usz@google.com

Llion Jones[*] Google Research llion@google.com

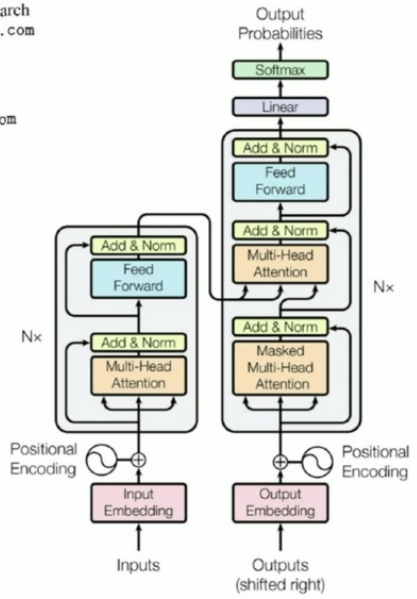
Aidan N. Gomez[*] [†] University of Toronto aidan@cs.toronto.edu

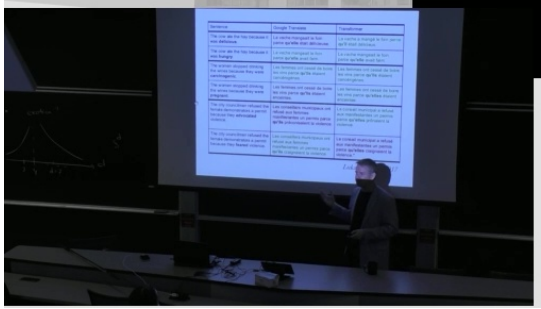
Łukasz Kaiser[*] Google Brain lukaszkaiser@google.com

Illia Polosukhin[*] [‡] illia.polosukhin@gmail.com

Abstract

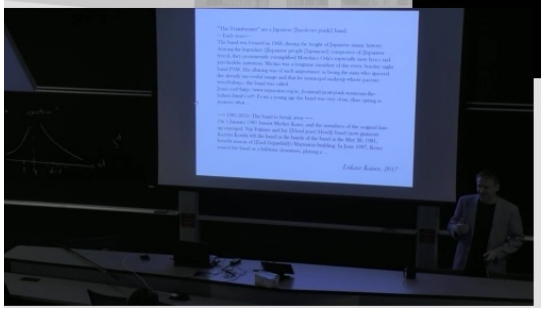
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.





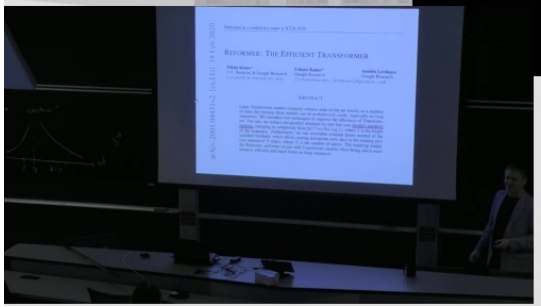
Sentence	Google Translate	Transformer
The cow ate the hay because it was **delicious**.	La vache mangeait le foin parce **qu'elle** était délicieuse.	La vache a mangé le foin parce **qu'il** était délicieux.
The cow ate the hay because it was **hungry**.	La vache mangeait le foin parce **qu'elle** avait faim.	La vache mangeait le foin parce **qu'elle** avait faim.
The women stopped drinking the wines because they were **carcinogenic**.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient cancérogènes.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient cancérigènes.
The women stopped drinking the wines because they were **pregnant**.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient enceintes.	Les femmes ont cessé de boire les vins parce **qu'elles** étaient enceintes.
The city councilmen refused the female demonstrators a permit because they **advocated** violence.	Les conseillers municipaux ont refusé aux femmes manifestantes un permis parce **qu'ils** préconisaient la violence.	Le conseil municipal a refusé aux manifestantes un permis parce **qu'elles** prônaient la violence.
The city councilmen refused the female demonstrators a permit because they **feared** violence.	Les conseillers municipaux ont refusé aux femmes manifestantes un permis parce **qu'ils** craignaient la violence	Le conseil municipal a refusé aux manifestantes un permis parce **qu'elles** craignaient la violence.*

