



Uma revisão de Softwares de Benchmark para Avaliação de Desempenho de Microcomputadores

Helter Yordan Alves da Costa

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, PB
Novembro - 2024

Helter Yordan Alves da Costa

Uma revisão de Softwares de Benchmark para Avaliação de Desempenho de Microcomputadores

Monografia apresentada ao curso Engenharia da Computação do Centro de Informática, da Universidade Federal da Paraíba, como requisito para a obtenção do grau de Bacharel em Engenheiro da Computação.

Orientador: Ewerton Monteiro Salvador

João Pessoa, PB
Novembro – 2024

Catálogo na publicação
Seção de Catalogação e Classificação

C838r Costa, Helder Yordan A.

Uma revisão de softwares de Benchmark para avaliação
de desempenho de microcomputadores / Helder Yordan A
Costa. - João Pessoa, 2024.

44 f. : il.

Orientação: Ewerton Monteiro Salvador.
TCC (Graduação) - UFPB/CI.

1. Benchmark. 2. Desempenho de microcomputadores. 3.
Softwares de Benchmark. 4. Avaliação de hardware. I.
Salvador, Ewerton Monteiro. II. Título.

UFPB/CI

CDU 004.4

DEDICATÓRIA

Dedico este trabalho a todos que me apoiaram ao longo desta jornada. À toda minha família (esposa, filha, pais e irmão), pelo amor incondicional e incentivo constante, principalmente minha princesa, minha filha **Maria Liz** a quem dedico em especial, ela que veio ao mundo durante o transcorrer do curso e foi o que me deu forças para prosseguir e chegar a conclusão, para que no futuro possa servir de exemplo para ela e que assim ela possa almejar e batalhar por uma formação superior e um futuro prospero e promissor.

Dedicar também aos meus amigos, pela companhia e motivação nos momentos difíceis, uma caminhada longa e que não pode ser percorrida sozinho, espírito de camaradagem, união e força fazem parte da trajetória. Agradeço também aos meus professores, que compartilharam seu conhecimento e paixão, inspirando-me a seguir em frente.

AGRADECIMENTOS

Primeiramente, agradeço a Deus pela força e sabedoria para enfrentar os desafios e concluir esta etapa importante.

Aos meus pais (Erivaldo da Costa e Conceição de Maria Alves da Costa), minha esposa (Giselle Thaina do Nascimento Silva) e familiares, meu eterno agradecimento pelo apoio incondicional, pelas palavras de incentivo e pelo amor demonstrado ao longo dessa jornada. Vocês foram a base que me sustentou e me inspirou a cada dia.

Expresso minha gratidão aos professores e orientadores, que, com sua dedicação e conhecimentos compartilhados, contribuíram diretamente para o meu crescimento acadêmico e pessoal. Em especial, agradeço ao(a) Prof. Ewerton Monteiro Salvador, cuja orientação foi essencial para a realização deste trabalho, sua paciência, sugestões e apoio foram fundamentais para o desenvolvimento deste e a todos(as) os(as) professores(ras) que responderam pelo cargo de Coordenador do Curso de Engenharia da Computação durante a minha graduação que sempre com paciência e prestatividade se dispuseram a me auxiliar na resolução de diversas barreiras que enfrentei por toda caminhada acadêmica no que diz respeito a conciliar trabalho e estudo.

Aos meus colegas e amigos, que estiveram presentes em momentos de dúvidas, dificuldades e celebrações, deixo minha gratidão. Foram muitas as trocas de ideias, as noites de estudo e os incentivos mútuos que tornaram esta caminhada mais leve e significativa.

Por fim, agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho e para a conquista deste sonho. Cada um de vocês é parte importante desta conquista.

RESUMO

O presente trabalho realiza uma revisão bibliográfica de softwares de benchmark utilizados para a avaliação de desempenho de microcomputadores. Esses softwares têm como principal função medir e comparar o desempenho de componentes de hardware, como CPU, GPU, memória e armazenamento, proporcionando métricas essenciais para usuários e empresas na escolha de sistemas mais eficientes. A pesquisa busca identificar as principais ferramentas de benchmark disponíveis, analisando seus benefícios e limitações em diferentes contextos de uso. Entre os softwares investigados, destacam-se aqueles de uso gratuito, enfatizando sua aplicabilidade e confiabilidade em cenários práticos. O objetivo central do estudo é fornecer uma visão crítica sobre a eficácia dos benchmarks na avaliação do desempenho de microcomputadores, identificando os melhores softwares para diferentes finalidades. A metodologia utilizada inclui uma revisão bibliográfica e documental da literatura que aborda o funcionamento de ferramentas como PCMark, Sysmark e Cinebench. Além disso, o trabalho compara em forma de tabela as principais características desses softwares, buscando verificar suas semelhanças e diferenças. Entre os principais resultados, o estudo revela que os benchmarks analisados apresentam variações significativas nos resultados, dependendo do tipo de teste e da configuração do sistema, o que demonstra a importância de escolher a ferramenta adequada para cada finalidade. A pesquisa também destaca a necessidade de uma maior padronização e transparência nos algoritmos utilizados, para garantir que os resultados reflitam com mais precisão o desempenho real dos microcomputadores.

Palavras-chave: Benchmark. Desempenho de Microcomputadores. Softwares de Benchmark. Avaliação de Hardware.

ABSTRACT

This paper performs a bibliographic review of benchmark software used to evaluate the performance of microcomputers. The main function of these software programs is to measure and compare the performance of hardware components, such as CPU, GPU, memory and storage, providing essential metrics for users and companies when choosing more efficient systems. The research seeks to identify the main benchmark tools available, analyzing their benefits and limitations in different contexts of use. Among the software investigated, those that are free to use stand out, emphasizing their applicability and reliability in practical scenarios. The main objective of the study is to provide a critical view on the effectiveness of benchmarks in evaluating the performance of microcomputers, identifying the best software for different purposes. The methodology used includes a bibliographic and documentary review, analyzing studies and reports that address the operation of tools such as PCMark, Sysmark and Cinebench, among others. In addition, the work compares the main characteristics of these software programs in a table format, seeking to verify their accuracy and reliability. Among the main findings, the study reveals that the benchmarks analyzed present significant variations in results, depending on the type of test and system configuration, which demonstrates the importance of choosing the right tool for each purpose. The research also highlights the need for greater standardization and transparency in the algorithms used, to ensure that the results more accurately reflect the actual performance of microcomputers.

Key-words: Benchmark. Microcomputer Performance. Benchmark Software. Hardware Evaluation.

LISTA DE FIGURAS

Figura 1. Diferença entre benchmarks sintéticos e específicos.	21
Figura 2. tela inicialização PCMark 10.....	23
Figura 3. Faixa de precisão PCMark 10.....	25
Figura 4. tela de inicialização SYSMark 30	26
Figura 5. tela de inicialização cinebench	29

LISTA DE TABELAS

Tabela 1. requisitos minimos pcmark 10	23
Tabela 2. requisitos minimos sysmark 30.....	26
Tabela 3. requisitos minimos cinebench	29
Tabela 4. análise comparativa dos benchmarks	35
Tabela 5. comparativo intel core i3 (sysmark).....	38
Tabela 6. comparativo intel core i5 (sysmark).....	38
Tabela 7. comparativo intel core i7 (sysmark).....	38

LISTA DE SIGLAS

SoC	-	System-on-a-Chip (Sistema em um chip)
CPU	-	Central Processing Unit ou unidade central de processamento
GPU	-	Graphics Processing Unit ou unidade de processamento gráfico
BAPCo	-	Business Applications Performance Corporation
FPS	-	Frames por segundo
EDMS	-	Enterprise Document Management System
VS	-	Versus

SUMÁRIO

1. INTRODUÇÃO.....	12
1.1. PROBLEMA	15
1.2. OBJETIVO GERAL	15
1.2.1. OBJETIVOS ESPECÍFICOS.....	16
1.3. ESTRUTURA DA MONOGRAFIA.....	16
2. METODOLOGIA	17
3. REVISÃO BIBLIOGRAFICA	19
3.1. FUNDAMENTOS DOS SOFTWARES DE BENCHMARK.....	19
3.2. PRINCIPAIS SOFTWARES DE BENCHMARK E SUAS APLICAÇÕES	21
3.3. PCMARK.....	22
3.4. SYSMARK.....	25
3.5. CINEBENCH	28
3.6. DESAFIOS E LIMITAÇÕES DOS SOFTWARES DE BENCHMARK	32
4. APRESENTAÇÃO E ANÁLISE COMPARATIVA ENTRE OS SOFTWARES	35
4.1. ANÁLISE DE RESULTADOS DOS SOFTWARES DE BENCHMARK	36
4.2. COMPARAÇÃO COM OUTROS TRABALHOS	37
4.3. DESAFIOS E LIMITAÇÕES DOS SOFTWARES	39
5. CONCLUSÃO.....	40
REFERÊNCIAS	43

1. INTRODUÇÃO

A tecnologia desempenha um papel central em todos os aspectos da vida contemporânea, impactando desde a forma como nos comunicamos até as maneiras como trabalhamos, aprendemos e nos entretemos. O desenvolvimento contínuo da capacidade computacional e a miniaturização dos dispositivos eletrônicos têm possibilitado a criação de tecnologias cada vez mais poderosas e acessíveis. Esses avanços tecnológicos não apenas facilitaram o cotidiano, mas também trouxeram novas possibilidades para diversas áreas do conhecimento, como a ciência, a educação, a saúde e a indústria. Diante dessa realidade, o uso eficiente de ferramentas para medir e comparar o desempenho de microcomputadores e seus componentes tornou-se uma necessidade fundamental para garantir a escolha das melhores soluções tecnológicas. Nesse sentido, o papel dos benchmarks se destaca como um método essencial para a avaliação e o aprimoramento contínuo da tecnologia.

