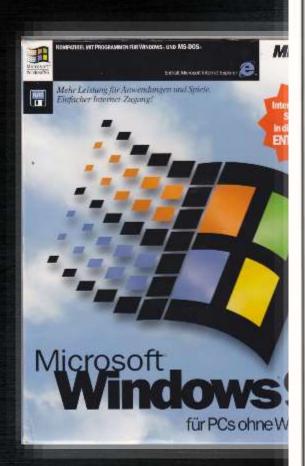
C++ ao resgate! Acelerando apps Android com instruções SIMD



Era uma vez...



Era uma vez...



Kaufpreises zurück.

Systemanforderungen für Microsoft Windows 95 :

- PC mit 386DX- oder h\u00f6herem Prozessor
- Mindestens 4 MB Arbeitsspeicher (8 MB oder mehr empfohlen)
- 55 MB freier Festplattenspeicher (tatsächlicher Speicherbedarf hängt von den installierten Komponenten ab)
- 8,89-cm-HD-Diskettenlaufwerk (entspricht 3,5")
- VGA- oder h\u00f6her aufl\u00f6sende Grafikkarte
- Microsoft Mouse oder kompatibles Zeigegerät

Optional:

- Modern/Faxmodern
- Audiokarte und Kopfhörer bzw. Lautsprecher

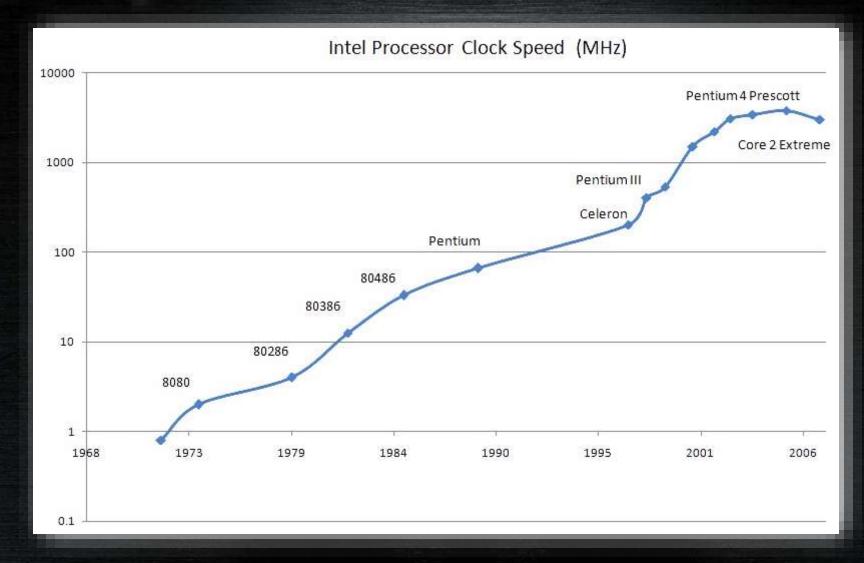
Hinwels:

- Systemanforderungen für Windows 95-basierte Programme können die für Windows 95 genannten Systemanforderungen überschreiten.
- Um präemptives Multitasking voll ausnutzen zu können, benötigen Sie ausschließlich auf Windows basierende 32-Bit-Programme, z. B. Produkte mit dem Logo "Entwickelt für Windows 95".

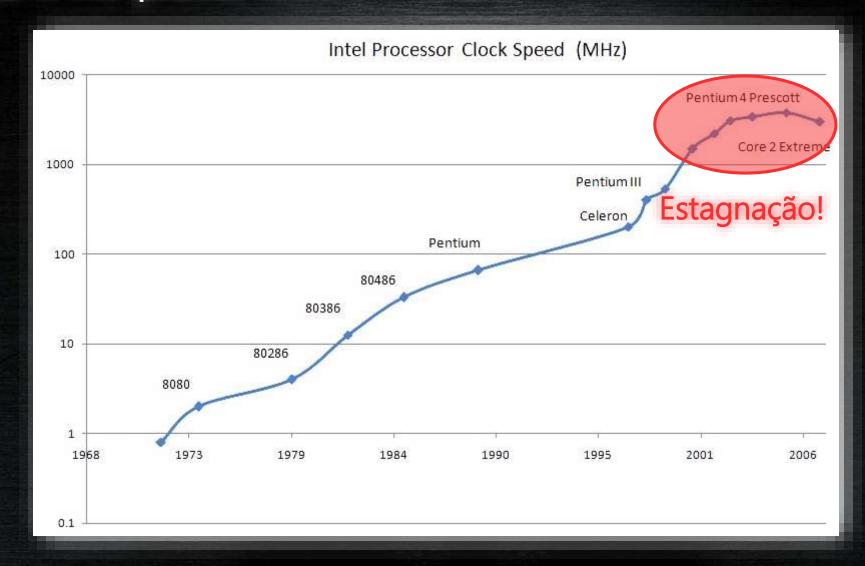
田 schon immer gewünscht haben. system Microson, Windows, 95, his gesamte Potential in a n neuen Möglichkeiten des In erosoft Internet Explorer Starter Microsoft r Point hard from suggest/held Selfs, Audio and Acceptorate internet Dollers Startes Kit enthal it Microsoft

Protomondandanument dis Mineralli

De lá para cá...