Lukasz Kaiser, 2017



"The Transformer" are a Japanese [[hardcore punk]] band.
==Early years==
The band was formed in 1968, during the height of Japanese music history. Among the legendary [[Japanese people|Japanese]] composers of [Japanese lyrics], they prominently exemplified Motohiro Oda's especially tasty lyrics and psychedelic intention. Michio was a longtime member of the every Sunday night band PSM. His alluring was of such importance as being the man who ignored the already successful image and that he municipal makeup whose parents were - the band was called Jenei.<ref>http://www.separatist.org/se_frontend/post-punk-musician-the-kidney.html</ref> From a young age the band was very close, thus opting to pioneer what ...
:
=== 1981-2010: The band to break away ===
On 1 January 1981 bassist Michio Kono, and the members of the original line-up emerged. Niji Fukune and his [[Head poet|Head]] band (now guitarist) Kazuya Kouda left the band in the hands of the band at the May 28, 1981, benefit season of [[Led Zeppelin]]'s Marmarin building. In June 1987, Kono joined the band as a full-time drummer, playing a ...

Lukasz Kaiser, 2017



Published as a conference paper at ICLR 2020

REFORMER: THE EFFICIENT TRANSFORMER

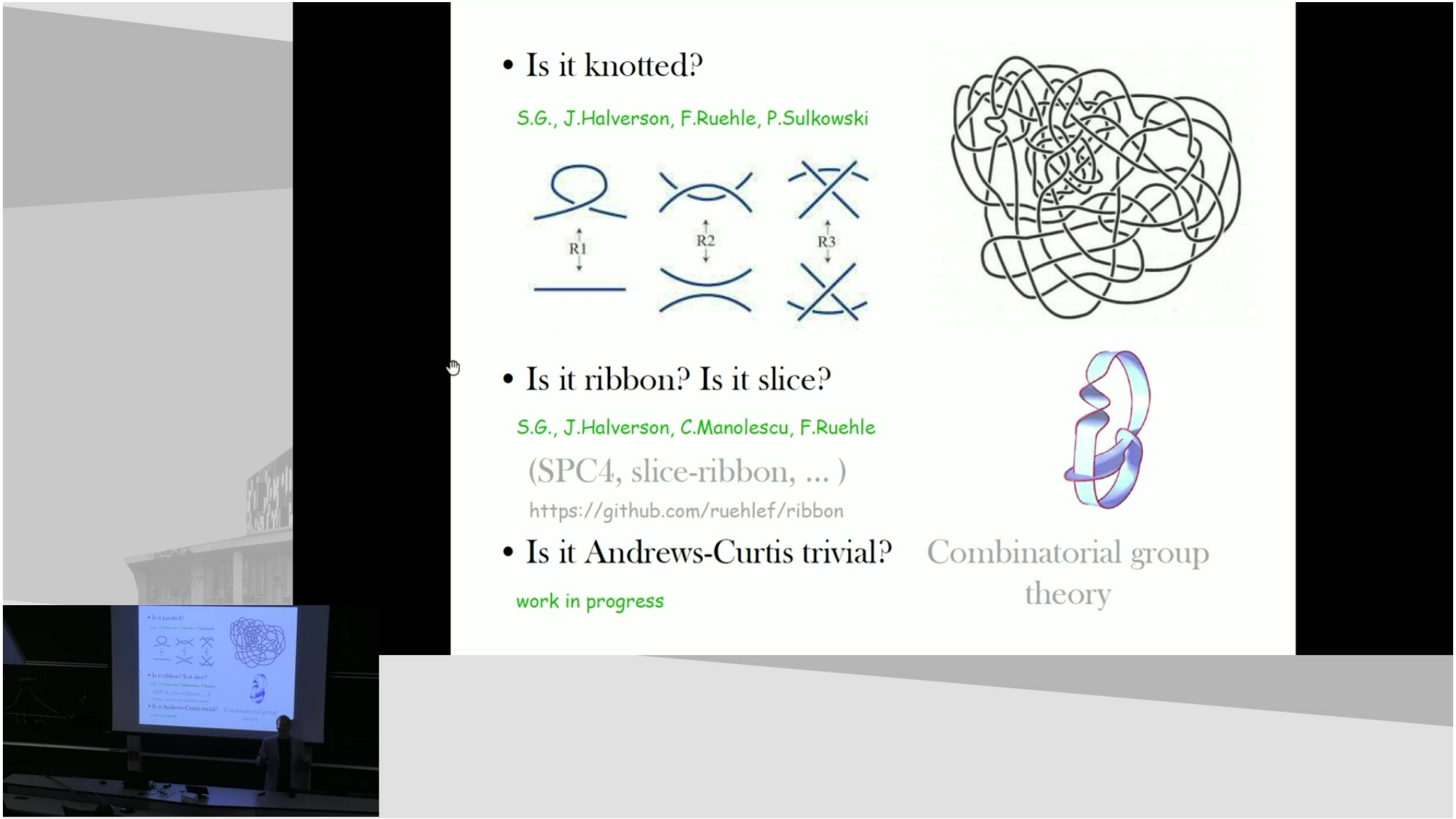
Nikita Kitaev*
U.C. Berkeley & Google Research
kitaev@cs.berkeley.edu

