Implementations of a model of physical sorting

N. MURPHY[1]*, T. J. NAUGHTON[1], D. WOODS[2], B. HENLEY[3],
K. MCDERMOTT[3], E. DUFFY[4], P. J. M. VAN DER BURGT[4], N. WOODS[4]

[1] *Dept. of Computer Science, National University of Ireland, Maynooth, Ireland*
[2] *Boole Centre for Research in Informatics, School of Mathematics, University College Cork, Ireland*
[3] *Dept. of Anatomy, BioSciences Institute, University College Cork, Ireland*
[4] *Dept. Experimental Physics, National University of Ireland, Maynooth, Ireland*

Received 1 December 2006; In final form 1 April 200x

We define a model of physical devices that have a parallel atomic operation that transforms an unordered list input such that the sorted output can be sequentially read off in linear time. We show that commonly used biology, chemistry, and physics laboratory techniques are model instances and provide implementations.

Key words: sorting, natural computation, chromatography, gel electrophoresis, mass spectrometry, optics, physical implementation.

1 INTRODUCTION

In recent years, there has been interest in identifying, analysing, and utilising computations performed in nature [1, 6, 7, 14, 17, 21, 23], in particular where they appear to offer interesting resource trade-offs when compared with the best-known sequential (e.g. Turing machine) equivalents. One such computation is sorting, for which several natural sorting algorithms have been proposed [2, 9, 19]. Independently, scientists routinely separate millions of particles based on their physical characteristics. For example, gel electrophoresis [20], chromatography [18], and mass spectrometry [12]. The common idea behind these techniques is that some physical force affects objects by

* email: {nmurphy,tomn}@cs.nuim.ie, d.woods@bcri.ucc.ie

an amount that is proportional to some physical property of the objects. In this paper (the full manuscript was published elsewhere [16]), we present a Model of Physical Sorting that computes a stable sorting [13] of its input list of natural numbers, and describe five implementations. Further, we physically implement four of the instances, define a Restricted Model for unstable sorting, and illustrate how to introduce stability into an existing unstable sort.

2 MODEL (AND RESTRICTED MODEL) OF PHYSICAL SORTING

The Model computes the *stable sorting* [13] of its input list of natural numbers and the Restricted Model computes an unstable sorting. A sort is stable if and only if sorted elements with the same value retain their original order. Not all sorting algorithms are stable; any sorting algorithm that does not preserve the original relative ordering of equal values in the input is not stable. Further, a sorting algorithm that relies on each element of its inputs being distinct, or that only outputs list elements rather than indices, is not regarded as stable. In the definitions below we let $\mathbb{N} = \{1, 2, 3, \ldots\}$.

2.1 The Model

Before formally describing the computation of the Model we give an informal description. The input list is transformed to a 2D matrix that has a number of rows equal to the input list length and a number of columns linear in the maximum allowable input value. The matrix is zero everywhere except where it is populated by the elements of the input list, whose row position in the matrix is their index in the input and whose column position is linearly proportional to their value. The values in the matrix are then read sequentially, column by column, and the row index of each nonzero value is appended to an output list. This output list of indices is a stable sorting of the input list.

Definition 1 *A Model of Physical Sorting is a tuple $S = (m, a, b) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}$, where m is an upper bound the values to be sorted, and a, b are scaling constants.*

The Model acts on a list $L = (l_1, l_2, \ldots, l_n)$ where $l_i \in \{1, 2, \ldots, m\}$ and m is some constant independent of n. Given L, and a Model of Physical Sorting S, we define the $n \times (am + b)$ matrix G (example in Figure 1) with elements

$$G_{i,j} = \begin{cases} l_i & \text{if } j = al_i + b \\ 0 & \text{otherwise}\,. \end{cases} \tag{1}$$

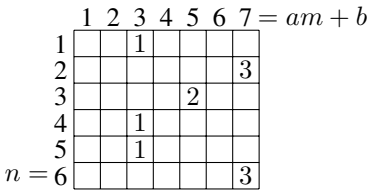

FIGURE 1
Graphical illustration of the matrix G for example model $S = (m, a, b) = (3, 2, 1)$ and for example input list $L = (1, 3, 2, 1, 1, 3)$.