De lá para cá...



Como amenizar o cenário?



Uma solução!



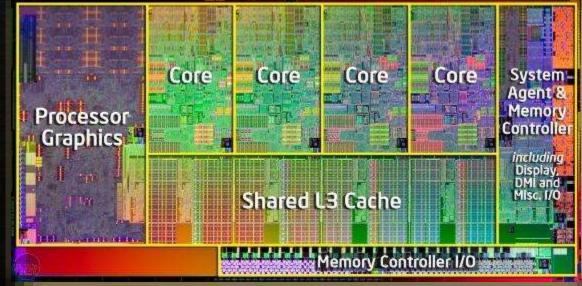
"Many hands make light work"

John Heywood

Uma solução!



Intel Core i7 que está no meu PC :)



O preço da solução



Antes, os ganhos de desempenho eram gratuitos, mas ficaram

caros!

Não só de multicore vive o mundo!

Existe algo entre o tradicional e o multicore...

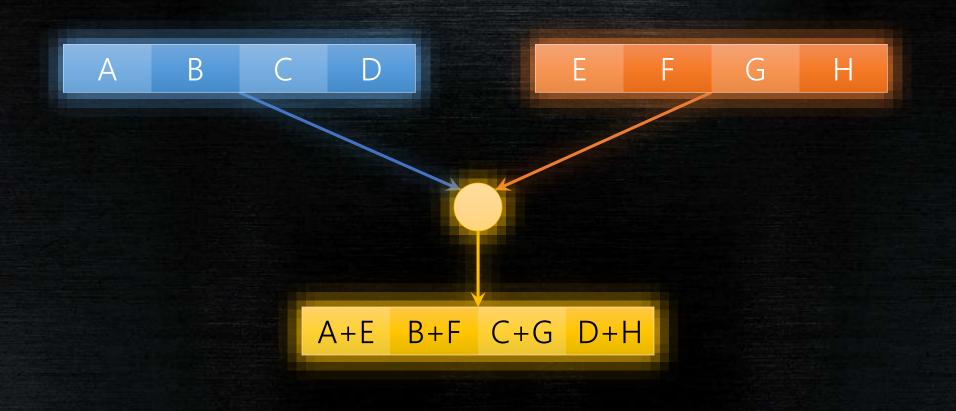
Não só de multicore vive o mundo!

Existe algo entre o tradicional e o multicore...



Single Instruction Multiple Data

SIMD = Cálculo com vetores



Um pequeno exemplo fictício...

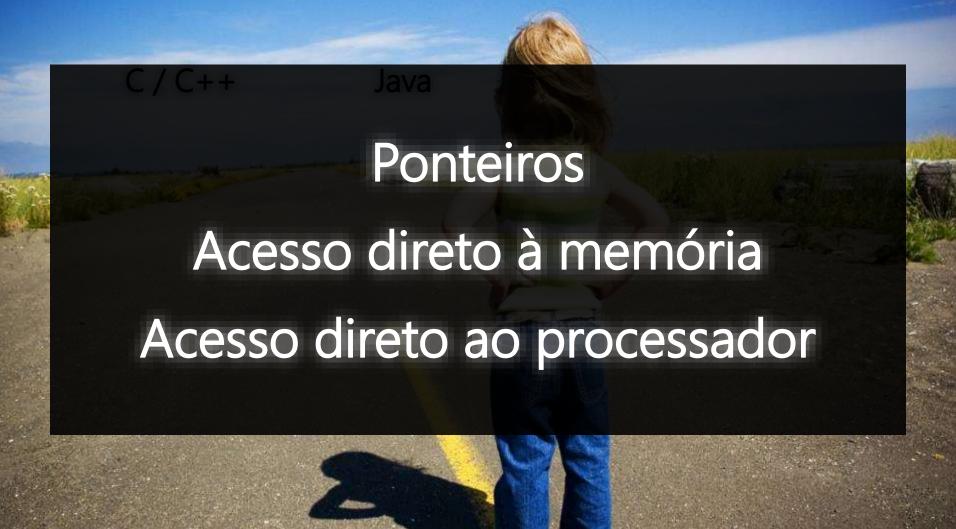
```
void dobro (int a[]) {
  a[0] = a[0] * 2;
  a[1] = a[1] * 2;
  a[2] = a[2] * 2;
  a[3] = a[3] * 2;
}
```

Sem SIMD

Com SIMD

Por que C / C++?!?! C/C++ Java

Por que C / C++?!?!



