AB GRAD CON23

1
00:00:13,749 --> 00:00:11,380
[Music]

2
00:00:17,210 --> 00:00:13,759
[Applause]

3
00:00:18,769 --> 00:00:17,220
hello everyone yeah so those you have

4
00:00:20,330 --> 00:00:18,779
who haven't met me yet yeah my name is

5
00:00:22,189 --> 00:00:20,340
Gage and I'm going to be talking about

6
00:00:24,710 --> 00:00:22,199
assembly Theory and in particular

7
00:00:26,570 --> 00:00:24,720
strings and hopefully by the end of this

8
00:00:28,250 --> 00:00:26,580
I will have convinced you that it is

9
00:00:29,570 --> 00:00:28,260
very useful for astrobiology in

10
00:00:32,389 --> 00:00:29,580
particular

11
00:00:34,130 --> 00:00:32,399
so in assembly Theory the primary

12
00:00:36,350 --> 00:00:34,140
quantity we're concerned with is called

13
00:00:38,270 --> 00:00:36,360

assembly index and it's a complexity

14
00:00:39,770 --> 00:00:38,280
measure that corresponds to the minimum

15
00:00:43,130 --> 00:00:39,780
number of joining operations that it

16
00:00:46,069 --> 00:00:43,140
takes to build an object so if my object

17
00:00:47,389 --> 00:00:46,079
is like X and I need at least three of

18
00:00:49,670 --> 00:00:47,399
these steps where I combine two things

19
00:00:52,069 --> 00:00:49,680
to make something new to get to X from

20
00:00:54,470 --> 00:00:52,079
some set of fundamental building blocks

21
00:00:55,549 --> 00:00:54,480
then we said the assembly index of X is

22
00:00:58,010 --> 00:00:55,559
three

23
00:00:59,510 --> 00:00:58,020
and so

24
00:01:01,130 --> 00:00:59,520
assembly index is kind of system

25
00:01:03,290 --> 00:01:01,140
dependent in much the same way that

26
00:01:04,789 --> 00:01:03,300
entropy is where for entropy I get to

27
00:01:07,010 --> 00:01:04,799
choose what my micro and macro states

28
00:01:08,929 --> 00:01:07,020
are and this influences exactly what it

29
00:01:10,490 --> 00:01:08,939
means and for assembly index these

30
00:01:12,289 --> 00:01:10,500
things that we get to choose are what

31
00:01:14,450 --> 00:01:12,299
our fundamental building blocks are and

32
00:01:16,730 --> 00:01:14,460
what the different joining operations

33
00:01:18,590 --> 00:01:16,740
that are allowable are

34
00:01:20,870 --> 00:01:18,600
and so

35
00:01:22,609 --> 00:01:20,880
we

36
00:01:24,710 --> 00:01:22,619
kind of philosophically what this thing

37
00:01:27,350 --> 00:01:24,720
means is that if I find some really

38
00:01:29,090 --> 00:01:27,360

complex object let's call it a and it

39
00:01:30,050 --> 00:01:29,100
has an assembly index of

40
00:01:46,010 --> 00:01:30,060
n

41
00:01:49,370 --> 00:01:46,020
true because

42
00:01:50,810 --> 00:01:49,380
for objects even as small as like modest

43
00:01:52,789 --> 00:01:50,820
size proteins there's not enough like

44
00:01:55,010 --> 00:01:52,799
matter on Earth to make all of them in

45
00:01:57,889 --> 00:01:55,020
any kind of meaningful abundance

46
00:02:00,109 --> 00:01:57,899
and so this approach kind of lives

47
00:02:01,010 --> 00:02:00,119
between two different worlds so one is

48
00:02:03,050 --> 00:02:01,020
like

49
00:02:04,609 --> 00:02:03,060
you know I might want to know the

50
00:02:06,469 --> 00:02:04,619
explicit physics or chemistry that's

51
00:02:07,910 --> 00:02:06,479
going on and know exactly how these

52
00:02:09,650 --> 00:02:07,920
things are actually built every time and

53
00:02:11,930 --> 00:02:09,660
understand those pathways

54
00:02:13,729 --> 00:02:11,940
but in assembly Theory we just take

55
00:02:15,290 --> 00:02:13,739
what's the minimal like hypothetical

56
00:02:17,150 --> 00:02:15,300
thing that we can't rule out as

57
00:02:18,410 --> 00:02:17,160
impossible and this is a massive

58
