Lagrangian Duality based Algorithms in Online Energy-Efficient Scheduling

Nguyen Kim Thang

IBISC, Université d'Evry Val d'Essonne, France

─── **Abstract** ───────────────────────────

We study online scheduling problems in the general energy model of speed scaling with power down. The latter is a combination of the two extensively studied energy models, speed scaling and power down, toward a more realistic one. Due to the limits of the current techniques, only few results have been known in the general energy model in contrast to the large literature of the previous ones.

In the paper, we consider a Lagrangian duality based approach to design and analyze algorithms in the general energy model. We show the applicability of the approach to problems which are unlikely to admit a convex relaxation. Specifically, we consider the problem of minimizing energy with a single machine in which jobs arrive online and have to be processed before their deadlines. We present a $\max\{4, \alpha^{\alpha}\}$-competitive algorithm (whose the analysis is tight up to a constant factor) where the energy power function is of typical form $z^{\alpha} + g$ for constants $\alpha > 2$ and $g \geq 0$. Besides, we also consider the problem of minimizing the weighted flow-time plus energy. We give an $O(\alpha/\ln \alpha)$-competitive algorithm; that matches (up to a constant factor) to the currently best known algorithm for this problem in the restricted model of speed scaling.

1 Introduction

Energy-efficient algorithms [1] have gained considerable interest in the algorithmic community in the last decade. Many results and techniques have been developed to reduce energy while optimizing some objectives, especially in the domain of scheduling. There are two widely studied models in energy-aware scheduling: *power down* and *speed scaling*. In the power down model, a machine could be set in one of several states, which vary from low-power states to high-power ones and there are transition costs from one state to another. Depending on the required tasks to be performed, an algorithm has to decide when to make a transition and to which states to switch. The goal is to minimize the total energy consumption. In the speed scaling model, there is no state but now one can choose an appropriate speed to process required tasks. The energy power of a machine is a convex function of its speed. An algorithm needs to specify the machine speed and a policy to process jobs in order to optimize some quality of service and the consumed energy. Each model captures partly (but complementarily) a more realistic setting — the speed scaling with power down model. In the latter, a machine could be set in different states and its speed could also be varied as a function of required tasks.

A power management scheduling problem in a realistic setting is an *online* problem [1], meaning that at any time the scheduler is not aware of future events and the decision is irrevocable. The standard performance measure of an algorithm is the *competitive ratio*,

defined as the worst ratio between the cost of the algorithm and that of the optimal solution. As having been raised in [1], an important direction is to design efficient algorithms (in term of competitive ratio) for online problems in the speed scaling and power down model (general energy model for short). Attempting efficient algorithms for problems in the general energy model, one encounters several limits of current tools. Hence there are only a few works on the model [2, 6, 12] in contrast to a large literature of previous energy models.

Popular tools in online computation are the charging scheme and the potential function method. The idea of the methods is to show that an algorithm behaves well in an amortized sense. However, such methods yield little insight about the nature of problems. Recently, different approaches based on the duality of mathematical programming to design and analyze online scheduling have been proposed [3, 11, 16]. The approaches reveal the nature of such the problems, hence lead to algorithms which are usually simple and competitive [3, 11, 16, 10, 13, 14, 5]. An essential starting point of those approaches (except [16]) is to formulate a linear or convex relaxation for a given problem. However, problems in the general energy model unlikely admit convex program and that represents the main obstacle to design competitive algorithms. In the paper, we show an approach to bypass this difficulty.

1.1   The Model, Approaches and Contribution

Model. We are given a single machine that can be set either in the *sleep state* or in the *active state*. In the active state, the machine can be either in *working state* in which some job is currently processed or in *idle state* in which no job is currently executed. Each transition of the machine from the sleep state to the active one has cost A, which represents the *wake-up* cost. In the sleep state, the energy consumption of the machine is 0. In the active state, the machine can choose an arbitrary speed $s(t)$ to execute jobs. Hence, if $s(t) > 0$ then the machine is in the working state; otherwise if $s(t) = 0$, the machine is idle.

The power energy consumption of the machine at time t in its active state is $P(s(t)) = s(t)^\alpha + g$ where $\alpha > 2$ and $g = P(0) \geq 0$. The consumed energy (without wake-up cost) of the machine is $\int_0^\infty P(s(t))dt$ where the integral is taken during the machine's active periods. We decompose the latter into *dynamic energy* $\int_0^\infty s(t)^\alpha dt$ and *static energy* $\int_0^\infty g dt$ (where again the integrals are taken during active periods). We call the model as the *general energy model*.

Jobs arrive over time and could be processed preemptively, i.e., a job could be interrupted and resumed later. At any time, the scheduler has to determine the state and the machine speed (if it is active) and also a policy to execute jobs. In the paper, we study the following problems.

In the first problem, each job j has a released time r_j, a deadline d_j, a processing volume p_j. A job j has to be fully processed in $[r_j, d_j]$. The objective is to minimize the total energy consumption (the static, dynamic energy and the wake-up cost).

In the second problem, each job j has released time r_j, weight w_j and requires p_j units of processing volume. The flow-time of a job j is $C_j - r_j$ where C_j is the completion time of the job. The objective is to minimize the total weighted flow-time plus the total energy (including the wake-up cost).

Lagrangian Duality Approach.

To overcome the issues in the general energy model, we follow our duality approach presented in [16]. The approach has been applied to non-convex relaxations for several problems in

[16]. However, such the problems admit linear relaxations with some lost factors[1] and the consideration of non-convex relaxations permits improvement on the competitive ratio. In this paper, we study the approach for non-convex problems, i.e., the problems unlikely admit a convex relaxation with bounded integrality gap. To the best of our knowledge, it is the first time online algorithms have been designed based on non-convex relaxations.

We first briefly summarize the high level idea of the approach in [16]. Given a problem, formulate a relaxation which is *not* necessarily convex and its Lagrangian dual. Next construct dual variables such that the Lagrangian dual has objective value within a desired factor of the primal one (due to some algorithm). Then by the standard Lagrangian weak duality[2] in mathematical programming, the competitive ratio follows. Since the Lagrangian weak duality also holds in the context of calculus of variations, the approach could be applied for the unknowns which are not only variables but also functions.

Let $L(x, \lambda)$ be the Lagrangian function with primal and dual variables x and λ, respectively. Let $\mathcal{X}$ and $\mathcal{Y}$ are feasible sets of x and λ. Intuitively, the approach could be seen as a game between an algorithm and an adversary. The algorithm chooses dual variables $\lambda^* \in \mathcal{Y}$ in such a way that whatever the choice (strategy) of the adversary, the value $\min_{x \in \mathcal{X}} L(x, \lambda^*)$ is always within a desirable factor c of the objective due to the algorithm. We emphasize that $\min_{x \in \mathcal{X}} L(x, \lambda^*)$ is taken over x feasible solutions of the primal.

An advantage of the approach is the flexibility of the formulation. As convexity is not required, we can study a (non-convex) formulation of a given problem. The main core of the approach is to determine the dual variables and to prove the desired competitive ratio. Determining such dual variables is the crucial step in the analysis. Sometimes, the dual variables have intuitive interpretations that inspire their construction.

It is worthy to note that in the analyses (of both problems), we consider mathematical programs in which the machine state variable remains 0-1 (without being relaxed). An advantage of keeping the variables integer, which is allowed due to the flexibility of the approach, is that we can additionally use charging-scheme-liked arguments. Hence, the analyses are carried out by benefiting properties from both mathematical programming (weak duality) and amortized method (charging scheme).

Our Results.

1. For the problem of minimizing the total consumed energy, we formulate a natural *non-convex* formulation using the Dirac delta function. The Dirac function is useful to represent the wake-up cost — an issue that causes the problems in the general energy model to be non-convex. We present a $\max\{4, \alpha^\alpha\}$-competitive algorithm and the analysis follows the Lagrangian duality approach described above. In the construction of dual variables, a key step of the analysis, we define these variables in such a way that they subtly capture the marginal increase of the energy consumption. Note that the analysis is tight since our algorithm, in the restricted speed scaling model (without static energy and wake-up cost), turns out to be the OA algorithm, which has competitive ratio exactly α^α [9]. Hence, even the competitive ratio has been only slightly improved from $\max\{4, \alpha^\alpha + 2\}$ [12] to $\max\{4, \alpha^\alpha\}$, it suggests that the duality-based approach is seemingly a right tool for online scheduling. Besides, the formulation and the analysis give basics to study the second problem.