Algumas tecnologias de mercado

Intel MMX
Intel SSE (2, 3, 4)
AMD 3DNow!

Veremos hoje!

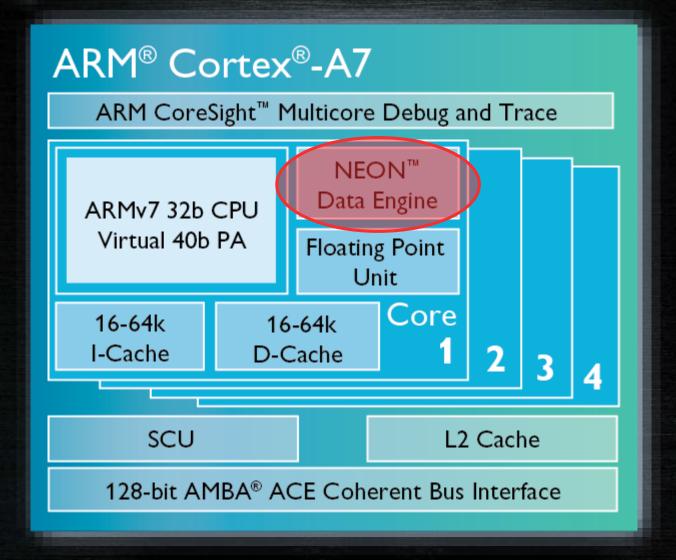
ARM NEON

O que é ARM NEON?

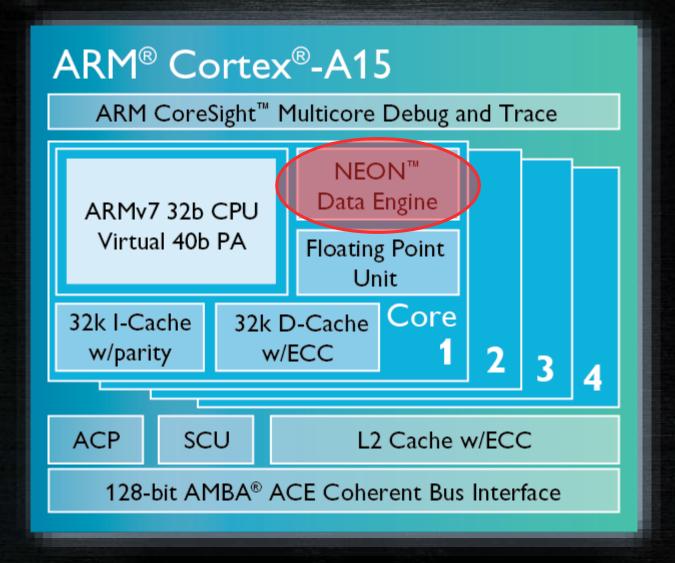
Extensão SIMD introduzida pela arquitetura ARMv7, como uma extensão opcional aos perfis ARMv7-A e ARMv7-R

Atualmente, é implementado em toda a série de processadores ARM Cortex-A

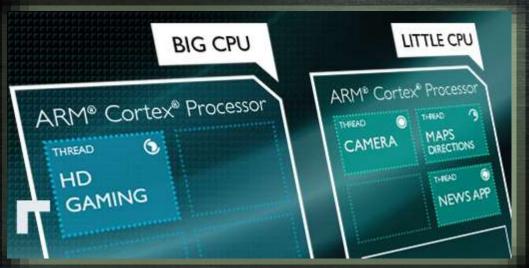
Exemplo: ARM Cortex-A7



Exemplo: ARM Cortex-A15



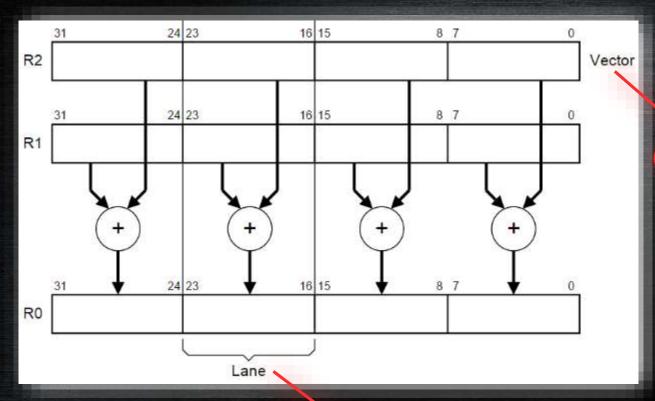
Misturando: ARM big.LITTLE





Automaticamente redireciona o processamento para a CPU adequada, com base na necessidade de processamento