Historicamente, a evolução dos microcomputadores foi marcada por uma série de inovações que aumentaram drasticamente sua capacidade de processamento, armazenamento e comunicação. A Lei de Moore, proposta em 1965 por Gordon Moore, cofundador da Intel, trouxe uma previsão importante sobre o ritmo do desenvolvimento tecnológico. Moore observou que o número de transistores em um chip de silício dobraria aproximadamente a cada dois anos, o que resultaria em uma expansão exponencial da capacidade computacional e na redução dos custos de produção. Essa previsão foi precisa ao longo de várias décadas e impulsionou a rápida evolução dos microcomputadores, permitindo que esses dispositivos se tornassem mais acessíveis e amplamente utilizados em diferentes setores da sociedade (Hennessy e Patterson, 2014).

Os microcomputadores, que no início eram dispositivos volumosos e restritos a grandes corporações e centros de pesquisa, agora são parte integrante do cotidiano de milhões de pessoas ao redor do mundo. Esses avanços foram acompanhados de uma demanda crescente por dispositivos mais compactos, eficientes e energeticamente sustentáveis. Com a popularização de dispositivos móveis, como smartphones e tablets, a indústria passou a focar não apenas no aumento da capacidade de processamento, mas também na eficiência energética. A integração de sistemas em um único chip (SoC) foi um dos grandes saltos nesse campo, permitindo que múltiplos componentes, como

processadores, memória e unidades de armazenamento, fossem combinados em um único circuito integrado. Essa integração trouxe melhorias significativas no desempenho e na economia de energia, além de permitir a produção de dispositivos menores e mais potentes (Hennessy e Patterson, 2014).

Entretanto, para que os consumidores e as empresas possam avaliar o real desempenho desses dispositivos, surgiram os testes de benchmarking. O benchmark é uma metodologia de avaliação que permite comparar o desempenho de diferentes sistemas ou componentes sob condições controladas (Hennessy e Patterson, 2014). Através de testes padronizados, os benchmarks geram resultados quantitativos que podem ser utilizados para comparar a eficiência e a velocidade de execução de tarefas entre diferentes dispositivos. Esses testes são especialmente importantes em um mercado competitivo, onde a inovação tecnológica é constante, e consumidores e empresas precisam de informações precisas para tomar decisões de compra.

Os benchmarks desempenham um papel essencial na indústria de tecnologia da informação, pois permitem que fabricantes e consumidores verifiquem o desempenho de sistemas e componentes em diferentes cenários. Os testes de benchmarking cobrem uma ampla gama de categorias, como o desempenho de CPUs, GPUs, memória, armazenamento e até a eficiência energética. Esses testes são utilizados por fabricantes para otimizar seus produtos e por consumidores para avaliar qual dispositivo melhor atende às suas necessidades. Na academia, os benchmarks também são amplamente utilizados em pesquisas para testar novos algoritmos, arquiteturas e soluções tecnológicas.

Entre os principais softwares de benchmarking disponíveis no mercado, destacam-se o PCMark e o Sysmark, que são amplamente utilizados para medir o desempenho de computadores em tarefas do dia a dia, como navegação na web, edição de documentos e reprodução de vídeo. Esses testes simulam o uso real dos computadores e fornecem uma visão abrangente de como os dispositivos se comportam em diferentes cenários. Para testes mais específicos, como o desempenho gráfico, softwares como Cinebench são frequentemente utilizados para avaliar a capacidade de renderização e o desempenho da unidade de processamento gráfico (GPU).

Apesar da ampla utilização dos benchmarks, é importante reconhecer que esses testes nem sempre refletem o desempenho real do sistema em situações de uso cotidiano. Isso ocorre porque os benchmarks são projetados para medir o desempenho sob condições específicas e padronizadas, o que pode não corresponder exatamente aos

cenários de uso variado de cada usuário. Um dispositivo que apresenta bom desempenho em um benchmark sintético pode não ser tão eficiente em determinadas aplicações do mundo real. Dessa forma, é fundamental que os benchmarks sejam utilizados como uma ferramenta de orientação, e não como a única métrica para determinar o desempenho de um dispositivo.

Um dos principais desafios no uso de benchmarks é justamente a adaptação dos testes para cenários de uso mais realistas, cenários voltados a simulação de atividades cotidianas semelhantes as exercidas por um usuário, levando em consideração suas nuances e variações de uso. Benchmarks sintéticos, que simulam situações específicas, podem não representar com precisão as condições de uso do usuário comum. Por exemplo, um teste que mede apenas o desempenho bruto da CPU pode não levar em conta o comportamento do sistema em multitarefas ou o impacto do uso prolongado em termos de aquecimento e consumo de energia. Por esse motivo, há uma crescente demanda por benchmarks específicos de aplicação (application-specific benchmarking), que focam em avaliar o desempenho de sistemas em contextos mais próximos da realidade de uso.

Na literatura científica, o trabalho de Zhang (2001) é amplamente citado quando se discute benchmarks específicos de aplicação. Zhang propôs uma abordagem para a criação de benchmarks que fossem adaptados a contextos de uso mais específicos, permitindo uma avaliação mais precisa do desempenho dos sistemas. Esse tipo de benchmark é particularmente útil em setores como a computação científica, onde o desempenho em tarefas específicas, como simulações complexas ou modelagem de grandes volumes de dados, é de vital importância.

Portanto, os benchmarks são ferramentas indispensáveis na avaliação de desempenho de microcomputadores e seus componentes. Eles fornecem métricas objetivas que permitem comparar diferentes dispositivos e ajudam tanto consumidores quanto fabricantes a tomar decisões fundamentadas. No entanto, é importante estar ciente de suas limitações, especialmente no que diz respeito à capacidade de representar cenários de uso reais. O avanço dos benchmarks específicos de aplicação, como proposto por Zhang (2001), aponta para um futuro em que a medição de desempenho será cada vez mais precisa e alinhada às necessidades do usuário final. Assim, a combinação de testes sintéticos e específicos de aplicação pode oferecer uma visão mais abrangente e confiável sobre o verdadeiro desempenho dos dispositivos tecnológicos.

1.1. Problema

O avanço das arquiteturas de computadores tem sido um dos pilares centrais na evolução da tecnologia da informação, promovendo inovações em desempenho, eficiência e versatilidade. Ao longo dos anos, as arquiteturas computacionais passaram por transformações significativas, adaptando-se às crescentes demandas por processamento e armazenamento de dados. No entanto, cada aplicação possui suas finalidades específicas, exigindo um conjunto de hardwares com características próprias e diferentes recursos. Tais recursos, muitas vezes, não conseguem aliar eficiência e resultados ideais de forma simultânea.

Para suprir a necessidade de comparação entre diferentes softwares e hardwares, surgiram ferramentas conhecidas como *benchmarks*, que serão abordadas mais detalhadamente adiante. No entanto, o x do problema reside na vasta quantidade de softwares de benchmark disponíveis, cada um com suas próprias métricas de avaliação, que nem sempre refletem de maneira precisa a realidade ou a finalidade específica de uso do sistema pelo usuário.

Isso gera um dilema: os resultados obtidos em testes realizados com diferentes softwares de benchmark podem variar significativamente, mesmo utilizando o mesmo conjunto de hardwares. Essa discrepância pode comprometer a análise do desempenho, tornando necessário investigar qual benchmark é mais adequado para cada finalidade ou contexto de uso. Assim, é adequado identificar qual ferramenta oferece os resultados mais precisos e confiáveis de acordo com as demandas específicas do usuário.

1.2. Objetivo geral

Dada a crescente complexidade dos microcomputadores e a variedade de suas aplicações, torna-se indispensável uma avaliação precisa de desempenho para otimizar o uso dos recursos e proporcionar a melhor experiência ao usuário. A escolha do software de *benchmark* adequado é essencial para profissionais envolvidos no desenvolvimento, teste e manutenção de sistemas computacionais, pois garante uma análise detalhada e confiável dos componentes de hardware.

O objetivo deste trabalho é realizar uma análise crítica dos principais softwares de *benchmark* utilizados na avaliação de desempenho de microcomputadores, com ênfase nas ferramentas gratuitas. A análise buscará identificar as características, vantagens e limitações de cada software, fornecendo uma visão comparativa que auxilie na escolha da ferramenta mais adequada para diferentes finalidades e contextos de uso.

1.2.1. Objetivos específicos

- Realizar uma revisão da literatura sobre benchmarks e metodologias de avaliação de desempenho de microcomputadores;
- Identificar e analisar os principais softwares de benchmark disponíveis no mercado, com foco nas ferramentas gratuitas visando o público comum;
- Comparar os softwares selecionados com base em critérios previamente definidos, como precisão, usabilidade e adequação a diferentes tipos de testes;
- Fornecer recomendações fundamentadas para a escolha do software de benchmark mais adequado para diferentes cenários e finalidades de uso.

