مبانی برنامه‌نویسی

(۱۳۹۹-۱۳-۱۱)

جلسه‌ی ششم

سیستم اعداد ۴

دانشگاه شهید بهشتی

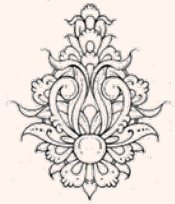
پاییز ۱۳۹۲

دانشکده‌ی مهندسی برق و کامپیوتر

احمد محمودی ازناوه

فهرست مطالب

- مروری بر جلسه‌ی پیش
  - چند مثال
- سرریز در اعداد علامت‌دار
- اعداد حقیقی
  - ممیز ثابت
  - ممیز شناور



تمرین کلاسی

تمرین ۱

	چهار بیتی	هشت بیتی
$(5)_{10}=(?)_2$	101	101
$(-5)_{10}=(?)_{1s'}$	1010	11111010
$(-5)_{10}=(?)_{2's}$	1011	11111011

تمرین ۲

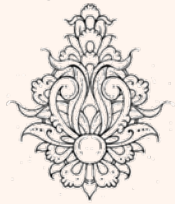
	بدون علامت	مکمل ۲ هشت بیتی	مکمل ۲ شانزده بیتی
$(11111011)_b$	251	-5	251

تمرین ۳

$-130=(?)_{2's(8\ bits)}$ در هشت بیت نمی‌گنجد

سرریز در مکمل ۲

- هنگامی که حاصل جمع(تفریق) در محدوده‌ی قابل نمایش نگنجد، گفته می‌شود که «سرریز» رخ داده است.
- برای اطمینان از درستی عملیات باید رخداد سرریز تشخیص داده شود.



سرریز (ادامه...)

$$\begin{array}{rr} (+7) & 0\,1\,1\,1 \\ +\,(+2) & +\;0\,0\,1\,0 \\ \hline (+9) & 1\,0\,0\,1 \end{array}$$

$c_4=0$

$c_3=1$

$$\begin{array}{rr} (+7) & 0\,1\,1\,1 \\ +\;(-2) & +\;1\,1\,1\,0 \\ \hline (+5) & 1\,0\,1\,0\,1 \end{array}$$

$c_4=1$

$c_3=1$

$$\begin{array}{rr} (-7) & 1\,0\,0\,1 \\ +\,(+2) & +\;0\,0\,1\,0 \\ \hline (-5) & 1\,0\,1\,1 \end{array}$$

$c_4=0$

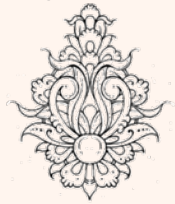
$c_3=0$

$$\begin{array}{rr} (-7) & 1\,0\,0\,1 \\ +\;(-2) & +\,1\,1\,1\,0 \\ \hline (-9) & 1\,0\,1\,1\,1 \end{array}$$

$c_4=1$

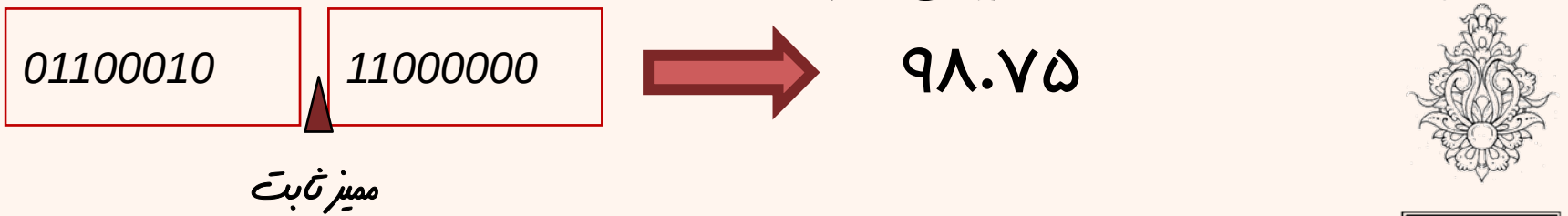
$c_3=0$

$$\text{Overflow} = c_3\bar{c}_4 + \bar{c}_3 c_4$$
$$= c_3 \oplus c_4$$



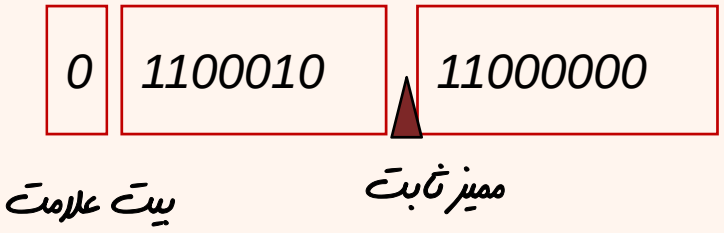
نمایش اعداد حقیقی

- برای نمایش همه‌ی اعداد حقیقی نمی‌توان دقت کامل داشت، از این رو هر عدد حقیقی به نزدیک‌ترین عدد قابل نمایش گرد می‌شود.
- برای نمایش اعداد اعشاری می‌توان از سیستم عددی «ممیز ثابت» استفاده کرد، در این شیوه جایگاه ممیز به صورت ثابت در نظر گرفته می‌شود.
- مثال: عدد حقیقی بدون علامت