Nomenclatura do NEON



Cada registrador armazena um vetor

Os elementos do vetor são chamados de "lanes" ("pistas" :P)

Registradores NEON



16 registradores de 128 bits (Q0 – Q15)

32 registradores de 64 bits (D0 – D31)

D0 e D1 compõe Q0 D2 e D3 compõe Q1

```
void dobro (float* v, int tamanho) {
  float32x4_t f;

for (int i = 0; i < tamanho; i += 4) {
  f = vld1q_f32(v + i);
  f = vaddq_f32(f, f);
  vst1q_f32(v + i, f);
}
}</pre>
```

Direto em C/C++ através de intrinsics

Funções intrínsecas parecem funções comuns, mas são substituídas diretamente por instruções Assembly pelo compilador!

Fonte:

http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0491c/BABFBJAA.html http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0491c/BABDFJCI.html http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0491c/BEHHCHAE.html

```
void dobro (float* v, int tamanho) {
  for (int i = 0; i < tamanho; i += 4) {
    asm volatile (
      "vld1q.32 {d0, d1} [%[v]] \n"
      "vaddq.f32 q0, q0, q0 \n"
      "vst1q.32 {d0, d1} [%[v]]! \n"
      : [v] "+r" (v)
      :
      : "q0");
  }
}</pre>
```

Mais avançado

Controle direto dos registradores e das instruções

Diversão garantida!!! :D

Inline Assembly!

Fonte:

http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0489c/CIHCADCI.html http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0489c/CIHJCAAG.html

A maioria das instruções Assembly do NEON utilizam três operandos, que podem ser vistos da seguinte forma:

vaddq.f32 q0, q1, q2

$$q0 = q1 + q2$$

O NDK do Android utiliza o GCC para compilar código nativo

Os links abaixo servem de referência sobre como o GCC trabalha com Inline Assembly dentro do C/C++

https://gcc.gnu.org/onlinedocs/gcc/Extended-Asm.html

http://www.ethernut.de/en/documents/arm-inline-asm.html

```
int32x4_t i;
float32x4_t f;

i = vcvtq_s32_f32(f);
```

Converte 4 float's em 4 int's

```
int32x4_t i;
float32x4_t f;

f = vcvtq_f32_s32(i);
```

Converte 4 int's em 4 float's

Idem a um typecast comum:

```
f = (float)i;
```

```
float32x4_t r, a, b, c;
r = vmlaq_f32(a, b, c);
```

$$r = a + (b * c)$$

No dia-a-dia, costuma-se utilizar:

