

# Speed Scaling for Weighted Flow Time

Nikhil Bansal \*

Kirk Pruhs †

Cliff Stein ‡

## Abstract

Intel’s SpeedStep and AMD’s PowerNOW technologies allow the Windows XP operating system to dynamically change the speed of the processor to prolong battery life. In this setting, the operating system must not only have a *job selection policy* to determine which job to run, but also a *speed scaling policy* to determine the speed at which the job will be run. We give an online speed scaling algorithm that is  $O(1)$ -competitive for the objective of weighted flow time plus energy. This algorithm also allows us to efficiently construct an  $O(1)$ -approximate schedule for minimizing weighted flow time subject to an energy constraint.

## 1 Introduction

In addition to the traditional goal of efficiently managing time and space, many computers now need to efficiently manage power usage. For example, Intel’s SpeedStep and AMD’s PowerNOW technologies allow the Windows XP operating system to dynamically change the speed of the processor to prolong battery life. In this setting, the operating system must not only have a *job selection policy* to determine which job to run, but also a *speed scaling policy* to determine the speed at which the job will be run. These policies must be online since the operating system does not, in general, have knowledge of the future. In current CMOS based processors, the speed satisfies the well known cube-root-rule, that the speed is approximately the cube root of the power [Mud01, BBS<sup>+</sup>00]. Thus, in this work, we make the standard generalization that the power used by a processor is equal to speed to some power  $\alpha \geq 1$ , where one should think of  $\alpha$  as being approximately 3 [YDS95, BKP07]. Energy is power integrated over time. An operating system is faced with a dual objective optimization problem as it both wants to conserve energy, and optimize some Quality of Service (QoS) measure of the resulting schedule.

By far the most commonly used QoS measure in the computer systems literature is *average response/flow time* or more generally *weighted response/flow time*. The flow time  $F_i$  of a job  $i$  is the time lag between when a job is released to the system and when the system completes that job. Pruhs, Uthaisombut, and Woeginger [PUW08] studied the problem of optimizing total flow time ( $\sum_i F_i$ ) subject to the constraint that the total energy does not exceed some bound, say the energy in the battery, and showed how to efficiently construct

---

\*IBM T.J. Watson Research, P.O. Box 218, Yorktown Heights, NY. nikhil@us.ibm.com.

†Computer Science Department. University of Pittsburgh. kirk@cs.pitt.edu. Supported in part by NSF grants CNS-0325353, CCF-0448196, CCF-0514058 and IIS-0534531.

‡Department of IEOR, Columbia University, New York, NY. Supported in part by NSF grant CCF-0728733 and an IBM faculty fellowship. cliff@ieor.columbia.edu.

offline an optimal schedule for instances with unit work jobs. For unit work jobs, all job selection policies that favor a partially executed job in favor of an unexecuted job are equivalent. Thus the job selection policy is essentially irrelevant.

If there is an upper bound on energy used, then there is no  $O(1)$ -competitive online speed scaling policy for total flow time. To understand intuitively why this is the case, consider the situation when the first job arrives. The scheduler has to allocate a constant fraction of the total energy to this job; otherwise, the scheduler would not be  $O(1)$ -competitive in the case that no more jobs arrive. However, if many more jobs arrive in the future, then the scheduler has wasted a constant fraction of its energy on only one job. By iterating this process, one obtains a bound of  $\omega(1)$  on the competitive ratio with respect to total flow time. (See Section 6.)

Albers and Fujiwara [AF07] proposed combining the dual objectives of energy and flow time into the single objective of energy used plus total flow time. Optimizing a linear combination of energy and total flow time has the following natural interpretation. Suppose that the user specifies how much improvement in flow time, call this amount  $\rho$ , is necessary to justify spending one unit of energy. For example, the user might specify to the Windows XP operating system that he is willing to spend 1 erg of energy from the battery for a decrease of 3 micro-seconds in response time. Then the optimal schedule, from this user's perspective, is the schedule that optimizes  $\rho = 3$  times the energy used plus the total flow time. By changing the units of either energy or time, one may assume without loss of generality that  $\rho = 1$ .

[PUW08] observe that in any locally-optimal normal schedule, each job  $i$  is run at a power proportional to the number of jobs that depend on  $i$ . Roughly speaking, *normal* means that no job completes exactly when another job is released. We say that a job  $j$  *depends* on a job  $i$  if delaying  $i$  would delay  $j$ . In the online setting, an obvious lower bound to the number of jobs that depend on the selected job is the number of *active* jobs, where an active job is one that has been released but has not yet completed. Thus Albers and Fujiwara [AF07] propose the natural online speed scaling algorithm that always runs at a power equal to the number of active jobs. They again only consider the case of unit work jobs. They do not actually analyze this natural algorithm, but rather analyze a batched variation, in which jobs that are released while the current batch is running are ignored until the current batch finishes. They show that this batched algorithm is  $8.3e^{(\frac{3+\sqrt{5}}{2})^\alpha}$ -competitive with respect to the objective of total flow time plus energy, and also give a dynamic programming algorithm to compute the offline optimal schedule for unit work jobs.

One reason that both [PUW08] and [AF07] consider only unit work jobs is that it seems that the optimal schedule for arbitrary work jobs is quite difficult to characterize.

## 1.1 Our Results

We give significantly stronger results for the problem of minimizing the objective of (weighted) flow time plus energy. We both improve the algorithm and analysis in the special case (unit jobs, no weights) considered previously [AF07] and then we give algorithms for the more general problem with weights and arbitrary work jobs.

First, we show that the natural online speed scaling algorithm proposed in [AF07] is 4-competitive for unit work jobs. This guarantee is independent of  $\alpha$ . In comparison, the competitive ratio  $8.3e^{(\frac{3+\sqrt{5}}{2})^\alpha}$  obtained in [AF07] is a bit over 400 when the cube-root rule holds ( $\alpha = 3$ ).

More importantly, we consider the case of arbitrary work jobs, and consider a much more general QoS measure: weighted flow. We assume that each job has a positive integer weight  $w_i$ . The weighted flow objective is then the weighted sum of flow times,  $\sum_i w_i F_i$ . Weighted flow generalizes both total flow time, and total/average stretch, which is another common QoS measure in the systems literature. The stretch/slow-

down of a job is the flow time divided by the work of the job. Many server systems, such as operating systems and databases, have mechanisms that allow the user or the system to give different priorities to different jobs. For example, Unix has the `nice` command. In our setting, the weight of a job is indicative of the flow time versus energy trade-off for it. The user may be willing to spend more energy to reduce the flow time of a higher priority job, than for lower priority jobs.

Our analysis consists of two steps. We first relax the objective function to be fractional weighted flow plus energy, instead of weighted flow plus energy. In the fractional weighted flow time measure, at each time step a job contributes its weight times the fraction of unfinished work to the objective (See Section 2 for details). In the second step we show how to modify our algorithm for fractional weighted flow plus energy to obtain results for weighted flow plus energy at the loss of a small factor in the competitive ratio. The main reason for this two step analysis is that fractional flow is substantially easier to analyze than total flow. For example, for a constant speed processor, computing the optimal weighted flow schedule is NP-hard [LLK84], but the simple algorithm Highest Density First (HDF) is optimal for fractional weighted flow. HDF is the algorithm that always runs the active job with the maximum *density*, where the density of a job is the ratio of its weight to its work. HDF is still the optimal job selection policy for fractional weighted flow when speed scaling is allowed.

