# The Perils of Implementing Code from PDFs

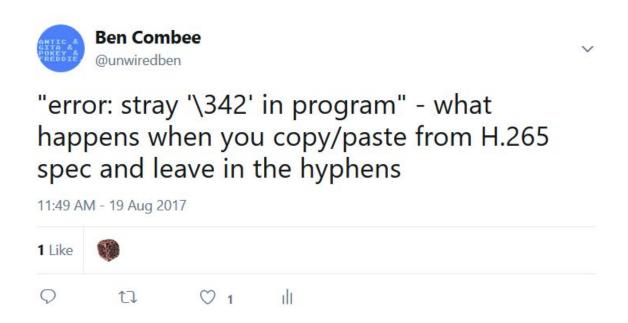
Ben Combee Sr. Software Developer, Roku

### I'm Ben

I work on streaming video and digital television technology for the Roku TV platform from our development office in Austin, TX.

We use a lot of C++ in our codebase.

#### This talk started with a tweet...



I needed to implement an algorithm from the H.265 spec used for 4K video to determine picture order.

If you don't get that correct, your closed captions can get all scrambled.

"hili tkes" instead of "like this"

The H.265 spec is a complex beast. It's the result of a very smart committee of Ph.Ds who precisely convert their working code into psuedo-English so it can be standardized.

## Here's a quote from 7.4.7.1 General slice segment header semantics

slice\_pic\_order\_cnt\_lsb specifies the picture order count modulo MaxPicOrderCntLsb for the current picture. The length of the slice\_pic\_order\_cnt\_lsb syntax element is log2\_max\_pic\_order\_cnt\_lsb\_minus4 + 4 bits. The value of the slice\_pic\_order\_cnt\_lsb shall be in the range of 0 to MaxPicOrderCntLsb - 1, inclusive. When slice\_pic\_order\_cnt\_lsb is not present, slice\_pic\_order\_cnt\_lsb is inferred to be equal to 0, except as specified in clause 8.3.3.1.

There's about 600 pages of this in Rec. ITU-T H.265 v4 (12/2016)

Some of it is in the form of tables and there are a few basic figures.

I started by pasting in data structure tables from the spec and turning that into bitstream reading code.

Ctrl-F (Find...) is your friend here.

The spec tends to use lots of unique identifiers with underscores.

Complex descriptions often are built on references to other parts of the specification, but it's up to you to build your own data dependency graph.

After getting all that figured out, it was time compile.

Hence the '\342' error.

'error: stray \342 in program`

0342 (oct) == 0xE2 (hex) == 'â'

But my code had a '-' in it. That's not a hyphen?

What gives?

### **UTF-8 FOR THE WIN**

Hyphen is U+2010

In UTF-8 representation, that's '0xE2 0x80 0x90'

But gcc doesn't understand UTF-8

### Other characters to check:

"Curly double quotes" 'Curly single quotes' em—dash & en–dash non-breaking spaces

Be careful if someone edits your code in Microsoft Word!

In the end, stray characters was only a minor problem, but it was a fun one.

I got my code working.

Want to know more about Roku? Find me on the 'cpplang' Slack or connect on Twitter at @unwiredben

