Formalisation de la norme IEEE754 dans Coq

L. Théry

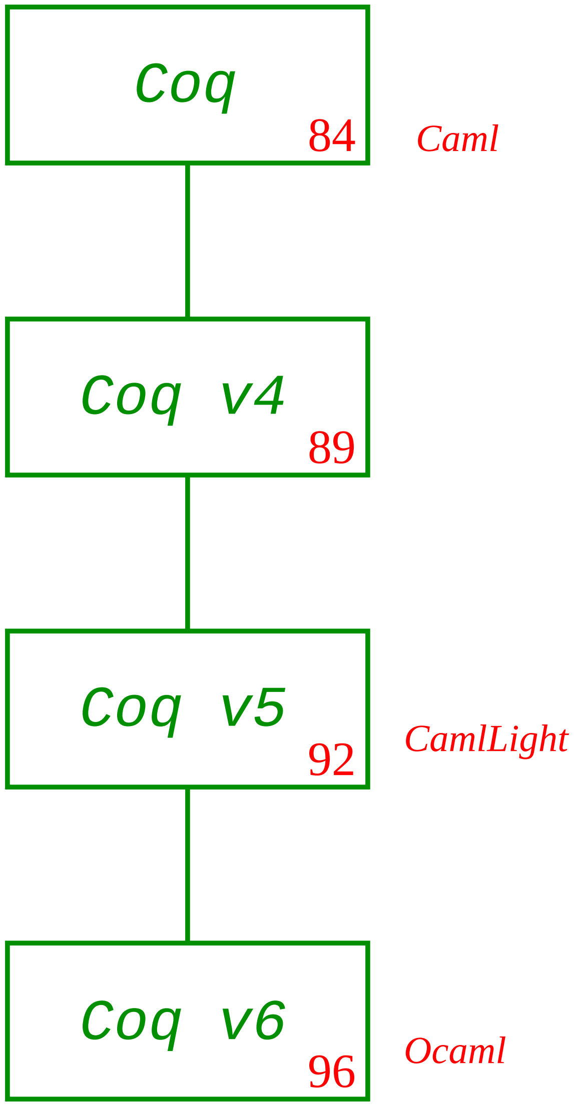
26 Juin 2000

1 Coq

Coq

Coq: *C*alculus *Of* *C*onstruction

T.Coquand, G. Huet, C. Paulin, ... INRIA



Coq (suite)

Définir des objets:

Inductive $list[A : Set] : Set :=$
$Nil : (list\ A)$
$|\ Cons : A \rightarrow (list\ A) \rightarrow (list\ A)$.

$$
\begin{array}{rcl}
(Nil\ \ A) & \longrightarrow & [] \\
(Cons\ \ A\ \ a\ \ L) & \longrightarrow & [a|L] \\
(Cons\ \ A\ \ a\ \ (Cons\ \ A\ \ b\ \ (Nil\ \ A))) & \longrightarrow & [a;b]
\end{array}
$$

Définir des fonctions:

Fixpoint $length\,[A : Set;\ L :\ (list\ \ A)] :\ \ (list\ \ A) :=$
Cases L **of**
$[] \Longrightarrow 0$
$|\ [x|L_1] \Longrightarrow 1 + (length\ \ L_1)$
end.

Coq (suite)

Définir des propriétés:

$$
\begin{array}{ll}
\textbf{Inductive} \equiv : (\mathit{list}\ A) \to (\mathit{list}\ A) \to \mathit{Prop} := & \\
\quad \mathit{nil} : [] \equiv [] & \\
\mid \quad \mathit{skip} : \forall a \colon A.\, \forall L_1, L_2 \colon (\mathit{list}\ A). & \\
\qquad\qquad L_1 \equiv L_2 \Rightarrow [a|L_1] \equiv [a|L_2] & \\
\mid \quad \mathit{swap} : \forall a, b \colon A.\, \forall L \colon (\mathit{list}\ A).\, [a; b|L] \equiv [b; a|L] & \\
\mid \quad \mathit{trans} : \forall L_1, L_2, L_3 \colon (\mathit{list}\ A). & \\
\qquad\qquad L_1 \equiv L_2 \Rightarrow L_2 \equiv L_3 \Rightarrow L_1 \equiv L_3. &
\end{array}
$$