Our algorithm is a natural generalization of the algorithm proposed in [AF07]. We define the algorithm  $A$  to be the one that uses HDF for job selection, and always runs at a power equal to the fractional weight of the active jobs. In Section 4 we consider the case of unit-work unit-weight jobs. We show that algorithm  $A$  is 2-competitive with respect to the objective of fractional flow plus energy. As a corollary to this, we show that the algorithm  $B$  (proposed by [AF07]), that runs at power equal to the number of unfinished jobs is 4-competitive for total flow plus energy. More generally, we show that algorithm  $A$  is 2-competitive for instances where all jobs have unit density. This leads to another algorithm  $B'$  that is 4-competitive for total flow plus energy in the unit density case.

In Section 5 we consider jobs with arbitrary work and arbitrary weights. Let  $\gamma = \max\left(2, \frac{2(\alpha-1)}{\alpha-(\alpha-1)^{1-1/(\alpha-1)}}\right)$ . We show that algorithm  $A$  is  $\gamma$ -competitive with respect to the objective of fractional flow plus energy. For any  $\alpha > 1$ , the value of  $\gamma \leq \max(2, 2\alpha - 2)$ . For large values of  $\alpha$ ,  $\gamma$  is approximately  $2\alpha / \ln \alpha$  (ignoring lower order terms) and for  $\alpha = 3$ ,  $\gamma \approx 2.52$ . We then use  $A$  to define an algorithm  $C_\epsilon$ , which is parameterized by  $\epsilon > 0$ , and is  $\gamma\mu_\epsilon$ -competitive with respect to the objective of total weighted flow plus energy, where  $\mu_\epsilon = \max(1 + \frac{1}{\epsilon}, (1 + \epsilon)^\alpha)$ . When the cube-root rule holds, by picking  $\epsilon = .463$ ,  $\mu_\epsilon$  is about 3.15, and the competitive ratio  $\gamma\mu_\epsilon$  is a bit less than 8. For large values of  $\alpha$ , picking  $\epsilon \approx \ln \alpha / \alpha$  implies that  $C_\epsilon$  is approximately  $2\alpha^2 / \ln^2 \alpha$  competitive.

The analysis in [AF07] was based on comparing the online schedule directly to the optimal schedule. However, even for the case of unit work jobs, the optimal schedule can be rather complicated [PUW08, AF07]. Our analyses of algorithm  $A$  are based on the use of a potential function, and does not require us to understand the structure of the optimal schedule. This approach has two advantages over previous methods. First, our analysis is simpler and tighter than the analysis in [AF07] in the case of unit-work unit-weight jobs. Second, we can extend our analysis to the case of arbitrary work and arbitrary weight jobs. We give an overview of our potential function analysis technique in Section 3.

Further, our results also give a way to compute the schedule that optimizes weighted flow subject to a constraint  $E$  on the energy used. It is not too hard to see that if we trace out all possible optimal weighted flow schedules for all energy bounds  $E$ , and we traced out all optimal weighted flow plus  $\rho$  times energy schedules for all possible factors  $\rho$ , that the resulting schedules are the same. That is, a schedule  $S$  is an optimal weighted flow schedule for some energy bound  $E$  if and only if it is an optimal weighted flow plus  $\rho$  time energy schedule for some factor  $\rho$ . Thus, by performing a binary search over the possible  $\rho$ , and

applying our algorithm  $C$ , one can find an  $O(1)$ -approximate weighted flow schedule subject to an energy bound of  $E$ .

## 1.2 Related Results

Theoretical investigations of speed scaling algorithms were initiated by Yao, Demers, and Shenker [YDS95], who considered the problem of minimizing energy usage when each task has to be finished on one machine by its deadline. Since [YDS95] the vast majority of the algorithmic literature has focused on problems where jobs have deadlines. One reason that much of the research has focused on speed scaling of jobs with deadlines is that in general, algorithm design and analysis are significantly more difficult in speed scaling problems than in the corresponding scheduling problem on a fixed speed processor, but deadlines help constrain how the energy can be distributed throughout the schedule, thus making scheduling problems with deadlines more amenable to analysis.

In many computational settings, however, most processes do not have natural deadlines associated with them. As one example of this point, observe that neither Linux nor Microsoft Windows use deadline based schedulers. This observation motivates our work, in which we consider flow rather than a deadline-based objective.

[YDS95] show that there is an optimal offline greedy algorithm to compute the energy optimal schedule subject to deadline feasibility constraints. Constant competitive online algorithms are given in [YDS95, BKP07, BBCP08, BCPRK]. In particular, the online algorithm Optimal Available (OA), proposed in [YDS95], was shown to be  $O(1)$ -competitive in [BKP07] using a potential function analysis. OA runs at the speed that would be optimal, given the current state and given that no more jobs arrive in the future. The speed scaling component of our algorithm  $A$  is similar in spirit as it can be described as: run at a constant factor  $\alpha - 1$  times the optimal speed given the current state, and given that no more jobs arrive in the future. The potential functions that we use to analyze  $A$  are reminiscent, but certainly not the same, as the one used in [BKP07] to analyze OA. Competitive algorithms for throughput and weighted throughput are given in [CCL<sup>+</sup>07]. There have been a couple papers in the literature on speed scaling with the makespan objective. The makespan can be thought of as a common deadline for all the jobs. [PvSU08] give a poly-log approximation algorithm for makespan on identical machines with precedence constraints give a bound on the available energy. [Bun06] gives an efficient algorithm to compute all Pareto optimal schedules for tasks on one machine with release dates, and for unit-work tasks on multiple machines with release dates.

We do not attempt to exhaustively survey the constantly growing literature on speed scaling. We refer the reader to a somewhat dated survey [IP05] for further background.

One recent negative result by Bunde [Bun06] shows that even for unit-work jobs, optimal flow time schedules can not generally be expressed with the standard four arithmetic operations, and the extraction of roots.

## 2 Definitions

An instance consists of  $n$  jobs, where job  $i$  has a release time  $r_i$ , a positive work  $y_i$ , and a positive integer weight  $\bar{w}_i$ . The density of job  $i$  is  $\frac{\bar{w}_i}{y_i}$  and the inverse density is  $\frac{y_i}{\bar{w}_i}$ . We assume, without loss of generality, that  $r_1 \leq r_2 \leq \dots \leq r_n$ . An online scheduler is not aware of job  $i$  until time  $r_i$ , and, at time  $r_i$ , learns  $y_i$  and weight  $\bar{w}_i$ . For each time, a schedule specifies a job to be run and a speed at which the job is run. A job  $i$  completes once  $y_i$  units of work have been performed on  $i$ . The speed is the rate at which work is completed; a job with work  $y$  run at a constant speed  $s$  completes in  $\frac{y}{s}$  seconds. The power consumed

when running at speed  $s$  is  $s^\alpha$ , where we assume that  $\alpha \geq 1$ . The energy used is power integrated over time. We assume that preemption is allowed, that is, a job may be suspended and later restarted from the point of suspension. A job is *active* at time  $t$  if it has been released but not completed at time  $t$ .