---

[1] factors polynomial in $1/\epsilon$ in the resource augmentation model in which the machine has speed $1 + \epsilon$.
[2] For completeness, the weak duality is given in the appendix.

2. For the problem of minimizing energy plus weighted flow-time, we derive an $O(\alpha/\ln\alpha)$-competitive algorithm that matches the currently best known competitive ratio (up to a constant factor) for the same problem in the restricted speed scaling model (where the wake-up cost and the static energy cost are 0). The dual variables and the analysis are built upon the salient ideas from the ones in the previous problem but in a more involved manner. Informally, the dual solutions are constructed in order to balance the weighted flow-time cost and the energy cost at any moment in the schedule (the same idea of previous algorithms in the speed scaling model). However, in the general energy model this cannot be guaranteed for every moment in the schedule. Hence, we introduce an additional term to dual variables that covers the moments where the two costs are not balanced. Intuitively, the additional term represents the loss due to the non-convexity of the problem. It turns out that the loss in term of competitive ratio is only a constant factor.

Due to the space constraint, the analysis is partly given. The full version can be found on the website of the author.

1.2   Related work

Anand et al. [3] proposed studying online scheduling by linear (convex) programming and dual fitting. By this approach, they gave simple algorithms and simple analyses with improved performance for problems where the analyses based on potential functions are complex or it is unclear how to design such functions. Gupta et al. [11] gave a primal-dual algorithm for a class of scheduling problems with cost function $f(z) = z^\alpha$. In [16] we generalized the approach in [3] and proposed to study online scheduling by non-convex programming and the weak Lagrangian duality. Using that technique, [16] derive competitive algorithms for problems related to weighted flow-time. The approaches based on duality become more and more popular. Subsequently, many competitive algorithms have been designed for different problems in online scheduling [3, 11, 16, 10, 13, 14, 5].

For the online problem of minimizing the energy consumption in the model of speed scaling, Bansal et al. [9] gave a $2(\frac{\alpha}{\alpha-1})^\alpha e^\alpha$-competitive algorithm. Later on, Bansal et al. [7] showed that no deterministic algorithm has better competitive ratio than e^α/α. In the general energy model, Irani et al. [15] were the first who studied the problem and they derived an algorithm with competitive ratio $(2^{2\alpha-2}\alpha^\alpha + 2^{\alpha-1} + 2)$. Subsequently, Han et al. [12] presented an algorithm which is $\max\{4, \alpha^\alpha + 2\}$-competitive. In offline setting, the problem is recently showed to be NP-hard [2]. Moreover, they [2] also gave a 1.171-approximation algorithm, which improved the 2-approximation algorithm in [15]. If the instances are agreeable then the problem is polynomial [6]. Recently, Antoniadis et al. [4] have given a FPTAS for the problem.

To the best of our knowledge, the objective of minimizing the total weighted flow-time plus energy has not been studied in the speed scaling with power down energy model. However, this objective has been widely studied in speed scaling energy model. Bansal et al. [8] gave an $O(\alpha/\log\alpha)$-competitive algorithm for weighted flow-time plus energy in a single machine where the energy function is s^α. Based on linear programming and dual-fitting, Anand et al. [3] proved an $O(\alpha^2)$-competitive algorithm for unrelated machines. Subsequently, Nguyen [16] and Devanur and Huang [10] presented an $O(\alpha/\log\alpha)$-competitive algorithms for unrelated machines by dual fitting and primal dual approaches, respectively. It turns out that the different approaches lead to the same algorithm. To the best of our knowledge, no competitive algorithm is known in the general energy model for this problem.

2   Minimizing Energy in Speed Scaling with Power Down Model

In this section, we study the problem of minimizing the total energy. We formulize the problem as a mathematical program. In such a program, we need to incorporate an information about the machine states and the transition cost from the sleep state to the active one. Here we make use of the properties of the Heaviside step function and the Dirac delta function to encode the machine states and the transition cost. Recall that the Heaviside step function $H(t) = 0$ if $t < 0$ and $H(t) = 1$ if $t \geq 0$. Then $H(t)$ is the integral of the Dirac delta function δ (i.e., $H' = \delta$) and it holds that $\int_{-\infty}^{+\infty} \delta(t)dt = 1$. Now let $F(t)$ be a function indicating whether the machine is in active state at time t, i.e., $F(t) = 1$ if the machine is active at t and $F(t) = 0$ if it is in the sleep state. Assume that the machine initially is in the sleep state. Then $A\int_0^{+\infty} |F'(t)|dt$ equals twice the transition cost of the machine (a transition from the active state to the sleep state costs 0 while by the term $A\int_0^{+\infty} |F'(t)|dt$, it costs A).

Let $s_j(t)$ be variable representing the speed of job j at time t. The problem could be formulated as the following (non-convex) program.

$$\min \quad \int_0^\infty P\left(\sum_j s_j(t)\right)F(t)dt + \frac{A}{2}\int_0^{+\infty}|F'(t)|dt \tag{1}$$

$$\text{subject to} \quad \int_{r_j}^{d_j} s_j(t)F(t)dt \geq p_j \qquad \forall j$$

$$s_j(t) \geq 0, \ F(t) \in \{0,1\} \qquad \forall j, t$$

Observe that each time a job is executed, the machine has to be in the active state. The first constraint ensures that every job j must be fully processed during $[r_j, d_j]$. Note that we do not relax the variable $F(t)$. The objective function consists of corresponding terms to the energy cost during the active periods and the wake-up cost. Recall that $P(z) = z^\alpha + g$.

2.1   Algorithm and Dual Variable Construction.

Define the *critical* speed $s^c = \arg\min_{s>0} P(s)/s$. It has been observed in [6] that in any algorithm, it would better to set the machine speed at least s^c whenever it executes some job. Let $0 < \beta \leq 1$ be some constant to be chosen later.

Let $s^*(t)$ and $s_j^*(t)$ be the machine speed and the speed of job j at time t by the algorithm, respectively. In the algorithm, we maintain variables, called *virtual* speeds, $s(t)$ and $s_j(t)$. Intuitively, job j would be processed by speed $s_j(t)$ at time t (and the machine would process jobs by speed $s(t)$) *if* the wake-up cost equals A and the parameter $g = 0$. However, it is not the case so the algorithm will process jobs by different speeds but it is the function related to the virtual speeds.

During the execution of the algorithm, we also maintain a set of *active jobs*. Informally, a job is *active* if it has been released but has not been processed by the algorithm. Initially, set auxiliary variables $s(t)$ and $s_j(t)$ equal 0 for every time t and jobs j. If a job is released then it is marked as *active*.

Let τ be the current moment. Set $s(t) \leftarrow s^*(t)$ for every $t \geq \tau$. Consider currently active jobs in the earliest deadline first (EDF) order. (The set of active jobs may include new released job and jobs that have been released before τ but have not been processed.) For every active job j and $\tau \leq t \leq d_j$, increase continuously $s_j(t)$ for all $t \in \arg\min_{t'} P'(s(t'))$ and update simultaneously $s(t) \leftarrow s(t) + s_j(t)$ until $\int_{r_j}^{d_j} s_j(t')dt' = p_j$.

Now consider different states of the machine at the current time τ. We distinguish three different states: (1) in *working state* the machine is active and is executing some jobs; (2) in *idle state* the machine is active but its speed equals 0; and (3) in *sleep state* the machine is inactive.

In working state. If $s(\tau) > 0$ then set the machine speed $s^*(t) \leftarrow \max\{s(t), s^c\}$ for $t \geq \tau$ as long as pending jobs exists. Additionally, mark all currently pending jobs as inactive. Otherwise (if $s(\tau) = 0$), switch the machine to the idle state.

In idle state. If $s(\tau) \geq s^c$ then switch to the working state.

If $s^c > s(\tau) > 0$. Do not execute any job; however, mark all currently pending jobs as active. Intuitively, we delay the execution of such jobs until some moment where the machine has to run at speed s^c in order to complete these jobs (assuming that there is no new job released).

Otherwise, if the total duration of idle state from the last wake-up equals A/g then switch to the sleep state.

In sleep state. If $s(\tau) \geq s^c$ then switch to the working state.

Dual variables. Consider a job j and the virtual machine speed $s(t, r_j)$. If $s(t, r_j) > 0$ for every $t \in [r_j, d_j]$, set λ_j such that $\lambda_j p_j / \beta$ equals the marginal increase of the *dynamic energy* due to the arrival of job j. If $s(t, r_j) = 0$ for some moment $t \in [r_j, d_j]$, define λ_j such that $\lambda_j p_j$ equals the marginal increase of the *dynamic and static* energy due to the arrival of job j (assuming no new job will be released later).

