دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی کامپیوتر

اعداد اعشاری ممیز شناور
IEEE 754

طراحی واحد منطق و حساب

Arithmetic logic unit (ALU) design

نمایش اعداد اعشاری ممیز شناور

- در این نمایش، از n بیت برای نمایش بخش صحیح و اعشار عدد اعشاری استفاده می‌شود که در آن محل ممیز با عدد مشخص میشود (یک بیت هم برای علامت لحاظ می‌شود).
- اعداد اعشاری قبل از نمایش، باید در مبنای ۲ بوده و بصورت نماد علمی (اصطلاحاً عدد هنجار شده) نمایش داده شوند.
  - عدد هنجار شده، عددی است که فقط یک رقم غیرصفر قبل از ممیز (بخش صحیح) داشته باشد.
  - تنها عددی که هنجار نمی‌شود، عدد صفر است.

ina hichkodoom hanjar shode nistan chon har kodom too ghesmate sahih bishtar az ye ragham daran.

- مثال از اعداد غیر هنجار در مبنای ۱۰:

$+34.7$ -147.25 $+40$ -0.125

- مثال از اعداد هنجار در مبنای ۱۰:

$+3.47 \times 10^{+1}$ $-1.4725 \times 10^{+2}$ $+4.0 \times 10^{+1}$ -1.25×10^{-1}

نمایش عدد اعشاری

- مثال از اعداد غیر هنجار در مبنای ۲:

$+1101.01$ -100.01 $+1100$ -0.001

- مثال از اعداد هنجار در مبنای ۲:

$+1.10101 \times 2^{+3}$ $-1.0001 \times 2^{+2}$ $+1.1000 \times 2^{+3}$ -1.0×2^{-3}

momayez 3 ta rafte jolo —> +3   momayez 3 ta rafte aghab —> -3

درحالت کلی عدد هنجار بصورت زیر است:

+ , -   اعشار   محل ممیز

$$(-1)^{S} \times 1.F \times 2^{E}$$

که در آن:

S : بیت علامت است.

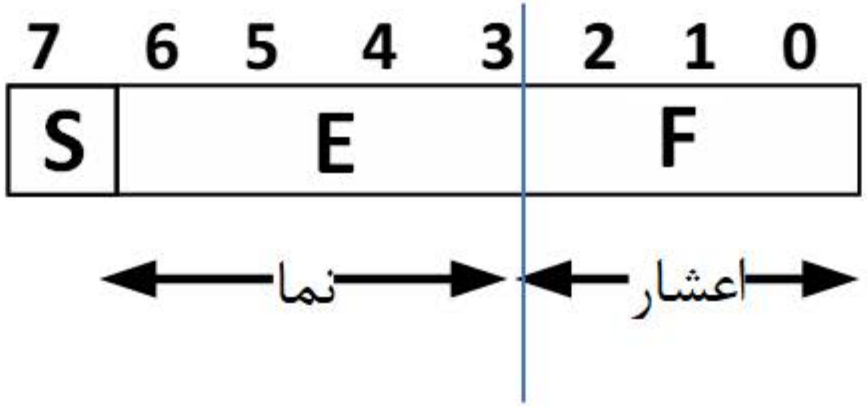
F : بخش اعشار یا Fraction یا مانتیس است و بصورت یک عدد بی‌علامت قابل نمایش است.

E : بخش نما یا Exponent است که بصورت یک عدد علامت‌دار (مکمل ۲ تغییریافته) قابل نمایش است.