Definition 2 *A Physical Sorting computation is a function* $c : \{1, 2, \ldots, m\}^n \rightarrow \{1, 2, \ldots, n\}^n$ *that maps a list* L *of values to a sorted list of indices* $c(l_1, l_2, \ldots, l_n) = (k_1, k_2, \ldots, k_n)$*, where* l_{k_p} *is the* p^{th} *non-zero element of* G *and where the elements of* G *are assumed to be ordered by column then by row.*

A Physical Sorting computation outputs a stable sorting of L: k_1 is the index of the first value in the stable sorting of L, k_2 is the index of the second, and so on. We assume the computation takes at most $(am + b)n + 1 = O(n)$ timesteps. Creation of G takes one parallel timestep and obtaining the indices of the nonzero values in G takes one sequential timestep per element of G.

2.2 Restricted Model

It is possible to restrict some physical instances of the Model in Section 3 to become instances of the Restricted Model. This restriction is achieved by removing the abilities to track indices and deal with repeated elements.

Definition 3 *A Restricted Model of Physical Sorting is a tuple* $S = (m, a, b) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}$*, where* m *is an upper bound on the values to be sorted, and* a, b *are scaling constants.*

The Restricted Model acts on multiset $T = \{t_1, t_2, \ldots, t_n\}$ where $t_i \in \{1, 2, \ldots, m\} \subset \mathbb{N}$ and m is independent of n. Given such a T and a Restricted Model of Physical Sorting S, we define the vector V of length $am+b$. Here V has elements

$$V_j = \begin{cases} t_i & \text{if } j = at_i + b \\ 0 & \text{otherwise .} \end{cases} \tag{2}$$

Definition 4 *A Restricted Physical Sorting computation is a function c that maps a multiset $T = \{t_1, t_2, \ldots, t_n\}$ where $t_i \in \{1, 2, \ldots, m\} \subset \mathbb{N}$ to a list $c(T) = (t_{k_1}, t_{k_2}, \ldots, t_{k_{n'}})$, where t_{k_p} is the p^{th} non-zero element of V and $n' \leq n$.*

It can be seen that $c(T)$ is a list of strictly increasing values, such that $t_{k_i} < t_{k_{i+1}}$ for all $i \in \{1, 2, \ldots, n-1\}$. The input to a Restricted Model of Physical Sorting is a multiset, however the output vector does not contain any duplicated elements (and since no index information is available it is not a stable sort). We assume that a Restricted Physical Sorting computation is computed in at most $(am+b)+1 = O(1)$ timesteps. The creation of vector V takes one timestep, and obtaining the sorted list takes one timestep per element of V.

3 PHYSICAL INSTANCES OF THE MODEL

In this section we give five example instances of the (Restricted) Model that arise in commonly used scientific laboratory techniques of gel electrophoresis, chromatography, the dispersion of light, optical tweezers, and mass spectrometry. Some details are omitted and are to be found elsewhere [16].

3.1 Gel Sort

Gel electrophoresis [20] is a tool of molecular biologists and is a standard technique for separating large molecules (e.g. DNA and RNA) by length. It utilises the differential movement of molecules of different sizes in a gel of a given density. Samples of DNA molecules are placed (separated from each other) into wells along one end of a rectangle of agarose gel. Electrodes apply a voltage in the orthogonal direction across the gel which provides a force upon the charged molecules causing them to be pulled towards the opposite end of the gel. Smaller molecules move through the gel more quickly and easily than larger molecules. This difference in velocity orders the molecular samples by number of base pairs. We refer to implementations of the model in Definition 1 and Definition 3 using gel electrophoresis as Gel Sort and Restricted Gel Sort, respectively.