As an algorithm runs, we will keep track of the total weight of jobs active at time  $t$ . There are actually two different notions of a job's weight. When we just use weight, we mean the original weight  $\bar{w}_i$  of the job. When we say *fractional weight* of a job  $i$  at time  $t$ , we mean the weight of the job times the percentage of work on the job that has not yet been finished. We will use an overbar for weight and omit it for fractional weight.

Let  $X$  be an arbitrary algorithm. Let  $\bar{w}_x(t)$  denote the weight of jobs active at time  $t$  for algorithm  $X$ . (Note that we make algorithms lower case in subscripts for typographic reasons.) Let  $w_x(t)$  denote the fractional weight of the active jobs at time  $t$  for algorithm  $X$ . Let  $s_x(t)$  be the speed at time  $t$  for algorithm  $x$ , and let  $p_x(t) = (s_x(t))^\alpha$  be the power consumed at time  $t$  by algorithm  $X$ . Let  $E_x(t) = \int_{k=0}^t p_x(k)dk$  be the energy spent up until time  $t$  by algorithm  $A$ .

Just as we defined weight and fractional weight, we can define weighted flow time and fractional weighted flow time analogously. We use the well-known observation that the total weighted flow time is the total weight of the set of active jobs, integrated over time. Let  $W_x(t) = \int_{k=0}^t w_x(k)dk$  be the fractional weighted flow up until time  $t$  for algorithm  $X$ . Let  $\bar{W}_x(t) = \int_{k=0}^t \bar{w}_x(k)dk$  be the weighted flow up until time  $t$  for algorithm  $X$ . Our objective function combines flow and energy and we let  $G_x(t) = W_x(t) + E_x(t)$  be the fractional weighted flow and energy up until time  $t$  for algorithm  $X$ , and  $\bar{G}_x(t) = \bar{W}_x(t) + E_x(t)$  be the weighted flow and energy up until time  $t$  for algorithm  $X$ . Let  $E_x = E_x(\infty)$ ,  $W_x = W_x(\infty)$ ,  $\bar{W}_x = \bar{W}_x(\infty)$ ,  $G_x = G_x(\infty)$  and  $\bar{G}_x = \bar{G}_x(\infty)$  be the energy, fractional weighted flow, weighted flow, fractional weighted flow plus energy, and weighted flow plus energy, respectively, for algorithm  $X$ . We use Opt to denote the offline adversary, and subscript a variable by "o" to denote the value of a variable for the adversary. So  $W_o$  is the fractional weighted flow for the adversary.

### 3 Amortized Local Competitiveness

A common notion to measure an on-line scheduling algorithm is *local competitiveness*, meaning roughly that the algorithm is competitive at all times during the execution. Local competitiveness is generally not achievable in speed scaling problems because the adversary may spend essentially all of its energy in some small period of time, making it impossible for any online algorithm to be locally competitive at that time. Thus, we will analyze our algorithms using *amortized local competitiveness*, which we now define. Let  $X$  be an arbitrary online scheduling algorithm, and  $H$  an arbitrary objective function. Let  $\frac{dH(t)}{dt}$  be the rate of increase of the objective  $H$  at time  $t$ . The online algorithm  $X$  is *amortized locally  $\gamma$ -competitive with potential function  $\Phi(t)$  for objective function  $H$*  if the following two conditions hold:

**Boundary Condition:**  $\Phi$  is initially 0, and finally nonnegative. That is,  $\Phi(0) = 0$ , and there exists some time  $t'$  such that for all  $t \geq t'$  it is the case that  $\Phi(t) \geq 0$ .

**General Condition:** For all times  $t$ ,

$$\frac{dH_x(t)}{dt} - \gamma \frac{dH_o(t)}{dt} + \frac{d\Phi(t)}{dt} \leq 0. \quad (1)$$

We break the general condition into three cases:

**Running Condition:** For all times  $t$  when no job arrives (1) holds.

**Job Arrival Condition:**  $\Phi$  does not increase when a new job arrives.

**Completion Condition:**  $\Phi$  does not increase when either the online algorithm or the adversary complete a job.

Observe that when  $\Phi(t)$  is identically zero, we have ordinary local competitiveness. It is well known that amortized local  $\gamma$ -competitiveness implies that when the algorithm completes, the total cost of the online algorithm is at most  $\gamma$  times the total cost of the optimal offline algorithm.

**Lemma 1** *If online algorithm  $X$  is amortized locally  $\gamma$ -competitive with potential function  $\Phi(t)$  for objective function  $H$ , then  $H_x \leq \gamma H_o$ .*

*Proof:* Let  $t_1, \dots, t_{3n}$  be the events that either a job is released, the online algorithm  $x$  completes a job, or the adversary completes a job. Let  $\Delta(\Phi(t_i))$  denote the change in potential in response to event  $t_i$ . Let  $t_0 = 0$  and  $r_{3n+1} = +\infty$ . Integrating equation 1 over time, we get that

$$H_x + \sum_{i=1}^{3n+1} \Delta(\Phi(t_i)) \leq \gamma H_o.$$

By the job arrival condition, and the completion condition, we can conclude that  $H_x + \Phi(\infty) - \Phi(0) \leq \gamma H_o$ , and finally, by the boundary condition, we can conclude that  $H_x \leq \gamma H_o$ . ■

Now consider the case that the objective function is  $G$ , the fractional weighted flow plus energy. Then  $\frac{dG(t)}{dt} = w(t) + p(t) = w(t) + s(t)^\alpha$ , and equation 1 is equivalent to:

$$w_x(t) + s_x(t)^\alpha - \gamma(w_o(t) + s_o(t)^\alpha) + \frac{d\Phi(t)}{dt} \leq 0. \quad (2)$$

For our purposes, we will always consider the algorithm  $A$ , where  $s_a(t)^\alpha = w_a(t)$ . Thus equation 2 is equivalent to:

$$2w_a(t) - \gamma(w_o(t) + s_o(t)^\alpha) + \frac{d\Phi(t)}{dt} \leq 0. \quad (3)$$

If  $w_o(t) + s_o(t)^\alpha = 0$  then equation 3 is equivalent to

$$\frac{d\Phi(t)}{dt} \leq -2w_a. \quad (4)$$

Thus, we are essentially required to pick a potential function satisfying equation 4. Note that if  $w_o(t) + s_o(t)^\alpha = 0$  then it must be the case that the adversary has no active jobs at time  $t$ , and  $w_o(t) = s_o(t)^\alpha = 0$ . If  $w_o(t) + s_o(t)^\alpha \neq 0$  then equation 3 can be rewritten as

$$\gamma \geq \frac{2w_a(t) + \frac{d\Phi(t)}{dt}}{w_o(t) + s_o(t)^\alpha}. \quad (5)$$

Since we want to choose  $\gamma$  to be as small as possible, while still satisfying inequality 5, the right side of this inequality will denote our competitive ratio.