قالب نمایش اعداد اعشاری

$$(-1)^{S} \times 1.F \times 2^{E}$$

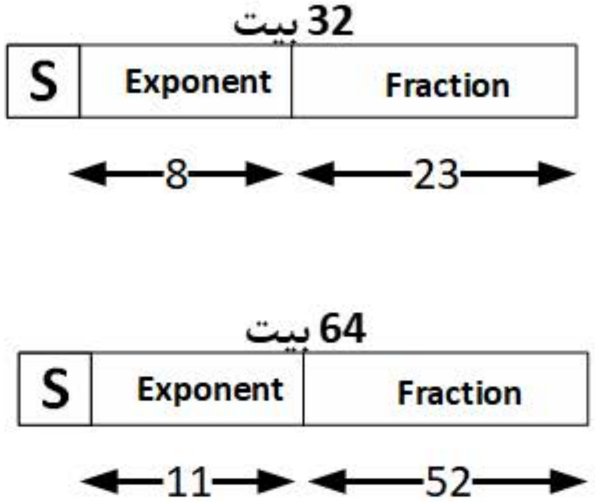
مثال



استاندارد IEEE

Single Precision

Double Precision

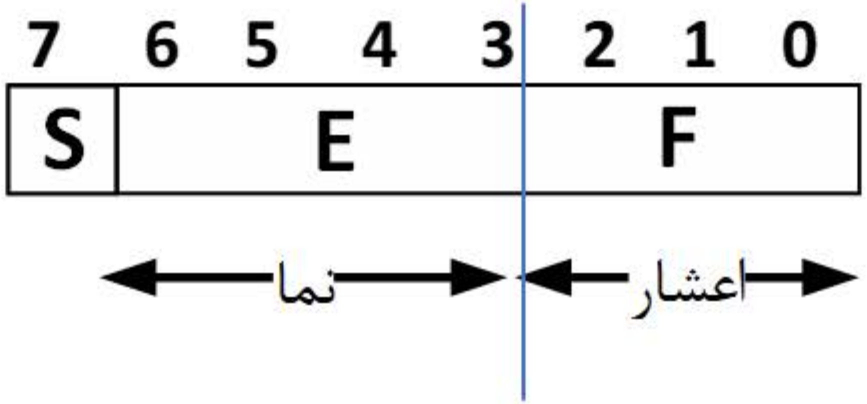


fazaye biti az 32 be 64 2 barabar shode vali namaye ma 2 be tavane 3 barabar yani 8 barabar shode . dalilesh ineke nama khodesh to tavane va darvaghe 2 be tavane 8 barabar shode.

نمایش اعشار و نما

- عدد اعشار بصورت بی علامت ذخیره می‌شود ولی عدد نما بصورت یک عدد مکمل ۲ تغییر یافته ذخیره میشود. علت تغییر یافتگی به دو دلیل است:
  - الف) مقایسه نماها ساده‌تر باشد
  - ب) نمایش عدد صفر (عدد ناهنجار)، طوری باشد که تمام بیت‌های آن صفر باشد.

- به مثال ذکر شده توجه کنید که در آن:
  - نما ۴ بیت
  - اعشار ۳ بیت دارد



$$(-1)^S \times 1.F \times 2^E$$

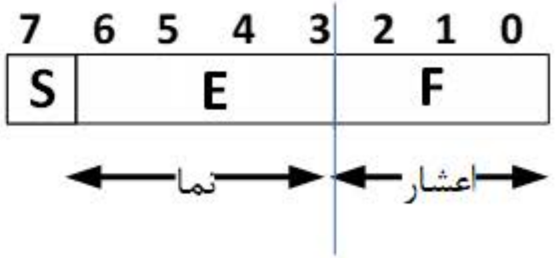
- کوچکترین عدد مثبت (اپسیلون)، عددی است که $S=0$، $E=E_{min}$ و $F=0$ است.

که در آن $E_{min} = -2^{e-1}=-8$ (اگر نما بصورت مکمل ۲ باشد بصورت 1000 نمایش داده می‌شود) یعنی 0 1000 000

از آنجا که عدد اپسیلون با صفر جایگزین میشود، خوب است که نمایشی داشته باشد که همه بیتهای آن صفر باشد.

نحوه نمایش نما در ۴ بیت (عدد علامت‌دار) بایاس ۱

Exponent	ده دهی	نمایش مکمل ۲	نمایش مکمل ۲ افزوده با +8
E_{max}	+7	0111	1111
	+6	0110	1110
	…	…	…
	+2	0010	1010
	+1	0001	1001
	0	0000	1000
	-1	1111	0111
	-2	1110	0110
	…	…	…
	-6	1010	0010
	-7	1001	0001
E_{min}	-8	1000	0000



+8 = B1

همیشه نماها با اندازه منفی‌ترین نما، جمع می‌شود.