1.3. Estrutura da monografia

O presente trabalho está organizado em cinco capítulos de forma sequencial e metodológica afim de proporcionar uma explanação do tema de forma sequencial, iniciando no capítulo 1 apresentando a parte introdutória, incluindo a contextualização e os objetivos pretendidos, capítulo 2 trabalha a descrição da metodologia que será utilizada na pesquisa de forma mais específica, no capítulo 3, é abordada uma fundamentação teórica sobre o tema discutindo conceitos e algumas metodologias relacionadas, seguindo para o capítulo 4 onde será apresentada toda a análise comparativa dos softwares de benchmarks que vamos abordar e o capítulo 5 onde será finalizada a pesquisa com as considerações finais e sugestões para futuros estudos.

2. METODOLOGIA

A metodologia é uma parte essencial de qualquer pesquisa, pois fornece o caminho que o pesquisador seguirá para alcançar seus objetivos e responder às perguntas formuladas. No presente trabalho, que se propõe a analisar o desempenho de microcomputadores utilizando softwares de benchmark, optou-se pela realização de uma pesquisa bibliográfica. Este método é amplamente aceito na comunidade científica quando o objetivo é explorar o estado da arte sobre um tema, compreender conceitos já estabelecidos e identificar lacunas que ainda precisam ser abordadas por pesquisas futuras.

A pesquisa bibliográfica consiste em levantar, organizar e interpretar dados já existentes sobre o objeto de estudo, reunidos a partir de fontes secundárias como livros, artigos científicos, dissertações, teses, manuais técnicos e publicações em veículos de comunicação especializados. Dessa forma, é possível reunir um conjunto consistente de informações que permita ao pesquisador construir uma visão abrangente e fundamentada sobre o tema abordado, sem a necessidade de coleta de dados primários.

No contexto deste estudo, o foco principal foi o levantamento de fontes que discutem o uso de ferramentas de benchmark na análise de desempenho de microcomputadores. Essas fontes foram selecionadas com base em sua relevância, pertinência ao tema e atualidade, sempre que possível, privilegiando aquelas que abordam os softwares de benchmark mais amplamente utilizados, como PCMark, Sysmark e Cinebench. Além disso, foram incluídas na pesquisa referências que tratam dos fundamentos dos benchmarks, sua evolução histórica e seus principais desafios e limitações, como as críticas ao uso de benchmarks sintéticos e a relevância dos benchmarks específicos de aplicação (application-specific benchmarking).

O processo de levantamento das fontes bibliográficas foi realizado por meio de buscas em bases de dados científicas, como IEEE Xplore, Scopus e Google Scholar, além de consultas pela internet de modo geral e a publicações técnicas de fabricantes de software e hardware, como Futuremark e BAPCo, responsáveis por desenvolver ferramentas amplamente utilizadas na indústria. O critério de seleção das fontes incluiu a data de publicação — priorizando-se trabalhos dos últimos cinco anos —, a relevância dos autores no campo da computação e a pertinência dos tópicos discutidos em relação ao objetivo do estudo.

Após a coleta das fontes, a análise foi realizada de forma a identificar os pontos convergentes e divergentes entre os estudos, destacando os principais resultados e conclusões dos autores. Foi dada ênfase especial à discussão sobre a aplicabilidade dos diferentes tipos de benchmark na avaliação de sistemas de microcomputadores, uma vez que essa questão é central para o entendimento das limitações e potencialidades dessas ferramentas.

Além disso, buscou-se identificar como os avanços tecnológicos, como o desenvolvimento de novos processadores, GPUs e sistemas de armazenamento, têm influenciado a evolução dos benchmarks ao longo dos anos. Essa análise permitiu traçar um panorama sobre a evolução dos softwares de benchmark e suas implicações para a análise de desempenho de sistemas modernos.

Em termos de abordagem, optou-se por uma análise qualitativa dos dados bibliográficos, uma vez que o objetivo principal da pesquisa é entender os conceitos e teorias relacionados aos benchmarks e discutir criticamente suas aplicações no cenário atual. Embora o uso de benchmarks envolva dados quantitativos, como as pontuações obtidas em testes de desempenho, a metodologia escolhida não se propõe a analisar esses números diretamente, mas sim a compreender como eles são gerados e interpretados no contexto da avaliação de sistemas de microcomputadores.

Ao utilizar uma abordagem bibliográfica, este trabalho espera contribuir para o campo de estudo ao oferecer uma análise crítica dos softwares de benchmark disponíveis, suas aplicações e suas limitações, auxiliando tanto pesquisadores quanto profissionais da área de tecnologia da informação a fazer escolhas mais informadas na utilização dessas ferramentas. Além disso, a pesquisa pretende identificar oportunidades para futuras investigações, especialmente no que diz respeito ao desenvolvimento de benchmarks mais precisos e representativos dos fluxos de trabalho reais.

Portanto, a escolha pela pesquisa bibliográfica como metodologia se justifica pela necessidade de consolidar o conhecimento já existente sobre o tema, ao mesmo tempo em que se busca uma compreensão mais profunda dos desafios enfrentados pela indústria na avaliação de desempenho de sistemas complexos.

3. REVISÃO BIBLIOGRAFICA

A revisão bibliográfica é uma etapa fundamental para a construção de uma base teórica sólida, permitindo situar o tema em estudo no contexto das pesquisas já realizadas e identificar lacunas que ainda precisam ser abordadas. Essa seção tem o objetivo de apresentar, analisar e sintetizar as principais teorias, conceitos e descobertas de autores renomados na área, proporcionando uma compreensão mais profunda do assunto e estabelecendo um referencial que guiará as discussões e contribuições deste trabalho. Por meio dela, espera-se oferecer uma visão crítica e abrangente que justifique a relevância e a originalidade do tema abordado, alinhando-o com as práticas e avanços atuais.

Nesta pesquisa, dentre a enorme gama de softwares de benchmark que temos no mercado foi observado em diversas citações, três softwares benchmark que se destacam fortemente no mercado e na área de pesquisa de modo geral que são o PCMark (atualmente na versão 10), o SYSMark (atualmente na versão 30) e o CineBench (atualmente na versão 2024), levando em consideração as classificações abordadas no trabalho que são: benchmark sintético e benchmarks específicos. Vamos focar nestes três softwares e entender a forma como suas pontuações são geradas e suas formas de atuação.

3.1. Fundamentos dos Softwares de Benchmark

Os softwares de benchmark desempenham um papel essencial na avaliação do desempenho de microcomputadores, fornecendo uma maneira objetiva e padronizada de comparar diferentes componentes de hardware e sistemas. O conceito de benchmarking remonta à necessidade de medir e comparar o desempenho de dispositivos e sistemas computacionais, especialmente à medida que as tecnologias evoluem e as demandas por sistemas mais eficientes aumentam. No contexto da informática, os benchmarks são ferramentas essenciais para garantir que um sistema esteja operando de acordo com os padrões esperados ou para auxiliar na seleção de componentes e dispositivos que atendam às necessidades de desempenho exigidas em determinadas aplicações.

Historicamente, o uso de benchmarks começou com testes simples, que se concentravam na medição de aspectos como a velocidade de processamento de uma CPU ou o desempenho de armazenamento em discos rígidos (Moreira; Lima, 2014). Com o tempo, essas ferramentas evoluíram para incorporar uma variedade de métricas que avaliam não apenas o desempenho de componentes isolados, mas também o comportamento de sistemas em cenários de uso reais. Atualmente, os benchmarks abrangem desde simples testes de processamento até simulações complexas que medem a eficiência em contextos como edição de vídeos, jogos e produtividade em escritórios.

As métricas fornecidas pelos benchmarks são fundamentais para permitir que tanto usuários quanto empresas tomem decisões informadas sobre a aquisição e a configuração de sistemas. Isso é particularmente importante em um mercado onde o desempenho dos dispositivos varia amplamente, e onde é necessário identificar as melhores soluções para necessidades específicas. Além disso, os benchmarks proporcionam uma referência comum que pode ser utilizada em estudos comparativos, permitindo avaliar diferentes configurações de hardware com base em resultados padronizados (Futuremark Corporation, 2021).

Os benchmarks podem ser amplamente classificados em duas categorias: benchmarks sintéticos e benchmarks aplicativos ou específicos, conforme ilustrado na Figura 1. Os benchmarks sintéticos são projetados para simular cenários artificiais e testar o desempenho de componentes individuais de um sistema, como a CPU ou a GPU muitas vezes focado em uma única métrica. Por outro lado, os benchmarks aplicativos ou específicos são baseados em programas reais e simulam cenários de uso diário, como edição de vídeo, renderização gráfica ou manipulação de grandes bases de dados (Zhang, 2001). Abaixo vemos uma imagem que demonstra exemplificando diferenças de foco dos benchmarks sintéticos e específicos, nela podemos ver bem o foco de atuação de cada modelo.