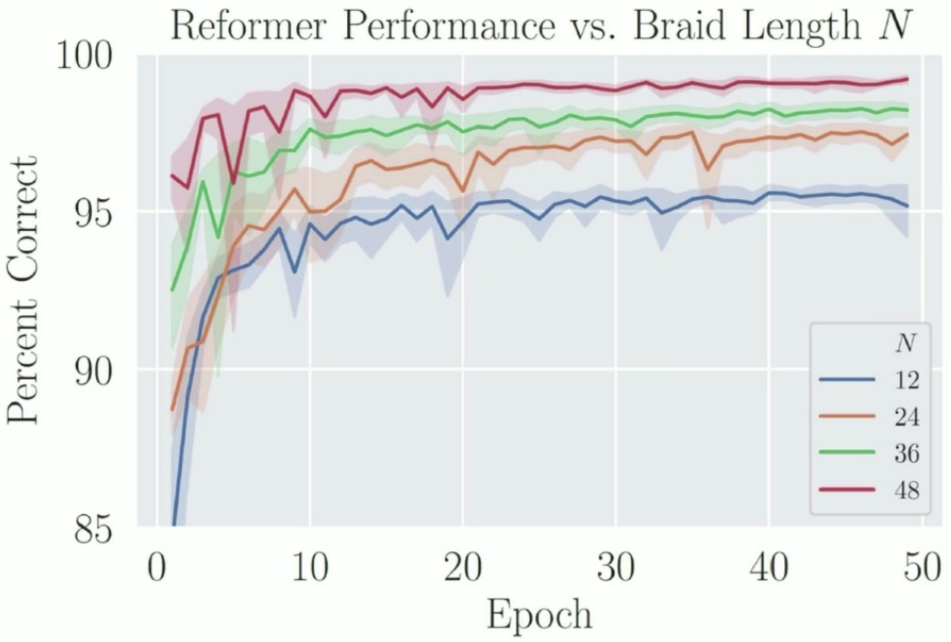
Łukasz Kaiser*
Google Research
{lukaszkaiser,levskaya}@google.com

Anselm Levskaya
Google Research

ABSTRACT

Large Transformer models routinely achieve state-of-the-art results on a number of tasks but training these models can be prohibitively costly, especially on long sequences. We introduce two techniques to improve the efficiency of Transformers. For one, we replace dot-product attention by one that uses locality-sensitive hashing, changing its complexity from $O(L^2)$ to $O(L \log L)$, where L is the length of the sequence. Furthermore, we use reversible residual layers instead of the standard residuals, which allows storing activations only once in the training process instead of N times, where N is the number of layers. The resulting model, the Reformer, performs on par with Transformer models while being much more memory-efficient and much faster on long sequences.