دو حسن دارد:

۱ – مقایسه نماها ساده تر است (بصورت بی علامت ذخیره می‌شود)

۲ – منفی ترین نما، نمایش تمام صفر دارد (موجب میشود عدد اپسیلون تماماً صفر نمایش داده شود)

نکته: در این روش، عدد اپسیلون از مجموعه اعداد قابل نمایش حذف میشود، به جای آن صفر اضافه می‌شود. توجه شود نمایش اپسیلیون (یا صفر) نباید در قالب ذکر شده نماد علمی بیان شود بلکه بصورت استثنا خارج از قالب بررسی می‌شود.

نکته:

- سوال ۱: نمایش باینری 0.000...0001 (دارای f بیت بعد از اعشار) معادل چه عدد دهدهی است؟
  - جواب: 2^{-f}
- سوال ۲: نمایش باینری 0.111...1111 (دارای f بیت بعد از اعشار) معادل چه عدد دهدهی است؟
  - جواب: $1-2^{-f}$
- سوال ۳: در صورتیکه نما e بیت باشد، کوچکترین نما (Emin) و بزرگترین نما (Emax)چقدر است؟
  - جواب: کوچکترین نما -2^{e-1} و بزرگترین نما $+2^{(e-1)}-1$ است.

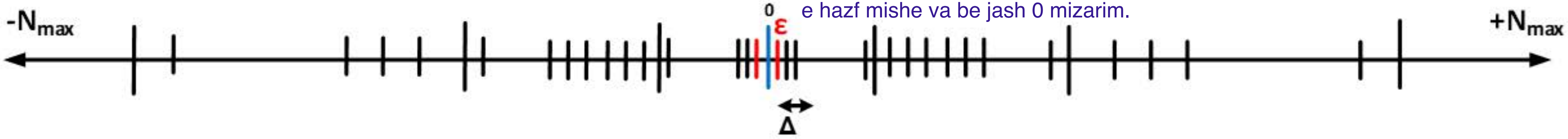
اکتاوها		نمایش	دهدهی	
اکتاو اول	کوچکترین عدد مثبت	$E=E_{min}$ F=0.000...00	$1*2^{Emin}$	اپسیلون
	دومین کوچکترین عدد مثبت	$E=E_{min}$ F=0.000...01	$(1+2^{-f}) * 2^{Emin}$	
	سومین کوچکترین عدد مثبت	$E=E_{min}$ F=0.000...1~~1~~ 0	$(1+2^{-f}+2^{-f}) * 2^{Emin}$	
	...	...	...	
	؟ مین کوچترین عدد مثبت	$E=E_{min}$ F=0.111...111	$(2-2^{-f}) * 2^{Emin}$	
اکتاو دوم	اولین کوچکترین عدد مثبت	$E=E_{min}+1$ F=0.000...00	$1*2^{Emin+1}$	
	دومین کوچکترین عدد مثبت	$E=E_{min}+1$ F=0.000...01	$(1+2^{-f}) * 2^{Emin+1}$	
	سومین کوچکترین عدد مثبت	$E=E_{min}+1$ F=0.000...11	$(1+2^{-f}+2^{-f}) * 2^{Emin+1}$	
	...	...	...	
	؟ مین کوچترین عدد مثبت	$E=E_{min}+1$ F=0.111...111	$(2-2^{-f}) * 2^{Emin+1}$	
...	...	...	...	...
اکتاو آخر (اکتاو چندم؟)	اولین کوچکترین عدد مثبت	$E=E_{max}$ F=0.000...00	$1*2^{Emax}$	
	...	...	...	
	؟ مین کوچترین عدد مثبت	$E=E_{max}$ F=0.111...111	$(2-2^{-f}) * 2^{Emax}$	N_{max}

