On the I/O Complexity of Stencil Computations on 2 Dimensional Grids

Philipp Hupp (hupp@in.tum.de), Riko Jacob (jacob@in.tum.de)

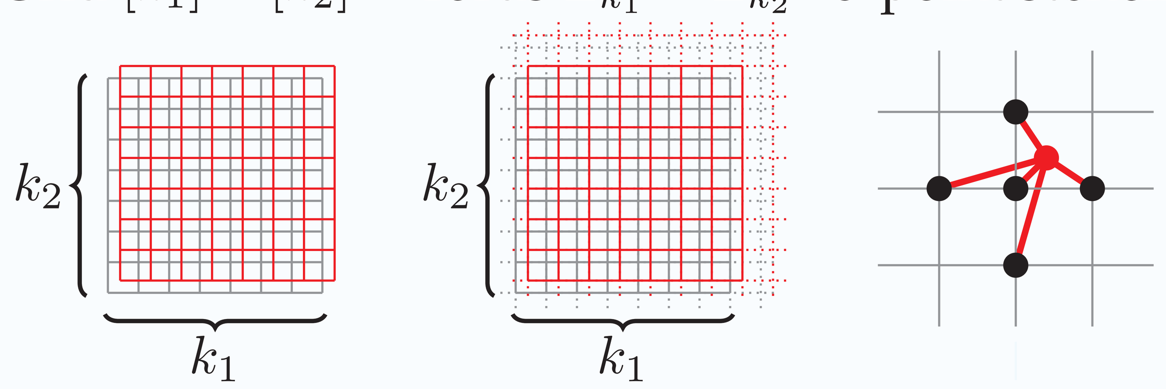
Technische Universität München

Background, Task and Results

Stencil computations are important kernels in **scientific computing**. We consider the memory access requirements of such tasks, in particular the first two terms of the **I/O complexity** for evaluating the 5-point stencil.

Computation Graphs

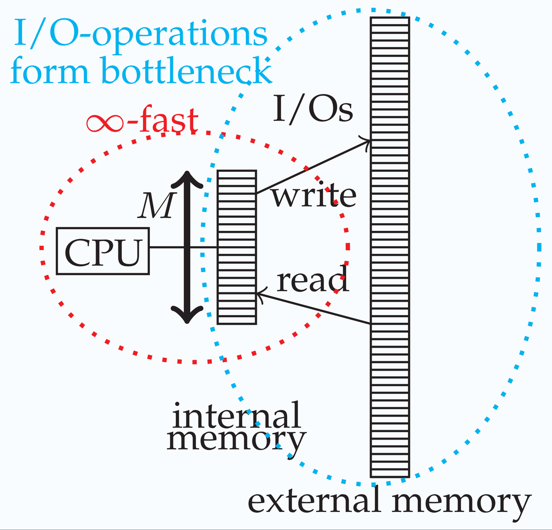
Grid $[k_1] \times [k_2]$ Torus $\mathbb{Z}_{k_1} \times \mathbb{Z}_{k_2}$ 5-point stencil



- input layer
- output layer

Compulsory I/Os, I/O Model [1, 3]

The I/O operations needed to initially read the input and write the output are called **compulsory** I/O operations. All other I/Os are called **non-compulsory**. For the grid there are $2k_1k_2$ compulsory I/Os.



Results

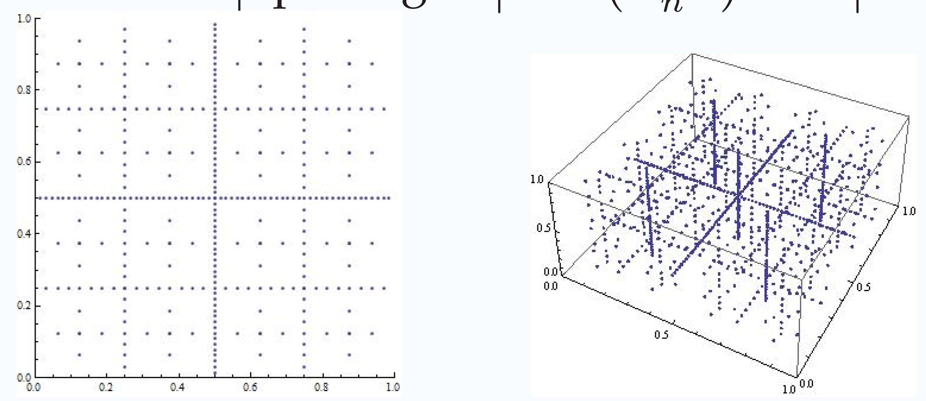
Assuming:
- **Internal memory size:** M
- **Grid:** $[k_1] \times [k_2]$
- **Grid size:** $k_1 \geq k_2$ and $\frac{k_2}{M} \to \infty$