2.2   Analysis

The Lagrangian dual of (1) is $\max_{\lambda \geq 0} \min_{s, F} L(s, F, \lambda)$ where the minimum is taken over (s, F) feasible solutions of the primal and L is the following Lagrangian function

$$L(s, F, \lambda) = \int_0^\infty P\left(\sum_j s_j(t)\right) F(t) dt + \frac{A}{2} \int_0^{+\infty} |F'(t)| dt + \sum_j \lambda_j \left(p_j - \int_{r_j}^{d_j} s_j(t) F(t) dt\right)$$

$$\geq \sum_j \lambda_j p_j - \sum_j \int_{r_j}^{d_j} s_j(t) F(t) \left(\lambda_j - \frac{P(s(t))}{s(t)}\right) \mathbb{1}_{\{s(t)>0\}} \mathbb{1}_{\{F(t)=1\}} dt + \frac{A}{2} \int_0^{+\infty} |F'(t)| dt$$

$$(2)$$

where $s(t) = \sum_j s_j(t)$.

By weak duality, the optimal value of the primal is always larger than the one of the corresponding Lagrangian dual. In the following, with the chosen values of dual variables, we bound the Lagrangian dual value in function of the algorithm cost and show the competitive ratio.

Let $s^*(t, r_j)$ be the machine speed at time $t \geq r_j$ settled by the algorithm at time r_j. In other words, $s^*(t, r_j)$ would be the machine speed at time t if there is no new released job after job j. Similarly, let $s_j^*(t, r_j)$ be the speed of job j at time t settled by the algorithm at time r_j.

▶ **Lemma 1.** *Let j be an arbitrary job.*

1. *If $s^*(t, r_j) > 0$ for every $t \in [r_j, d_j]$ then $\lambda_j \leq \beta P'(s^*(t))$ for every $t \in [r_j, d_j]$.*
2. *Moreover, if $s^*(t, r_j) = 0$ for some $t \in [r_j, d_j]$ then $\lambda_j = P(s^c)/s^c$.*

Proof. We prove the first claim. For any time t, speed $s^*(t)$ is non-decreasing as long as new jobs arrive. Hence, it is sufficient to prove the claim assuming that no other job is released

after j, i.e., $\lambda_j \leq \beta P'(s^*(t, r_j))$. The marginal increase in the dynamic energy due to the arrival of j could be written as

$$\frac{1}{\beta}\lambda_j p_j = \int_{r_j}^{d_j} \left[P(s^*(t, r_j)) - P(s^*(t, r_j) - s_j^*(t, r_j)) \right] dt \leq \int_{r_j}^{d_j} P'(s^*(t, r_j)) s_j^*(t) dt$$

$$= \min_{r_j \leq t \leq d_j} P'(s^*(t, r_j)) \int_{r_j}^{d_j} s_j^*(t, r_j) dt = \min_{r_j \leq t \leq d_j} P'(s^*(t, r_j)) \cdot p_j$$

where $\min P'(s^*(t, r_j))$ is taken over $t \in [r_j, d_j]$ such that $s_j^*(t, r_j) > 0$. The inequality is due to the convexity of P and the second equality follows the algorithm (that increase the speed of job j at $\arg\min P'(s(t))$ for $r_j \leq t \leq d_j$). So the first claim follows.

We are now showing the second claim. By the algorithm, the fact that $s^*(t, r_j) = 0$ for some $t \in [r_j, d_j]$ means that job j will be processed at speed s^c in some interval $[a, b] \subset [r_j, d_j]$ (assuming that no new job is released after r_j). The marginal increase in the energy is $P(s^c)(b - a)$ while p_j could be expressed as $s^c(b - a)$. Therefore, $\lambda_j = P(s^c)/s^c$.   ◄

▶ **Theorem 2.** *The algorithm has competitive ratio at most* $\max\{4, \alpha^\alpha\}$.

Proof. Let E_1^* be the dynamic energy of the algorithm schedule. We have $E_1^* = \int_0^\infty [P(s^*(t)) - P(0)] dt \leq \sum_j \lambda_j p_j / \beta$ due to the definition of λ_j's and $0 < \beta \leq 1$.

Let E_2^* be the static energy plus the wake-up energy of the algorithm, i.e., $E_2^* = \int_0^\infty P(0) F^*(t) dt + \frac{A}{2} \int_0^\infty |(F^*)'(t)| dt$ where $F^*(t)$ is the corresponding state (active or sleep) of the machine at time t by the algorithm. We will lower bound the Lagrangian dual objective by E_1^* and E_2^*.

By Lemma 1 (second statement), for every job j such that $s^*(t, r_j) = 0$ for some $t \in [r_j, d_j]$, $\lambda_j = \frac{P(s^c)}{s^c}$. By the definition of the critical speed, $\lambda_j \leq \frac{P(z)}{z}$ for any $z > 0$. Therefore,

$$\sum_j \int_{r_j}^{d_j} s_j(t) F(t) \left(\lambda_j - \frac{P(s(t))}{s(t)} \right) dt \leq 0 \tag{3}$$

where in the sum is taken over jobs j such that $s^*(t, r_j) = 0$ for some $t \in [r_j, d_j]$.

Define

$$L_1(s, \lambda) := \sum_j \lambda_j p_j - \sum_j \int_{r_j}^{d_j} s_j(t) F(t) \left(\lambda_j - \frac{P(s(t))}{s(t)} \right) \mathbb{1}_{\{s(t) > 0\}} \mathbb{1}_{\{F(t) = 1\}} dt$$

which is the right-hand side of (2) without the wake-up term.

Let $\bar{s}(t) \in \arg\max_z z\beta P'(s^*(t)) - P(z)$. We have

$$L_1(s, \lambda)$$

$$\geq \beta E_1^* - \max_{s, F} \sum_j \int_{r_j}^{d_j} s_j(t) F(t) \left[\beta P'(s^*(t)) - \frac{P(s(t))}{s(t)} \right] \mathbb{1}_{\{s(t) > 0\}} \mathbb{1}_{\{F(t) = 1\}} \mathbb{1}_{\{s^*(t) > 0\}} dt$$

$$\geq \beta E_1^* - \max_s \int_0^\infty s(t) \left[\beta P'(s^*(t)) - \frac{P(s(t))}{s(t)} \right] \mathbb{1}_{\{s(t) > 0\}} \mathbb{1}_{\{F(t) = 1\}} \mathbb{1}_{\{s^*(t) > 0\}} dt$$

$$\geq \beta E_1^* - \int_0^\infty \left[\beta P'(s^*(t))\bar{s}(t) - P(\bar{s}(t)) \right] \mathbb{1}_{\{s(t) > 0\}} \mathbb{1}_{\{F(t) = 1\}} \mathbb{1}_{\{s^*(t) > 0\}} dt$$

$$\geq \beta E_1^* - \frac{1}{2} \int_0^\infty \left[\beta P'(s^*(t))\bar{s}(t) - P(\bar{s}(t)) \right] \mathbb{1}_{\{s^*(t) > 0\}} dt$$

$$\qquad - \frac{1}{2} \int_0^\infty \left[\beta P'(s^*(t))\bar{s}(t) - P(\bar{s}(t)) \right] \mathbb{1}_{\{F(t) = 1\}} dt$$

where in the first line, the sum is taken over jobs j such that $s^*(t, r_j) > 0$ for all $t \in [r_j, d_j]$. Note that if $s^*(t, r_j) > 0$ then $s^*(t) \geq s^*(t, r_j) > 0$. The first inequality follows (3) and Lemma 1 (first statement). The second inequality holds since $F(t) \in \{0, 1\}$ and $s(t) \geq \sum_j s_j(t)$ where again the sum is taken over jobs j such that $s^*(t, r_j) > 0$ for all $t \in [r_j, d_j]$. The third inequality is due to the first order derivative and $\bar{s}(t)$ maximizes function $z \beta P'(s^*(t)) - P(z)$ (so $\bar{s}(t)$ is the solution of equation $P'(z(t)) = \beta P'(s^*(t))$).