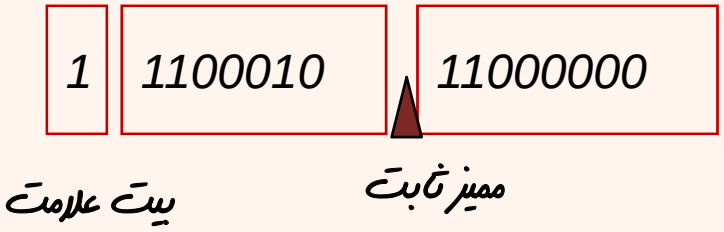
- برای اعداد حقیقی علامت‌دار از تمام شیوه‌های گفته شده می‌توان استفاده کرد.

اعداد ممیز ثابت علامت‌دار

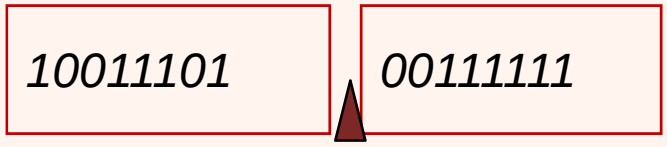


۹۸.۷۵

علامت و مقدار

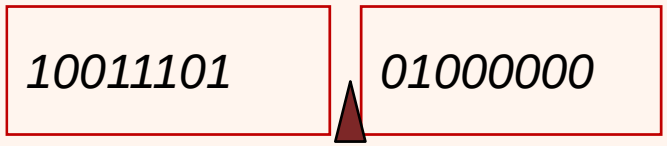


-۹۸.۷۵



-۹۸.۷۵

مکمل ۱

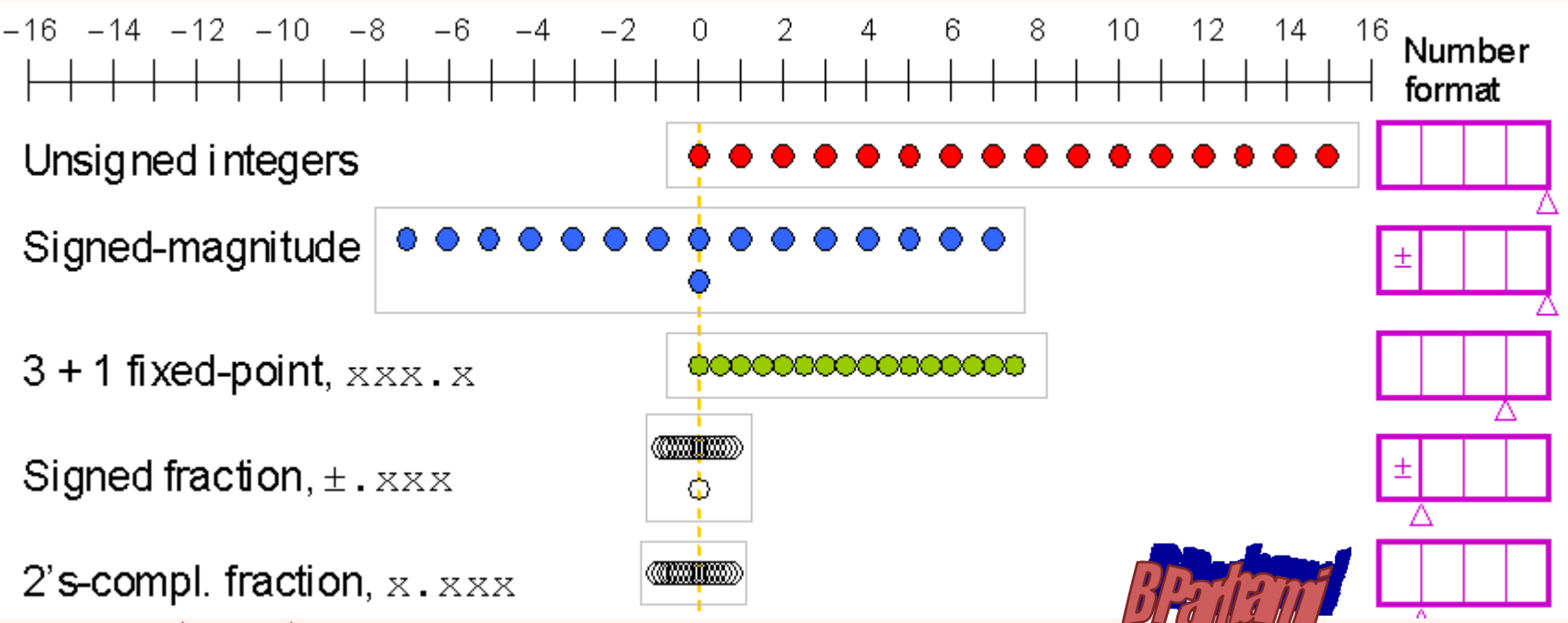


-۹۸.۷۵

مکمل ۲

دقت در ممیز ثابت

- دقت (precision) در ممیز ثابت، وابسته به تعداد بیت‌هایی است که بخش کسری اختصاص داده می‌شود، هر چه دقت را افزایش دهیم، بازه‌ی قابل نمایش کاهش می‌یابد و با افزایش بازه با کاهش دقت مواجه خواهیم شد.



اعداد کسری

- در کاربردهای علمی استفاده از اعداد بسیار بزرگ و یا اعداد بسیار کوچک لازم است.
- با استفاده از سیستم عددی ممیز ثابت دسترسی همزمان به بازه‌ی قابل نمایش گسترده همراه با دقت بالا امکان‌پذیر نمی‌باشد.
- استفاده از «ممیز شناور»، برای چنین کاربردهایی راهگشاست.
  - ۳.۱۴۱۵۹۲۶۵
  - ۲.۷۱۸۲۸۱۸۲۸۴۵۹۰
  - $۰.۰۰۰۰۰۰۰۰۱ = ۰.۱ \times ۱۰^{-۹}$

$$Mantissa \times Base^{Exponent}$$

نمایش نمایی

- تمام اعداد زیر نمایش عدد 1234 می‌باشند.

$123,400.0 \times 10^{-2}$

$12,340.0 \times 10^{-1}$

$1,234.0 \times 10^{0}$

123.4×10^{1}

12.34×10^{2}

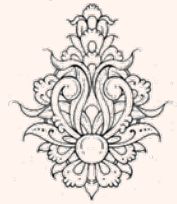
1.234×10^{3}

0.1234×10^{4}

با تغییر همزمان توان و جایگاه ممیز، نمایش‌های متفاوتی برای یک عدد به دست می‌آید.

ممیز شناور

- وجود نمایش‌های متعدد برای یک عدد در عمل معادل از دست دادن دقت یا دامنه‌ی نمایش اعداد است.
