

Alto Desempenho

Marco Dimas Gubitoso

2016



IME-USP

Alto Desempenho

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão



Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade? quanto?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Introdução

- Para que precisamos de um computador?
 - Conveniência
 - velocidade
 - muita velocidade?
- Como conseguir velocidade?
- Como saber se conseguimos?
- Até que ponto vale a pena?



Sumário

- 1 Introdução
- 2 A corrida pela velocidade**
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão



Lei de Moore

Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje
- As empresas lutam por isso.
- Qual o limite?



Lei de Moore

Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje
- As empresas lutam por isso.
- Qual o limite?



Lei de Moore

Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje
- As empresas lutam por isso.
- Qual o limite?



Lei de Moore

Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje?
- As empresas lutam por isso.
- Qual o limite?



Lei de Moore

Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje?
- As empresas lutam por isso.
- Qual o limite?



Lei de Moore

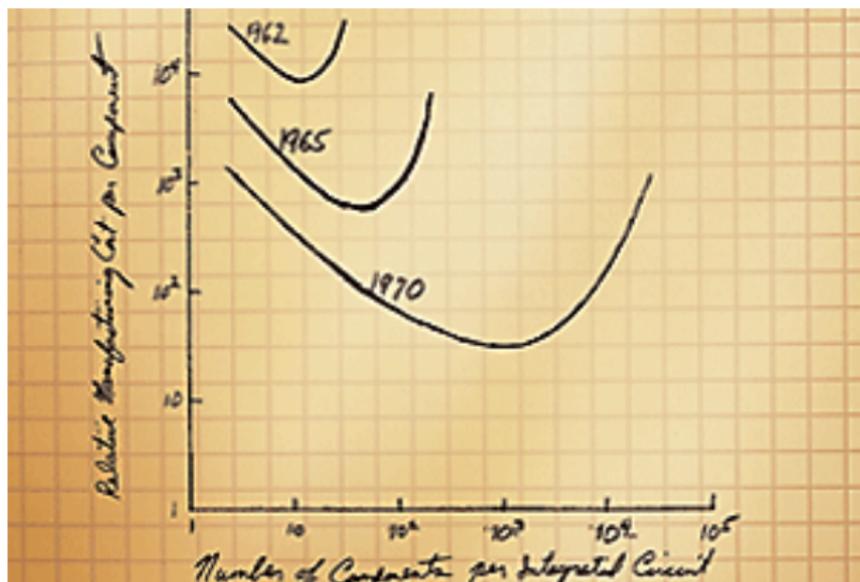
Moore em 1965

A capacidade dos processadores deve dobrar a cada 2 anos, aproximadamente.

- Continua válida até hoje?
- As empresas lutam por isso.
- Qual o limite?



Gráfico original



Até onde vai?

- Em pouco tempo (600 anos) seríamos capazes de simular todo o Universo em um único *chip*.
- A miniaturização vai atingir o nível atômico antes disso.
- Dissipação de calor e consumo tornam-se problemas críticos.
- Convergência



Até onde vai?

- Em pouco tempo (600 anos) seríamos capazes de simular todo o Universo em um único *chip*.
- A miniaturização vai atingir o nível atômico antes disso.
- Dissipação de calor e consumo tornam-se problemas críticos.
- Convergência



Até onde vai?

- Em pouco tempo (600 anos) seríamos capazes de simular todo o Universo em um único *chip*.
- A miniaturização vai atingir o nível atômico antes disso.
- Dissipação de calor e consumo tornam-se problemas críticos.
- Convergência



Até onde vai?

- Em pouco tempo (600 anos) seríamos capazes de simular todo o Universo em um único *chip*.
- A miniaturização vai atingir o nível atômico antes disso.
- Dissipação de calor e consumo tornam-se problemas críticos.
- Convergência



Até onde vai?

- Em pouco tempo (600 anos) seríamos capazes de simular todo o Universo em um único *chip*.
- A miniaturização vai atingir o nível atômico antes disso.
- Dissipação de calor e consumo tornam-se problemas críticos.
- Convergência



Capacidade computacional

1 The accelerating pace of change ...



2 ... and exponential growth in computing power ...

