

| Rank | Citations | Year | Title (★ means it won the ISCA Influential Paper Award) | First Author + HOF Authors | Type | Topic |
|------|-----------|------|---|--|-------|-----------------------|
| 1 | 5351 | 1995 | The SPLASH-2 programs: Characterization and methodological considerations | <u>Stephen Woo</u> , <i>Anoop Gupta</i> | Tool | Benchmark |
| 2 | 4214 | 2017 | In-datacenter performance analysis of a Tensor Processing Unit | <i>Norm Jouppi, David Patterson</i> | Arch | Machine Learning |
| 3 | 3834 | 2000 | ★ Wattch: A framework for architectural-level power analysis and optimizations | <i>David Brooks, Margaret Martonosi</i> | Tool | Power |
| 4 | 3386 | 1993 | ★ Transactional memory: Architectural support for lock-free data structures | <u>Maurice Herlihy</u> | Micro | Parallelism |
| 5 | 2690 | 2016 | EIE: Efficient inference engine on compressed deep neural network | <u>Song Han</u> , <i>Bill Dally, Mark Horowitz</i> | Arch | Machine Learning |
| 6 | 2620 | 2007 | ★ Power provisioning for a warehouse-sized computer | <u>Xiaobo Fan</u> , <i>Luiz Barroso</i> | Micro | Power |
| 7 | 2507 | 1992 | Active messages: a mechanism for integrated communication and computation | <u>Thorsten von Eiken</u> | Micro | Parallelism |
| 8 | 2391 | 2011 | Dark silicon and the end of multicore scaling | <i>Hadi Esmaeilzadeh, Doug Burger, Karthikeyan Sankaralingam</i> | Micro | Parallelism |
| 9 | 2352 | 1995 | ★ Simultaneous multithreading: Maximizing on-chip parallelism | <i>Dean Tullsen, Susan Eggers, Hank Levy</i> | Micro | Parallelism |
| 10 | 2243 | 1990 | ★ Improving direct-mapped cache performance by the addition of a small fully-associative cache and prefetch buffers | <i>Norm Jouppi</i> | Micro | Cache |
| 11 | 1801 | 2009 | Architecting phase change memory as a scalable DRAM Alternative | <u>Benjamin Lee</u> , <i>Doug Burger, Engin Ipek, Onur Mutlu</i> | Micro | NV RAM |
| 12 | 1790 | 1990 | Memory consistency and event ordering in scalable shared-memory multiprocessors | <i>Kourosh Gharachorloo, Anoop Gupta, John Hennessy</i> | Micro | Consistency/Coherence |
| 13 | 1769 | 2009 | Scalable high performance main memory system using phase-change memory technology | <i>Moinuddin Qureshi</i> | Micro | NV RAM |
| 14 | 1659 | 2016 | ISAAC: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars | <u>Ali Shafiee</u> , <i>Rajeev Balasubramonian, Naveen Muralimanohar</i> | Arch | Machine Learning |
| 15 | 1643 | 2003 | ★ Temperature-aware microarchitecture | <u>Kevin Skandron</u> | Micro | Power |
| 16 | 1557 | 2016 | Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks | <u>Yu-Hsin Chen</u> , <i>Joel Emer</i> | Micro | Machine Learning |
| 17 | 1420 | 2016 | Prime: A novel processing-in-memory architecture for neural network computation in ReRAM-based main memory | <u>Ping Chi</u> , <i>Yuan Xie</i> | Arch | Machine Learning |
| 18 | 1401 | 2014 | A reconfigurable fabric for accelerating large-scale datacenter services | <u>Andrew Putnam</u> , <i>Hadi Esmaeilzadeh</i> | Micro | Interconnect |
| 19 | 1374 | 1992 | The turn model for adaptive routing | <u>Christopher Glass</u> | Micro | Interconnect |
| 20 | 1350 | 1995 | Multiscalar processors | <i>Guri Sohi, T. N. Vijaykumar</i> | Micro | Parallelism |
| 21 | 1302 | 2000 | Memory access scheduling | <u>Andrew Putnam</u> , <i>Bill Dally</i> | Micro | Parallelism |
| 22 | 1284 | 1997 | ★ Complexity-effective superscalar processors | <u>Subbarao Palacharla</u> , <i>Norm Jouppi, Jim Smith</i> | Micro | Parallelism |
| 23 | 1221 | 2002 | ★ Drowsy caches: simple techniques for reducing leakage power | <u>Krisztián Flautner</u> , <i>Nam Sung Kim, Trevor Mudge</i> | Micro | Power |
| 24 | 1210 | 1996 | ★ Exploiting choice: Instruction fetch and issue on an implementable simultaneous multithreading processor | <i>Hank Levy, Susan Eggers, Joel Emer, Dean Tullsen</i> | Micro | Parallelism |
| 25 | 1201 | 1997 | A Study of Branch Prediction Strategies | <i>Jim Smith</i> | Micro | Parallelism |