Viney and Fenton [22] provide a linear equation that describes the physics of gel electrophoresis and satisfies Equations (1) and (2). Given a list L to be sorted, we encode each element as a sample of identical molecules each with a number of base pairs proportional to the element value. Each sample is placed (in the same order as in L) in the wells at one end of the gel. A voltage is applied for a time and the molecules move through the gel at a rate

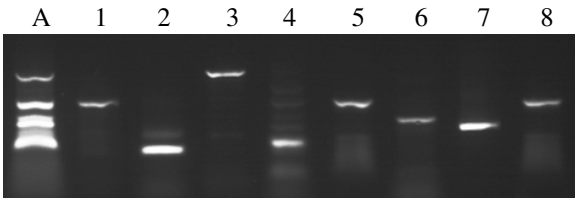

FIGURE 2
Implementation of Gel Sort (electrophoresis of DNA in a 1% agarose gel) with list $L = (550, 162, 650, 200,550, 350, 323, 550)$. Lanes 1 to 8 the DNA molecule lengths are respectively 550, 162, 650, 200,550, 350, 323, and 550 base pairs. Lane A is an instance of Restricted Gel Sort (a non-stable sort) which contains DNA with the same numbers of base pairs as those in lanes 1 to 8.

inversely proportional to their length. When the voltage is removed the gel is a representation of the matrix G (Definition 1). We read the list of sorted indices first by which traveled the least and then in order of their index. The resulting list is in decreasing order. Restricted Gel Sort is similar to Gel Sort except that all the samples of molecules are placed in the same well.

3.2 Optomechanical Sort

The movement of small transparent particles by light alone is an effect most commonly employed in optical tweezers [5] for biologists to manipulate micro-scale objects. Several methods of ordering particles using this technology have been proposed [8, 11]. We, however, propose a novel method that is an instance of the model of Physical Sorting.

Transparent objects experience a force when a beam of light passes through them [4]. This force is caused by the beam's path being refracted by the object. A change in light beam direction causes a change in the beam's momentum, and momentum is only conserved if there is an equal but opposite change of momentum for the object. This momentum change has a component in the same direction as the direction of the beam and a component in the direction of the increasing intensity gradient of the beam. Ashkin [3] provides a linear equation describing the behaviour of objects smaller than the wavelength of the light beam that satisfies Equation (1).

We propose the use of optical tweezers technology as an implementation of the model in Definition 1 to sort objects and we refer to this as Optomechanical Sort. In Optomechanical Sort, all of the input objects are arranged in a straight line in a medium (e.g. water). There is a barrier that prevents the

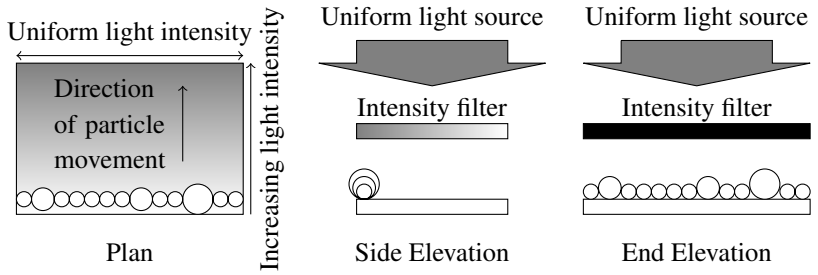

FIGURE 3
The initial configuration of Optomechanical Sort. The circles represent the particles to be sorted.

objects from moving in the direction of the beam. A light source, constant in time, and with a strictly increasing intensity gradient perpendicular to the axis of the input objects is applied (see Figure 3). This intensity gradient is achieved by modulating a uniform light field with an intensity filter variable in one direction only. The objects with a larger volume move more quickly in the direction of increasing intensity than those of a smaller volume. This movement orders the objects by volume.

3.3 Chromatography Sort

Chromatography is a collection of many different procedures in analytical chemistry [15] which behave similarly (e.g. gas, liquid, ion exchange, and thin layer chromatography). It is commonly used to separate the components in a mixture. It separates the input chemicals (analytes) over time in two phases (mobile and stationary). The mobile phase is a solvent for the analytes and filters through the stationary phase. The stationary phase resists the movement of the analytes to different degrees based on their chemical properties. This causes the analytes to separate over time. We refer to the use of chromatography to sort substances by their average velocity through the stationary phase as Chromatography Sort.