Computer technology, shown here climbing dramatically by powers of 10, is now progressing more each hour than it did in its entire first 90 years

COMPUTER RANKINGS
 By calculations per second per \$1,000



Analytical engine
 Never fully built, Charles Babbage's invention was designed to solve computational and logical problems



Colossus
 The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WW II

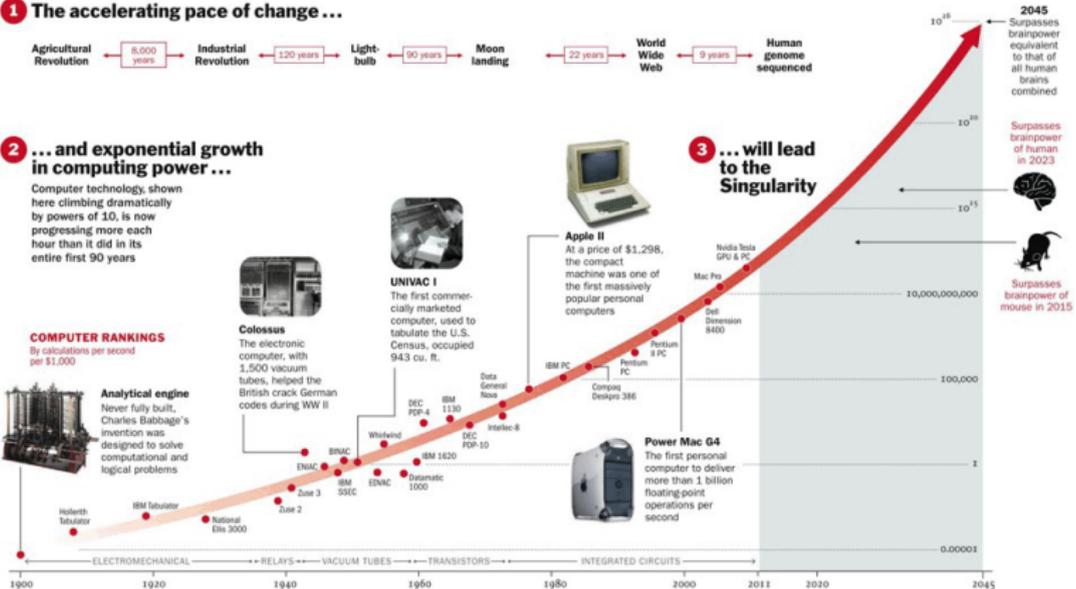


UNIVAC I
 The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.



Apple II
 At a price of \$1,298, the compact machine was one of the first massively popular personal computers

3 ... will lead to the Singularity



IME-USP

Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas**
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Problemas

O número de aplicações que dependem de altíssima velocidade é crescente:

- Simulação
- Previsão do tempo
- Bioinformática
- LHC
- *Data mining*



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Alguns casos

Wall Street

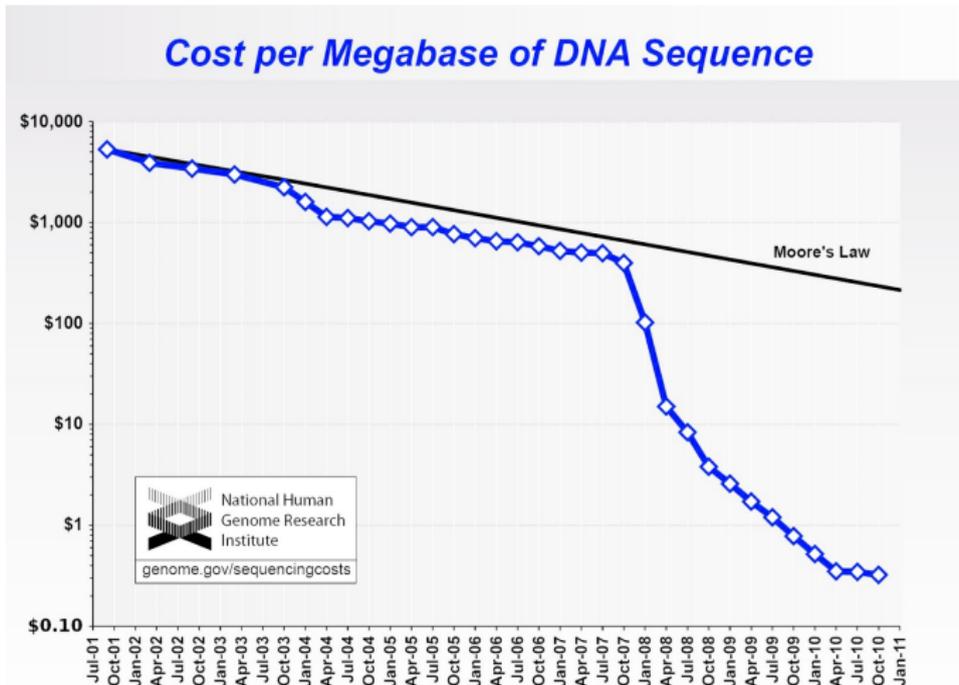
- O volume de transações é enorme e decisões devem ser tomadas muito rapidamente.
- O banco de dados é maior do que o da NASA.
- $5\mu\text{s}$ por transação custa muito dinheiro.
- 1 *click* do *mouse* corresponde a $500000\mu\text{s}$

Sequenciamento de DNA

Os métodos para produção de sequencias estão se tornando mais eficientes e gerando muito mais dados. Dá para processar em tempo hábil?