Prouver des propriétés:

$$\forall L_1, L_2 \colon (\mathit{list}\ A).\, L_1 \equiv L_2 \Rightarrow (\mathit{length}\ L_1) = (\mathit{length}\ L_2)$$

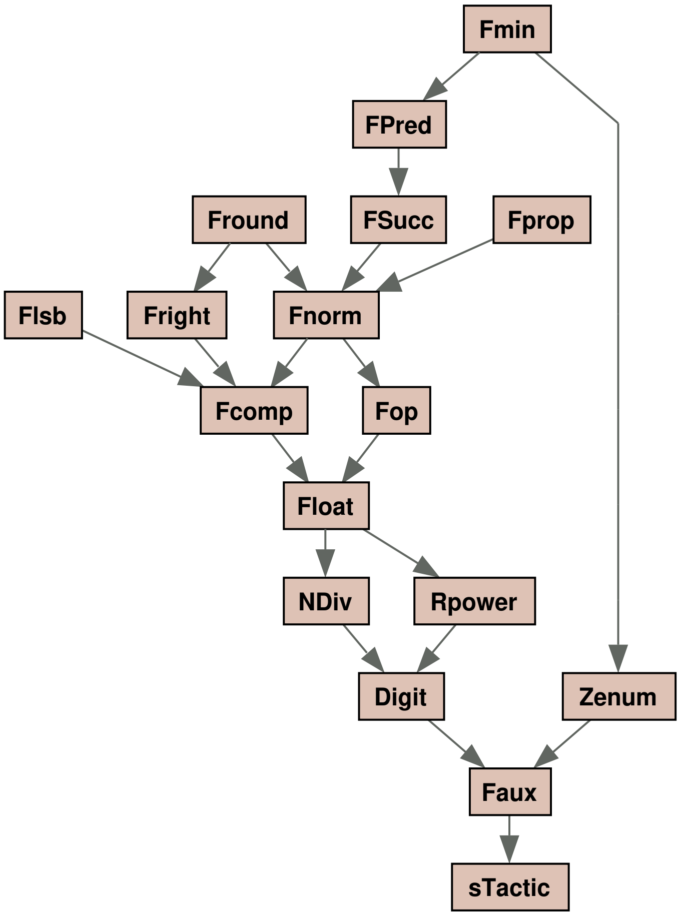
Coq (suite)

Construction par couches

Librairies: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$

Formalisation

Stratégie: on privilègie les propriétés sur les définitions

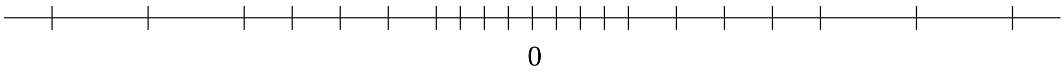


5000 lignes, 68 définitions, 340 theorèmes.

http://www-sop.inria.fr/lemme/AOC/coq/

6 Formalisation

Flottants



$$m * r^e$$

Représentation choisie des flottants:

$$\{Fnum : \mathbb{Z}, Fexp : \mathbb{Z}\} : float$$

$$Float : \mathbb{Z} \to \mathbb{Z} \to float$$

$$Fnum : float \to \mathbb{Z}$$

$$Fexp : float \to \mathbb{Z}$$

$$p = q \iff$$

$$(Fnum\ p) = (Fnum\ q) \wedge (Fexp\ p) = (Fexp\ q)$$

Flottants (suite)

Alternatives possibles:

$$\{Fsign : bool, Fnum : \mathbb{N}, Fexp : \mathbb{Z}\} : float$$

$$\{Fsign : bool, Fnum : \mathbb{N}, Fexp : \mathbb{N}\} : float$$

Flottants (suite)

Évaluation: dépendant de la base $r : \mathbb{N}$

Hypothèse: $1 < r$

Définition:

$$FtoR : float \rightarrow \mathbb{R}$$

$$(FtoR\ p) = (Fnum\ p) * r^{(Fexp\ p)}$$

Égalité sur $\mathbb{R}$

$p == q$ compris comme $(FtoR\ p) == (FtoR\ q)$

Float (suite)