00:02:19,610 --> 00:02:18,420
simplification so we don't have to

59
00:02:21,530 --> 00:02:19,620
understand all the microphysical

60
00:02:24,050 --> 00:02:21,540
properties but we can still put lower

61
00:02:27,050 --> 00:02:24,060
bounds on how much it can take to make

62
00:02:28,250 --> 00:02:27,060
these objects on the other end of the

63
00:02:30,710 --> 00:02:28,260

spectrum it kind of varies from

64
00:02:33,110 --> 00:02:30,720
algorithmic information Theory where I

65
00:02:35,449 --> 00:02:33,120
would instead say you know something is

66
00:02:37,250 --> 00:02:35,459
so simple based off of how small of a

67
00:02:38,930 --> 00:02:37,260
program I can write that creates it in a

68
00:02:40,970 --> 00:02:38,940
computer language but in assembly Theory

69
00:02:42,290 --> 00:02:40,980
we're requiring it every step that that

70
00:02:44,710 --> 00:02:42,300
everything that's happening is

71
00:02:46,850 --> 00:02:44,720
physically possible

72
00:02:49,790 --> 00:02:46,860
and so

73
00:02:51,710 --> 00:02:49,800
uh strings are biologically relevant as

74
00:02:53,270 --> 00:02:51,720
as we are all probably quite familiar

75
00:02:55,070 --> 00:02:53,280
with it can be encoded in like two

76
00:02:58,490 --> 00:02:55,080
different alphabets so I can write it in

77
00:03:02,330 --> 00:02:58,500
like nucleotides or in amino acids

78
00:03:04,790 --> 00:03:02,340
and like many other fields of physics

79
00:03:06,530 --> 00:03:04,800
setting strings is like studying

80
00:03:08,210 --> 00:03:06,540
symmetries and so this is kind of a

81
00:03:10,130 --> 00:03:08,220
diagram here of this top string of where

82
00:03:11,149 --> 00:03:10,140
different symmetric sub strings are

83
00:03:13,009 --> 00:03:11,159
within it

84
00:03:15,350 --> 00:03:13,019
and when you look at this these

85
00:03:16,490 --> 00:03:15,360
correspond to meaningful properties in

86
00:03:18,649 --> 00:03:16,500
the ways you can build it in the

87
00:03:21,830 --> 00:03:18,659
shortest paths so

88
00:03:24,290 --> 00:03:21,840

one feature I want to point out is that

89
00:03:26,509 --> 00:03:24,300
so in this bottom left bit here I have

90
00:03:28,550 --> 00:03:26,519
this one two segment that I'm doubling

91
00:03:30,649 --> 00:03:28,560
with itself into one two one two and we

92
00:03:32,030 --> 00:03:30,659
call this kind of innovation-like in

93
00:03:33,830 --> 00:03:32,040
that it's like I have the blueprints for

94
00:03:36,649 --> 00:03:33,840
it and I don't need to rebuild it from

95
00:03:38,990 --> 00:03:36,659
scratch every time

96
00:03:40,369 --> 00:03:39,000
so this has a lot of really nice

97
00:03:42,589 --> 00:03:40,379
mathematical properties and it's a

98
00:03:45,110 --> 00:03:42,599
really rich subject so like

99
00:03:46,729 --> 00:03:45,120
this thing on the bottom looks like like

100
00:03:49,490 --> 00:03:46,739
ice cream with sprinkles in it but it's

101
00:03:51,530 --> 00:03:49,500
it's a some kind of weird graph which

102
00:03:53,110 --> 00:03:51,540
actually corresponds to like how you

103
00:03:55,550 --> 00:03:53,120
reconcile all these symmetries together

104
00:03:57,350 --> 00:03:55,560
uh to calculate like what the shortest

105
00:03:58,729 --> 00:03:57,360
build path is and it has a really nice

106
00:04:01,190 --> 00:03:58,739
relation to some like really classic

107
00:04:02,809 --> 00:04:01,200
problems in computation theory that tell

108
00:04:04,729 --> 00:04:02,819
us about how complicated it is to

109
00:04:07,309 --> 00:04:04,739
calculate these things

110
00:04:10,250 --> 00:04:07,319
um but much more Salient to astrobiology

111
00:04:12,830 --> 00:04:10,260
is thinking about uh

112
00:04:15,470 --> 00:04:12,840