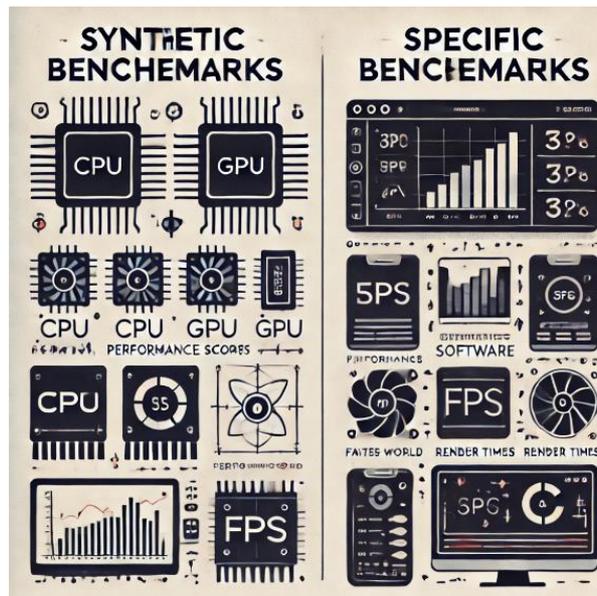


Figura 1. Diferença entre benchmarks sintéticos e específicos.

Ambos os tipos têm seus méritos e limitações. Enquanto os benchmarks sintéticos podem fornecer uma visão detalhada sobre o desempenho de componentes específicos, eles muitas vezes não refletem com precisão o desempenho em situações do mundo real. Já os benchmarks aplicativos oferecem uma visão mais realista do desempenho do sistema, mas podem ser limitados em termos de variabilidade dos resultados.

3.2. Principais Softwares de Benchmark e Suas Aplicações

Os softwares de benchmark são ferramentas imprescindíveis para avaliar o desempenho de sistemas e componentes de microcomputadores, sendo amplamente utilizados tanto por entusiastas quanto por profissionais da área de tecnologia. Dentre os principais softwares disponíveis no mercado, destacam-se o PCMark, Sysmark e Cinebench, cada um especializado em medir diferentes aspectos de desempenho, como o processamento da CPU, a capacidade gráfica da GPU, o uso de memória e o desempenho de armazenamento. A confiabilidade e a aplicação dessas ferramentas dependem do cenário em que são empregadas, e cada uma delas possui características específicas que as tornam adequadas para determinados tipos de avaliação.

Além dos softwares mencionados, é importante destacar que há uma variedade

de outros benchmarks no mercado, cada um com suas próprias características e áreas de especialização. Alguns, como o 3DMark, são focados no desempenho gráfico e são amplamente utilizados por gamers e entusiastas de hardware que buscam otimizar suas máquinas para jogos. Outros, como o PassMark, oferecem uma avaliação mais generalista, medindo o desempenho de todos os principais componentes do sistema, como CPU, GPU, memória e armazenamento, oferecendo uma visão equilibrada do desempenho geral do computador (Zhang, 2001).

No entanto, a escolha do software de benchmark mais adequado depende das necessidades específicas de cada usuário ou organização. Para aqueles que necessitam avaliar o desempenho de seus sistemas em tarefas diárias e gerais, o PCMark oferece uma avaliação abrangente e confiável. Para empresas que buscam otimizar seus sistemas para cenários corporativos, o Sysmark é a escolha ideal. Já para profissionais criativos ou desenvolvedores que dependem de um alto desempenho gráfico, o Cinebench se destaca como a ferramenta mais apropriada.

Os softwares de benchmark são ferramentas poderosas para avaliar o desempenho de sistemas em diferentes cenários. Enquanto o PCMark foca em um uso mais generalista e o Sysmark oferece uma visão detalhada de fluxos de trabalho empresariais, o Cinebench se especializa em medir a eficiência gráfica e de processamento em tarefas de renderização 3D. Cada uma dessas ferramentas tem suas vantagens e limitações, e a escolha do software mais adequado depende diretamente das necessidades específicas do usuário ou da organização (Zhang, 2001; Futuremark Corporation, 2021).

3.3. PCMark

Um exemplo clássico de benchmark aplicativo é o PCMark, atualmente na versão 10 e possuindo versão de avaliação de forma gratuita, é uma ferramenta amplamente utilizada para avaliar o desempenho de sistemas de uso pessoal em uma variedade de cenários, como produtividade em escritórios, criação de conteúdo digital e jogos. A Figura 2 demonstra a tela inicial do software.

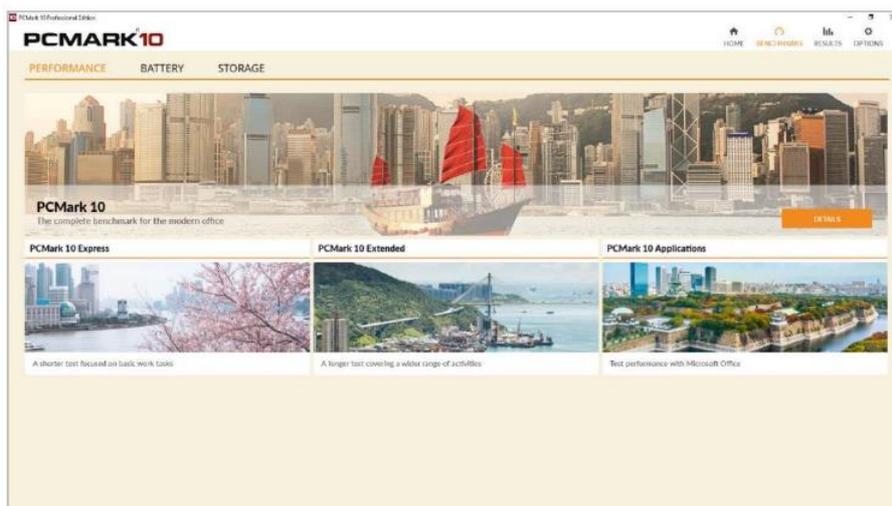


Figura 2. tela inicialização PCMark 10

Desenvolvido pela Futuremark Corporation, o PCMark 10 se destaca por seu foco em medir o desempenho geral de um sistema em condições de uso do dia a dia, o que o torna uma escolha popular tanto para consumidores quanto para empresas que desejam avaliar sistemas com base em sua eficiência em tarefas reais (Futuremark Corporation, 2021). A Tabela 1 demonstra o quadro de requisitos mínimos solicitados pela desenvolvedora para o software.

Tabela 1. requisitos minimos pemark 10

	PCMARK 10
OS	Windows® 7 SP1 64-bit
Processador	Dual core processor
Memória RAM	4 GB
Gráficos	DirectX 11 GPU
Resolução	1920 × 1080
Fator de escala de exibição	100%
Armazenamento	6 GB de espaço livre

A ferramenta oferece uma série de testes que abrangem desde a execução de planilhas até a renderização de gráficos complexos, permitindo que os usuários identifiquem pontos fortes e fracos de seus sistemas em diferentes contextos. Ele é

dividido em grupos de testes principais:

- **Essentials:** Testa tarefas básicas como navegação na web e videoconferência.
- **Productivity:** Avalia atividades de escritório como o uso de planilhas e escrita de documentos.
- **Digital Content Creation:** Mede o desempenho em edição de fotos, vídeos e renderização.
- **Gaming** (no PCMark 10 Extended): Testa capacidades gráficas e de processamento para jogos.

Cada grupo possui diferentes cargas de trabalho e gera uma pontuação específica, que é combinada para obter uma pontuação geral do sistema. Essas pontuações são úteis para comparar a capacidade de diferentes sistemas em atender às demandas de trabalho.

A pontuação final é baseada no desempenho do hardware em cada carga de trabalho e em cada grupo de testes, permitindo uma avaliação ampla das capacidades de um PC para um ambiente de escritório ou necessidades de criação digital. Esses resultados são divididos em níveis alto, médio e baixo para facilitar a interpretação dos dados.

A pontuação no PCMark 10 é baseada em uma combinação geométrica dos resultados de vários testes, agrupados em categorias que simulam atividades comuns de uso de PC. A pontuação total é gerada pela fórmula:

$$\text{Pontuação total} = K \times \text{média Geométrica (Se,Sp,Sd)}$$

K = constante de escala

Se = Pontuação do grupo “Essentials”

Sp = Pontuação do grupo “Productivity”

Sd = Pontuação do grupo “Digital content Creation”

Para determinar a pontuação de cada grupo, o PCMark 10 utiliza a média geométrica dos testes individuais. Por exemplo, no grupo *Essentials*, os testes medem o desempenho de tarefas como navegação na web, videochamadas e tempo de inicialização de aplicativos. Já o grupo *Productivity* avalia desempenho em atividades de planilhas e redação. No grupo *Digital Content Creation*, os testes incluem edição de

foto e vídeo e renderização. Cada um desses grupos representa uma parte importante das atividades típicas de um usuário de PC, tornando a pontuação do PCMark 10 uma medida representativa para avaliar o desempenho geral do sistema.

A precisão das pontuações do PCMark 10 geralmente é melhor do que 3% ao seguir as etapas descritas em seu guia de testes. Isso significa que executar o benchmark repetidamente em um sistema igual e bem controlado produzirá pontuações que se enquadram em uma faixa de 3% de erro entre elas. (Technical Guide. 2021)

Na Figura 3 vemos o gráfico que mostra a variação da taxa de precisão em relação ao valor referência, que é o valor obtido por testes passados.