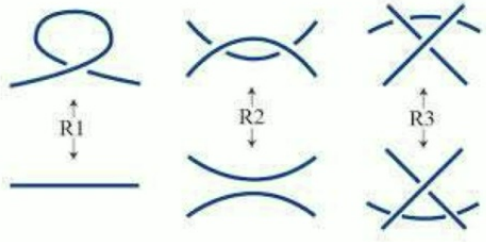

• Is it knotted?
S.G., J.Halverson, F.Ruehle, P.Sulkowski
R1
R2
R3
• Is it ribbon? Is it slice?
S.G., J.Halverson, C.Manolescu, F.Ruehle
(SPC4, slice-ribbon, ...)
https://github.com/ruehlef/ribbon
• Is it Andrews-Curtis trivial?
work in progress
Combinatorial group theory

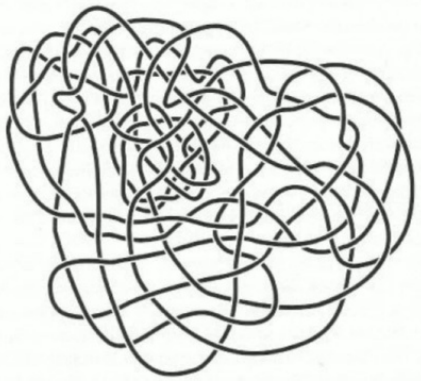


Reformer performance on UNKNOT as function of braid length. Performance increases with N.

- Is it knotted?

S.G., J.Halverson, F.Ruehle, P.Sulkowski





- Is it ribbon? Is it slice?

S.G., J.Halverson, C.Manolescu, F.Ruehle

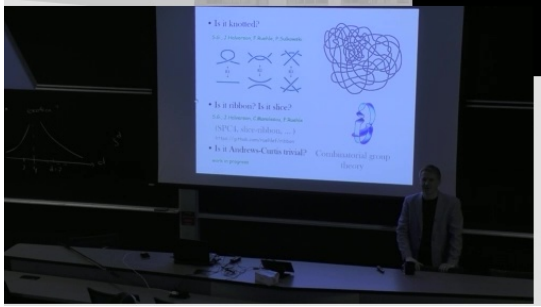
(SPC4, slice-ribbon, ...)