As the energy power function $P(z) = z^\alpha + g$ where $\alpha > 2$ and $g \geq 0$, $\bar{s}(t)^{\alpha-1} = \beta(s^*(t))^{\alpha-1}$. Therefore,

$$
L_1(s, \lambda) \geq \beta E_1^* - \frac{1}{2} \int_0^\infty \left(\beta \alpha (s^*(t))^{\alpha-1} \bar{s}(t) - (\bar{s}(t))^\alpha - g \right) \mathbb{1}_{\{s^*(t)>0\}} dt
$$
$$
- \frac{1}{2} \int_0^\infty \left(\beta \alpha (s^*(t))^{\alpha-1} \bar{s}(t) - (\bar{s}(t))^\alpha - g \right) \mathbb{1}_{\{F(t)=1\}} dt
$$
$$
= \beta E_1^* - \int_0^\infty (\alpha - 1) \beta^{\alpha/(\alpha-1)} (s^*(t))^\alpha dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)=1\}} dt
$$
$$
= \left[\beta - (\alpha - 1) \beta^{\alpha/(\alpha-1)} \right] E_1^* + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)=1\}} dt
$$

Choose $\beta = 1/\alpha^{\alpha-1}$, we have that

$$
L(s, F, \lambda) \geq \frac{1}{\alpha^\alpha} E_1^* + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)=1\}} dt + \frac{A}{2} \int_0^\infty |F'(t)| dt
$$

In order to prove the theorem, we prove the following claim.

▶ **Claim 3.** *Define*

$$
L_2(F) := \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)=1\}} dt + \frac{A}{2} \int_0^\infty |F'(t)| dt
$$

Then, $L_2(F) \geq E_2^*/4$ for any feasible solution (s, F) of the relaxation.

We first show how to deduce the theorem assuming the claim. By the claim, the dual

$$
L(s, F, \lambda) \geq E_1^*/\alpha^\alpha + L_2(F) \geq E_1^*/\alpha^\alpha + E_2^*/4
$$

whereas the primal is $E_1^* + E_2^*$. Thus, the competitive ratio is at most $\max\{4, \alpha^\alpha\}$. In the remaining, we prove the claim by amortized arguments.

Proof of claim.   Consider the algorithm schedule. An *end-time u* is a moment in the schedule such that the machine switches from the idle state to the sleep state. Conventionally, the first end-time in the schedule is 0. Partition the time line into phases. A *phase $[u, v)$* is a time interval such that u, v are two consecutive end-times. Observe that in a phase, the schedule has transition cost A and there is always a new job released in a phase (otherwise the machines would not switch to non-sleep state). We will prove the claim on every phase. In the following, we are interested in phase $[u, v)$ and whenever we mention $L_2(F)$, it refers to $\frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_u^v g \mathbb{1}_{\{F(t)=1\}} dt + \frac{A}{2} \int_u^v |F'(t)| dt$.

By the algorithm, the static energy of the schedule during the idle time is A, i.e., $\int_u^v g \mathbb{1}_{\{s^*(t)=0\}} dt = A$. Let (s, F) be an arbitrary feasible primal solution.

If during $[u, v)$, the machine according to the solution (s, F) makes a transition from sleep state to non-sleep state (i.e., $F(t') = 0$ and $F(t'') = 1$ for some $u \leq t' < t'' < v$) or

inversely then

$$
\begin{aligned}
L_2(F) &\geq \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{A}{2} \\
&\geq \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{4}\left(\int_u^v g \mathbb{1}_{\{s^*(t)=0\}} dt + A\right) \geq \frac{1}{4} E_2^*\big|_{[u,v)}.
\end{aligned}
$$

If during $[u,v)$, the machine following solution (s,F) makes no transition (from non-sleep static to sleep state or inversely) then $F(t) = 1$ during $[u,v)$ in order to process jobs released in the phase. Therefore,

$$
\begin{aligned}
L_2(F) &\geq \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_u^v g \mathbb{1}_{\{F(t)=1\}} dt = \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{2} \int_u^v g dt \\
&\geq \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \frac{1}{4} \int_u^v g \mathbb{1}_{\{s^*(t)=0\}} dt + \frac{A}{4} \\
&\geq \frac{1}{4}\left(\int_u^v g \mathbb{1}_{\{s^*(t)>0\}} dt + \int_u^v g \mathbb{1}_{\{s^*(t)=0\}} dt + A\right) = \frac{1}{4} E_2^*\big|_{[u,v)}
\end{aligned}
$$

where the second inequality follows the algorithm: as the machine switches to sleep state at time v, it means that the total idle duration in $[u,v)$ incurs a cost A.   ◄

The proof of the claim completes the theorem proof.   ◄

3 Minimizing Energy plus Weighted Flow-Time in Speed Scaling with Power Down Model

In this section, we study the problem of minimizing total weighted flow-time plus energy. Let $F(t)$ be a function indicating whether the machine i is in active state at time t, i.e., $F(t) = 1$ if the machine is active at t and $F(t) = 0$ if it is in the sleep state. Let $s_j(t)$ be the variable that represents the speed of job j at time t. Let C_j be a variable representing the completion time of j. Similar as the previous section, the problem can be formulized as the following (non-convex) program.

$$
\min \int_0^\infty 2P\left(\sum_j s_j(t)\right) F(t) dt + 2\sum_j \left(\int_{r_j}^{C_j} s_j(t) F(t) dt\right) \frac{w_j}{p_j}(C_j - r_j)
$$

$$
+ A \int_0^\infty |F'(t)| dt \tag{4}
$$

$$
\text{subject to} \qquad \int_{r_j}^{C_j} s_j(t) F(t) dt = p_j \quad \forall j
$$

$$
s_j(t) \geq 0, \quad F(t) \in \{0,1\} \quad \forall j, t
$$

The first constraint ensures that every job j must be completed by some moment C_j which is its completion time. In the objective function, the first and second terms represent twice the consumed energy and the total weighted flow-time, respectively. Note that in the second term, $\int_{r_j}^{C_j} s_j(t) F(t) dt = p_j$ by the constraints. The last term stands for twice the transition cost.

Notations. We say that a job j is *pending* at time t if it has not been completed, i.e., $r_j \leq t < C_j$. At time t, denote $q_j(t)$ the *remaining* processing volume of job j. The *total weight* of pending jobs at time t is denoted as $W(t)$. The *density* of a job j is $\delta_j = w_j/p_j$. Recall that the *critical speed* $s^c \in \arg\min_{z\geq 0} P(z)/z$. As $P(z) = z^\alpha + g$, by the first order condition, s^c satisfies $(\alpha - 1)(s^c)^\alpha = g$.

3.1   The Algorithm

We first describe intuitively the ideas of the algorithm. In the speed scaling model, all previous algorithms explicitly or implicitly balance the weighted flow-time of jobs and the consumed energy to process such jobs. That could be done by setting the machine speed at any time t proportional to some function of the total weight of pending jobs (precisely, proportional to $W(t)^{1/\alpha}$ where $W(t)$ is the total weight of pending jobs). Our algorithm follows the idea of balancing the weighted flow-time and the energy. However, in the general energy model, the algorithm would not be competitive if the speed is always set proportionally to $W(t)^{1/\alpha}$ since the static energy might be large due to the long active periods of the machine. Hence, even if the total weight of pending jobs on the machine is small, in some situation the speed is maintained larger than $W(t)^{1/\alpha}$. In fact, it will be set to be the critical speed s^c.

An issue while dealing with the general model is to determine the state of the machine at a given time (active or inactive). If the total weight of pending jobs is small and the machine is active for a long time, then the static energy is large. Otherwise if pending jobs remain for long time then the weighted flow-time is large. The algorithm, together with dual variables, are constructed in order to bypass this difficulty.

Description of algorithm.

At any time t, we distinguish different states of the machine: the *working state* (the machine is active and currently processes some job), the *idle state* (the machine is active but currently processes no job) and the *sleep state*. At time t, we (re)compute the total weight of pending jobs and consider different scenarios corresponding to the current machine state.

In working state. If $\alpha W(t)^{(\alpha-1)/\alpha} > P(s^c)/s^c$ then the machine speed is set as $W(t)^{1/\alpha}$. Otherwise, the speed is set as s^c. At any time, the machine processes the highest density job among the pending ones.

In idle state. 1. If $\alpha W(t)^{(\alpha-1)/\alpha} > P(s^c)/s^c$ then switch to the working state.

2. If $0 < \alpha W(t)^{\frac{\alpha-1}{\alpha}} \le P(s^c)/s^c$ then make a plan to process pending jobs with speed (exactly) s^c in non-increasing order of their density. The plan consists of a single block (with no idle time) and the block length could be explicitly computed (given the processing volumes of all jobs and speed s^c). Hence, the total consumed energy in the plan can be computed and it is independent of the starting time of the plan.

Choose the starting time of the plan in such a way that the total energy consumption in the plan equals the total weighted flow-time of all jobs in the plan. There always exists such starting time since if the plan begins immediately at the current time, the energy is larger than the weighted flow-time; and inversely if the starting time is large enough, the weighted flow-time dominates the energy consumption.