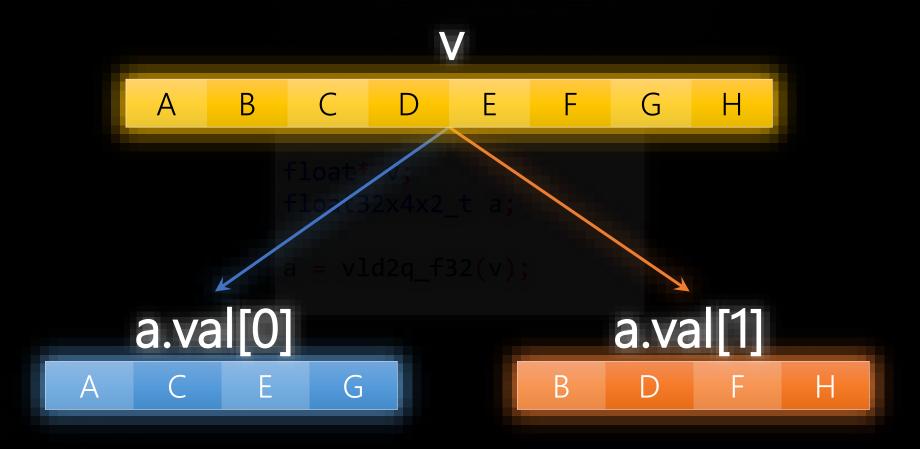
$$a = vmlaq_f32(a, b, c);$$

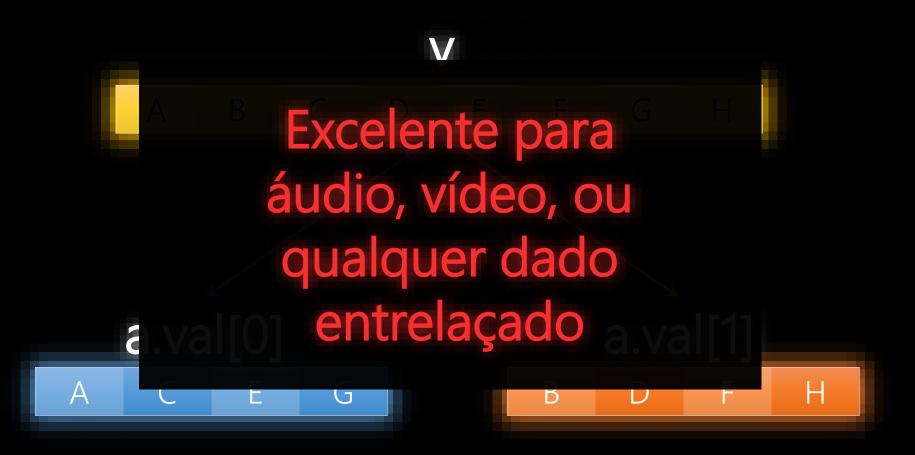
```
if (b > c) {
float32x4_t a, b, c;
a = vmaxq_f32(b, c);
float32x4_t a, b, c;
a = b;
} else {
a = c;
}
```

No dia-a-dia de C/C++/Java/JavaScript, costuma-se utilizar a forma ternária em vez do if:

$$a = ((b > c) ? b : c);$$

```
float* v;
float32x4x2_t a;
a = vld2q_f32(v);
```





Cuidados com NEON - Alinhamento

Para que o processador possa trabalhar com mais rapidez, devemos tomar cuidado com *onde* os vetores estão alocados na memória

É o endereço do vetor na RAM!!!

Cuidados com NEON - Alinhamento

Para variáveis globais (vetores estáticos), utilizar o alinhamento do GCC

Em vez de: float v[256];

Aloca o vetor em um endereço múltiplo de 16

Utilizar:

float v[256] __attribute__((aligned(16)));

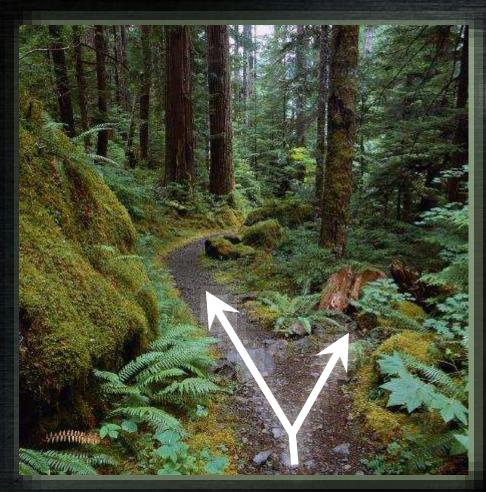
Cuidados com NEON - Alinhamento

Para vetores estáticos criados dinamicamente

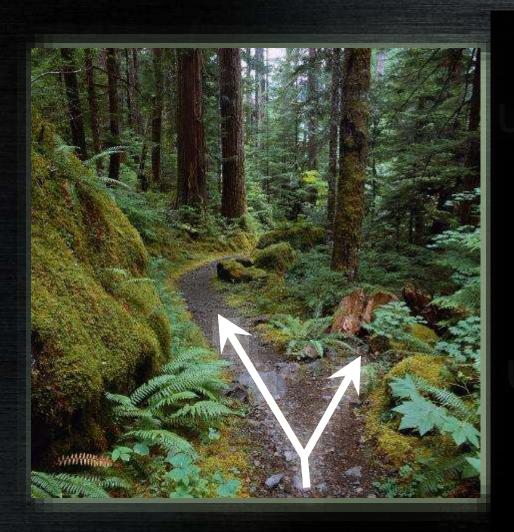
Em vez de: new, malloc ...

Utilizar: aligned_alloc, memalign ...

NEON e Condicionais



Não é possível utilizar condicionais com instruções SIMD do mesmo modo com que utiliza condicionais com instruções comuns



Não é possível

Uma instrução não pode tomar dois rumos distintos ao mesmo tempo!

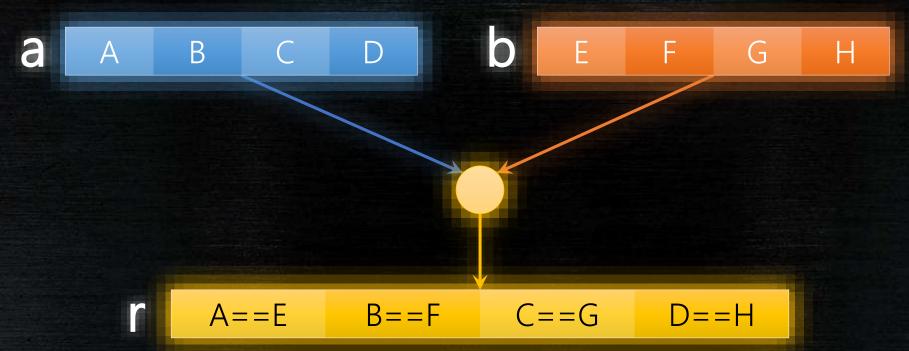
comuns

Então, para que servem instruções de comparação, como vceqq_f32?