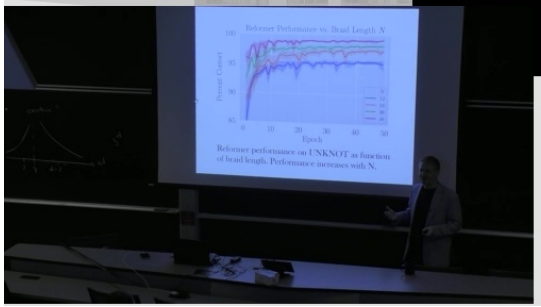
https://github.com/ruehlef/ribbon

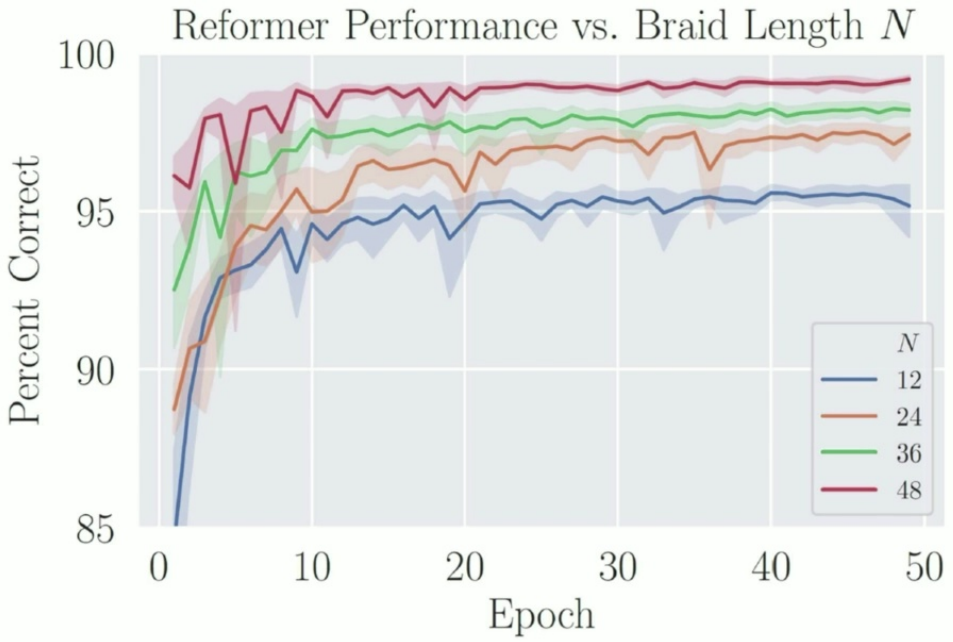
- Is it Andrews-Curtis trivial?