Propriétés:

$$p == 0 \iff (Fnum\ x) = 0$$

$$p == q \wedge (Fexp\ p) = (Fexp\ q) \Rightarrow p = q$$

$$p == q \wedge (Fnum\ p) = (Fnum\ q) \Rightarrow p = q$$

$\boxed{\text{Comparaison}}$

$p < q$ compris comme $(FtoR\ p) < (FtoR\ q)$

$p \leq q$ compris comme $(FtoR\ p) \leq (FtoR\ q)$

Propriétés:

$$0 < (Fnum\ x) \iff 0 < x$$

$$0 \leq (Fnum\ x) \iff 0 \leq x$$

Si $(Fexp\ x) = (Fexp\ y)$

$$x < y \iff (Fnum\ x) < (Fnum\ y)$$

$$x \leq y \iff (Fnum\ x) \leq (Fnum\ y)$$

$\boxed{\text{Shift}}$

Définition:

$$Fshift : \mathbb{N} \rightarrow float \rightarrow float$$

$$(Fshift\ n\ p) = \\ (Float\ ((Fnum\ p) * r^n)\ ((Fexp\ p) - n))$$

Propriétés:

$$(Fshift\ n\ p) == p$$

$$p == q \iff \\ \exists n, m \colon \mathbb{N}.\ (Fshift\ n\ p) = (Fshift\ m\ q)$$

Opposé

Définition:

$$Fopp : float \rightarrow float$$

$$(Fopp\ p) = (Float\ \ -(Fnum\ p)\ \ (Fexp\ p))$$

Propriétés:

$$(Fopp\ p) == -p$$

$$(Fopp\ (Fopp\ p)) = p$$

Valeur absolue

Définition:

$$Fabs : float \rightarrow float$$

$$(Fabs\ p) = (Float\ \ |(Fnum\ p)|\ \ (Fexp\ p))$$

Propriété:

$$(Fabs\ p) == |p|$$

$\boxed{\text{Addition}}$

Définition:

$$Fplus : float \rightarrow float \rightarrow float$$

soit $m = (min\ (Fexp\ p)\ (Fexp\ q))$

$$
\begin{array}{l}
(Fplus\ p\ \ q) = \\
\quad (Float \\
\quad\quad (Fnum\ p) * r^{(Fexp\ p)-m} + \\
\quad\quad\quad (Fnum\ q) * r^{(Fexp\ q)-m} \\
\quad\quad m \\
\quad)
\end{array}
$$

Propriété:

$$(Fplus\ p\ \ q) == p + q$$

Soustraction

Définition:

$$Fminus : float \rightarrow float \rightarrow float$$

soit $m = (min\ (Fexp\ p)\ (Fexp\ q))$

$$\begin{array}{l}(Fminus\ p\ \ q) = \\ \quad (Float \\ \qquad (Fnum\ p) * r^{(Fexp\ p)-m} - \\ \qquad\quad (Fnum\ q) * r^{(Fexp\ q)-m} \\ \qquad m \\ \quad)\end{array}$$

Propriétés:

$$(Fminus\ p\ \ q) == p - q$$

$$(Fopp\ (Fminus\ p\ q)) = (Fminus\ q\ p)$$

Multiplication

Définition:

$$Fmult : float \rightarrow float \rightarrow float$$

$$\begin{aligned}
&(Fmult\ p\ \ q) = \\
&\quad (Float \\
&\qquad (Fnum\ p) * (Fnum\ q) \\
&\qquad (Fexp\ p) + (Fexp\ q) \\
&\quad)
\end{aligned}$$

Propriété:

$$(Fmult\ p\ \ q) == p * q$$

Flottants bornés

Borne:

$$\{vNum : \mathbb{N}, dExp : \mathbb{N}, uExp : \mathbb{N}\} : Fbound$$

Être borné:

$$Fbounded : Fbound \to float \to Prop$$

$$\begin{array}{l} (Fbounded\ b\ p) \iff \\ \quad -(vNum\ b) \leq (Fnum\ p) \leq (vNum\ b) \\ \qquad\qquad\qquad \wedge \\ \quad -(dExp\ b) \leq (Fexp\ p) \leq (uExp\ b) \end{array}$$