I/O complexity of the 5-point stencil $C(k_1,\ k_2)$:

$$C(k_1,\ k_2) = 2k_1k_2 + 4\frac{k_1k_2}{M-4} + \begin{cases} +\mathcal{O}(k_1) \\ -\mathcal{O}(\frac{k_1k_2}{M^2}) - \mathcal{O}(k_1) \end{cases}$$

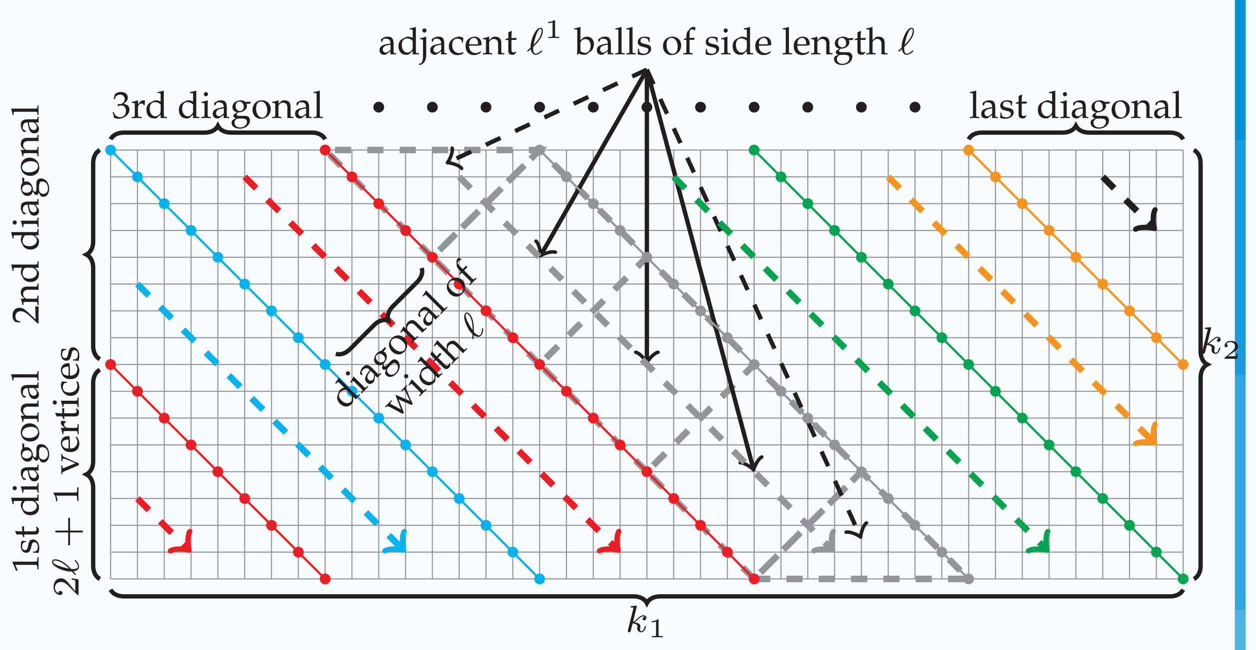
Outlook

Deduce tight I/O bounds for **sparse grids** in **high dimensions.** $|\text{Sparse grid}| = \mathcal{O}(\frac{\log n}{n})^{d-1} \cdot |\text{Full grid}|$