The standard equations of analytical chemistry [18] used to calculate the distance traveled by an analyte in a particular mobile phase and stationary phase are linear and satisfy Equation (1). We proceed assuming known relative velocities for analytes in our apparatus. The apparatus is either wide enough to accommodate many analytes side by side or is made of several

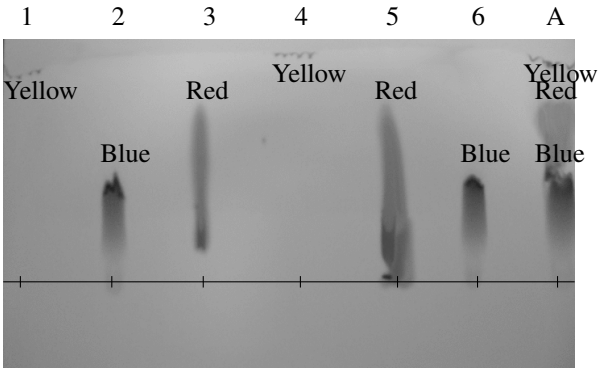

FIGURE 4
Chromatography of household food dye in water on a thin layer plate. Elements of the list to be sorted $L = (3, 1, 2, 3, 2, 1)$ were proportionally encoded as analytes with average speeds proportional to their value: 1 was encoded as blue (0.00001ms^{-1}), 2 as red (0.00002ms^{-1}), 3 as yellow (0.00003ms^{-1}). These are placed in lanes 1 to 6. Lane A is a non-stable sort of L which contains each of the dyes in lanes 1 to 6.

identical setups which allow side by side comparisons.

Given a list L of numbers to be sorted, we encode each element of L as a sample of analyte with a relative velocity proportional to the element value. Each analyte is placed in the chromatography apparatus in the same order as in L. When the process commences the analytes move along the stationary medium at a rate proportional to their relative velocity. When the process is halted the apparatus is a representation of the matrix G (from Equation (1)). We then read off the list of sorted indices by recording the index of each element in order of those which traveled the most and then in order of their index position in L.

Restricted Chromatography Sort is similar to Chromatography Sort except that all analytes are mixed together and placed in the apparatus as one sample.

3.4 Rainbow Sort

Rainbow Sort was first described by Schultes [19] as an unstable sort but by a simple generalisation it becomes an instance of the Model. Rainbow Sort utilises the phenomenon of dispersion, where light beams of longer wavelengths are refracted to a lesser degree than beams of a shorter wavelength. In Rainbow Sort, as described by Schultes [19], each element of a list L is

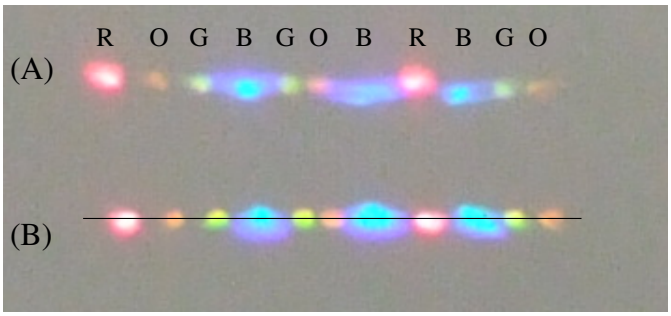

FIGURE 5
An implementation of Generalised Rainbow Sort. $L = (635, 592, 513, 426, 513, 592, 426, 635, 426, 513, 592)$ were encoded as beams of light of proportional wavelength. (A) Beams that have been refracted by a distance proportional to their wavelength. Beams that are lower down have lesser wavelengths than those higher up. (B) The status of the beams before they were refracted by the prism. There is some blurring in the image introduced by the lenses. The labels B,G,O,R refer to blue(426), green(513), orange(592), and red(635), respectively.

encoded as a distinct wavelength proportional to its value. A beam of light containing only the wavelengths to be sorted is passed through a prism. The component wavelengths are refracted at different angles and so emerge from the prism as separate beams, and in an order dictated by their wavelengths. A light measurement device is positioned to sequentially read the ordered component beams. This is an unstable sort as it does not output repeated input elements or return a list of indices.