Flottants bornés (suite)

Propriétés:

$$(Fbounded\ b\ p) \iff (Fbounded\ b\ (Fopp\ p))$$

$$(Fbounded\ b\ p) \iff (Fbounded\ b\ (Fabs\ p))$$

soit $maxFloat = (Float\ (vNum\ b)\ (uExp\ b))$

$$(Fbounded\ b\ p) \Rightarrow -maxFloat \leq p \leq maxFloat$$

Flottants bornés (suite)

Propriétés (suite):

Si $(Fbounded\ b\ p)$, $(Fbounded\ b\ q)$
et $(Fexp\ p) = (Fexp\ q)$
$0 \leq p \wedge 0 \leq q \Rightarrow (Fbounded\ b\ (Fminus\ p\ q))$

Preuve:

Cas $0 \leq q \leq p$
- $(Fnum\ (Fminus\ p\ q)) = (Fnum\ p) - (Fnum\ q)$

donc $0 \leq (Fnum\ (Fminus\ p\ q)) \leq (Fnum\ p)$
$-(vNum\ b) \leq (Fnum\ (Fminus\ p\ q)) \leq (vNum\ b)$
- $(Fexp\ (Fminus\ p\ q)) = (Fexp\ p)$

$-(dExp\ b) \leq (Fexp\ (Fminus\ p\ q)) \leq (uExp\ b)$

Flottants bornés (suite)

Propriétés (suite):

Cas $0 \leq p < q$

Par le cas suivant $(Fbounded\ b\ (Fminus\ q\ p))$

donc $(Fbounded\ b\ (Fopp\ (Fminus\ q\ p)))$

donc $(Fbounded\ b\ (Fminus\ p\ q))$

Sterbenz

Propriétés (suite):

Si $(Fbounded\ b\ p)$ et $(Fbounded\ b\ q)$
$q/2 \leq p \leq 2 * q \Rightarrow (Fbounded\ b\ (Fminus\ p\ q))$

Preuve:

Cas $q \leq p \leq 2 * q$
soit $m = (min\ (Fexp\ p)\ (Fexp\ q))$
soit $p' = (Fshift\ ((Fexp\ p) - m)\ p)$
soit $q' = (Fshift\ ((Fexp\ q) - m)\ q)$
on a $q' \leq p' \leq 2 * q'$
$(Fnum\ q') \leq (Fnum\ p') \leq 2 * (Fnum\ q')$
on a $(Fnum\ (Fminus\ p\ q)) =$
$(Fnum\ p') - (Fnum\ q')$
$0 \leq (Fnum\ (Fminus\ p\ q)) \leq (Fnum\ q')$
$-(vNum\ b) \leq (Fnum\ (Fminus\ p\ q)) \leq (vNum\ b)$

Sterbenz (suite)

on a soit $(Fexp\ (Fminus\ p\ q)) = (Fexp\ p)$

soit $(Fexp\ (Fminus\ p\ q)) = (Fexp\ q)$

$$-(dExp\ b) \leq (Fexp\ (Fminus\ p\ q)) \leq (uExp\ b)$$

$\boxed{\text{Nombre de chiffres}}$

Sur les entiers:

$$digit : \mathbb{N} \rightarrow \mathbb{N}$$

Propriétés:

$$(digit\ 0) = 0$$

$$r^{q-1} \leq p < r^q \Rightarrow (digit\ p) = q$$

$$p \leq q \Rightarrow (digit\ p) \leq (digit\ q)$$

$$(digit\ p) < (digit\ q) \Rightarrow p < q$$

Nombre de chiffres (suite)