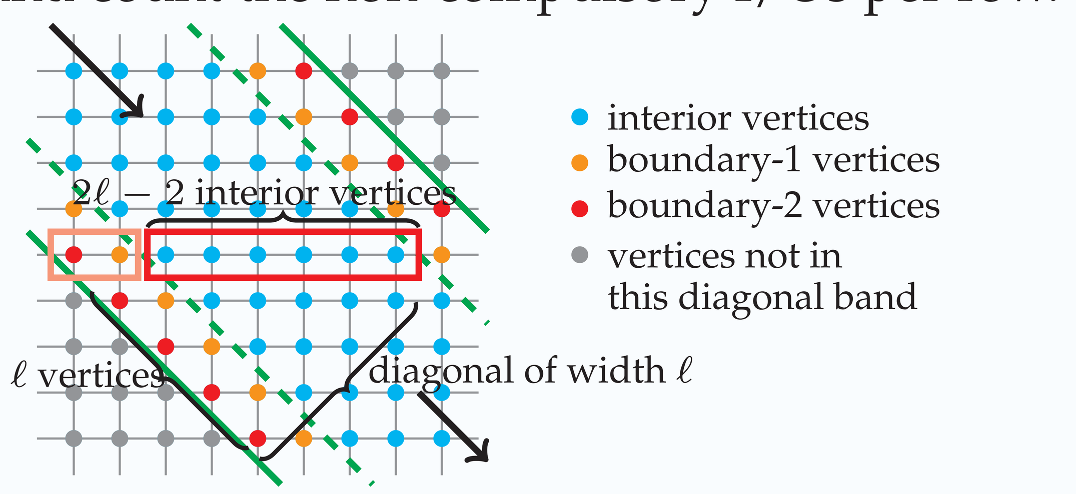


Algorithm

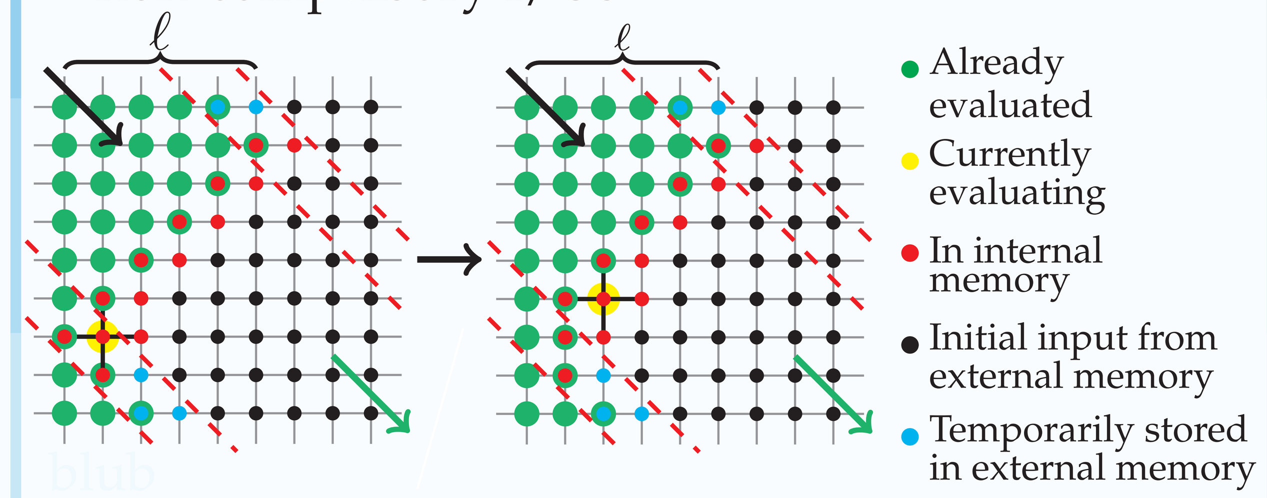
- Work in adjacent (overlapping) ℓ^1 balls ...



- ... and count the non-compulsory I/Os per row.



- Diagonal of width ℓ: 2 boundary vertices are followed by $2\ell - 2$ interior vertices.
- For $\ell = \frac{M}{2} - 2$ we can sweep a diagonal without non-compulsory I/Os.