## 4 Unit Work and Unit Weight Jobs

In this section we consider jobs with unit work and unit weight. We first show that the speed scaling algorithm  $A$ , where  $s_a(t) = w_a(t)^{1/\alpha}$  is 2-competitive for the objective function of fractional flow time plus energy. We then show how to modify  $A$  to obtain a 4-competitive algorithm for the objective function of (integral) flow time plus energy. Later we extend the analysis to the case when all jobs have unit density.

We first recall the following classic inequality and its corollary [HLP52](Section 8.3), that we use throughout this paper.

**Theorem 2 (Young's Inequality)** *Let  $f$  be a real-valued, continuous, and strictly increasing function on  $[0, c]$  with  $c > 0$ . If  $f(0) = 0$ , and  $a, b$  such that  $a \in [0, c]$ , and  $b \in [0, f(c)]$ , then*

$$\int_0^a f(x)dx + \int_0^b f^{(-1)}(x)dx \geq ab ,$$

where  $f^{(-1)}$  is the inverse function of  $f$ .

**Corollary 3** *For positive reals  $a, b, \mu, p$  and  $q$  such that  $1/p + 1/q = 1$ , the following holds:*

$$\mu \frac{a^p}{p} + \left(\frac{1}{\mu}\right)^{q/p} \frac{b^q}{q} \geq ab.$$

Note that for  $\mu = 1$ , this is the classic Hölder's inequality.

We now prove the main result of this section.

**Theorem 4** *Assume that all jobs have unit work and unit weight. The speed scaling algorithm  $A$ , where  $s_a(t) = w_a(t)^{1/\alpha}$  is 2-competitive with respect to the objective  $G$  of fractional flow plus energy.*

*Proof:* We prove that algorithm  $A$  is amortized locally 2-competitive using the potential function

$$\Phi(t) = \frac{2\alpha}{(\beta + 1)} (\max(0, w_a(t) - w_o(t)))^{\beta+1} ,$$

where  $\beta = (\alpha - 1)/\alpha$ .

We first need to verify the boundary condition. Clearly  $\Phi(0) = 0$ , as  $w_a(0) = w_o(0)$ , and  $\Phi(t)$  is always non-negative.  $\Phi$  satisfies the job completion condition since the fractional weight of a job approaches zero continuously as the job nears completion and there is no discontinuity in  $w_a(t)$  or  $w_o(t)$  when a job completes.  $\Phi$  satisfies the job arrival condition since both  $w_a(t)$  and  $w_o(t)$  increase simultaneously by 1 when a new job arrives.

We are left to establish the running condition. We now break the argument into two cases. In the first case assume that that  $w_a(t) < w_o(t)$ . This case is simpler, since the offline adversary has large fractional weight. Here  $\Phi(t) = 0$  and  $\frac{d\Phi(t)}{dt} = 0$  by the definition of  $\Phi$ . Since we know that  $w_a(t) < w_o(t)$ , it must be the case that  $w_o(t) + s_o(t)^\alpha \neq 0$ , and then that the right side of equation 5 is clearly at most 2.

We now turn to the interesting case that  $w_a(t) \geq w_o(t)$ . For notational ease, we will drop the time  $t$  from the notation, since all variables are understood to be functions of  $t$ . We consider  $d\Phi/dt$ .

$$\frac{d\Phi}{dt} = \frac{2\alpha}{(\beta + 1)} \frac{d\left((w_a - w_o)^{\beta+1}\right)}{dt} = 2\alpha(w_a - w_o)^\beta \frac{d(w_a - w_o)}{dt} . \quad (6)$$

Since jobs have unit density, the rate at which the fractional weight decreases is exactly the rate at which unfinished work decreases, which is just the speed of the algorithm. Thus  $\frac{dw}{dt} = -s$ . Moreover since  $s_a = w_a^{1/\alpha}$ , by the definition of  $A$ , equation 6 can be written as

$$\frac{d\Phi}{dt} = -2\alpha(w_a - w_o)^\beta(s_a - s_o) = -2\alpha(w_a - w_o)^\beta(w_a^{1/\alpha} - s_o). \quad (7)$$

Since  $w_a \geq w_a - w_o$ , it follows that  $-w_a^{1/\alpha} \leq -(w_a - w_o)^{1/\alpha}$  and as  $\beta + 1/\alpha = 1$  by definition of  $\beta$ , equation 7 implies that

$$\frac{d\Phi}{dt} \leq -2\alpha(w_a - w_o) + 2\alpha(w_a - w_o)^\beta s_o. \quad (8)$$

Applying Young's inequality (cf. Corollary 3) with  $\mu = 1$ ,  $a = s_o$ ,  $p = \alpha$ ,  $b = (w_a - w_o)^\beta$ , and  $q = \frac{1}{\beta}$ , we obtain that  $(w_a - w_o)^\beta s_o \leq \beta(w_a - w_o) + s_o^\alpha/\alpha$ . Thus equation 8 can be written as

$$\frac{d\Phi}{dt} \leq -2\alpha(w_a - w_o) + 2\alpha\beta(w_a - w_o) + 2s_o^\alpha = -2(w_a - w_o) + 2s_o^\alpha. \quad (9)$$

If  $w_o + s_o^\alpha = 0$  then equation 9 implies that  $\frac{d\Phi}{dt} \leq -2w_a$ , and equation 4 holds. If  $w_o + s_o^\alpha \neq 0$  then, plugging equation 9 into the right side of equation 5, we obtain a bound on the competitive ratio of

$$\frac{2w_a + \frac{d\Phi}{dt}}{w_o + s_o^\alpha} \leq \frac{2w_a + (-2w_a + 2w_o + 2s_o^\alpha)}{w_o + s_o^\alpha} = \frac{2w_o + 2s_o^\alpha}{w_o + s_o^\alpha} = 2. \quad (10)$$

■

We now modify the algorithm  $A$  to handle integral flow time. Consider the algorithm  $B$  that at all times runs a partially finished job if one exists (there will be at most one), and otherwise runs an arbitrary job. Further  $B$  runs at power equal to the (integral) weight of unfinished jobs. That is  $s_b(t) = \bar{w}_b^{1/\alpha}$ . To analyze algorithm  $B$ , we relate  $B$  to algorithm  $A$ . For the optimum algorithm we use its total fractional flow time plus energy as a lower bound to the integral objective. We begin by observing that under any algorithm each job incurs a flow time plus energy of at least 1.