work in progress

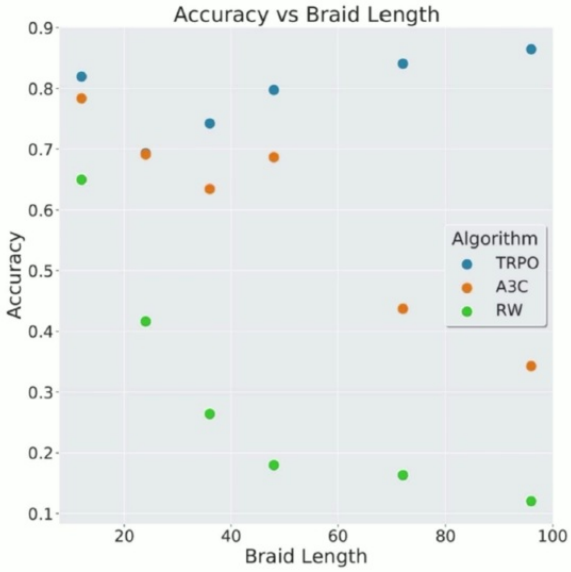
Combinatorial group theory



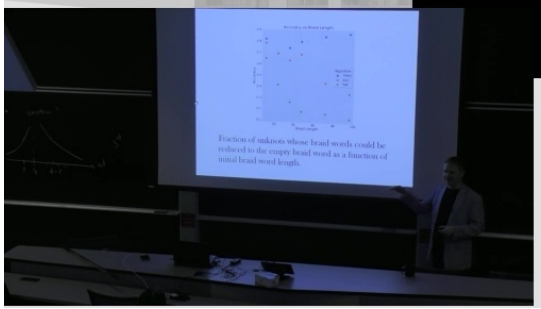


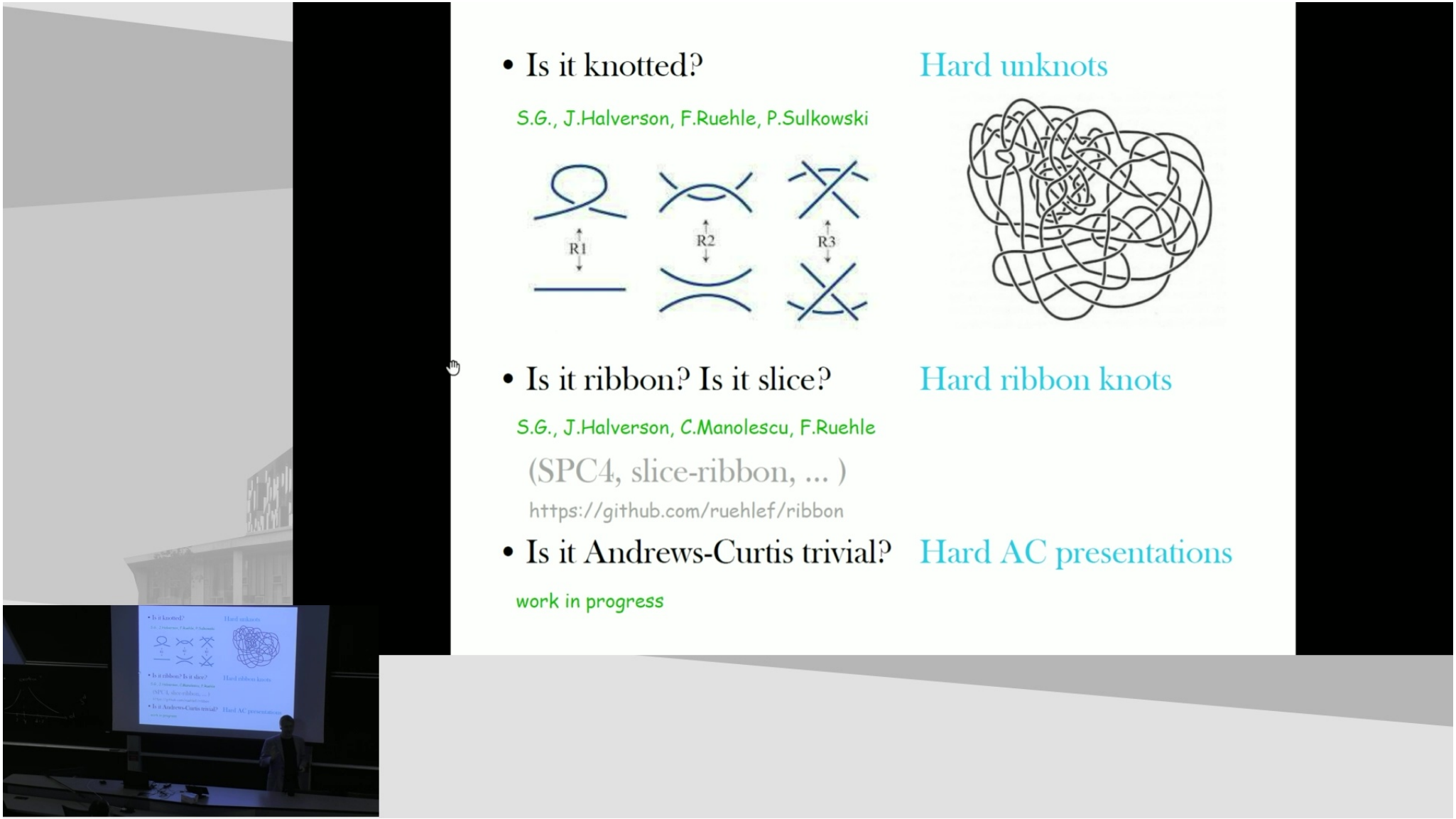


Reformer performance on UNKNOT as function of braid length. Performance increases with N.



Fraction of unknots whose braid words could be reduced to the empty braid word as a function of initial braid word length.




• Is it knotted?
Hard unknots
S.G., J.Halverson, F.Ruehle, P.Sulkowski
R1
R2
R3
• Is it ribbon? Is it slice?
Hard ribbon knots
S.G., J.Halverson, C.Manolescu, F.Ruehle
(SPC4, slice-ribbon, ...)
https://github.com/ruehlef/ribbon
• Is it Andrews-Curtis trivial?
Hard AC presentations
work in progress