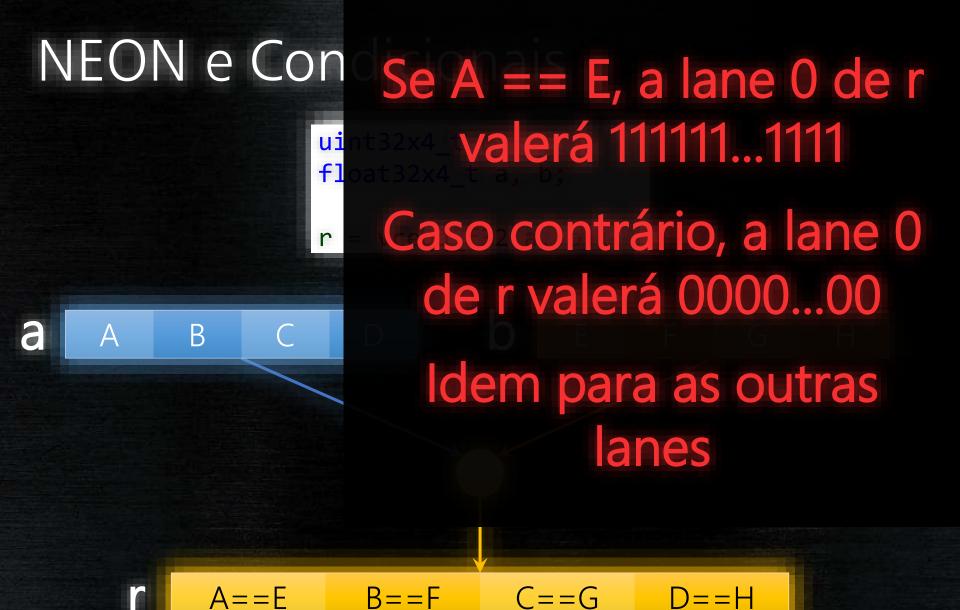


```
uint32x4_t r;
float32x4_t a, b;

r = vceqq_f32(a, b);
```



Fonte: http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0491c/BABGDDDH.html



Depois de feita a comparação, normalmente usa-se esse vetor r para "limpar" valores de outros vetores através de funções como vandq_u32

*Lembrando da tabela verdade do E;)

noi Em breve veremos um para exemplo ao vivo! :D es atra de la comparação para para exemplo ao vivo! :D es atra de la comparação para para exemplo ao vivo! :D es atra de la comparação para para exemplo ao vivo! :D es atra de la comparação para para exemplo ao vivo! :D es atra de la comparação para para exemplo ao vivo! :D es atra de la comparação para exemplo a comparação para exemplo a comparação para exemplo actual exemplo a comparação para

*Lembrando da tabela verdade do E;)

```
float m = ...

const float old = fft[i];

if (m < old)
  m = (coefNew * m) + (coefOld * old);

fft[i] = m;</pre>
```

Versão em C++

```
vcgeq.f32 resCmp, m, old
```

```
vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m
```

vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old

Os nomes dos registradores foram trocados por nomes mais significativos

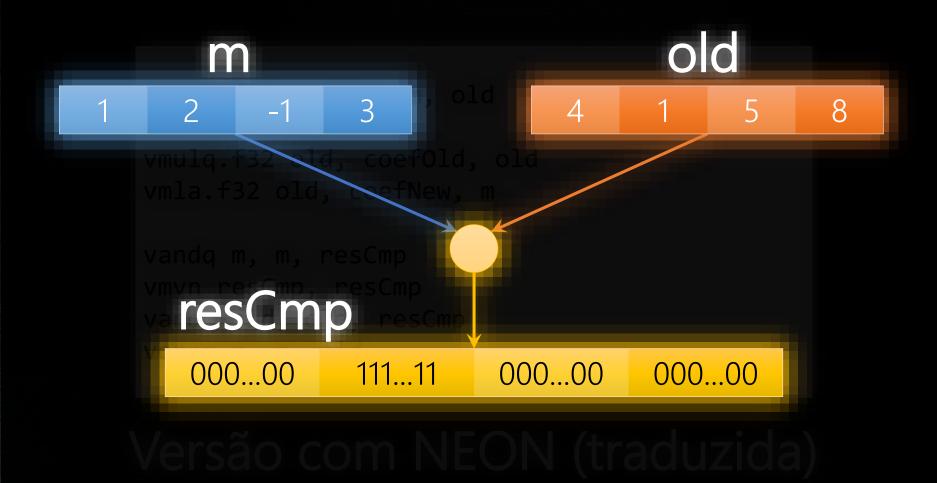
vandq m, m, resCmp
vmvn resCmp, resCmp

Os valores utilizados a seguir são exemplos fictícios!

```
vcgeq.f32 resCmp, m, old
```