- Every boundary vertex causes 2 non-compulsory I/Os.

$$NC(k_1,\ k_2) \leq k_2 \cdot 2 \cdot 2 \cdot \left\lceil \frac{k_1}{M-4} \right\rceil \leq 4\frac{k_1k_2}{M-4} + 4k_2$$

The Lower Bound

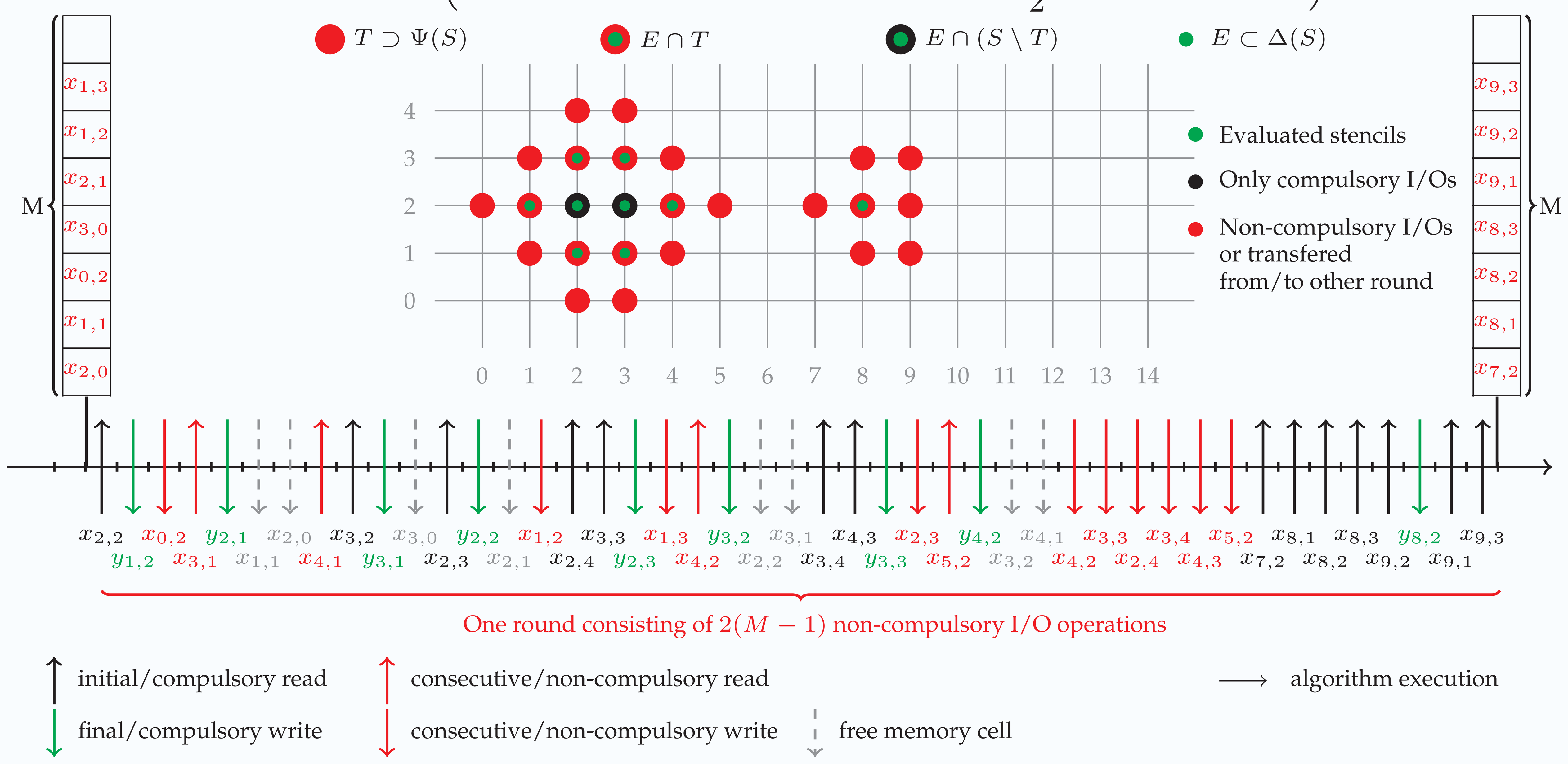
Partition Arbitrary Algorithm into Rounds & Bound Work per Round