what we could call Prime strings or I

113
00:04:17,030 --> 00:04:15,480

would argue also materially efficient

114
00:04:19,729 --> 00:04:17,040
techno signatures

115
00:04:21,409 --> 00:04:19,739
so the example I want to kind of

116
00:04:22,790 --> 00:04:21,419
describe this as is it's just a series

117
00:04:25,310 --> 00:04:22,800
of numbers but what would it mean in

118
00:04:29,150 --> 00:04:25,320
maybe a material is if I restrict myself

119
00:04:30,770 --> 00:04:29,160
to working with polymer chains and my

120
00:04:33,170 --> 00:04:30,780
alphabet size is basically how many

121
00:04:36,050 --> 00:04:33,180
different monomers I have access to

122
00:04:38,210 --> 00:04:36,060
I want to find how do I efficiently

123
00:04:39,409 --> 00:04:38,220
create something that demonstrates I

124
00:04:40,550 --> 00:04:39,419
know how to compute a lot of different

125
00:04:42,890 --> 00:04:40,560
reactions

126
00:04:44,990 --> 00:04:42,900
then I need something that has as few

127
00:04:46,430 --> 00:04:45,000
symmetries in it as possible and there's

128
00:04:47,749 --> 00:04:46,440
actually really explicit ways to

129
00:04:52,610 --> 00:04:47,759
construct these things such that they

130
00:04:54,230 --> 00:04:52,620
have absolutely no I I can construct

131
00:04:56,150 --> 00:04:54,240
there's many different ways to build

132
00:04:58,550 --> 00:04:56,160
these strings but none of them reuse the

133
00:04:59,930 --> 00:04:58,560
exact same reaction twice by virtue of

134
00:05:01,670 --> 00:04:59,940
just what they are

135
00:05:02,990 --> 00:05:01,680
and that's a property introduced to

136
00:05:04,550 --> 00:05:03,000
these objects and this also kind of

137
00:05:06,350 --> 00:05:04,560
highlights the difference between this

138
00:05:07,670 --> 00:05:06,360

and algorithmic information Theory

139
00:05:10,129 --> 00:05:07,680
because like

140
00:05:11,930 --> 00:05:10,139
if I asked anybody to like make the next

141
00:05:13,010 --> 00:05:11,940
one you'd probably all get it right on

142
00:05:14,450 --> 00:05:13,020
the first try there's a lot of

143
00:05:15,890 --> 00:05:14,460
regularity in what these objects look

144
00:05:17,570 --> 00:05:15,900
like and so algorithmically they're not

145
00:05:19,010 --> 00:05:17,580
that complicated but in terms of the

146
00:05:22,249 --> 00:05:19,020
physical reactions you need to make them

147
00:05:27,469 --> 00:05:23,770
so

148
00:05:29,570 --> 00:05:27,479
future and that what I'm really excited

149
00:05:32,150 --> 00:05:29,580
to do with this is have a dynamics of

150
00:05:33,950 --> 00:05:32,160
how life can explore the chemical space

151
00:05:36,350 --> 00:05:33,960
of what's possible

152
00:05:38,330 --> 00:05:36,360
and so the picture I want to illustrate

153
00:05:40,249 --> 00:05:38,340
these are not real data this is just to

154
00:05:43,909 --> 00:05:40,259
illustrate a point is I want to think

155
00:05:46,249 --> 00:05:43,919
about some space of objects and I think

156
00:05:48,950 --> 00:05:46,259
a good example is all the proteins that

157
00:05:50,749 --> 00:05:48,960
a species creates say so we take all

158
00:05:52,570 --> 00:05:50,759
these proteins and I want to assign

159
00:05:54,890 --> 00:05:52,580
distances between them all and

160
00:05:55,969 --> 00:05:54,900
qualitatively I mean there's an example

161
00:05:58,490 --> 00:05:55,979
of how to do that on the top but

162
00:06:01,670 --> 00:05:58,500
qualitatively what we should think of is

163
00:06:04,010 --> 00:06:01,680

the distance between X and Y is the

164
00:06:05,990 --> 00:06:04,020
number of joining operations that I need

165
00:06:07,969 --> 00:06:06,000
to change to go from one to the other