- از طرفی دیگر، موجب پیچیدگی عملیات ریاضی (به عنوان مثال مقایسه) می‌شود.
- در سیستم نمایش ممیز شناور اعداد معمولا به صورت «نماد علمی نرمال» نمایش داده می‌شوند.
- در مبنای دو، نماد علمی نرمال عددی به فرم $1.F\times 2^{E}$ است.



روش‌های مختلف نمای‌ش ممیز شناور

- سازندگان مختلف هریک شیوه‌ای برای نمایش اعداد ممیز شناور در پیش گرفتند.

Copyright 2004 Koren

	IBM/370	DEC/VAX	Cyber 70
Word length (double)	32 (64) bits	32 (64) bits	60 bits
Significand+{hidden bit}	24 (56) bits	$23+1$ $(55+1)$ bits	48 bits
Exponent	7 bits	8 bits	11 bits
Bias	64	128	1024
Base	16	2	2
Range of M	$\frac{1}{16} \leq M < 1$	$\frac{1}{2} \leq M < 1$	$1 \leq M < 2$
Representation of M	Signed-magnitude	Signed-magnitude	One's complement
Approximate range	$16^{63} \approx 7 \cdot 10^{75}$	$2^{127} \approx 1.9 \cdot 10^{38}$	$2^{1023} \approx 10^{307}$
Approximate resolution	$2^{-24} \approx 10^{-7}$ (10^{-17})	$2^{-24} \approx 10^{-7}$ (10^{-17})	$2^{-48} \approx 10^{-14}$

- این تنوع نمایش به معنای دقت‌های متفاوت در سیستم‌های مختلف است، از این رو نتیجه‌ی اجرای دستورهای یکسان، متفاوت خواهد بود.

استاندارد ممیز شناور

- برای فراهم آوردن سازگاری، نمایش ممیز شناور توسط IEEE استاندارد شده است.

IEEE standard 754

- دو شیوه‌ی زیر در این استاندارد ارائه شده است:
  - single-precision
  - double precision
- برای نمایش عدد در این استاندارد ابتدا آن را به صورت $1.F \times 2^E$ در می‌آوریم.