TS 18661-3 AS ANNEX FOR C2X

N2374
WG 14 - London
April 29 – May 3, 2019

C FP group
TS 18661-3

Types and functions to support IEC 60559 interchange and extended formats
IEC 60559 interchange formats

• IEC 60559:2011 specifies a “tower” of *interchange* formats
• Arbitrarily large widths (32x)
• For binary and decimal
• Balanced precision and range determined by width
• For exchange of FP data
• binary16, for GPU data, etc.
• Formats may be supported as
  • Arithmetic – with all standard operations
  • Non-arithmetic – with conversion operations
IEC 60559 extended formats

- IEEE specifies *extended* formats that extend its basic formats: binary32|64|128 and decimal64|128
- Have at least a specified precision and range
- For explicit wide evaluation
- Not for data exchange
TS 18661-3

- Three features
  - Interchange floating types
  - Extended floating types
  - Support for non-arithmetic interchange formats
- Full language and library support for interchange and extended floating types
- Conversion operations for non-arithmetic interchange formats represented in unsigned char arrays
TS 18661-3 – type structure extensions

interchange floating types: _FloatN, _DecimalN
extended floating types: _FloatNx, _DecimalNx

real floating types
  standard floating types: float, double, long double
  binary floating types: _FloatN, _FloatNx
  decimal floating types: _DecimalN, _DecimalNx

complex types
  float _Complex, double _Complex, long double _Complex
  _FloatN _Complex, _FloatNx _Complex

Imaginary types
  float _Imaginary, double _Imaginary, long double _Imaginary
  _FloatN _Imaginary, _FloatNx _Imaginary
TS 18661-3 – type structure unchanged

floating types
  real floating types
  complex types
  imaginary types

real types
  integer types
  real floating types

arithmetic types
  integer types
  floating types
TS 18661-3

- Standard binding for extension floating types with IEC 60559 formats, which are common extensions (e.g., float16, float128, float80)
- Facilitates exchange of FP data, without full support type
- Enables explicit wide evaluation, for robustness
Publication

TS 18661-3 as annex

- TS 18661-3 written as changes to C, on top of the changes in TS parts 1 and 2
- Recast as annex for C2X: N2342
- Per WG14 direction
- Original skepticism because TS 18661-3 affects so many parts of C, but …
Advantages of annex

• Shorter specification: 38 vs 73 pages
• More comprehensible: direct specification vs changes to changes to changes
• Avoids extensive changes to C, just

In 6.10.8.3#1, add:

__STDC_IEC_60559_TYPES__ The integer constant \texttt{20yymmL}, intended to indicate conformance to the specification in Annex X (IEC 60559 interchange and extended types).
Changing to annex

• Requirement: achieve equivalent specification

• Some direct specification, e.g.,
  Types designated _Float32x, _Float64x, _Float128x, _Decimal64x, and _Decimal128x support the corresponding IEC 60559 extended formats and are collectively called the extended floating types. …

• Some broadening of definitions, e.g.,
  The real floating types are broadened to include all interchange floating types and extended floating types, as well as standard floating types.

• Some additions to C, e.g.,
  This subclause expands floating-suffix (6.4.4.2) to also include:
    fN Fn fNx FNx dN DN dNxDNx

• Reorganized to better match C organization and for general order
Annex outline - language

- X.1 Introduction
- X.2 Types
  - X.2.1 Interchange floating types
  - X.2.2 Non-arithmetic interchange formats
  - X.2.3 Extended floating types
  - X.2.4 Classification of real floating types
  - X.2.5 Complex types
  - X.2.6 Imaginary types
- X.3 Characteristics in <float.h>
- X.4 Conversions
  - X.4.1 Real floating and integer
  - X.4.2 Usual arithmetic conversions
  - X.4.3 Arithmetic and non-arithmetic formats
- X.5 Lexical elements
  - X.5.1 Keywords
  - X.5.2 Constants
- X.6 Expressions
- X.7 Declarations
Annex outline - library

- X.8 Identifiers in standard headers
  - X.8.1 <float.h>
  - X.8.2 <complex.h>
  - X.8.3 <math.h>
  - X.8.4 <stdlib.h>
- X.9 Complex arithmetic <complex.h>
- X.10 Floating-point environment <fenv.h>
- X.11 Mathematics <math.h>
  - X.11.1 Macros
  - X.11.2 Function prototypes
  - X.11.3 Encoding conversion functions
    - X.11.3.1 Encode and decode functions
    - X.11.3.2 Encoding-to-encoding conversion functions
- X.12 Numeric conversion functions in <stdlib.h>
  - X.12.1 String from floating
  - X.12.2 String to floating
  - X.12.3 String from encoding
  - X.12.4 String to encoding
- X.13 Type-generic macros <tgmath.h>
Implementation

• GCC supports _FloatN and _FloatNx types (including with _Complex) on multiple systems.
• Several C implementations have provided additional floating-point types as extensions. For examples, HPUX C/C++ has a fourth type with the IEC 60559 double64-extended format, and LCC supports float128_t and qfloat.