Custo do DNA × Lei de Moore



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes?
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes?
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Outros aspectos

- Visualização — como apresentar uma massa de dados correspondente a petabytes? (Quanto é isso?)
 - Processamento adicional
 - Apresentação iterativa
- *Data mining* — como achar informação relevante no meio de tantos dados?
 - Redes sociais
 - Resultados experimentais de LHC, Bioinfo, Astronomia, etc.
- Filtragem e tratamento de dados.



Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte**
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão

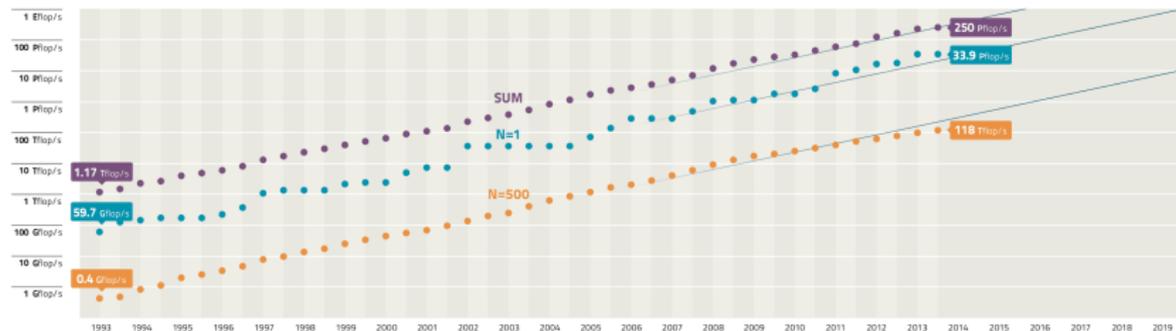


Velocidade (Top 500)

#	Site	Manufacturer	Computer	Country	Cores	Rmax (Flops)	Power (MW)
1	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	3,120,000	33.9	17.8
2	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	560,640	17.6	8.21
3	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	1,572,864	17.2	7.89
4	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	795,024	10.5	12.7
5	Argonne National Laboratory	IBM	Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	786,432	8.59	3.95
6	Swiss National Supercomputing Centre (CSCS)	Cray	Piz Daint Cray XC30, Xeon E5 8C 2.6GHz, Aries, NVIDIA K20x	Switzerland	115,984	6.27	2.33
7	Texas Advanced Computing Center/UT	Dell	Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi	USA	462,462	5.17	4.51
8	Forschungszentrum Juelich (FZJ)	IBM	JuQUEEN BlueGene/Q, Power BQC 16C 1.6GHz, Custom	Germany	458,752	5.01	2.30
9	Lawrence Livermore National Laboratory	IBM	Vulcan BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	393,216	4.29	1.97
10	Leibniz Rechenzentrum	IBM	SuperMUC iDataPlex DX360M4, Xeon E5 8C 2.7GHz, Infiniband FDR	Germany	147,456	2.90	3.52



Tendências



Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios**
- 6 Análise de desempenho
- 7 Conclusão



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- **Uso de multicore. Como aproveitar?**
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Desafios

- Uso de multicore. Como aproveitar?
- Precisão numérica.
- Balanceamento de carga.
- Ajuste a cenários (otimização dinâmica).
- Escalonamento e alocação de recursos.
- Tolerância a falhas.
- Desempenho na nuvem.
- Computação verde.



Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho**
- 7 Conclusão



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Análise de Desempenho

- Modelagem matemática.
- Modelagem estatística.
- Simulação.
- Modelos mistos.

Saber qual abordagem usar depende muito do problema e da plataforma.



Sumário

- 1 Introdução
- 2 A corrida pela velocidade
- 3 Problemas
 - Outros aspectos
- 4 Estado da arte
- 5 Desafios
- 6 Análise de desempenho
- 7 Conclusão



Conclusão

Tanto a evolução das máquinas, como a complexidade dos problemas tratados, crescem exponencialmente.
Isto é, estamos sempre no começo!



Conclusão

Tanto a evolução das máquinas, como a complexidade dos problemas tratados, crescem exponencialmente.
Isto é, estamos sempre no começo!



Conclusão

Dá para se divertir muuuuito!



OBRIGADO!!!