**Lemma 5** *If all jobs have unit work and unit weight, then for any instance with  $n$  jobs,  $\bar{G}_o \geq n$ .*

*Proof:* Suppose a job has flow time  $f$ , then by convexity of the power function its energy is minimized if it is run at speed  $1/f$  throughout, and hence it uses at least  $f \cdot (1/f)^\alpha = (1/f)^{\alpha-1}$  amount of energy. It suffices to show that  $f + (1/f)^{\alpha-1} \geq 1$  for any  $f > 0$ . Clearly, this is true if  $f \geq 1$ . If  $f \leq 1$ , then  $(1/f) \geq 1$  and hence  $(1/f)^{\alpha-1} \geq 1$  as  $(\alpha - 1) \geq 0$ . ■

**Lemma 6** *Assume that all jobs have unit work and unit weight. The algorithm  $B$ , where  $s_b(t) = \bar{w}(t)^{1/\alpha}$ , is 4-competitive with respect to the objective  $\bar{G}$  of total flow plus energy.*

*Proof:* Consider  $B$  and  $A$  running on the same input instance. At any time  $t$ , the fractional weight  $w_b(t)$  under  $B$  never exceeds that under  $A$ , since if they were ever equal, then algorithm  $B$  must run at least as fast as algorithm  $A$ .  $B$  has at most one partially executed job at any time, which implies that  $\bar{w}_b(t) \leq w_a(t) + 1$ . Since  $B$  runs at speed at least 1 when it is not idling, it follows that  $\bar{W}_b \leq W_a + n$ . Since  $E_b = \bar{W}_b$  and  $E_a = W_a$  it follows that  $\bar{G}_b = \bar{W}_b + E_b = 2\bar{W}_b \leq 2W_a + 2n = G_a + 2n$ . By Theorem 4, we have that  $G_a \leq 2G_o$  and since  $G_o \leq \bar{G}_o$ , it follows that  $\bar{G}_b \leq G_a + 2n \leq 2\bar{G}_o + 2n \leq 4\bar{G}_o$ . The last step follows as  $\bar{G}_o \geq n$  by Lemma 5. ■



## 4.1 Unit Density Jobs

We first note that algorithm  $A$  is 2-competitive for instances with unit density jobs.

**Theorem 7** *Assume that all jobs have unit density. The algorithm  $A$ , where  $s_a(t) = \bar{w}(t)^{1/\alpha}$ , is 2-competitive with respect to the objective  $G$  of total fractional flow plus energy.*

*Proof:* We choose the same potential function as in the proof of Theorem 4 and show that the same proof works. First, it is easily seen that the boundary conditions and the arrival conditions hold, and hence it remains to show the running condition. For unit density jobs, the rate at which fractional weight decreases is equal to the rate at which work decreases which is equal to the speed, i.e.  $dw/dt = -s$ . The running condition follows by observing that this is only property used in the analysis of Theorem 4. ■

We now modify  $A$  to handle integral flow time. Consider the algorithm  $B$  introduced in the last subsection, that gives preference to partially run jobs, and that runs at power equal to total integer weight of unfinished jobs. The following is an easy lower bound of optimal.

**Lemma 8** *Any algorithm must incur a total flow plus energy of  $\sum_i y_i^{2-1/\alpha}$ , where  $y_i$  is the size of job  $i$ .*

*Proof:* We consider each job separately. If a job of size  $y$  is executed at average speed  $s$ , its weighted flow time is exactly  $(y/s) \cdot y$  (as the job has unit density, its weight is also  $y$ ), and its energy consumption is at least  $(y/s) \cdot s^\alpha = ys^{\alpha-1}$ . If  $s \leq y^{1/\alpha}$ , the first term is at least  $y^{2-1/\alpha}$ , otherwise, if  $s \geq y^{1/\alpha}$ , the second term is at least  $y^{2-1/\alpha}$ . ■

We can now show that algorithm  $B$  is 4-competitive.

**Lemma 9** *Assume that all jobs have unit density. The algorithm  $B$ , where  $s_b(t) = \bar{w}(t)^{1/\alpha}$ , is 4-competitive with respect to the objective  $\bar{G}$  of total flow plus energy.*

*Proof:* Clearly, at any time the fractional weight under  $B$  never exceeds that under  $A$ . At any time under  $B$ , if the (unique) partially executed job has size  $y$ , then  $\bar{w}_b(t) \leq w_b(t) + y \leq w_a(t) + y$ . Since  $B$  runs at speed at least  $y^{1/\alpha}$  when the partially executed job has size  $y$ , it follows that  $\bar{W}_b \leq W_a + \sum_i y_i^{2-1/\alpha}$ . Since  $E_b = \bar{W}_b$  and  $E_a = W_a$ , it follows that  $\bar{G}_b = 2\bar{W}_b \leq 2W_a + 2\sum_i y_i^{2-1/\alpha} = G_a + 2\sum_i y_i^{2-1/\alpha}$ . By Theorem 7, we have that  $G_a \leq 2G_o$  and since  $G_o \leq \bar{G}_o$ , together with Lemma 8 it follows that  $\bar{G}_b \leq 2\bar{G}_o + 2\sum_i y_i^{2-1/\alpha} \leq 4\bar{G}_o$ . ■

## 5 Arbitrary Size and Weight Jobs

In this section we consider jobs with arbitrary work and arbitrary weight. We first show that the algorithm  $A$  is  $\gamma = \max(2, \frac{2(\alpha-1)}{\alpha-(\alpha-1)^{1-1/(\alpha-1)}})$  competitive with respect to fractional weighted flow plus energy. Later we show how to use  $A$  to obtain an algorithm for (integral) weighted flow time plus energy. This algorithm will be parameterized by  $\epsilon$  and denoted as  $C_\epsilon$ . The competitive ratio of  $C_\epsilon$  will be  $\gamma\mu_\epsilon$  where  $\mu_\epsilon = \max(1 + \frac{1}{\epsilon}, (1 + \epsilon)^\alpha)$ . We first state a simple algebraic fact, the proof of which can be found in [BKP07].

**Lemma 10** *Let  $q, r, \delta \geq 0$  and  $\mu \geq 1$ . Then  $(q + \delta)^{\mu-1}(q - \mu r - (\mu - 1)\delta) - q^{\mu-1}(q - \mu r) \leq 0$ .*

We now show the main result of this section.

**Theorem 11** *The speed scaling algorithm A, where the job selection policy is HDF and  $s_a(t) = w(t)^{1/\alpha}$ , is  $\gamma = \max\left(2, \frac{2(\alpha-1)}{\alpha-(\alpha-1)^{1-1/(\alpha-1)}}\right)$  competitive with respect to the objective G of fractional weighted flow plus energy. In particular,  $\gamma = 2$  for  $1 < \alpha \leq 2$ , and  $\gamma \leq 2(\alpha - 1)$  for  $\alpha > 2$ . For  $\alpha \geq 2 + e$  we have  $\gamma \leq \alpha - 1$ , and finally, for large  $\alpha$ ,  $\gamma \approx 2\alpha / \ln \alpha$ .*

*Proof:* For technical reasons it will be convenient to work with inverse density which is defined as the ratio of the work of a job divided by its weight. In this terminology the algorithm HDF is the one that works on the job with the least inverse density at any time.

Let  $w_a(h)$  and  $w_o(h)$  be functions of time  $t$  denoting the *total fractional weight of the active jobs which have an inverse density of at least  $h$* , for algorithm A and some fixed optimum algorithm *Opt* respectively. Note that for  $h = 0$ , these terms simply correspond to the total fractional weight at time  $t$ . We will prove that A is amortized locally  $\gamma$ -competitive using the potential function

$$\Phi(t) = \eta \int_{h=0}^{\infty} \left( w_a(h)^\beta (w_a(h) - (\beta + 1)w_o(h)) \right) dh, \quad (11)$$

where  $\beta = (\alpha - 1)/\alpha$ , and  $\eta$  is some constant that we will set later.