166
00:06:09,529 --> 00:06:07,979
and so things that are intuitively

167
00:06:13,610 --> 00:06:09,539
similar will be closer

168
00:06:15,230 --> 00:06:13,620
and just yeah as you might expect and so

169
00:06:18,469 --> 00:06:15,240
this is really useful because when we

170
00:06:20,150 --> 00:06:18,479
think about what we have as a generative

171
00:06:22,550 --> 00:06:20,160
model for like a null hypothesis like

172
00:06:25,070 --> 00:06:22,560
what should life be doing we know that

173
00:06:26,990 --> 00:06:25,080
mutations are things that really only

174
00:06:28,490 --> 00:06:27,000
change assembly index by one either I'm

175
00:06:30,050 --> 00:06:28,500
inserting some large segment that

176
00:06:31,790 --> 00:06:30,060
already existed and so I didn't have to

177
00:06:33,650 --> 00:06:31,800
build it from scratch or I'm making some

178
00:06:36,529 --> 00:06:33,660
really small change and in any case this

179
00:06:37,730 --> 00:06:36,539
is a very a change in assembly index of

180
00:06:40,010 --> 00:06:37,740
just one

181
00:06:42,110 --> 00:06:40,020
and so we have this kind of

182
00:06:44,330 --> 00:06:42,120
thick surface of the adjacent possible

183
00:06:46,969 --> 00:06:44,340
like things you can mutate into and

184
00:06:49,430 --> 00:06:46,979
because it's reasonable to imagine that

185
00:06:51,529 --> 00:06:49,440
this is happening kind of randomly we

186
00:06:53,330 --> 00:06:51,539
can have some kind of no model for what

187
00:06:55,670 --> 00:06:53,340
we expect life to look like like what's

188
00:06:56,990 --> 00:06:55,680

the shape of this object as it grows in

189
00:06:58,490 --> 00:06:57,000
time

190
00:07:00,529 --> 00:06:58,500
and

191
00:07:02,210 --> 00:07:00,539
so I like to think of it as kind of

192
00:07:04,129 --> 00:07:02,220
spherical but actually it's probably a

193
00:07:06,469 --> 00:07:04,139
little bit noisy just because it is so

194
00:07:08,510 --> 00:07:06,479
Random but in reality I don't think

195
00:07:10,189 --> 00:07:08,520
we'll really observe this because some

196
00:07:12,230 --> 00:07:10,199
mutations will be selected out and will

197
00:07:14,090 --> 00:07:12,240
never observe them and so the way I want

198
00:07:16,430 --> 00:07:14,100
to illustrate this is this you have this

199
00:07:18,830 --> 00:07:16,440
kind of noisy sphere on the left and you

200
00:07:21,589 --> 00:07:18,840
have a much noisier version of it on the

201
00:07:23,629 --> 00:07:21,599
right and you know if I observe this for

202
00:07:25,249 --> 00:07:23,639
two different things the interpretation

203
00:07:27,230 --> 00:07:25,259
of this is we're seeing something about

204
00:07:29,089 --> 00:07:27,240
like the fitness landscape like this is

205
00:07:31,670 --> 00:07:29,099
kind of the shadow that makes it through

206
00:07:33,770 --> 00:07:31,680
right so like there's lots of voids here

207
00:07:38,270 --> 00:07:33,780
of things that are either not functional

208
00:07:40,610 --> 00:07:38,280
or even hurtful to the system and

209
00:07:42,529 --> 00:07:40,620
by studying like the geometry of these

210
00:07:43,909 --> 00:07:42,539
kinds of objects and looking for the

211
00:07:47,270 --> 00:07:43,919
ways they scale we can learn things

212
00:07:49,610 --> 00:07:47,280
about the ergodicity of life meaning

213
00:07:50,950 --> 00:07:49,620

if I were to restart us from like early

214
00:07:52,969 --> 00:07:50,960
Earth times

215
00:07:55,129 --> 00:07:52,979
how similar would I expect the

216
00:07:57,409 --> 00:07:55,139
biochemistry to be today that we observe

217
00:07:58,610 --> 00:07:57,419
just by virtue of how many ways does it

218
00:08:01,129 --> 00:07:58,620