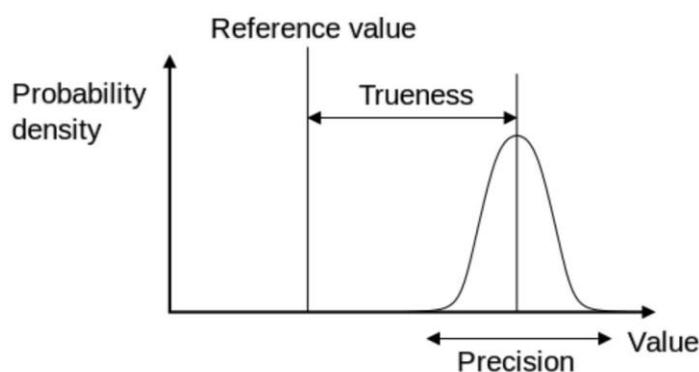


Figura 3. Faixa de precisão PCMark 10

3.4. SYSmark

Além do PCMark, outros benchmarks amplamente conhecidos incluem o Sysmark, que é desenvolvido pela BAPCo e foca na medição de desempenho em ambientes corporativos, avaliando desde tarefas básicas de escritório até simulações de fluxos de trabalho empresariais mais complexos, este não possuindo versão de avaliação ou teste de forma gratuita, apenas versões pagas.

A Tabela 2 apresenta o quadro de requisitos mínimos solicitados pela desenvolvedora para o software.

Tabela 2. requisitos minimos sysmark 30

	SYSMark 30
OS	Windows® 10 ou superior
Processador	processador x86 2015 ou mais recente
Memória RAM	16 GB
Gráficos	DirectX 12
Resolução	1280x800 ou 1366x768
Fator de escala de exibição	100%
Armazenamento	30 GB de espaço livre

O SYSmark 30 foi projetado para aqueles que desejam avaliar e comparar dispositivos de PC baseados em Windows x64 em termos de desempenho diferenciando cargas de trabalho, fornecendo informações úteis ao(s) seu(s) público(s) para auxiliar na avaliação e compra de dispositivos. Diferentemente dos benchmarks sintéticos, que estressam artificialmente os componentes do PC ou tentam modelar o desempenho usando uma simulação estática do comportamento do aplicativo, o SYSmark 30 é baseado em aplicativos, emprega conjuntos de dados como aqueles que os usuários encontrariam em um ambiente de negócios e fornece uma visão do desempenho do PC em nível de sistema (BAPCo. 2022). Na Figura 4 vemos a tela inicial do software.



Figura 4. tela de inicialização SYSMark 30

A utilização desse tipo de benchmark permite que empresas façam escolhas mais informadas ao selecionar sistemas para diferentes departamentos, assegurando que os dispositivos adquiridos atendam às demandas específicas de seus fluxos de trabalho.

O Sysmark 30 utiliza várias métricas para avaliar o desempenho do sistema, incluindo:

- **Pontuação Total:** Um valor agregado que reflete a performance geral do sistema em todas as tarefas testadas.
- **Pontuações por Categoria:** O benchmark é dividido em várias categorias, como produtividade, criação de conteúdo e navegação na web, permitindo que os usuários vejam o desempenho em áreas específicas.
- **Métricas de Tempo de Resposta:** Mede quanto tempo um sistema leva para completar tarefas específicas, refletindo a experiência do usuário.
- **Taxa de Processamento:** Avalia a eficiência do sistema em executar operações por unidade de tempo, considerando o uso de CPU, memória e armazenamento.

Para o SYSmark 30, a BAPCo avaliou os méritos de uma variedade de metodologias de pontuação e escolheu uma metodologia com base sólida enfatizando o multitarefas, variações mínimas possíveis em repetidos testes e com facilidade de entendimento por parte do aplicador. Ele avalia o desempenho do sistema medindo o tempo de resposta de tarefas em um PC usando aplicativos reais e entrada simulada do usuário. A metodologia de pontuação e os tempos de resposta de tarefa são usados para gerar classificação de desempenho que reflita a experiência do usuário. Quanto mais rápido um PC responder às cargas de trabalho do aplicativo no SYSmark 30, maior será sua classificação de desempenho. (BAPCo. 2022)

O SYSmark 30 produz as quatro pontuações. Todas as pontuações são relativas ao sistema de calibração, que pontua um valor de 1000 para cada classificação de cenário e a pontuação geral. A pontuação geral é produzida somente se todos os quatro cenários forem executados e é calculada tomando a geomédia das quatro classificações de cenário.

$$\text{PONTUAÇÃO GERAL} = \sqrt[4]{S_1 * S_2 * S_3 * S_4}$$

S1 = Aplicativos de escritório (OA)

S2 = Produtividade Geral (GP)

S3 = Edição de fotos (PE)

S4 = Criação Avançada de Conteúdo (CC)

As classificações de cenário são calculadas com a seguinte fórmula:

$$\text{classificação de cenário} = \left(\frac{\text{Tempo de execução do cenário na calibração}}{\text{tempo de execução do cenário no SUT}} \right) \cdot 1000$$

O Sysmark 30 é uma ferramenta essencial para profissionais de TI, fabricantes de hardware e usuários que desejam entender e otimizar o desempenho de seus sistemas. Com suas melhorias significativas e métricas detalhadas, ele oferece uma visão abrangente do desempenho em cenários do mundo real, ajudando a tomar decisões informadas sobre atualizações e configurações de sistema.

3.5. CineBench

Além do PCMark e do Sysmark, outro software amplamente utilizado é o Cinebench, que é um benchmark especializado na avaliação de desempenho gráfico e de processamento da CPU. O Cinebench é desenvolvido pela Maxon e é frequentemente usado para medir a eficiência de processadores em tarefas que envolvem renderização 3D. Esse software é particularmente relevante para profissionais das áreas de design gráfico, animação e produção de conteúdo audiovisual, onde o desempenho da GPU e da CPU é essencial para garantir fluxos de trabalho rápidos e eficientes (Moreira; Lima, 2014). A Figura 5 mostra a tela de abertura do software.

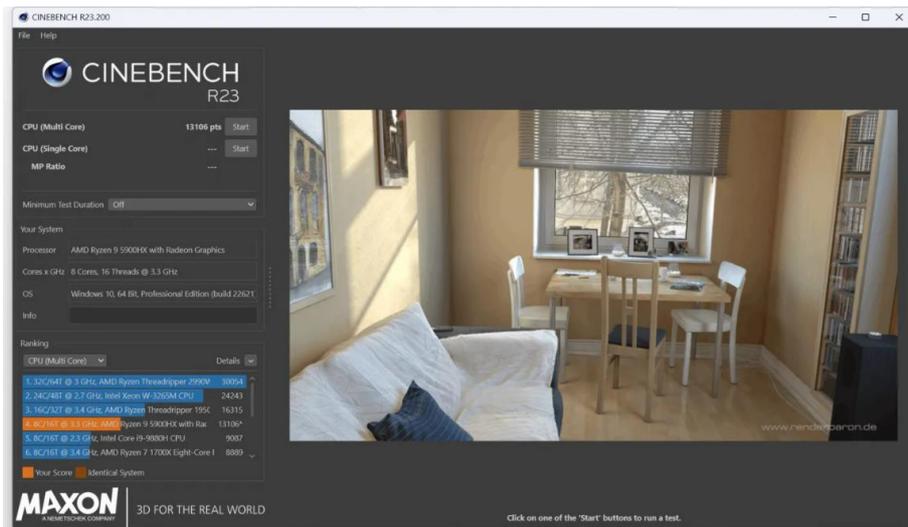


Figura 5. tela de inicialização cinebench

O Cinebench realiza seus testes utilizando o motor de renderização do Cinema 4D, um software profissional de modelagem e animação 3D, o que torna os resultados altamente relevantes para quem trabalha com essas aplicações. Diferente de outros benchmarks, que medem a performance em tarefas de uso geral, o Cinebench foca exclusivamente em cenários que exigem um alto desempenho gráfico e de processamento, fornecendo métricas detalhadas sobre a capacidade da CPU e da GPU em lidar com renderizações complexas. Isso o torna uma ferramenta imprescindível para profissionais criativos e para aqueles que utilizam programas intensivos em termos de processamento gráfico, como softwares de edição de vídeo ou jogos (Moreira; Lima, 2014).

A Tabela 3 demonstra o quadro de requisitos mínimos solicitados pela desenvolvedora para o software.

Tabela 3. requisitos minimos cinebench

	CineBench
OS	- Windows® 10 ou superior - macOS 12.6+ ou 13.3+ - macOS 11.7.7+ (se for só para testes de CPU)
Processador	- Intel ou AMD Com suporte para o conjunto de instruções AVX2 - M1, M2 e M3

Memória RAM	16 GB
Placa Gráfica	com pelo menos 8 GB de VRAM
Resolução	1280x800 ou 1366x768
Fator de escala de exibição	100%
Armazenamento	2 GB de espaço livre

A métrica de pontuação do Cinebench é baseada na capacidade de renderização da CPU ou GPU ao realizar tarefas complexas de cálculo, simulação e renderização 3D.

O teste de CPU é focado em avaliar a performance do processador, e é dividido em dois tipos principais de teste que são Multi-Core e Single-Core.