Sur les flottants:

$$Fdigit : float \rightarrow \mathbb{N}$$

$$(Fdigit\ p) = (digit\ |(Fnum\ p)|)$$

Propriétés:

si $\neg(p = 0)$

$$(Fdigit\ (Fshift\ n\ p)) = (Fdigit\ p) + n$$

$$p == q \wedge (Fdigit\ p) = (Fdigit\ q) \Rightarrow p = q$$

Précision

Précision:

$$p_r : \mathbb{N}$$

Hypothèses:

$$0 < p_r$$

$$(vNum\ \ b) = r^{p_r} - 1$$

Propriété:

$$(Fdigit\, p) \leq p_r \iff$$
$$-(vNum\, b) \leq (Fnum\, p) \leq (vNum\, b)$$

Normalisés

Définition:

$$Fnormal : float \rightarrow Prop$$

$$(Fnormal\ p) \Longleftrightarrow (Fbounded\ p) \wedge (Fdigit\ p) = p_r$$

Propriétés:

$$(Fnormal\ p) \Rightarrow \neg p == 0$$

$$(Fnormal\ p) \iff (Fnormal\ (Fopp\ p))$$

$$(Fnormal\ p) \iff (Fnormal\ (Fabs\ p))$$

$$(Fnormal\ p) \wedge (Fnormal\ q) \wedge p == q \Rightarrow p = q$$

$$\begin{array}{l} (Fnormal\ p) \wedge (Fnormal\ q) \wedge 0 \leq p < q \Rightarrow \\ \quad (Fexp\ p) < (Fexp\ q) \vee \\ \quad (Fexp\ p) = (Fexp\ q) \wedge (Fnum\ p) < (Fnum\ q) \end{array}$$

Sous-normalisés

Définition:

$$Fsubnormal : float \rightarrow Prop$$

$$(Fsubnormal\ p) \iff$$
$$(Fbounded\ p) \wedge$$
$$(Fdigit\ p) < p_r \wedge (Fexp\ p) = -(dExp\ b)$$

Propriétés:

$$(Fsubnormal\ p) \iff (Fsubnormal\ (Fopp\ p))$$

$$(Fsubnormal\ p) \iff (Fsubnormal\ (Fabs\ p))$$

$$\neg((Fnormal\ p) \wedge (Fsubnormal\ p))$$

$$(Fsubnormal\ p) \wedge (Fsubnormal\ q) \wedge p == q \Rightarrow$$
$$p = q$$

$$(Fsubnormal\ p) \wedge (Fsubnormal\ q) \wedge p < q \Rightarrow$$
$$(Fnum\ p) < (Fnum\ q)$$

Canonique

Définition:

$$Fcanonic : float \rightarrow Prop$$

$$(Fcanonic\ p) \Longleftrightarrow (Fsubnormal\ p) \vee (Fnormal\ p)$$

Propriétés:

$$(Fcanonic\ p) \quad \Longleftrightarrow \quad (Fcanonic\ (Fopp\ p))$$

$$(Fcanonic\ p) \quad \Longleftrightarrow \quad (Fcanonic\ (Fabs\ p))$$

$$(Fcanonic\ p) \wedge (Fcanonic\ q) \wedge p == q \Rightarrow p = q$$

$$\begin{aligned}&(Fcanonic\ p) \wedge (Fcanonic\ q) \wedge 0 \leq p < q \Rightarrow\\ &\qquad (Fexp\ p) < (Fexp\ q) \vee\\ &\qquad (Fexp\ p) = (Fexp\ q) \wedge (Fnum\ p) < (Fnum\ q)\end{aligned}$$

Normalisation

Définition:

$$Fnormalize : float \rightarrow float$$

$$\begin{array}{l}
(Fnormalize\ p) = \\
\quad \texttt{if}\ p == 0 \\
\quad \texttt{then}\ (Float\ 0\ \ -(dExp\ b)) \\
\quad \texttt{else}\ (Fshift\ (min\ (p_r - (Fdigit\ p)) \\
\qquad\qquad\qquad\qquad\qquad ((dExp\ b) + (Fexp\ p)) \\
\qquad\qquad\qquad\quad) \\
\qquad\qquad\quad p)
\end{array}$$