نمایش روی محور اعداد

- همانطور که میدانید، تعداد اعداد اعشاری در هر بازه بسیار کوچک، بینهایت است. لذا قاعدتاً با **n** بیت قادر به نمایش همه اعداد نخواهیم بود.
  - در نمایش اعداد اعشاری، وجود خطا در نمایش اجتناب ناپذیر است.
- هدف آن است که 2^n عدد اعشاری قابل نمایش طوری باشد که بیشترین اعداد کاربر را بتواند نشان دهد.
  - تجربه نشان داده است نمایش ممیزشناور، می‌تواند بیشترین اعداد اعشاری مدنظر کاربر را نشان دهد زیرا پخش شدگی اعداد بصورت نمایی است (نه یکنواخت)، اعداد نزدیک صفر بسیار فشرده و نزدیک بهم هستند، اعداد دور از صفر، با فاصله های نمایی و تنک شده هستند)

توی ممیز ثابت

e hazf mishe va be jash 0 mizarim.



محاسن و معایب محاسبات اعداد اعشاری ممیز شناور

معایب

- نمایش پیچیده و ثقیل‌الفهم برای طراحان.
- ساخت مدارات محاسباتی بسیار پر هزینه (در مقایسه با روش ممیز ثابت).

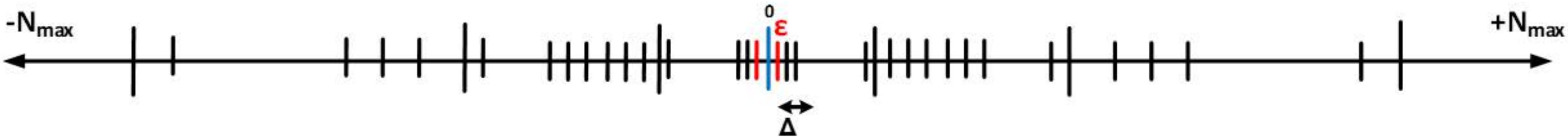
محاسن

- در نمایش اعداد، کمترین خطای قابل توجه از نگاه کاربران داشته باشد.
- مناسب برای محاسباتی که بخش صحیح بسیار بزرگ و اعشار کم لازم دارند، است. اکتاو های بزرگ
- مناسب برای محاسباتی که بخش صحیح کوچک ولی اعشار زیاد لازم دارند (مثل احتمال)، است.

اکتاو های کوچک
هر چی اکتاو پایین تر باشه دقتمون هم بهره ! پس این ۲ حالتی که گفتیم به بهترین شکل حل شدن .

سوال ۴: در حالت بایاس ~~۲~~ به سوالات زیر پاسخ دهید:

۱



- به سوالات زیر برای اعداد اعشاری ممیز شناور ۸ بیتی و نیز حالت کلی **n** بیتی پاسخ دهید:
  - الف) تعداد کل اعداد قابل نمایش؟
  - ب) بزرگترین عدد اعشاری قابل نمایش **(+Nmax)**؟
  - ج) کوچکترین عدد اعشاری مثبت قابل نمایش **(Epsilon)**؟
  - د) تفکیک‌پذیری یا کوچکترین فاصله بین دو عدد متوالی **(Resolution)**؟

سرریز شدن در محاسبات اعشاری

- از آنجا که امکان نمایش همه اعداد اعشاری توسط **n** بیت وجود ندارد، سرریز شدن در محاسبات اعشاری گریز ناپذیر است.
- در نتیجه، سرریز شدن **(overflow)** در محاسبات اعشاری مطرح نخواهد بود، به جای آن از کلمه "خطای محاسبات" استفاده می‌شود.
- کاربر باید بداند که محاسبات اعشاری با دقت ۱۰۰ درصد تامین نخواهد شد و ممکن است برخی مواقع خطای محاسباتی داشته باشیم.
- هدف آن است که محاسبات طوری باشد که خطای محاسبات کمترین مقدار ممکن باشد.
- تنها در یک حالت سرریز داریم و آن هنگام تقسیم بر صفر است.