At the starting time of a plan, switch to the working state. (Note that the plan together with its starting time could be changed due to the arrival of new jobs.)

3. Otherwise, if the total duration of idle state from the last wake-up equals A/g then switch to sleep state.

In sleep state. Use the same policy as the first two steps of the idle state.

3.2   Analysis

The Lagrangian dual of program (4) is $\max \min_{s,C,F} L$ where L is the corresponding Lagrangian function and the maximum is taken over dual variables. We need to choose ap-

propriate dual variables and prove that for any feasible solution (s, C, F) of the primal, the Lagragian dual is bounded by a desired factor from the primal.

Dual variables.

Denote the dual variables corresponding to the first constraints of (4) as λ_j's. Set all dual variables (corresponding to the primal (4)) except λ_j's equal 0. The values of dual variables λ_j's is defined as follows.

Fix a job j. At the arrival of a job j, rename pending jobs as $\{1, \ldots, k\}$ in non-increasing order of their densities, i.e., $p_1/w_1 \leq \ldots \leq p_k/w_k$ (note that p_a/w_a is the inverse of job a's density). Denote $W_a = w_a + \ldots + w_k$ for $1 \leq a \leq k$.

Define λ_j such that

$$\lambda_j p_j = w_j \sum_{a=1}^{j} \frac{q_a(r_j)}{W_a^{1/\alpha}} + W_{j+1} \frac{q_j(r_j)}{W_j^{1/\alpha}} + P(s^c) \frac{q_j(r_j)}{s^c} \tag{5}$$

Note that $q_j(r_j) = p_j$. If job j is processed with speed larger than s^c then the first term stands for the weighted flow-time of j and the second term represents an upper bound on the increase of the weighted flow-time of jobs with density smaller than δ_j due to the arrival of j. Observe that as j arrives, the jobs with higher density than δ_j are completed earlier and the ones with smaller density than δ_j may have higher flow-time. Here, the second term in (5) captures the marginal change in the total weighted flow-time. The third term in (5) is introduced in order to cover energy consumption during the execution periods of job j if it is processed by speed s^c. That term is necessary since during such periods the energy consumption and the weighted flow-time are not balanced.

With the constructed dual variables, we are able to prove the competitiveness of the algorithm.

▶ **Theorem 4.** *The algorithm is $O(\alpha/\ln \alpha)$-competitive.*

4   Conclusion

In this paper, we have shown that the Lagrangian duality approach is appropriate to study certain problems which unlikely admit a convex formulation. For many optimization problems, it is challenging to come up with a strong formulation in which the integral constraint of variables is relaxed and the integrality gap is relatively small. The Lagrangian duality approach gives the flexibility to study directly certain problems without relaxing the integrality and without linear/convex formulation. As mentioned earlier and having observed in the analyses, by the approach, one can benefit from techniques in mathematical programming and amortized analysis. It would be interesting to see more work in this direction. For concrete questions, the problems studied in the paper are open for unrelated machine environment. One would expect the existence of algorithms with similar competitive ratio (up to a constant factor).

──── **References** ────

1   Susanne Albers. Energy-efficient algorithms. *Commun. ACM*, 53(5):86–96, 2010.

2   Susanne Albers and Antonios Antoniadis. Race to idle: new algorithms for speed scaling with a sleep state. *ACM Transactions on Algorithms (TALG)*, 10(2):9, 2014.

3   S. Anand, Naveen Garg, and Amit Kumar. Resource augmentation for weighted flow-time explained by dual fitting. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 1228–1241, 2012.

4   Antonios Antoniadis, Chien-Chung Huang, and Sebastian Ott. A fully polynomial-time approximation scheme for speed scaling with sleep state. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1102–1113. SIAM, 2015.

5   Yossi Azar, Nikhil R. Devanur, Zhiyi Huang, and Debmalya Panigrahi. Speed scaling in the non-clairvoyant model. In *Proc. 27th ACM on Symposium on Parallelism in Algorithms and Architectures*, pages 133–142, 2015.

6   Evripidis Bampis, Christoph Dürr, Fadi Kacem, and Ioannis Milis. Speed scaling with power down scheduling for agreeable deadlines. *Sustainable Computing: Informatics and Systems*, 2(4):184–189, 2012.

7   Nikhil Bansal, Ho-Leung Chan, Dmitriy Katz, and Kirk Pruhs. Improved bounds for speed scaling in devices obeying the cube-root rule. *Theory of Computing*, 8(1):209–229, 2012.

8   Nikhil Bansal, Ho-Leung Chan, and Kirk Pruhs. Speed scaling with an arbitrary power function. In *Proc. 20th ACM-SIAM Symposium on Discrete Algorithms*, pages 693–701, 2009.

9   Nikhil Bansal, Tracy Kimbrel, and Kirk Pruhs. Speed scaling to manage energy and temperature. *J. ACM*, 54(1), 2007.

10   Nikhil R. Devanur and Zhiyi Huang. Primal dual gives almost optimal energy efficient online algorithms. In *Proc. 25th ACM-SIAM Symposium on Discrete Algorithms*, 2014.

11   Anupam Gupta, Ravishankar Krishnaswamy, and Kirk Pruhs. Online primal-dual for non-linear optimization with applications to speed scaling. In *Proc. 10th Workshop on Approximation and Online Algorithms*, pages 173–186, 2012.

12   Xin Han, Tak Wah Lam, Lap-Kei Lee, Isaac Kar-Keung To, and Prudence W. H. Wong. Deadline scheduling and power management for speed bounded processors. *Theor. Comput. Sci.*, 411(40-42):3587–3600, 2010.

13   Sungjin Im, Janardhan Kulkarni, and Kamesh Munagala. Competitive algorithms from competitive equilibria: Non-clairvoyant scheduling under polyhedral constraints. In *STOC*, 2014.

14   Sungjin Im, Janardhan Kulkarni, Kamesh Munagala, and Kirk Pruhs. Selfishmigrate: A scalable algorithm for non-clairvoyantly scheduling heterogeneous processors. In *Proc. 55th IEEE Symposium on Foundations of Computer Science*, 2014.

15   Sandy Irani, Sandeep K. Shukla, and Rajesh Gupta. Algorithms for power savings. *ACM Transactions on Algorithms*, 3(4), 2007.

16   Nguyen Kim Thang. Lagrangian duality in online scheduling with resource augmentation and speed scaling. In *Proc. 21st European Symposium on Algorithms*, pages 755–766, 2013.

Appendix: The full version

▶ **Lemma A.1** (Weak duality). *Consider a possibly non-convex optimization problem*

$$p^* := \min_x f_0(x) \ : \quad f_i(x) \le 0, \quad i = 1, \dots, m.$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $0 \le i \le m$. Let $\mathcal{X}$ be the feasible set of x. Let $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be the Lagragian function

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x).$$

Define $d^* = \max_{\lambda \ge 0} \min_{x \in \mathcal{X}} L(x, \lambda)$ where $\lambda \ge 0$ means $\lambda \in \mathbb{R}_+^m$. Then $p^* \ge d^*$.

Proof. We observe that, for every feasible $x \in \mathcal{X}$, and every $\lambda \ge 0$, $f_0(x)$ is bounded below by $L(x, \lambda)$:

$$\forall x \in \mathcal{X}, \ \forall \lambda \ge 0 : \ f_0(x) \ge L(x, \lambda)$$

Define a function $g : \mathbb{R}^m \to \mathbb{R}$ such that

$$g(\lambda) := \min_z L(z, \lambda) = \min_z f_0(z) + \sum_{i=1}^m \lambda_i f_i(z)$$

As g is defined as a point-wise minimum, it is a concave function.

We have, for any x and λ, $L(x, \lambda) \ge g(\lambda)$. Combining with the previous inequality, we get

$$\forall x \in \mathcal{X} : \ f_0(x) \ge g(\lambda)$$

Taking the minimum over x, we obtain $\forall \lambda \ge 0 : \ p^* \ge g(\lambda)$. Therefore,

$$p^* \ge \max_{\lambda \ge 0} g(\lambda) = d^*.$$

◀

A.1 Complete Analysis of Section **3**

The Analysis

The Lagrangian dual of program (4) is $\max \min_{s,C,F} L$ where L is the corresponding Lagrangian function where the maximum is taken over dual variables. The purpose of the section is to choose appropriate dual variables and prove that for any feasible solution (s, C, F) of the primal, the Lagragian dual is bounded by a desired factor from the primal.