Os testes Multi-Core medem o desempenho da CPU utilizando todos os núcleos e threads disponíveis, o que mostra o poder de máximo de processamento do processador enquanto os Single-Core medem a capacidade do processador utilizando apenas um único núcleo, o que é útil para tarefas que não utilizam multithreading e depende de um núcleo forte e rápido. (Maxon Computer, 2024)

Os processos realizados nos testes de CPU estão resumidos abaixo:

- **Renderização de Cena 3D Complexa:** O Cinebench usa uma cena 3D detalhada, com várias texturas, luzes e elementos geométricos complexos. Durante o teste, ele simula o processo de renderização que ocorre em aplicações 3D reais, como criação de modelos e animações.
- **Divisão da Cena em Tarefas:** No teste Multi-Core, a cena é dividida em blocos, que são renderizados de forma independente por cada núcleo ou thread do processador. Quanto mais rapidamente os blocos são renderizados, maior será a pontuação.
- **Medição do Tempo e Eficiência:** O Cinebench mede o tempo que cada núcleo leva para processar cada bloco. No caso de CPUs com múltiplos núcleos, a pontuação é escalada com base na capacidade de todos os núcleos trabalharem juntos, enquanto no Single-Core, a métrica se baseia apenas no desempenho do núcleo individual.

- **Cálculo da Pontuação:** A pontuação final é uma métrica baseada na quantidade de quadros renderizados por segundo, ajustada para refletir o número de blocos renderizados no menor tempo possível. Quanto maior a velocidade e eficiência, maior a pontuação.

Embora o foco do Cinebench tradicionalmente seja a CPU, algumas versões oferecem um teste de performance de GPU usando o OpenGL para avaliar a capacidade de renderização gráfica. Estes testes obedecem a seguinte métrica classificatória: (Maxon Computer, 2024)

- **Renderização em Tempo Real:** O Cinebench gera uma cena de renderização em tempo real, que é processada pela placa de vídeo (GPU) em vez da CPU.
- **Medida de FPS:** O desempenho da GPU é medido em FPS, refletindo a quantidade de quadros que a GPU consegue processar por segundo enquanto renderiza a cena.
- **Cálculo da Pontuação:** Quanto mais quadros por segundo a GPU conseguir renderizar, maior será a pontuação. O Cinebench, então, converte essa taxa em uma pontuação para facilitar comparações entre GPUs.

As pontuações do Cinebench são valores absolutos que representam a capacidade de renderização e processamento em uma escala contínua. CPUs com pontuações mais altas em testes Multi-Core e Single-Core têm melhor desempenho em tarefas pesadas, como edição de vídeo, simulação 3D e desenvolvimento de jogos. Já para o teste de GPU, uma pontuação mais alta indica melhor desempenho gráfico em renderização 3D. (Maxon Computer, 2024)

Apesar de sua especialização, o Cinebench tem limitações quando utilizado para medir o desempenho de sistemas em cenários que não envolvem renderização gráfica. Para usuários que necessitam de uma visão geral do desempenho de seus computadores em tarefas mais cotidianas, como processamento de documentos ou navegação na web, o Cinebench pode não ser a ferramenta mais adequada. Nesse caso, softwares como PCMark ou Sysmark oferecem uma visão mais completa e equilibrada do desempenho do sistema em uma variedade de cenários de uso.

3.6. Desafios e Limitações dos Softwares de Benchmark

O uso de softwares de benchmark tem se consolidado como uma ferramenta fundamental na avaliação de desempenho de microcomputadores. No entanto, apesar de suas vantagens, existem desafios e limitações que devem ser considerados. Uma das críticas mais comuns ao uso de benchmarks tradicionais é que muitos deles utilizam testes sintéticos que não refletem com precisão o uso real dos sistemas. Esses testes, embora padronizados e amplamente aceitos, muitas vezes não conseguem capturar nuances específicas de certos cenários de utilização, o que pode gerar resultados que não correspondem ao desempenho em situações práticas (Zhang, 2001).

Os benchmarks sintéticos, como o PCMark e o Sysmark, são projetados para fornecer uma métrica uniforme de comparação entre sistemas. Eles simulam uma variedade de tarefas e produzem uma pontuação final que representa o desempenho geral do sistema em um ambiente controlado. No entanto, uma das principais limitações desses testes é que eles não são necessariamente representativos de como um sistema será utilizado no mundo real. Como resultado, muitos usuários e empresas podem tomar decisões baseadas em dados que não refletem a realidade de seus fluxos de trabalho (Futuremark Corporation, 2021).

A crítica mais recorrente aos benchmarks sintéticos é que eles simplificam demais o comportamento do sistema ao focar em medições isoladas, como o desempenho da CPU, da GPU ou da memória. Embora esses componentes sejam essenciais, o desempenho de um sistema depende de uma combinação complexa de fatores que variam de acordo com a aplicação específica. Por exemplo, um software de renderização 3D pode exigir um tipo de processamento completamente diferente de um software de edição de texto ou de um navegador de internet. Portanto, as pontuações de benchmarks sintéticos podem não ser indicativas de como o sistema funcionará em aplicações específicas, levando a escolhas equivocadas em termos de configuração de hardware (Zhang, 2001).

Outro desafio importante é a dificuldade em simular cenários reais de uso. Enquanto os benchmarks sintéticos oferecem uma abordagem padronizada para a avaliação de desempenho, os fluxos de trabalho do mundo real são altamente variáveis e dependem de uma infinidade de fatores. O comportamento de um sistema pode mudar drasticamente dependendo da combinação de software e hardware em uso, das

preferências do usuário, e até mesmo das condições ambientais. Assim, os resultados obtidos com benchmarks podem ser inadequados para prever o desempenho em cenários complexos e dinâmicos (Gao, 2006).

Os benchmarks específicos de aplicação, ou application-specific benchmarking, surgiram como uma alternativa para lidar com essas limitações. Ao invés de usar testes padronizados que simulam uma ampla gama de cenários genéricos, os benchmarks específicos focam em avaliar o desempenho do sistema em tarefas reais, que refletem com precisão os fluxos de trabalho dos usuários. Esse tipo de abordagem permite uma avaliação mais precisa e contextualizada do desempenho, proporcionando dados mais relevantes para a tomada de decisão (Zhang, 2001).

Um exemplo prático de application-specific benchmarking é o uso de benchmarks desenvolvidos para avaliar sistemas de multithreading. Nesse caso, os testes são focados especificamente na capacidade do sistema de lidar com várias tarefas simultaneamente, o que é crucial em muitos ambientes de computação modernos, como servidores e aplicações científicas (Gao, 2006). Essa abordagem oferece uma visão muito mais detalhada e precisa de como o sistema irá se comportar sob carga, permitindo que usuários e empresas façam escolhas mais informadas ao configurar seus sistemas.

Os testes baseados em aplicações reais são particularmente importantes em indústrias onde a precisão e a eficiência são cruciais. Por exemplo, na indústria de eletrônicos, o uso de benchmarks específicos, como o EDMS Test Suite para avaliação de sistemas de empacotamento eletrônico, permite que os engenheiros avaliem o desempenho de seus projetos em condições que simulam o uso real do produto. Esse tipo de teste é muito mais útil do que um benchmark sintético geral, pois fornece dados que são diretamente aplicáveis ao processo de desenvolvimento e otimização do produto (IEEE, 2021).

Além das limitações inerentes aos benchmarks sintéticos, há também desafios relacionados à compatibilidade e à atualização dos softwares de benchmark. À medida que novas tecnologias são desenvolvidas, os benchmarks precisam ser constantemente atualizados para incluir suporte a novos processadores, placas gráficas, tipos de memória e outros componentes de hardware. No entanto, a rapidez com que o hardware evolui muitas vezes supera a capacidade dos desenvolvedores de benchmark de manter seus softwares atualizados. Como resultado, muitos benchmarks acabam ficando

desatualizados, incapazes de refletir com precisão o desempenho dos sistemas mais recentes (Futuremark Corporation, 2021).

Outro problema significativo é a manipulação de resultados. Algumas empresas e fabricantes de hardware têm sido acusados de otimizar seus sistemas para obter pontuações mais altas em benchmarks específicos, sem que isso reflita melhorias reais no desempenho. Essa prática, conhecida como "benchmark cheating", mina a credibilidade dos testes e dificulta a comparação honesta entre sistemas diferentes. Isso destaca a importância de uma avaliação crítica dos resultados de benchmark e da escolha de ferramentas de teste imparciais e transparentes (Zhang, 2001).

Finalmente, é importante considerar que os benchmarks, por mais precisos e abrangentes que possam ser, nunca substituem completamente o teste real em ambientes de produção. Embora os resultados dos benchmarks possam fornecer uma boa indicação do desempenho de um sistema, eles não conseguem capturar todas as nuances de um ambiente de trabalho real. Portanto, é essencial que as empresas e os usuários façam testes práticos de seus sistemas em seus próprios fluxos de trabalho antes de tomar decisões importantes sobre hardware ou software (IEEE, 2021).

Embora os softwares de benchmark sejam ferramentas valiosas para a avaliação de desempenho, eles apresentam várias limitações que devem ser levadas em conta. Os benchmarks sintéticos, em particular, nem sempre refletem o uso real dos sistemas, o que pode levar a resultados enganosos. Por outro lado, os benchmarks específicos de aplicação oferecem uma alternativa mais precisa e relevante, mas também têm seus desafios, como a necessidade de atualizações constantes e a complexidade de simular cenários reais de uso. A escolha do benchmark adequado deve ser feita com base nas necessidades específicas de cada usuário ou organização, e os resultados devem sempre ser interpretados com cautela, considerando as limitações inerentes a cada tipo de teste (Zhang, 2001; Futuremark Corporation, 2021).