That  $\Phi$  satisfies the boundary condition follows easily since  $w_a(h) = w_o(h) = 0$  for all values of  $h$  at time  $t = 0$  and as time approaches infinity. Similarly,  $\Phi$  satisfies the job completion condition since the fractional weight of a job approaches zero as the job nears completion, and there are no discontinuities.

Now consider the arrival condition (which is somewhat non-trivial in this case). Suppose a job  $i$  with inverse density  $h_i$  and weight  $w_i$  arrives at time  $t$ . If  $h \leq h_i$  then both  $w_a(h)$  and  $w_o(h)$  increase simultaneously by  $w_i$ . If  $h > h_i$  then both  $w_a(h)$  and  $w_o(h)$  remain unchanged. Thus after the arrival of job  $i$  the change in the potential function is

$$\eta \int_{h=0}^{h_i} \left[ (w_a(h) + w_i)^\beta (w_a(h) - (\beta + 1)w_o(h) - \beta w_i) - w_a(h)^\beta (w_a(h) - (\beta + 1)w_o(h)) \right] dh. \quad (12)$$

The fact that each summand

$$(w_a(h) + w_i)^\beta (w_a(h) - (\beta + 1)w_o(h) - \beta w_i) - w_a(h)^\beta (w_a(h) - (\beta + 1)w_o(h)) \quad (13)$$

in the integral above is not positive follows from Lemma 10 by setting  $q = w_a(h)$ ,  $r = w_o(h)$ ,  $\delta = w_i$  and  $\mu = \beta + 1$ . Thus the arrival condition holds.

We now consider the running condition. Let  $m_a$  and  $m_o$  denote the minimum inverse density of a job that is alive under A and *Opt* at time  $t$ , respectively. Assume algorithm A is running job  $i$  with inverse density  $h_i = m_a$ . Then for  $h \leq m_a$ , let us consider the rate at which  $w_a(h)$  changes with  $t$ . The remaining work decreases at rate  $-s_a$  and hence the fractional weight decreases at rate  $-s_a \cdot (w_i/y_i)$  where  $y_i$  is the (original) work of this job. Thus,

$$\frac{dw_a(h)}{dt} = -s_a \cdot \frac{w_i}{y_i} = -\frac{s_a}{m_a},$$

and if  $h > m_a$  then  $\frac{dw_a(h)}{dt} = 0$ . Similarly,  $\frac{dw_o(h)}{dt} = -\frac{s_o}{m_o}$  if  $h \leq m_o$  and is 0 otherwise. We now evaluate  $\frac{d\Phi}{dt}$ .

$$\begin{aligned} \frac{d\Phi}{dt} &= \eta \int_{h=0}^{\infty} \left( \frac{d(w_a(h)^{\beta+1})}{dt} - (\beta + 1) \frac{d((w_a(h)^\beta w_o(h)))}{dt} \right) dh \\ &= \eta(\beta + 1) \left[ \int_{h=0}^{\infty} w_a(h)^\beta \frac{dw_a(h)}{dt} dh - \int_{h=0}^{\infty} w_a(h)^\beta \frac{dw_o(h)}{dt} dh \right. \\ &\quad \left. - \beta \int_{h=0}^{\infty} w_a(h)^{\beta-1} w_o(h) \frac{dw_a(h)}{dt} dh \right]. \end{aligned} \quad (14)$$

We now focus on the first integral in equation 14. Since  $A$  works on job with minimum inverse density  $m_a$ , there is non-zero contribution only when  $h \in [0, m_a]$ . Further, for  $h \in [0, m_a)$ , it is the case that  $w_a(h) = w_a(0) = w_a$ . Thus,

$$\int_{h=0}^{\infty} w_a(h)^\beta \frac{dw_a(h)}{dt} dh = \int_{h=0}^{m_a} w_a(h)^\beta \frac{dw_a(h)}{dt} dh = m_a \left( w_a^\beta \left( \frac{-s_a}{m_a} \right) \right) = -w_a^\beta s_a = -w_a. \quad (15)$$

The fact that  $-w_a^\beta s_a = w_a$  follows by the definition of  $A$ , as  $s_a = w_a^{1/\alpha}$ .

We now focus on the second integral in equation 14. We have,

$$- \int_{h=0}^{\infty} w_a(h)^\beta \frac{dw_o(h)}{dt} dh = \int_{h=0}^{m_o} w_a(h)^\beta \frac{s_o}{m_o} dh \leq \int_{h=0}^{m_o} w_a^\beta \frac{s_o}{m_o} dh = w_a^\beta s_o. \quad (16)$$

The inequality 16 follows since  $w_a(h)$  is non-increasing function of  $h$ , and  $w_a(0) = w_a$ .

We now focus on the third integral in equation 14. Recall that  $w_a(h) = w_a$  for  $h \in [0, m_a)$ . Because the algorithm is working on the highest density job,  $\frac{dw_a(h)}{dt} = -\frac{s_a}{m_a}$  for  $h \in [0, m_a)$  and  $\frac{dw_a(h)}{dt} = 0$  for  $h > m_a$ . Then recalling that  $w_o(h)$  is non-increasing with  $h$ , we get,

$$- \int_{h=0}^{\infty} w_a(h)^{\beta-1} w_o(h) \frac{dw_a(h)}{dt} dh = \int_{h=0}^{m_a} w_a(h)^{\beta-1} w_o \frac{s_a}{m_a} dh \leq \int_{h=0}^{m_a} w_a^{\beta-1} w_o \frac{s_a}{m_a} dh = w_a^{\beta-1} w_o s_a = w_o. \quad (17)$$

Combining equations 14, 15, 16 and 17 we obtain

$$\frac{d\Phi}{dt} \leq (\beta + 1)\eta(-w_a + \beta w_o + w_a^\beta s_o). \quad (18)$$

We now consider two cases depending on whether  $\alpha \in [1, 2]$  or whether  $\alpha > 2$ . For  $\alpha \leq 2$ , we apply Young's inequality (cf. Corollary 3), with  $a = w_a^\beta$ ,  $b = s_o$ ,  $p = 1/\beta$ , and  $q = \alpha$ , and  $\mu = 1$ , which yields that

$$w_a^\beta s_o \leq \beta w_a + \frac{s_o^\alpha}{\alpha}. \quad (19)$$

Plugging this in the inequality 18 we obtain that

$$\frac{d\Phi}{dt} \leq \eta(\beta + 1)((\beta - 1)w_a + \beta w_o + \frac{s_o^\alpha}{\alpha}). \quad (20)$$

Now consider the subcase that  $w_o(t) = s_o(t)^\alpha = 0$ . Then, equation 20 implies that

$$\frac{d\Phi}{dt} \leq \eta(\beta + 1)(\beta - 1)w_a. \quad (21)$$

Setting  $\eta = 2\alpha/(\beta + 1)$  we get that  $\frac{d\Phi}{dt} \leq -2w_a$  as required by equation 4. Now consider the subcase that  $w_o(t) + s_o(t)^\alpha \neq 0$ . Plugging the bound on  $\frac{d\Phi}{dt}$  in equation 20 into the right side of equation 5, and regrouping terms, we obtain a bound on the competitive ratio of

$$\frac{(2 - \eta(\beta + 1)(1 - \beta))w_a + \eta\beta(\beta + 1)w_o + \eta(\beta + 1)s_o^\alpha/\alpha}{w_o + s_o^\alpha}. \quad (22)$$

Recalling that  $\eta = 2\alpha/(\beta + 1)$ , and substituting this equality into equation 22, we eliminate the  $w_a$  term. Furthermore, observing that  $\eta\beta(\beta + 1) = 2(\alpha - 1) \leq 2$  and that  $\eta(\beta + 1)/\alpha = 2$ , we obtain the desired competitive ratio of 2 for this case.