seem like there are to actually move in

219
00:08:03,050 --> 00:08:01,139
this plate in this space and by

220
00:08:05,629 --> 00:08:03,060
extension we can make inferences about

221
00:08:08,450 --> 00:08:05,639
just life in general and how

222
00:08:11,870 --> 00:08:08,460
constricted it must be and how diverse

223
00:08:14,990 --> 00:08:11,880
will potential alien biochemistries be

224
00:08:17,809 --> 00:08:15,000
so with that I'll leave you with a

225
00:08:24,420 --> 00:08:17,819
fractal that reminds me of the building

226
00:08:39,170 --> 00:08:34,120
[Music]

227
00:08:41,290 --> 00:08:39,180
so how can this be employed to sort of

228
00:08:43,490 --> 00:08:41,300
create a Criterion for what constitutes

229
00:08:47,090 --> 00:08:43,500
abiogenesis based off the assembly

230
00:08:48,769 --> 00:08:47,100
complexes so there is an excellent

231
00:08:52,490 --> 00:08:48,779
nature comes paper I want to say from

232
00:08:54,650 --> 00:08:52,500
2020 it's Marshall's the first author uh

233
00:08:57,490 --> 00:08:54,660
where they talk about so you can do this

234
00:08:59,570 --> 00:08:57,500
for molecules and they find a really big

235
00:09:01,670 --> 00:08:59,580
disparity between what kinds of things

236
00:09:04,190 --> 00:09:01,680
abiotic sources can produce versus

237
00:09:06,110 --> 00:09:04,200
biotic and this is something that you

238
00:09:08,509 --> 00:09:06,120

can measure quite accurately with like

239
00:09:11,509 --> 00:09:08,519
Mass Spec and some other methods too and

240
00:09:14,090 --> 00:09:11,519
so it just for a chemical biosignature

241
00:09:20,810 --> 00:09:14,100
way there's a concrete thing that

242
00:09:24,230 --> 00:09:22,790
hi um great talk and I was also just

243
00:09:26,329 --> 00:09:24,240
curious when you're evaluating the

244
00:09:27,769 --> 00:09:26,339
assembly theory is this based off of

245
00:09:30,949 --> 00:09:27,779
entirely the sequence or do you also

246
00:09:36,590 --> 00:09:33,470
I don't like 17 folding or something

247
00:09:38,509 --> 00:09:36,600
ah yes so this is kind of embedded in

248
00:09:39,889 --> 00:09:38,519
like our choice of the definition of the

249
00:09:41,389 --> 00:09:39,899
joining operation right so like the

250
00:09:43,490 --> 00:09:41,399
simplest Choice doesn't count the

251
00:09:45,110 --> 00:09:43,500
protein folding but you can imagine that

252
00:09:46,730 --> 00:09:45,120
I Define the joining operations such

253
00:09:49,850 --> 00:09:46,740
that I allow myself to consider those

254
00:09:52,370 --> 00:09:49,860
like high order constructions and so

255
00:10:01,730 --> 00:09:52,380
yeah it's a choice of your

256
00:10:06,050 --> 00:10:03,769
yeah uh yeah uh this is very interesting

257
00:10:07,490 --> 00:10:06,060
I'm wondering how do you consider the

258
00:10:08,750 --> 00:10:07,500
starting amino acids you have because

259
00:10:10,430 --> 00:10:08,760
there's a lot of theories on like what

260
00:10:12,110 --> 00:10:10,440
amino acids were like first or like

261
00:10:13,490 --> 00:10:12,120
which involved later and like are you

262
00:10:15,110 --> 00:10:13,500
only using the canonical ones that we

263
00:10:17,329 --> 00:10:15,120

have and like the current genetic code

264
00:10:20,810 --> 00:10:17,339
or like or any like theoretical amino

265
00:10:26,930 --> 00:10:22,790
I'm not uh there's someone at your table

266
00:10:30,650 --> 00:10:26,940
who does his name's Thomas

267
00:10:32,090 --> 00:10:30,660
um but I to to relate it to this I think

268
00:10:33,769 --> 00:10:32,100
this is kind of embedded in your choice

269
00:10:45,170 --> 00:10:33,779
of like what the fundamental building

270
00:10:45,180 --> 00:10:49,310
any other questions for cage