نمایش ناعددی ها در ممیز شناور

- همانطور که می‌دانید، صفر، عدد ناهنجاری بود که با اپسیلون جایگزین شد.
- پس به نظر می‌رسد، هر عدد محبوب ناهنجار را می‌توان با ترفند مشابه می‌توان به مجموعه اضافه کرد.
- اعدادی که هنجار نمی‌شوند اصطلاحاً ناعددی **(Not a Number = NaN)** گفته می‌شود.
- ناعددی‌های محبوب زیاد هستند که اعشاری بوده و خوب است بتوان در مجموعه نمایش داد. مثل:

$\pm\pi$ $\qquad$ $\pm e$ $\qquad$ $\pm\infty$ $\qquad$ $\pm\sqrt{2}$

$\frac{1}{0} =$

- در نتیجه، مشابه صفر، ترفند زیر اعمال می‌شود:

نحوه نمایش نما در ۴ بیت (عدد علامت‌دار) بایاس ۲

یعنی اگه نما تماما ۱ بود ، یعنی یک ناعدد است پس ۲ به توان f ناعدد میتونیم داشته باشیم.

Exponent	ده دهی	نمایش مکمل ۲	نمایش مکمل ۲ افزوده با +7
E_{max}	+7	0111	1110
	+6	0110	1101
	...	...	...
	+2	0010	1001
	+1	0001	1000
	0	0000	0111
	-1	1111	0110
	-2	1110	0101
	...	...	...
	-6	1010	0001
E_{min}	-7	1001	0000
Nan	-8	1000	1111

اکتاو آخر رو گذاشتن برای استثناعات

7	6	5	4	3	2	1	0
S	E				F		

نما ← → اعشار

همیشه نماها با یکی کمتر از اندازه منفی‌ترین نما، جمع می‌شود.

سه حسن دارد:

۱ – مقایسه نماها ساده تر است (بصورت بی علامت ذخیره می‌شود)

۲ – منفی ترین نما، نمایش تمام صفر دارد (موجب میشود عدد اپسیلون تماماً صفر نمایش داده شود)

۳ – ناعددی ها قابل نمایش است (نمایش نما بصورت تماماً یک است)

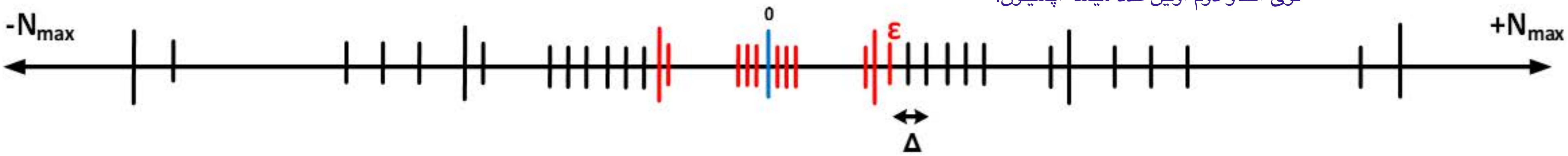
نکته: در این روش دو استثناء وجود دارد:

۱ – نمایش تمام صفر (عدد اپسیلون) از مجموعه اعداد قابل نمایش حذف میشود، به جای آن صفر جایگزین می‌شود.

۲ – نمایش نما تمام یک، ناعددی تلقی شده و حالتهای مختلف در اعشار حالتهای مختلف ناعددی‌ها را نشان می‌دهد.

سوال ۵: در حالت بایاس ۲ به سوالات زیر پاسخ دهید:

توی اکتاو دوم اولین عدد میشه اپسیلون.



- به سوالات زیر برای اعداد اعشاری ممیز شناور ۸ بیتی و نیز حالت کلی **n** بیتی پاسخ دهید:
  - الف) تعداد کل اعداد قابل نمایش؟
  - ب) بزرگترین عدد اعشاری قابل نمایش **(+Nmax)**؟
  - ج) کوچکترین عدد اعشاری مثبت قابل نمایش **(Epsilon)**؟
  - د) تفکیک‌پذیری یا کوچکترین فاصله بین دو عدد متوالی **(Resolution)**؟

سوال؟