We now consider the case of  $\alpha > 2$ . Applying Young's inequality (cf. Corollary 3), with  $a = w_a^\beta$ ,  $b = s_0$ ,  $p = 1/\beta$ , and  $q = \alpha$ , and  $\mu = (\alpha - 1)^{-1/(\alpha-1)}$ , we get that

$$w_a^\beta s_0 \leq \mu\beta w_a + \left(\frac{1}{\mu}\right)^{\alpha\beta} \left(\frac{s_0^\alpha}{\alpha}\right).$$

As  $\alpha\beta = \alpha - 1$  and plugging in the value of  $\mu$ , we have that  $\mu^{-\alpha\beta} = \alpha - 1 = \alpha\beta$  and hence,

$$w_a^\beta s_0 \leq \mu\beta w_a + \beta s_0^\alpha. \quad (23)$$

Plugging equation 23 into equation 18 we get

$$\frac{d\Phi}{dt} \leq \eta [-(\beta + 1)w_a + \beta(\beta + 1)w_o + \beta(\beta + 1)\mu w_a + \beta(\beta + 1)s_0^\alpha]. \quad (24)$$

Now consider the subcase that  $w_o(t) = s_o(t)^\alpha = 0$ . Equation 24 is then equivalent to

$$\frac{d\Phi}{dt} \leq \eta [-(\beta + 1)w_a + \beta(\beta + 1)\mu w_a]. \quad (25)$$

Setting  $\eta = \frac{-2}{(\beta+1)(\mu\beta-1)}$  we get that this is equivalent to  $\frac{d\Phi}{dt} \leq -2w_a$ , as required by equation 4. Now consider the subcase that  $w_o(t) + s_o(t)^\alpha \neq 0$ . Plugging the bound on  $\frac{d\Phi}{dt}$  in equation 24 into the right side of equation 5, and regrouping terms, we obtain a bound on the competitive ratio of

$$\frac{(2 + \eta(\beta + 1)(\mu\beta - 1))w_a + \eta\beta(\beta + 1)(w_o + s_o^\alpha)}{w_o + s_o^\alpha}.$$

Recalling that  $\eta = \frac{-2}{(\beta+1)(\mu\beta-1)}$ , we can eliminate the  $w_a$  term. Thus we obtain a bound on the competitive ratio of

$$\frac{2\beta}{1 - \mu\beta} = \frac{2(\alpha - 1)}{\alpha - (\alpha - 1)^{1-1/(\alpha-1)}} = \gamma. \quad (26)$$

■

To obtain a guarantee for integral weighted flow time, we define the algorithm  $C_\epsilon$  to be the one that uses HDF for job selection, and whenever there is unfinished work, it runs at a power equal to  $(1 + \epsilon)$  times the power that algorithm  $A$  would run at. Note that  $C_\epsilon$  must simulate algorithm  $A$ , and is not the same algorithm as run at power  $(1 + \epsilon)$  times the fractional weight of the active jobs.

**Corollary 12** *Let  $\mu_\epsilon = \max((1 + \frac{1}{\epsilon}), (1 + \epsilon)^\alpha)$ . The algorithm  $C_\epsilon$ , where  $s_{c_\epsilon}(t) = (1 + \epsilon)s_a(t)$ , is  $\mu_\epsilon\gamma$ -competitive with respect to the objective  $\overline{G}$  of weighted flow plus energy. For large values of  $\alpha$ , choosing  $\epsilon \approx \ln \alpha / \alpha$  optimizes  $\mu_\epsilon \approx \alpha / \ln \alpha$ .*

*Proof:* Given an input instance, and any arbitrary speed function  $s_a(t)$ , consider the following two schedules. The first is obtained by running HDF at speed  $s_a(t)$  and the second is obtained by running HDF at speed  $(1 + \epsilon)s_a(t)$ . A simple inductive argument (a formal proof can be found in [BLMSP06]) shows that at any time  $t$ , if some job  $j$  has received  $x$  amount of service under the first schedule, then it has received at least  $\min(p_j, (1 + \epsilon)x)$  amount of service under the second schedule, where  $p_j$  is the size of job  $j$ .

This implies that if job  $j$  is alive under  $C_\epsilon$ , then  $j$  has at least an  $\epsilon/(1 + \epsilon)$  fraction of its weight unfinished under  $A$ . Thus  $W_a \geq (\epsilon/(1 + \epsilon))\overline{W}_{c_\epsilon}$  or equivalently,  $\overline{W}_{c_\epsilon} \leq (1 + \frac{1}{\epsilon})W_a$ . Moreover  $E_{c_\epsilon} \leq (1 + \epsilon)^\alpha E_a$  as the speed under  $C_\epsilon$  is always within  $(1 + \epsilon)$  times that of  $A$ . Together with Theorem 11, the result follows.

■

## 6 An Online Lower Bound

In this section, we show that the problem of minimizing flow time subject to a fixed energy bound online has no constant competitive algorithm. This records into the literature a fact that was generally known by researchers in this area.

**Theorem 13** *Assume that there is some fixed energy bound  $E$ , and the objective is to minimize total flow time. Then there is no  $O(1)$ -competitive online speed scaling algorithm even for unit work and unit weight instances.*

*Proof:* We will give an adversarial strategy for generating an input. The jobs are divided into batches  $B_1, \dots, B_\ell$ . Batch  $B_i$  contains  $n_i = ((2 - 1/\alpha)^\alpha E)^{1/(2\alpha-1)} 2^{i/(2\alpha-1)}$  jobs that arrive together at some time after both the online algorithm and the optimal schedule have finished all the jobs in batch  $B_{i-1}$ .