Schultes provides a possible technique to sort lists with repeated elements with Rainbow Sort [19]. We suggest our own method that follows from the Model of Physical Sorting called Generalised Rainbow Sort. This is similar to Rainbow Sort except it utilises the full geometry of the prism and is an instance of the Model. It also returns the indices of the sorted list elements, thus guaranteeing stability. Each element of the list L is encoded as a beam of light of a distinct wavelength proportional to its value. Each beam is then passed through the prism at a different depth in the prism. We then read the list of sorted indices by recording the index of each refracted beam in the order of those which where refracted the most, and were there are multiple beams refracted to the same degree, in order of their index.

3.5 Mass Spectrometry Sort

Mass spectrometry [12] is a technique used for separating ions by their mass-to-charge ratio and is most commonly used to identify unknown compounds and to clarify the structure and chemical properties of molecules. Of the several types of mass spectrometry we describe here the "time of flight" method. Gaseous sample particles are ionised by a short pulse of electrons and accelerated to a speed that is inversely proportional to their mass and directly proportional to their charge by a series of high voltage electric fields towards a long field free vacuum tube known as the field-free drift region. Here each ion moves at its entry velocity as they travel along the vacuum tube in a constant high voltage. At the opposite end of the tube there is a detector to record the arrival of the ions. Since different ions all travel the same distance but with characteristic velocities they arrive at the detector at different times. Using the time of arrival (time of flight) we identify the ions. Gross [12] provides a linear equation to describe the time of flight of an ion based on its mass-to-charge ratio which satisfies Equation (1).

We refer to implementations of the model in Definition 3 using mass spectrometry for sorting as Mass Spectrometry Sort. Given a list L of numbers to be sorted, we encode each element of L as a sample of molecules with a time of flight proportional to the element value. The samples of molecules are fired simultaneously by a mass spectrometer. The time of flight of each element is then recorded as it arrives at the sensor. This is the sorted list. Since we cannot record or distinguish multiple instances in the input Mass Spectrometry Sort is an unstable sort.

Our usual technique for introducing stability is to run identical instances that sort each element in parallel. However due to the high cost of mass spectrometers we regard this as this as unfeasible.

4 CONCLUSION

In this paper we have proposed a Model of Physical Sorting that computes a stable sorting of its input list of natural numbers. This model has a constant-time parallel atomic operation that transforms the input list into a matrix, where only one dimension of the matrix is dependent on the input list length. The list of stable sorted indices is then generated in linear time (it would be possible to parallelise this operation to be more efficient, e.g. logarithmic time). We have provided five physical instances of the Model that are well-known laboratory techniques. Nature is providing a simple mechanism to

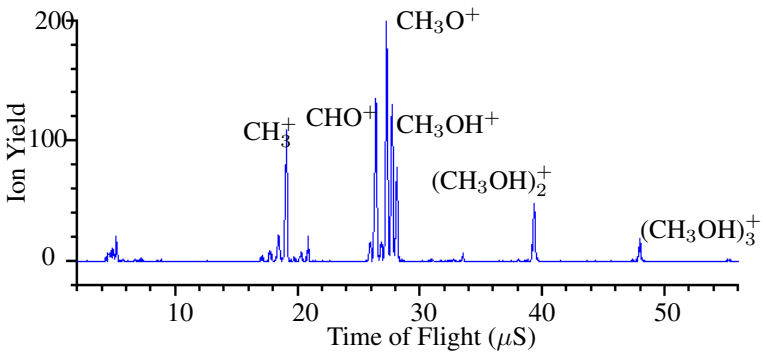

FIGURE 6
A mass spectrum [10] representing the sorting of a multiset $T = (19, 24, 9, 13, 19, 9, 14, 13)$. We map the ion CH_3^+ to 9, CHO^+ to 13, CH_3O^+ to 14, CH_3OH^+ to 14, $(CH_3OH)_2^+$ to 19 and $(CH_3OH)_3^+$ to 24 based on root mean charge ratio. Since the original order of the elements was lost along with multiplicity of the elements, the sort here is unstable. Unlabeled peaks are not involved in the sort.

order our data in such a way that it takes only linear time for us to sequentially generate a sorted list. We showed how the relationship between the Model and a Restricted Model naturally suggests how to introduce stability into an existing physics-inspired sort, i.e. Rainbow Sort. Several of the implementations have the potential to rapidly sort millions or more items. Other candidate instances of the model that we have not considered here include centrifugal separation and fractional distillation [15].