Split arbitrary algorithm into rounds of $2(M-1)$ non-compulsory I/Os. In addition $2(M-1)$ vertices are transfered to or from that round in the internal memory. Fix one round:

- S: Vertices in internal memory at some point.
- $T \subset S$: Vertices also available in other rounds.
- $E \subset S$: Stencils computed in the current round.
- $|T| \leq 4(M-1)$ vertices available in other rounds
- $S \setminus T$ must have pathwidth $\leq M-1$
- $\Psi(S) \subset T$ and $E \subset \Delta(S)$ ⟷

Because of limited pathwidth [4] the torus behaves like an infinitely large grid:

$$\Longrightarrow \quad \left(\quad |T| \leq 4(M-1) \quad \Rightarrow \quad |E| \leq \frac{1}{2}(M-1)^2 + 1 \quad \right)$$



An Isoperimetric Inequality [2]

- Fractional system f on $\mathbb{Z}_k^n$: $f : \mathbb{Z}_k^n \to [0,\ 1]$
- Weight $w(f)$ of a system f: $w(f) = \sum_{x \in \mathbb{Z}_k^n} f(x)$
- Closure ∂f of a system f:

$$\partial f(x) = \begin{cases} 1, & \text{if } f(x) > 0 \\ \max\{f(y) : d(x,y) = 1\}, & \text{if } f(x) = 0 \end{cases}$$

- Fractional ball b^v of weight v: $\exists (r, \alpha) \in \mathbb{N} \times [0,\ 1]$ s.t.:

$$w(b^v) = v \quad \text{and} \quad b^v(x) = \begin{cases} 1, & \text{if } d(x,0) < r \\ \alpha, & \text{if } d(x,0) = r \\ 0, & \text{if } d(x,0) > r \end{cases}$$

Theorem 1 (Isoperimetric inequality - discrete torus [2]). *For $k \geq 2$ even, f a system on $\mathbb{Z}_k^n$:*

$$w(\partial f) \geq w(\partial b^{w(f)})$$

- Inner core:

$$\Delta f(x) = \begin{cases} 0, & \text{if } f(x) < 1 \\ \min\{f(y) : d(x,y) = 1\}, & \text{if } f(x) = 1 \end{cases}$$

- Inner 2-core: $\odot f(x) = \Delta(\Delta f)(x)$
- Inner 2-boundary $\Psi f(x) = (f - \odot f)(x)$

Theorem 1 translates into

$$w(\Delta(f)) \leq w(\Delta(b^{w(f)}))$$
$$w(\Psi(f)) \geq w(\Psi(b^{w(f)}))$$

Deducing the Lower Bound

Lower bound on torus:

$$NC(k_1, k_2) \geq M - 1 + \left(\left\lfloor \frac{k_1k_2}{\frac{1}{2}(M-1)^2+1} \right\rfloor \right) 2(M-1)$$

By reduction - **Lower bound** on the **grid**:

$$NC(k_1, k_2) \geq 4\frac{k_1k_2}{M-4} - \mathcal{O}\left(\frac{k_1k_2}{M^2}\right) - \mathcal{O}(k_1)$$

References

[1] Alok Aggarwal and Jeffrey S. Vitter. The input/output complexity of sorting and related problems. *Communications of the ACM*, 31(9):1116–1127, 1988.

[2] Béla Bollobás and Imre Leader. An isoperimetric inequality on the discrete torus. *SIAM J. Discret. Math.*, 3:32–37, January 1990.

[3] Hong, Jia-Wei and H. T. Kung. I/O complexity: The red-blue pebble game. In *Proceedings of STOC '81*, pages 326–333, New York, NY, USA, 1981. ACM.

[4] P. D. Seymour and Robin Thomas. Graph searching and a min-max theorem for tree-width. *J. Comb. Theory Ser. B*, 58:22–33, May 1993.