We first consider the adversary's schedule. Scheduling batch  $B_i$  is equivalent to the off-line problem without release dates. [PUW08] shows that the optimal strategy to minimize flow time is to run the  $j^{\text{th}}$  job to finish at power equal to  $\sigma_j = \rho(n_i - (j - 1))$  for some constant  $\rho$ . The job run at speed  $\sigma_j^{1/\alpha}$  takes time  $\sigma_j^{-1/\alpha}$  to finish. The total energy  $E_i$  expended on batch  $B_i$  is then

$$E_i = \sum_{j=1}^{n_i} \sigma_j^{1-1/\alpha} = \rho^{1-1/\alpha} \sum_{j=1}^{n_i} \sigma(n_i - j + 1)^{1-1/\alpha} = \rho^{1-1/\alpha} \sum_{j=1}^{n_i} j^{1-1/\alpha},$$

which implies that

$$\rho = \left( \frac{E_i}{\sum_{j=1}^{n_i} j^{1-1/\alpha}} \right)^{\alpha/(\alpha-1)}.$$

The optimal flow time for batch  $B_i$  is then

$$\sum_{j=1}^{n_i} (n_i - (j - 1)) \sigma_j^{-1/\alpha} = \rho^{-1/\alpha} \sum_{j=1}^{n_i} j^{1-1/\alpha} = \frac{\left( \sum_{j=1}^{n_i} j^{1-1/\alpha} \right)^{\alpha/(\alpha-1)}}{E_i^{1/(\alpha-1)}}.$$

We now approximate the sum by an integral (the error is negligible for these calculations) and obtain that the optimal flow time for batch  $B_i$  is

$$\left( \frac{n_i^{2\alpha-1}}{((2 - 1/\alpha)^\alpha E_i)} \right)^{1/(\alpha-1)}. \quad (27)$$

Plugging our choice of  $n_i = ((2 - 1/\alpha)^\alpha E)^{1/(2\alpha-1)} 2^{i/(2\alpha-1)}$  into equation (27), we get that the optimal flow time for batch  $B_i$  is

$$\left( \frac{2^i E}{E_i} \right)^{1/(\alpha-1)} \quad (28)$$

The adversary, who knows  $\ell$ , could set

$$E_i = \frac{(\ell + 1 - i)^{2\alpha-2} 2^i E}{\sum_{k=1}^{\ell} (\ell + 1 - k)^{2\alpha-2} 2^k}. \quad (29)$$

With this choice, we clearly have that  $\sum_{i=1}^{\ell} E_i = E$ , as desired. The adversary's flow time of batch  $i$  is then

$$\left(\frac{2^i E}{E_i}\right)^{1/(\alpha-1)} = \left(\frac{\sum_{k=1}^{\ell} (\ell+1-k)^{2\alpha-2} 2^k}{(\ell+1-i)^{2\alpha-2}}\right)^{1/(\alpha-1)}. \quad (30)$$

The quantity  $\sum_{k=1}^{\ell} (\ell+1-k)^{2\alpha-2} 2^k$  is  $O(2^{\ell})$ . Thus the adversary's flow time of each batch is at most

$$O\left(\frac{(2^{\ell})^{1/(\alpha-1)}}{(\ell+1-i)^2}\right). \quad (31)$$

Hence the adversary's total flow time, which is the sum of the flow time of the batches, is  $O(2^{\ell/(\alpha-1)})$ . Note that the last block  $\ell$  consumes a constant fraction of the total energy.

We now consider the online algorithm. Assume that the online algorithm was  $O(1)$ -competitive for flow time. Consider batch  $B_i$ , and let us graciously assume that the online algorithm has spent no energy on the jobs in the previous batches, and yet has accumulated no flow time for these batches. If the online algorithm were to allocate energy  $E_i$  to batch  $B_i$ , then according to equation (28), its flow time for batch  $B_i$  is at least  $\left(\frac{2^i E}{E_i}\right)^{1/(\alpha-1)}$ . By the previous calculations, we know that, in the case that  $B_i$  is the last batch, the adversary's total flow time is  $O(2^{i/(\alpha-1)})$ . Thus, to be  $O(1)$  competitive, the online algorithm needs that  $\left(\frac{2^i E}{E_i}\right)^{1/(\alpha-1)} = O(2^{i/(\alpha-1)})$ . This is equivalent to  $\frac{E_i}{E} = \Omega(1)$ . Thus, to be  $O(1)$ -competitive, the online algorithm needs to allocate a constant fraction of its energy to each batch, and it can only do this for a constant number of batches, which is a contradiction. ■

## References

- [AF07] Susanne Albers and Hiroshi Fujiwara. Energy-efficient algorithms for flow time minimization. *ACM Transactions on Algorithms*, 3(4):49, 2007.
- [BBCP08] Nikhil Bansal, David P. Bunde, Ho-Leung Chan, and Kirk Pruhs. Average rate speed scaling. In *Latin American Theoretical Informatics Symposium*, pages 240–251, 2008.
- [BBS<sup>+</sup>00] David M. Brooks, Pradip Bose, Stanley E. Schuster, Hans Jacobson, Prabhakar N. Kudva, Alper Buyuktosunoglu, John-David Wellman, Victor Zyuban, Manish Gupta, and Peter W. Cook. Power-aware microarchitecture: Design and modeling challenges for next-generation microprocessors. *IEEE Micro*, 20(6):26–44, 2000.
- [BCPRK] Nikhil Bansal, Ho-Leung Chan, Kirk Pruhs, and Dmitriy Rogozhnikov-Katz. Improved bounds for speed scaling in devices obeying the cube-root rule. IBM Technical Report.
- [BKP07] Nikhil Bansal, Tracy Kimbrel, and Kirk Pruhs. Speed scaling to manage energy and temperature. *Journal of the ACM*, 54(1):1, 2007.
- [BLMSP06] Luca Becchetti, Stefano Leonardi, Alberto Marchetti-Spaccamela, and Kirk Pruhs. Online weighted flow time and deadline scheduling. *Journal of Discrete Algorithms*, 4(3):339–352, 2006.
- [Bun06] David Bunde. Power-aware scheduling for makespan and flow. In *2006 ACM Symposium on Parallel Algorithms and Architectures*, pages 190–196, 2006.

- [CCL<sup>+</sup>07] Ho-Leung Chan, Wun-Tat Chan, Tak-Wah Lam, Lap-Kei Lee, Kin-Sum Mak, and Prudence W. H. Wong. Energy efficient online deadline scheduling. In *ACM-SIAM Symposium on Discrete algorithms*, pages 795–804, 2007.
- [HLP52] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, 1952.
- [IP05] Sandy Irani and Kirk R. Pruhs. Algorithmic problems in power management. *SIGACT News*, 36(2):63–76, 2005.
- [LLLK84] J. Labetoulle, E. Lawler, J.K. Lenstra, and A. Rinnooy Kan. Preemptive scheduling of uniform machines subject to release dates. *Progress in Combinatorial Optimization*, pages 245–261, 1984.
- [Mud01] Trevor Mudge. Power: A first-class architectural design constraint. *Computer*, 34(4):52–58, 2001.
- [PUW08] Kirk Pruhs, Patchrawat Uthaisombut, and Gerhard J. Woeginger. Getting the best response for your erg. *ACM Transactions on Algorithms*, 4(3), 2008.
- [PvSU08] Kirk Pruhs, Rob van Stee, and Patchrawat Uthaisombut. Speed scaling of tasks with precedence constraints. *Theory Comput. Syst.*, 43(1):67–80, 2008.
- [YDS95] F. Yao, A. Demers, and S. Shenker. A scheduling model for reduced cpu energy. In *IEEE Symposium on Foundations of Computer Science*, page 374, 1995.