5 ACKNOWLEDGMENTS

Murphy and Woods were funded by the Irish Research Council for Science, Engineering and Technology. We acknowledge Charles Markham, William (Bill) Lanigan, and Fiachra Matthews for their help implementing Generalised Rainbow Sort. This paper is dedicated to the memory of Dr. Lanigan.

REFERENCES

[1] L. Adleman. (1994). Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024.

[2] J. J. Arulanandham, C. S. Calude, and M. J. Dinneen. (2002). Beadsort: A natural sorting algorithm. *Bulletin of EATCS*, 76:153–162.

[3] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. (1986). Observation of a single-beam gradient force optical trap for dielectric particles. *Optics Letters*, 11:288–290.

[4] A. Ashkin. (1970). Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24:156–159.

[5] A. Ashkin. (2000). History of optical trapping and manipulation of small-neutral particle, atoms, and molecules. *IEEE Journal on Selected Topics in Quantum Electronics*, 6:841–856.

[6] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, and E. Shapiro. (2001). Programmable and autonomous computing machine made of biomolecules. *Nature*, 414:430–434.

[7] C. S. Calude and G. Păun. (2001). *Computing with Cells and Atoms*. Taylor & Francis, London.

[8] P. Y. Chiou, A. T. Ohta, and M. C. Wu. (2005). Massively parallel manipulation of single cells and microparticles using optical images. *Nature*, 436:370–372.

[9] A. K. Dewdney. (1984). On the spaghetti computer and other analog gadgets for problem solving. *Scientific American*, 250:19–26.

[10] E. M. Duffy. (2005). Studying a supersonic beam of clusters using a mass spectrometer. Master's thesis, National University of Ireland, Maynooth.

[11] J. Glückstad. (2004). Microfluidics: Sorting particles with light. *Nature Materials*, 3:9–10.

[12] J. H. Gross. (2004). *Mass Spectrometry: A Textbook*. Springer, Berlin.

[13] D. E. Knuth. (1998). *The Art of Computing Programming: Sorting and Searching*, volume 3. Addison-Wesley, Reading, MA, second edition.

[14] C. Mead. (1989). *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA.

[15] C. E. Meloan. (1999). *Chemical Separations: principles, techniques, and experiments*. Wiley, New York.

[16] N. Murphy, T. J. Naughton, D. Woods, B. Henley, K. McDermott, E. Duffy, P. J. M. van der Burgt, and N. Woods. (2006). Implementations of a model of physical sorting. In A. Adamatzky and C. Teuscher, editors, *From Utopian to Genuine Unconventional Computers*, pages 79–99. Luniver Press, Frome, UK.

[17] T. Naughton, Z. Javadpour, J. Keating, M. Klíma, and J. Rott. (1999). General-purpose acousto-optic connectionist processor. *Optical Engineering*, 38:1170–1177.

[18] C. F. Poole and S. A. Schuette. (1984). *Contemporary practice of chromatography*. Elsevier, Amsterdam.

[19] D. Schultes. (2006). Rainbow sort: Sorting at the speed of light. *Natural Computing*, 5:67–82.

[20] C. F. Simpson and M. Whittaker, editors. (1983). *Electrophoretic techniques*. Academic Press, London.

[21] T. Toffoli. (1992). What are nature's 'natural' ways of computing? In *PhysComp '92 – Proceedings of the Workshop on Physics of Computation*, pages 5–9.

[22] C. Viney and R. A. Fenton. (1998). Physics and gel electrophoresis: using terminal velocity to characterize molecular weight. *European Journal of Physics*, 19:575–580.

[23] D. Woods and T. J. Naughton. (April 2005). An optical model of computation. *Theoretical Computer Science*, 334:227–258.