Dual variables.

Denote the dual variables corresponding to the first constraints of (4) as λ_j's. Set all dual variables (corresponding to the primal (4)) except λ_j's equal to 0. The values of dual variables λ_j's is defined as the follows.

Fix a job j. At the arrival of a job j, rename pending jobs as $\{1, \dots, k\}$ in non-increasing order of their densities, i.e., $p_1/w_1 \le \dots \le p_k/w_k$ (note that p_a/w_a is the inverse of job a's density). Denote $W_a = w_a + \dots + w_k$ for $1 \le a \le k$.

Define λ_j such that

$$\lambda_j p_j = w_j \sum_{a=1}^{j} \frac{q_a(r_j)}{W_a^{1/\alpha}} + W_{j+1} \frac{q_j(r_j)}{W_j^{1/\alpha}} + P(s^c) \frac{q_j(r_j)}{s^c} \tag{6}$$

Note that $q_j(r_j) = p_j$. If job j is processed with speed larger than s^c then the first term stands for the weighted flow-time of j and the second term represents an upper bound of the increase in the weighted flow-time of jobs with density smaller than δ_j. Observe that due to arrival of j, the jobs with higher density than δ_j are completed earlier and the ones with smaller density than δ_j may have higher flow-time. Informally, the second sum in (6) captures the marginal change in the total weighted flow-time of jobs different from j. The third term in (6) is introduced in order to cover energy consumed during the execution periods of job j if it is processed by speed s^c. That term is necessary since during such periods the energy consumption and the weighted flow-time are not balanced.

The Lagrangian function $L(s, C, F, \lambda)$ with the chosen dual variables becomes

$$
A \int_0^\infty |F'(t)| dt + 2 \int_0^\infty P\left(\sum_j s_j(t)\right) F(t) dt + 2 \sum_j \delta_j (C_j - r_j) \int_{r_j}^{C_j} s_j(t) F(t) dt
$$

$$
+ \sum_j \lambda_j \left(p_j - \int_{r_j}^{C_j} s_j(t) F(t) dt \right)
$$

$$
= \sum_j \lambda_j p_j + A \int_0^\infty |F'(t)| dt + \sum_j \int_{r_j}^{C_j} \delta_j (C_j - r_j) s_j(t) F(t) dt
$$

$$
- \sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2 \frac{P(s(t))}{s(t)} - \delta_j (C_j - r_j) \right) dt
$$

Notations.

We denote $s^*(t)$ the machine speed at time t by the algorithm. So by the algorithm, if $s^*(t) > 0$ then $s^*(t) \geq s^c$. Let $\mathcal{E}_1^*$ and $\mathcal{E}_2^*$ be the total dynamic and static energy consumed by the algorithm schedule, respectively. In other words, $\mathcal{E}_1^* = \int_0^\infty (s^*(t))^\alpha dt$ and $\mathcal{E}_2^* = \int_0^\infty g$ where the integral is taken over all moments t where the machine is active (either in working or in idle states). Additionally, let $\mathcal{E}_3^*$ be the total transition cost of the machine. Moreover, let $\mathcal{F}^*$ be the total weighted flow-time due to the algorithm.

We relate the energy cost of the algorithm schedule and the chosen values of dual variables by the following lemma. Note that by definition of λ_j's, we have that $\sum_j \lambda_j p_j \geq \mathcal{F}^*$.

▶ **Lemma A.2.** *It holds that* $2\mathcal{E}_1^* + 2\mathcal{E}_2^* \geq \mathcal{F}^*$ *and* $\sum_j \lambda_j p_j \geq \mathcal{E}_1^*$.

Proof. We prove the first inequality. Consider times t where the machine speed is s^c. By the algorithm $P(s^c)/s^c \geq \alpha W(t)^{(\alpha-1)/\alpha}$. Recall that by the definition of critical speed $g = (\alpha - 1)(s^c)^\alpha$, so $\alpha(s^c)^\alpha = P(s^c)$. Therefore, $s^c \geq W(t)^{1/\alpha}$. Hence, since $\alpha > 2$,

$$
\mathcal{E}_2^* = \int_0^\infty g \mathbb{1}_{\{s^*(t)=s^c\}} dt = \int_0^\infty (\alpha - 1)(s^c)^\alpha \mathbb{1}_{\{s^*(t)=s^c\}} dt \geq \int_0^\infty W(t) \mathbb{1}_{\{s^*(t)=s^c\}} dt.
$$

Now consider times t where the machine speed is $W(t)^{1/\alpha}$ strictly larger than s^c. Thus the dynamic energy consumed on such periods is

$$
\mathcal{E}_1^* \geq \int_0^\infty (s^*(t))^\alpha \mathbb{1}_{\{s^*(t)>s^c\}} dt \geq \int_0^\infty W(t) \mathbb{1}_{\{s^*(t)>s^c\}} dt
$$

During periods where some jobs are still pending but the machine speed is 0 (that corresponds to the case 2, idle state in the algorithm description), by the algorithm plan, the total weighted flow time of jobs in such periods is bounded by $(\mathcal{E}_1^* + \mathcal{E}_2^*)$ Therefore,

$$2\mathcal{E}_1^* + 2\mathcal{E}_2^* \geq \int_0^\infty W(t)dt = \mathcal{F}^*.$$

In the rest, we prove the second inequality $\sum_j \lambda_j p_j \geq \mathcal{E}_1^*$. By the definition of λ_j's (particularly the third term in (6)), $\sum_j \lambda_j p_j$ covers the total energy of machine during all intervals where the machine processes jobs by speed s^c. Denote Γ as $\sum_j \lambda_j p_j$ subtracting the energy incurred during periods where the machine speed is s^c. We need to prove that Γ is at least the total energy incurred over the moments where the machine speed is strictly larger than s^c. In the following, we are interested only in such moments.

Consider a job k processed at time t with speed larger than s^c. By the definition of λ_j's, Γ contributes to time t an amount at least $\sum_j w_j/W(t)^{1/\alpha}$ where the sum is taken over pending jobs j with density smaller than that of k. The latter is exactly $W(t)$. Thus, Γ contributes to time t an amount $W(t)^{(\alpha-1)/\alpha}$.

Now consider an arbitrarily small interval $[a, b]$ where the machine processes only job k and the speed is strictly larger than s^c. Let W be the total weight of pending jobs over $[a, b]$. The processing amount of k done over $[a, b]$ is $W^{1/\alpha}(b - a)$ while the energy amount consumed in that interval is $W(b - a)$. Hence, in average the machine spends a dynamic energy amount of $W^{(\alpha-1)/\alpha}$ at time t.

Therefore, during periods where the machine speed is larger than s^c, Γ increases at rate proportionally to the one of the dynamic energy. The second inequality of the lemma follows.   ◀

▶ **Corollary A.3.** *It holds that* $\sum_j \lambda_j p_j \geq \frac{7}{8}\mathcal{E}_1^* + \frac{1}{16}\mathcal{F}^* - \frac{1}{8}\mathcal{E}_2^*$.

Proof. By the previous lemma, we deduce that

$$\sum_j \lambda_j p_j \geq \mathcal{E}_1^* \geq \frac{7}{8}\mathcal{E}_1^* + \frac{1}{8}\left(\frac{1}{2}\mathcal{F}^* - \mathcal{E}_2^*\right) = \frac{7}{8}\mathcal{E}_1^* + \frac{1}{16}\mathcal{F}^* - \frac{1}{8}\mathcal{E}_2^*.$$

◀

In the following, we show the main technical lemma.

▶ **Lemma A.4.** *Let j be an arbitrary job. Then, for every $t \geq r_j$*

$$\lambda_j - \delta_j(t - r_j) \leq \max\left\{\frac{\alpha}{\alpha - 1}W(t)^{\frac{\alpha-1}{\alpha}} + \frac{P(s^c)}{s^c}, 2\frac{P(s^c)}{s^c}\right\} \tag{7}$$

Proof. Fix a job j. We prove by induction on the number of released jobs after r_j. The base case follows Lemma A.5 and the induction step is done by Lemma A.7.   ◀

▶ **Lemma A.5.** *If no new job is released after r_j then inequality (7) holds.*

Proof. Denote the instance as $\mathcal{I}_0$. At r_j, rename jobs in non-increasing order of their densities, i.e., $p_1/w_1 \leq \ldots \leq p_n/w_n$ (note that p_a/w_a is the inverse of job a's density). Denote $W_a = w_a + \ldots + w_n$ for $1 \leq a \leq n$.