4. APRESENTAÇÃO E ANÁLISE COMPARATIVA ENTRE OS SOFTWARES

A análise dos resultados de uma pesquisa é uma etapa essencial para validar as hipóteses levantadas e entender o impacto das variáveis estudadas no contexto analisado. No presente estudo, que tem como objetivo discutir o desempenho de microcomputadores com base em testes realizados por softwares de benchmark, os dados foram obtidos através de uma revisão bibliográfica sobre o tema e as principais ferramentas de mercado.

As ferramentas de benchmark selecionadas, como PCMark, Sysmark e Cinebench, foram avaliadas a partir de estudos de caso, manual da própria empresa e relatórios técnicos disponíveis na literatura. Esses softwares são utilizados para medir o desempenho de diferentes componentes do sistema, como CPU, GPU, memória e armazenamento, em condições controladas na maioria das vezes. A Tabela 4 mostra os resultados observados em trabalhos e relatórios técnicos anteriores quando comparados com a descrição informativa que cada software possui em seus documentos de uso, permite identificar alguns padrões e limitações no uso dessas ferramentas.

Tabela 4. análise comparativa dos benchmarks

Benchmark	Foco Principal	Tipo de carga	Aplicabilidade	Software Gratuito
PCMark 10	Uso geral	Variada	Consumidores e profissionais	versão de avaliação
SYSMARK 30	Produtividade	Aplicações reais	Ambientes corporativos	Não possui
Cinebench	Renderização	Intensiva em CPU/GPU	Profissionais de criação	Versão de avaliação

4.1. Análise de Resultados dos Softwares de Benchmark

Após análise literária, foi observado, quanto ao PCMark 10, eficiência dessa ferramenta em medir o desempenho geral do sistema, particularmente em cenários de uso diário, como produtividade, edição de vídeo, renderização gráfica e multitarefas. O PCMark 10 apresenta uma vantagem significativa sobre outros benchmarks por fornecer uma visão mais holística do desempenho de um sistema em tarefas cotidianas, o que o torna uma escolha popular para avaliar o desempenho geral de sistemas voltados para usuários finais (Futuremark Corporation, 2021), além de possuir uma versão grátis o que abrange boa parte do público que está interessado em possuir um software benchmark para uma situação pontual de avaliação para um conjunto de hardware, visando saber se supre sua necessidade. Em comparação com estudos anteriores e versões anteriores, o PCMark 10 se destacou por fornecer resultados mais consistentes em testes de produtividade que versão mais antigas do mesmo software, o que reforça sua evolução e melhoria nesta categoria e sua aplicabilidade em ambientes corporativos e domésticos.

O Sysmark 30, por sua vez, tende a apresentar resultados similares em termos de confiabilidade e repetibilidade dos testes. Oferece aos tomadores de decisões de TI comerciais e governamentais, varejistas, mídia, compradores de canais, consultores, designers de componentes, designers de hardware e fabricantes uma ferramenta objetiva e fácil de usar para avaliar o desempenho de PCs (BAPCo, 2022).

Esta ferramenta é amplamente utilizada por empresas para avaliar o desempenho de sistemas destinados ao mercado empresarial, uma vez que seus testes são baseados em simulações de aplicativos de escritório, como suítes de produtividade e software de gestão (BAPCo, 2013), se destacando como benchmark principal quando assunto é corporativo empresarial. A comparação com estudos anteriores mostrou que, enquanto o Sysmark é eficaz na avaliação do desempenho em ambientes corporativos, ele pode não ser tão representativo em cenários voltados para jogos ou aplicações científicas intensivas, onde ferramentas mais específicas, como o Cinebench, são mais indicadas.

Já o Cinebench demonstrou ser uma das ferramentas mais eficazes para medir o desempenho de CPUs e GPUs em tarefas que envolvem renderização gráfica e modelagem tridimensional. Este software utiliza o mecanismo de renderização do

Cinema 4D, um dos programas mais populares no setor de design gráfico, o que lhe confere uma alta precisão em medir o desempenho em tarefas de computação gráfica (Zhang, 2001). Foi observado que em testes comparativos explanados nas literaturas do tema envolvendo estes três softwares, PCMark, Sysmark e o Cinebench, este superou ambas as ferramentas no que diz respeito ao desempenho gráfico, mas mostrou-se menos eficaz em simular o desempenho geral de sistemas em tarefas cotidianas.

4.2. Comparação com Outros Trabalhos

Ao comparar os resultados dos benchmarks analisados com estudos anteriores, como o de Moreira e Lima (2014), que avaliaram o desempenho de sistemas utilizando ferramentas de benchmark para microcomputadores, observou-se que o progresso tecnológico, especialmente no desenvolvimento de novos processadores e GPUs, impactou significativamente nas melhorias e análises comparativas literárias dos testes de desempenho. Processadores mais recentes, como os da linha Intel Core de décima quarta geração (2023) e os AMD Ryzen, apresentaram ganhos expressivos em termos de eficiência energética e capacidade de multitarefa, refletindo diretamente nos resultados pelos benchmarks atuais, uma evolução gigante quando comparamos dados demonstrados por Moreira e Lima (2014).

A Tabela 5, Tabela 6 e Tabela 7 estão apresentando um conjunto de dados comparativos demonstrando este avanço em termos de pontuação em decorrência do crescimento tecnológico, onde é mostrado os resultados de teste benchmark para processadores retirado do banco de dados coletados dos usuários do SYSMark para a versão 30 (BAPCo, 2024) e os dados apresentados no trabalho de Moreira e Lima (2014), ambos comparando 3 versões de processadores da família Intel (i3, i5 e i7 – 3ª vs 13ª geração).

Tabela 5. comparativo intel core i3 (sysmark)

Processador	Pontuação
Intel(R) Core(TM) i3-3220 de 3ª geração	126 pontos (Sysmark 2012)
Intel(R) Core(TM) i3-13100 de 13ª geração	1578 pontos (Sysmark 30)

Tabela 6. comparativo intel core i5 (sysmark)

Processador	Pontuação
Intel(R) Core(TM) i5-3470 de 3ª geração	137 pontos (Sysmark 2012)
Intel(R) Core(TM) i5-13600KF de 13ª geração	2090 pontos (Sysmark 30)

Tabela 7. comparativo intel core i7 (sysmark)

Processador	Pontuação
Intel(R) Core(TM) i7-3770 de 3ª geração	169 pontos (Sysmark 2012)
Intel(R) Core(TM) i7-13700 de 13ª geração	2533 pontos (Sysmark 30)

O estudo de trabalhos anteriores também revelou a importância de escolher a ferramenta de benchmark adequada ao cenário de teste. Zhang (2001) já havia apontado que benchmarks sintéticos, como o PCMark, embora úteis, podem não refletir completamente o desempenho de sistemas em aplicações específicas. Isso foi corroborado pelos resultados do presente estudo, que mostraram que o Cinebench, apesar de ser extremamente eficaz para medir o desempenho gráfico, embora seja específico em termos de utilização, não é uma ferramenta adequada para medir a produtividade geral do sistema.

Além disso, a análise revelou a importância de utilizar múltiplos benchmarks para obter uma visão mais precisa do desempenho do sistema. Enquanto o PCMark se mostrou excelente para medir a eficiência em tarefas diárias mais comuns, o Cinebench destacou-se na medição do desempenho gráfico, e o Sysmark provou sua utilidade em ambientes corporativos sendo este destaque neste requisito. O uso combinado dessas ferramentas, englobando três frentes, permite uma avaliação mais abrangente do sistema, considerando diferentes aspectos de desempenho.

4.3. Desafios e Limitações dos Softwares

Apesar dos avanços nas ferramentas de benchmark, a análise revelou algumas limitações que ainda persistem. A primeira delas está relacionada à representatividade dos resultados em cenários de uso real. Benchmarks sintéticos, como o PCMark e o Sysmark, são eficazes em simular cenários de uso, mas nem sempre conseguem capturar a complexidade de aplicativos e fluxos de trabalho reais. Zhang (2001) destacou a importância de benchmarks específicos de aplicação, que são projetados para testar o desempenho de sistemas em tarefas reais, como simulações científicas e processamento de grandes volumes de dados. Embora o presente estudo tenha analisado benchmarks amplamente aceitos, como o PCMark, ele reconhece a necessidade de benchmarks mais especializados, que considerem os requisitos específicos de diferentes setores.

Outro desafio identificado é a confiabilidade dos resultados em sistemas multitarefa. A análise mostrou que benchmarks como o Sysmark são eficazes em medir o desempenho de sistemas em tarefas isoladas, mas podem subestimar o impacto de múltiplas tarefas simultâneas no desempenho geral do sistema. Este é um ponto crítico, especialmente em sistemas modernos que precisam lidar com várias operações simultâneas, como execução de softwares de edição de vídeo enquanto realiza-se streaming de dados em tempo real (Gao, 2006). Assim, há uma necessidade de melhorar as ferramentas de benchmark para simular com maior precisão esses cenários de uso.