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old



```
vcgeq.f32 resCmp, m, old
```

```
vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m
```

vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old

```
vcgeq.f32 resCmp, m, old

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

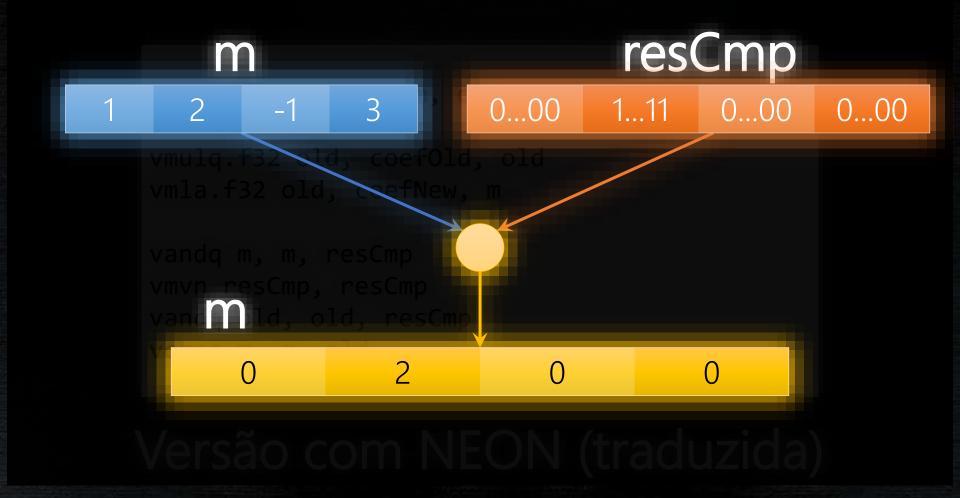
= (coefNew * m) + (coefOld * old)

varidq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old
```

```
vcgeq.f32 resCmp, m, old
```

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

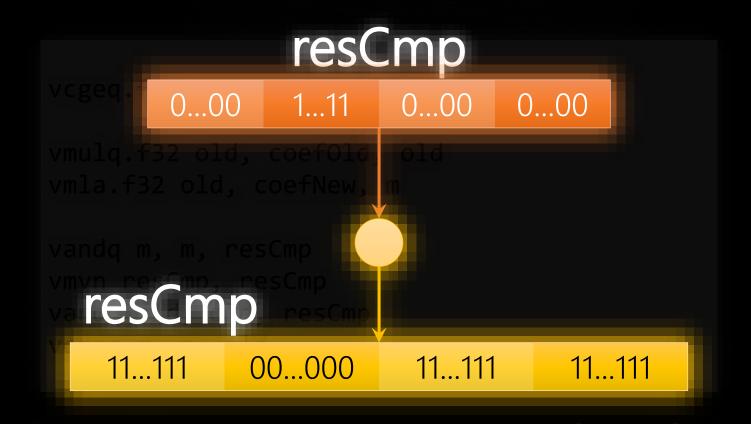
vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old



```
vcgeq.f32 resCmp, m, old

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

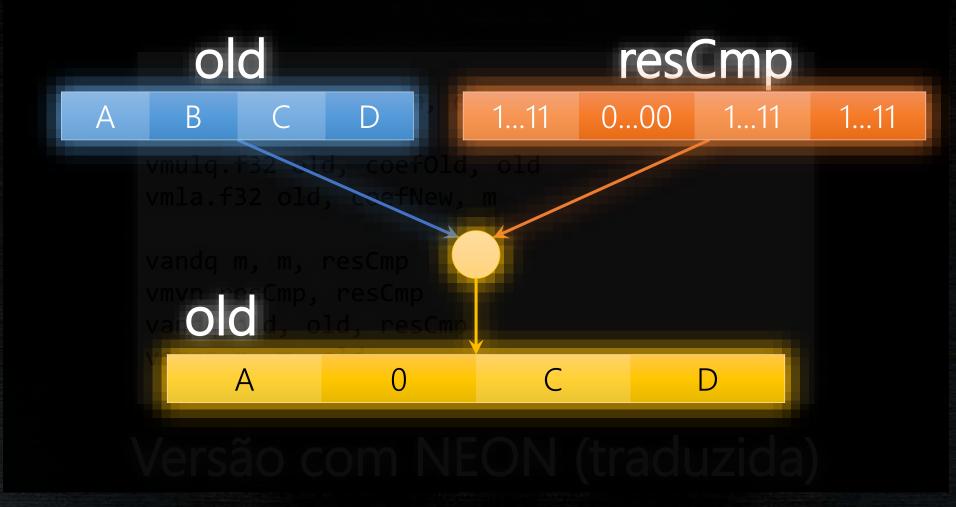
vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old
```



```
vcgeq.f32 resCmp, m, old

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

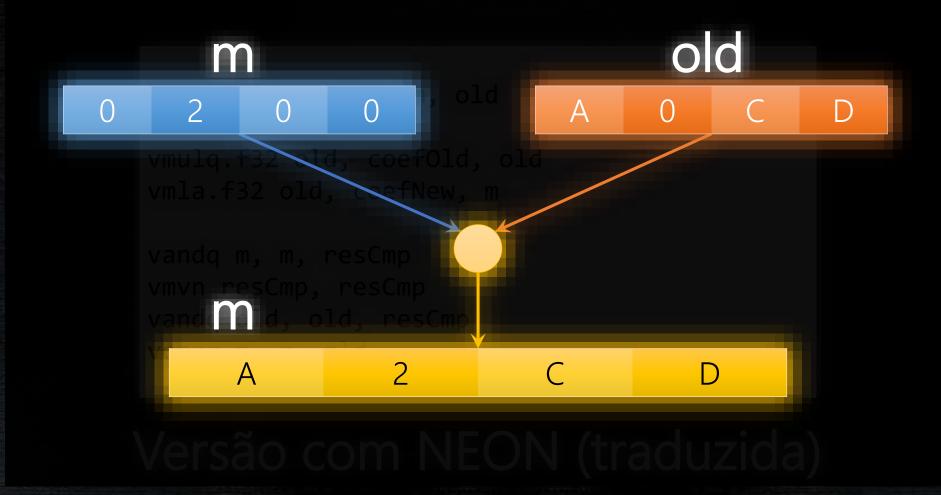
vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old
```