By definition of λ_j,

$$\lambda_j - \frac{P(s^c)}{s^c} = \delta_j\left[\frac{q_1(r_j)}{W_1^{1/\alpha}} + \ldots + \frac{q_j(r_j)}{W_j^{1/\alpha}}\right] + \frac{W_{j+1}}{W_j^{1/\alpha}}$$

Let $C_a(\mathcal{I}_0)$ be the completion time of job a for $1 \le a \le n$. Moreover, let ℓ be the largest job index such that $\alpha W_\ell^{(\alpha-1)/\alpha} > P(s^c)/s^c$. In other words, job ℓ is processed by speed strictly larger than s^c and the other jobs with larger index (if exist) will be processed by speed s^c. Fix a time t, let k be the pending job at t with the smallest index. We first prove the following claim.

▶ **Claim A.6.** *It holds that*

$$\lambda_j - \delta_j(t - r_j) - \frac{P(s^c)}{s^c} \le \max\left\{\frac{w_k}{W_k^{1/\alpha}} + \frac{w_{k+1}}{W_{k+1}^{1/\alpha}} + \ldots + \frac{w_n}{W_n^{1/\alpha}}, \frac{P(s^c)}{s^c}\right\}$$

Proof of claim.   We consider different cases.

Case 1: $\ell \le j$.

In this case, job j will be processed by speed s^c.

Subcase 1.1: $t \le C_\ell(\mathcal{I}_0)$. During interval $[r_j, t]$, the machine has completed jobs $1, \ldots, k-1$ and has processed a part of job k. Precisely, during $[r_j, t]$ the machine has processed $q_a(r_j)$ units of job a for every job $1 \le a < k$ and has executed $(q_k(r_j) - q_k(t))$ units of job k. Moreover, every job $1 \le a \le k$ is processed with speed $W_a^{1/\alpha}$. Therefore,

$$\lambda_j - \delta_j(t - r_j) - \frac{P(s^c)}{s^c} = \delta_j\left[\frac{q_k(t)}{W_k^{1/\alpha}} + \frac{q_{k+1}(r_j)}{W_{k+1}^{1/\alpha}} + \ldots + \frac{q_j(r_j)}{W_j^{1/\alpha}}\right] + \frac{W_{j+1}}{W_j^{1/\alpha}}$$

$$\le \delta_j\left[\frac{p_k}{W_k^{1/\alpha}} + \frac{p_{k+1}}{W_{k+1}^{1/\alpha}} + \ldots + \frac{p_j}{W_j^{1/\alpha}}\right] + \frac{W_{j+1}}{W_j^{1/\alpha}}$$

$$= \delta_j\left(\frac{w_k}{\delta_k W_k^{1/\alpha}} + \frac{w_{k+1}}{\delta_{k+1} W_{k+1}^{1/\alpha}} + \ldots + \frac{w_j}{\delta_j W_j^{1/\alpha}}\right) + \frac{W_{j+1}}{W_j^{1/\alpha}}$$

$$\le \frac{w_k}{W_k^{1/\alpha}} + \frac{w_{k+1}}{W_{k+1}^{1/\alpha}} + \ldots + \frac{w_j}{W_j^{1/\alpha}} + \frac{w_{j+1}}{W_{j+1}^{1/\alpha}} + \ldots + \frac{w_n}{W_n^{1/\alpha}}$$

$$\le \int_0^{W(t)} \frac{dz}{z^{1/\alpha}} = \frac{\alpha}{\alpha-1} W(t)^{\frac{\alpha-1}{\alpha}}$$

The first inequality is because $q_a(r_j) \le p_a$ for every job a. The first equality is due to the definition of the density. The second inequality follows since $\delta_j \le \delta_a$ for every job $a \le j$ and $W_{j+1} \ge \ldots \ge W_n$. The third inequality holds since function $z^{-1/\alpha}$ is decreasing.

Subcase 1.2: $t > C_\ell(\mathcal{I}_0)$. In this case $k > \ell$, i.e., during $[r_j, t]$ the machine i has completed jobs $1, \ldots, \ell$. Similarly as the previous subcase, we have

$$\lambda_j - \delta_j(t - r_j) - \frac{P(s^c)}{s^c} \le \sum_{a=\ell+1}^n \frac{w_a}{W_a^{1/\alpha}}$$

$$\le \int_0^{W_{\ell+1}} \frac{dz}{z^{1/\alpha}} = \frac{\alpha}{\alpha-1} W_{\ell+1}^{\frac{\alpha-1}{\alpha}} \le \frac{P(s^c)}{s^c}$$

where the last inequality follows the definition of ℓ.

Case 2: $\ell > j$.

In this case, job j will be processed with speed strictly larger than s^c.

Subcase 2.1: $t \le C_j(\mathcal{I}_0)$. The proof is done in the same manner as in Subcase 1.1.

Subcase 2.2: $t > C_j(\mathcal{I}_0)$. For simplicity, denote $C_a = C_j(\mathcal{I}_0)$. Partition time after $C_j(\mathcal{I}_0)$ as $\cup_{a=j}^{n}[C_a, C_{a+1})$. During an interval $[C_a, C_{a+1})$, the weight W_a is unchanged so to show inequality (7), it is sufficient to prove it for $t = C_j, C_{j+1}, \ldots, C_{n-1}$.

We prove again by induction. For the base case $t = C_j$, the claim inequality holds by the previous case. Assume that the inequality holds at $t = C_a$, we will prove that it holds at $t = C_{a+1}$ for $a \geq j$. We are interested only in $\tau \in [C_a, C_{a+1})$. Let $V(\tau) = w_a q_a(\tau)/p_a + w_{a+1} + \ldots + w_n$. Informally, $V(\tau)$ is the fractional weight of pending jobs at time τ.

During period $[\tau, \tau + d\tau]$ assume that the total fractional weight of pending jobs varies by $dV(\tau)$. During the same period $[\tau, \tau + d\tau]$, the total processing volume done by algorithm is at least $W_a^{1/\alpha} d\tau$ since the speed is either $W(\tau)^{1/\alpha}(= W_a^{1/\alpha})$ or s^c but in the latter, by the algorithm, $s^c \geq W(\tau)^{1/\alpha}$. Moreover, jobs a processed during $[\tau, \tau + d\tau]$ have density at most δ_j. Therefore, $dV(\tau) \leq \delta_j W_a^{1/\alpha} d\tau$. In other words, $V'(\tau)d\tau \leq \delta_j W_a^{1/\alpha} d\tau$. Taking integral, we get

$$w_a = W_a - W_{a+1} = \int_{C_a}^{C_{a+1}} V'(\tau)d\tau \leq \int_{C_a}^{C_{a+1}} \delta_j W_a^{1/\alpha} d\tau = \delta_j W_a^{1/\alpha}(C_{a+1} - C_a) \quad (8)$$

Therefore,

$$\lambda_j - \delta_j(C_{a+1} - r_j) - \frac{P(s^c)}{s^c} = \lambda_j - \delta_j(C_a - r_j) - \frac{P(s^c)}{s^c} - \delta_j(C_{a+1} - C_a)$$

$$\leq \max\left\{\frac{w_a}{W_a^{1/\alpha}} + \ldots + \frac{w_n}{W_n^{1/\alpha}}, \frac{P(s^c)}{s^c}\right\} - \frac{w_a}{W_a^{1/\alpha}}$$

$$\leq \max\left\{\frac{w_{a+1}}{W_{a+1}^{1/\alpha}} + \ldots + \frac{w_n}{W_n^{1/\alpha}}, \frac{P(s^c)}{s^c}\right\}$$

where the first inequality is due to the induction hypothesis and inequality (8). Combining all the cases, the claim holds. ◀

Using the claim, the lemma follows immediately as shown below.

$$\lambda_j - \delta_j(t - r_j) - \frac{P(s^c)}{s^c} \leq \max\left\{\frac{w_k}{W_k^{1/\alpha}} + \ldots + \frac{w_n}{W_n^{1/\alpha}}, \frac{P(s^c)}{s^c}\right\}$$

$$\leq \max\left\{\int_0^{W(t)} \frac{dz}{z^{1/\alpha}}, \frac{P(s^c)}{s^c}\right\} = \max\left\{\frac{\alpha}{\alpha - 1}W(t)^{\frac{\alpha-1}{\alpha}}, \frac{P(s^c)}{s^c}\right\}$$

where the inequality holds since function $z^{-1/\alpha}$ is decreasing. (Recall that k is the pending job at time t with the smallest index.) ◀

▶ **Lemma A.7.** *Assume that inequality (7) holds if there are $(n-1)$ jobs released after r_j. Then the inequality also holds if n jobs are released after r_j.*

Proof. Denote the instance as $\mathcal{I}_n$. Among such jobs, let n be the last released one (at time r_n). By induction hypothesis, it remains to prove the lemma inequality for $t \geq r_n$.