Propriétés:

$$(Fnormalize\ p) == p$$

$$(Fbounded\ p) \Rightarrow (Fcanonic\ (Fnormalize\ p))$$

Successeur

Définition:

$$FSucc : float \to float$$

$$
\begin{array}{l}
(FSucc\ p) = \\
\quad \texttt{if}\ (Fnum\ x) = (vNum\ b) \\
\quad \texttt{then}\ (Float\ \ r^{p_r - 1}\ \ (Fexp\ p) + 1) \\
\quad \texttt{else if}\ (Fnum\ x) = -r^{p_r - 1} \\
\qquad \texttt{then if}\ (Fexp\ p) = -(dExp\ b) \\
\qquad\quad \texttt{then}\ (Float\ \ (Fnum\ p) + 1\ \ (Fexp\ p)) \\
\qquad\quad \texttt{else}\ (Float\ \ -(vNum\ b)\ \ (Fexp\ p) - 1) \\
\qquad \texttt{else}\ (Float\ \ (Fnum\ p) + 1\ \ (Fexp\ p))
\end{array}
$$

Successeur (suite)

Propriétés:

$$(Fcanonic\ p) \wedge \neg(p = maxFloat) \Rightarrow \\ \quad (Fcanonic\ (FSucc\ p))$$

$$(Fcanonic\ p) \wedge (Fcanonic\ q) \Rightarrow \\ \quad p < q \Rightarrow (FSucc\ p) \leq q$$

$\boxed{\text{Prédécesseur}}$

Définition:

$$FPred : \mathit{float} \rightarrow \mathit{float}$$

$$
\begin{array}{l}
(FPred\ p) = \\
\quad \texttt{if}\ (Fnum\ x) = -(vNum\ b) \\
\quad \texttt{then}\ (Float\ \ -r^{p_r-1}\ \ (Fexp\ p)+1) \\
\quad \texttt{else if}\ (Fnum\ x) = r^{p_r-1} \\
\qquad \texttt{then if}\ (Fexp\ p) = -(dExp\ b) \\
\qquad\quad \texttt{then}\ (Float\ \ (Fnum\ p)-1\ \ (Fexp\ p)) \\
\qquad\quad \texttt{else}\ (Float\ \ (vNum\ b)\ \ (Fexp\ p)-1) \\
\qquad \texttt{else}\ (Float\ \ (Fnum\ p)-1\ \ (Fexp\ p))
\end{array}
$$

Prédécesseur (suite)

Propriétés:

$$
\begin{array}{l}
(Fcanonic\ p) \wedge \neg(p = -maxFloat) \Rightarrow \\
\quad (Fcanonic\ (FPred\ p))
\end{array}
$$

$$
\begin{array}{l}
(Fcanonic\ p) \wedge (Fcanonic\ q) \Rightarrow \\
\quad p < q \Rightarrow p \leq (FPred\ q)
\end{array}
$$

Arrondi "inférieur"

Définition:

$$Fmin : float \rightarrow (Option\ float)$$

$$\begin{array}{ll} \textbf{Inductive}\ Option[A : Set] : Set := & \\ \quad None : (Option\ A) & \\ \mid \quad Some : A \rightarrow (Option\ A). & \end{array}$$

Propriétés:

Si $(Fmin\ p) = None$

$$(Fbounded\ q) \Rightarrow p < q$$

Si $(Fmin\ p) = (Some\ q)$

$$(Fcanonic\ q)$$

$$q \leq p$$

$$(Fbounded\ t) \wedge t \leq p \Rightarrow t \leq q$$

$$(Fcanonic\ q) \wedge q \leq p \wedge p < (FSucc\ q) \Rightarrow (Fmin\ p) = (Some\ q)$$

Arrondi "supérieur"

Définition:

$$Fmax : float \rightarrow (Option\ float)$$

Propriétés:

Si $(Fmax\ p) = None$

$$(Fbounded\ q) \Rightarrow q < p$$

Si $(Fmin\ p) = (Some\ q)$

$$(Fcanonic\ q)$$

$$p \leq q$$

$$(Fbounded\ t) \wedge p \leq t \Rightarrow q \leq t$$

$$(Fcanonic\ q) \wedge (FPred\ q) < p \wedge p \leq q \Rightarrow (Fmax\ p) = (Some\ q)$$

Conclusions

Encore loin de couvrir la norme

Couches basses

Nouvelles métriques:

- Sterbenz: relativement facile
- $error(a+b)$: beaucoup plus dur

37 Formalisation

$\boxed{\text{end}}$