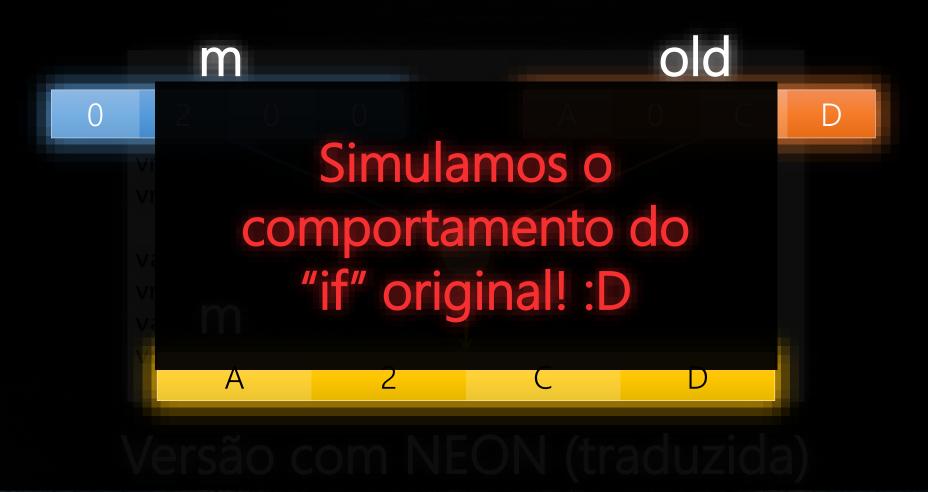


```
vcgeq.f32 resCmp, m, old

vmulq.f32 old, coefOld, old
vmla.f32 old, coefNew, m

vandq m, m, resCmp
vmvn resCmp, resCmp
vandq old, old, resCmp
vorrq m, m, old
```





```
...
APP_ABI := armeabi armeabi-v7a x86
...
```

Application.mk

```
ifeq ($(TARGET_ARCH_ABI),armeabi-v7a)
LOCAL_SRC_FILES += MeuArquivo.cpp.neon
LOCAL_ARM_NEON := true
endif
```

Android.mk

```
#include <jni.h>
#ifdef __ARM_NEON__ Para utilizar a mesma

#include <errno.h>
#include <fcntl.h>
#include <arm_neon.h>
#endif

Para usar as funções
intrínsecas do NEON
```

MeuArquivo.cpp

```
#include <jni.h>
#ifdef __ARM_NEON__
```

#include <cpu-features.h>

Para utilizar uma técnica mais simplificada para detectar o NEON

```
#include <arm_neon.h>
#endif
```

Para usar as funções intrínsecas do NEON

MeuArquivo.cpp

Informações úteis

Links exibidos nos rodapés dos slides ;)

http://www.mikusite.de/pages/vfp_neon.htm (Coleção de bons links!)

http://carlosrafaelgn.com.br (onde está a apresentação)

Exemplo hello-neon do NDK do Android: ...ndk\samples\hello-neon