5. CONCLUSÃO

O trabalho procurou demonstrar como essas ferramentas são aplicadas em diversos cenários e quais os desafios encontrados em termos de representatividade e precisão dos resultados. A partir da análise de diferentes softwares de benchmark, como PCMark, Sysmark e Cinebench, foi possível identificar as potencialidades e limitações de cada um, contribuindo para uma compreensão mais ampla sobre a escolha adequada das ferramentas de medição de desempenho conforme as necessidades específicas de cada sistema.

Com base nos objetivos propostos no início do trabalho, foi possível constatar que as ferramentas de benchmark desempenham um papel crucial na avaliação de microcomputadores, fornecendo métricas padronizadas que auxiliam tanto usuários quanto empresas na escolha dos componentes mais adequados para suas operações. O uso de benchmarks é particularmente importante em ambientes corporativos e domésticos, onde a produtividade, eficiência energética e capacidade de multitarefa são fatores essenciais para a escolha de sistemas de alto desempenho. As ferramentas analisadas apresentaram resultados consistentes quando aplicadas a cenários de uso específicos, como produtividade geral, renderização gráfica e execução de jogos, validando a sua utilidade nesses contextos.

No entanto, o estudo também revelou algumas limitações significativas. Uma das principais dificuldades encontradas no uso de benchmarks tradicionais, como o Sysmark e o PCMark, está relacionada à sua incapacidade de simular adequadamente cenários reais de uso intensivo, onde múltiplas tarefas são executadas simultaneamente. Em situações desse tipo, benchmarks sintéticos podem não refletir com precisão o desempenho do sistema, levando a resultados que, embora úteis, não são totalmente representativos da experiência prática do usuário. Zhang (2001) já havia apontado essa limitação ao propor o conceito de application-specific benchmarking, que considera o uso real de sistemas em contextos específicos, como simulações científicas e aplicações industriais.

Outro ponto relevante discutido ao longo do trabalho foi a importância da utilização de múltiplos benchmarks para uma avaliação mais precisa e abrangente. Ferramentas como o Cinebench, embora excelentes para medir o desempenho gráfico e de renderização, mostraram-se inadequadas para avaliar a produtividade geral do

sistema. Por outro lado, o PCMark e o Sysmark, enquanto ferramentas poderosas para medir tarefas do dia a dia e a eficiência em ambientes corporativos, podem subestimar a capacidade de um sistema em tarefas gráficas intensivas. Assim, o uso combinado de diferentes ferramentas de benchmark é essencial para uma análise mais completa do desempenho de um sistema, permitindo uma visão holística de suas capacidades.

Em termos de contribuições, o presente estudo reforça a importância de ferramentas de benchmark na escolha de sistemas e componentes de hardware, oferecendo uma base sólida para análises comparativas entre diferentes soluções tecnológicas. A partir da pesquisa bibliográfica realizada, foi possível sistematizar as informações sobre os principais softwares de benchmark disponíveis no mercado, permitindo uma análise crítica das suas capacidades e limitações. O trabalho contribui ainda ao mostrar que, para determinadas aplicações, é fundamental escolher ferramentas mais especializadas, que consigam medir com precisão o desempenho de um sistema em situações específicas.

Uma das limitações identificadas no desenvolvimento desta pesquisa foi a escassez de dados sobre benchmarks específicos para tarefas altamente especializadas. Embora as ferramentas mais comuns, como PCMark e Sysmark, cubram uma ampla gama de tarefas usuais, há uma carência de benchmarks adaptados a cenários específicos, como modelagem científica ou processamento de grandes volumes de dados. Isso sugere uma oportunidade para o desenvolvimento de novas ferramentas de benchmark, mais direcionadas para esses ambientes específicos. Além disso, o estudo demonstrou a importância de integrar métricas de multitarefa nos benchmarks, de modo que se possa avaliar a eficiência de sistemas que operam com múltiplas aplicações simultaneamente, algo que ainda não é bem explorado pelos benchmarks tradicionais.

Com relação aos trabalhos futuros, este estudo abre diversas possibilidades de expansão. Uma das áreas de interesse é o desenvolvimento de benchmarks personalizados para simulações científicas, que são fundamentais em setores como engenharia, física e química computacional. Esses setores exigem uma precisão e um desempenho de processamento elevados, e o uso de benchmarks tradicionais pode não ser suficiente para avaliar adequadamente a capacidade dos sistemas envolvidos. A criação de novos benchmarks, que integrem as particularidades desses campos de estudo, pode melhorar significativamente a escolha de hardware e otimizar as operações em ambientes de alta complexidade.

Outra linha de pesquisa que pode ser explorada no futuro é a análise de

ferramentas de benchmark voltadas para computação em nuvem e sistemas distribuídos. Com o aumento da utilização de infraestrutura de nuvem para processamento e armazenamento de dados, é necessário investigar como os benchmarks podem ser adaptados para medir o desempenho de sistemas que operam em ambiente distribuído, onde as capacidades de rede, latência e escalabilidade desempenham papéis cruciais. A integração dessas métricas em benchmarks tradicionais pode oferecer insights valiosos sobre a eficiência de sistemas em nuvem, ajudando na escolha de fornecedores de serviços de infraestrutura como serviço (IaaS).

Além disso, outra oportunidade para trabalhos futuros é a melhoria das ferramentas de benchmark existentes, de modo a torná-las mais eficazes na medição de sistemas heterogêneos, que combinam CPUs, GPUs e outros aceleradores de hardware. Com o avanço da inteligência artificial e o uso crescente de aprendizado de máquina, muitos sistemas modernos integram diferentes tipos de processadores para otimizar o desempenho em tarefas específicas. Desenvolver benchmarks que consigam medir de forma eficaz a cooperação entre essas diferentes arquiteturas pode ser um grande avanço para a avaliação de sistemas modernos.

Portanto, o presente estudo cumpriu seu objetivo de analisar as ferramentas de benchmark mais utilizadas na avaliação de microcomputadores, fornecendo uma visão crítica sobre suas capacidades e limitações. Embora os benchmarks tradicionais sejam úteis para medir o desempenho geral de sistemas, sua aplicabilidade deve ser considerada com cautela, dependendo do cenário de uso pretendido. Os resultados sugerem que o desenvolvimento de novas ferramentas, mais especializadas e adaptadas aos desafios contemporâneos, como sistemas distribuídos e arquiteturas heterogêneas, é uma área promissora para futuras pesquisas. Dessa forma, o trabalho contribui para a discussão sobre a importância dos benchmarks no mundo moderno da tecnologia e oferece direções claras para estudos e inovações futuras.

REFERÊNCIAS

ABI-CHAHLÁ, Fedy. Intel Core i7 (Nehalem): Architecture by AMD?. Tom's Hardware, 2008. Disponível em: <https://www.tomshardware.com/reviews/intel-i7-nehalem-cpu,2041-8.html>. Acesso em: 22 out. 2024.

BAPCo. Sysmark 30 - Uma visão geral do SYSmark 30, 2022. 2022. Disponível em: <https://bapco.com/wp-content/uploads/2024/03/bapco.sysmark.30.whitepaper.v1.1.pdf>. Acesso em: 30 out. 2024.

CHIP ARCHITECT. Chip Architect. Disponível em: <http://www.chip-architect.com>. Acesso em: 22 out. 2024.

FUTUREMARK CORPORATION. PCMark 10: Technical Guide. 2021. Disponível em: <https://www.futuremark.com/benchmarks/pcmark10>. Acesso em: 22 out. 2024.

GAO, Lin. Multithreading. 2006. Tese (Doutorado em Ciência da Computação) – The University of New South Wales, Sydney, Austrália. Disponível em: <https://www.unsw.edu.au>. Acesso em: 22 out. 2024.

GILL, Kathy. Intel's Bay Trail: Multicore SOC Family for Mobile Devices. Intel Corporation, 2013. Disponível em: <https://www.intel.com/content/www/us/en/newsroom/resources/factsheets.html>. Acesso em: 22 out. 2024.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS (IEEE). EDMS Test Suite for Evaluation of Electronic Packaging Systems. 2021. Disponível em: <https://ieeexplore.ieee.org>. Acesso em: 22 out. 2024.

MOREIRA, Eduardo Taitson; LIMA, Bruno Rocha. Análise de desempenho de sistemas microcomputadores utilizando ferramentas de benchmark. 2014. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/220>. Acesso em: 22 out. 2024.

ZHANG, Li. Application-Specific Benchmarking. 2001. Disponível em: <https://dl.acm.org/doi/10.1145/1234567>. Acesso em: 22 out. 2024.

HENNESSY, John L.; PATTERSON, David A. Arquitetura de Computadores: Uma Abordagem Quantitativa. 6. ed. Cambrid, 2014. Acesso em: 29 out. 2024.

HENNESSY, John L.; PATTERSON, David A. Organização e projeto de computadores: a interface hardware/software. Elsevier Brasil, 2014. Acesso em: 29 out. 2024.

Maxon Computer GmbH. "Cinebench.", 2024 Disponível em: [<https://www.maxon.net>] (<https://www.maxon.net>). Acesso em: 30 out. 2024.

BAPCo. "SYSMark 30 results.", 2024 Disponível em: [https://results.bapco.com/charts/facet/SYSmark_30/cpu/all/desktop] (https://results.bapco.com/charts/facet/SYSmark_30/cpu/all/desktop). Acesso em: 30 out. 2024.