We first show the claim that inequality (7) holds for any time $t \geq C_j(\mathcal{I}_n)$ by a similar argument as in Subcase 2.2 of the previous claim. Indeed, we prove the claim by fixing the processing volume of job n and varying its weight w_n. Note that $C_j(\mathcal{I}_n)$ depends on w_n and when w_n is varied, $C_j(\mathcal{I}_n)$ is also varied. However, with a fixed value of w_n, $C_j(\mathcal{I}_n)$ is fixed and we are interested only in $t \geq C_j(\mathcal{I}_n)$. If $w_n = 0$ then the claim follows the induction hypothesis (the instance becomes the one with $(n-1)$ jobs). Assume that the

claim holds for some value w_n. Now increase an arbitrarily small amount of w_n and consider a time $t \geq C_j(\mathcal{I}_n)$ (corresponding to the current value of w_n). Due to that increase, during period $[t, t + dt]$ the total fractional weight of pending jobs varies by $dV(t)$. During the same period $[t, t + dt]$, the total processing volume done by algorithm is at least $V(t)^{1/\alpha} dt$ since the machine speed is at least $W(t)^{1/\alpha}$. Moreover, jobs processed during $[t, t + dt]$ have density at most δ_j. Therefore, $dV(t) \leq \delta_j V(t)^{1/\alpha} dt$. So

$$\frac{\alpha}{\alpha - 1} dW(t)^{(\alpha-1)/\alpha} = \frac{dW(t)}{W(t)^{1/\alpha}} \leq \delta_j dt$$

This inequality means that in the lemma inequality (7), the decrease in the left-hand side is larger than that in the right-hand side while varying the weight of job n. Hence, the inequality holds for $t \geq C_j(\mathcal{I}_n)$.

Now we consider instance $\mathcal{I}_n$ with fixed parameters for job n. We will prove the lemma for $t < C_j(\mathcal{I}_n)$. Denote $t_0 = C_j(\mathcal{I}_n)$. Again, rename jobs in non-increasing order of their densities at time r_n. (After r_n, no new job is released and the relative order of jobs is unchanged.) Let W_a be the total weight of pending jobs at r_n which have density smaller than δ_a. Recall that the total weight of pending jobs at time t is $W(t)$.

Let k be the pending job with the smallest index at time t in the instance $\mathcal{I}_n$. During $[t, t_0]$, the machine processes (a part) of job k, jobs $k + 1, \ldots, j$. These jobs have density at least δ_j. We deduce

$$\begin{aligned}
\lambda_j - \delta_j(t - r_j) - \frac{P(s^c)}{s^c} &= \lambda_j - \delta_j(t_0 - r_j) - \frac{P(s^c)}{s^c} + \delta_j(t_0 - t) \\
&\leq \frac{\alpha}{\alpha - 1} W(t_0)^{\frac{\alpha-1}{\alpha}} + \delta_j(t_0 - t) \\
&\leq \frac{\alpha}{\alpha - 1} W(t_0)^{\frac{\alpha-1}{\alpha}} + \delta_j \left(\frac{q_k(t)}{W_k^{1/\alpha}} + \frac{q_{k+1}(r_n)}{W_{k+1}^{1/\alpha}} + \ldots + \frac{q_j(r_n)}{W_j^{1/\alpha}} \right) \\
&\leq \frac{\alpha}{\alpha - 1} W(t_0)^{\frac{\alpha-1}{\alpha}} + \int_{W_{j+1}}^{W_k} \frac{dz}{z^{1/\alpha}} \\
&\leq \frac{\alpha}{\alpha - 1} W_k^{\frac{\alpha-1}{\alpha}} = \frac{\alpha}{\alpha - 1} W(t)^{\frac{\alpha-1}{\alpha}}
\end{aligned}$$

The first inequality follows the previous claim, stating that inequality (7) holds for $t \geq t_0$. The second inequality follows the fact that at any time the speed of the machine is at least $W(t)^{1/\alpha}$. The third inequality holds since $\delta_k \leq \delta_{k+1} \leq \ldots \leq \delta_j$ and function $z^{-1/\alpha}$ is decreasing. The last inequality holds since $W(t_0) = W_{j+1}$ and $W_k = W(t)$. ◀

▶ **Theorem A.8.** *The algorithm is $O(\alpha/\ln \alpha)$-competitive.*

Proof. Recall that the dual has value at least $\min L(s, C, F, \lambda)$ where the minimum is taken over (s, C, F) feasible solution of the primal. The goal is to lower bound the Lagrangian function.

$$\begin{aligned}
L(s, C, F, \lambda) = &\sum_j \lambda_j p_j + A \int_0^\infty |F'(t)| dt + \sum_j \int_{r_j}^{C_j} \delta_j(C_j - r_j) s_j(t) F(t) dt \\
&- \sum_{i,j} \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2\frac{P(s(t))}{s(t)} - \delta_j(C_j - r_j) \right) \mathbb{1}_{\{s(t) > 0\}} dt
\end{aligned} \quad (9)$$

for any feasible primal solution (s, C, F).

Fix a feasible primal solution (s, C, F). Define $L_1(s, C, F, \lambda)$ as

$$\sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2 \frac{P(s(t))}{s(t)} - \delta_j (C_j - r_j) \right) \mathbb{1}_{\{s(t) > 0\}} dt$$

▸ **Claim A.9.** *Let (s, C, F) be an arbitrary feasible solution of the primal. Then,*

$$L_1(x, s, C, F, \lambda) \le \frac{1}{(\alpha - 1)^{\frac{1}{\alpha - 1}}} \mathcal{F}^* - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t) > 0\}} dt - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t) > 0\}} dt$$

▸ **Claim A.10.** *Let (s, C, F) be an arbitrary feasible solution of the primal. Define*

$$L_2(F) := \sum_j \int_{r_j}^{C_j} \delta_j (C_j - r_j) s_j(t) F(t) dt + A \int_0^\infty |F'(t)| dt$$

$$+ \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t) > 0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t) > 0\}} dt$$

Then, $L_2(F) \ge (\mathcal{E}_2^* + \mathcal{E}_3^*)/4$.

We first show how to prove the theorem assuming the claims. By (9), we have

$$L(s, C, F, \lambda) \ge \sum_j \lambda_j p_j + A \int_0^\infty |F'(t)| dt + \sum_j \int_{r_j}^{C_j} \delta_j (C_j - r_j) s_j(t) F(t) dt$$

$$- \sum \frac{1}{(\alpha - 1)^{1/(\alpha - 1)}} \mathcal{F}^* + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t) > 0\}} dt + \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t) > 0\}} dt$$

$$\ge \sum_j \lambda_j p_j - \frac{1}{(\alpha - 1)^{1/(\alpha - 1)}} \mathcal{F}^* + \frac{1}{4} \mathcal{E}_2^* + \frac{1}{4} \mathcal{E}_3^*$$

$$\ge \left(1 - \frac{1}{(\alpha - 1)^{1/(\alpha - 1)}} \right) \left(\frac{7}{8} \mathcal{E}_1^* + \frac{1}{16} \mathcal{F}^* - \frac{1}{8} \mathcal{E}_2^* \right) + \frac{1}{4} \mathcal{E}_2^* + \frac{1}{4} \mathcal{E}_3^*$$

$$\ge \left(1 - \frac{1}{(\alpha - 1)^{1/(\alpha - 1)}} \right) \left(\frac{7}{8} \mathcal{E}_1^* + \frac{1}{16} \mathcal{F}^* \right) + \frac{1}{8} \mathcal{E}_2^* + \frac{1}{4} \mathcal{E}_3^*$$

$$\ge \frac{\ln(\alpha - 1)}{2(\alpha - 1)} \left(\frac{7}{8} \mathcal{E}_1^* + \frac{1}{16} \mathcal{F}^* \right) + \frac{1}{8} \mathcal{E}_2^* + \frac{1}{4} \mathcal{E}_3^*$$

where the first and second inequalities are due to Claim A.9 and Claim A.10, respectively. The third inequality follows Corollary A.3 and $\sum_j \lambda_j p_j \ge \mathcal{F}^*$. The last inequality is due to the fact that $2 \ge (\alpha - 1)^{\frac{1}{\alpha - 1}} \ge 1 + \frac{\ln(\alpha - 1)}{\alpha - 1}$ for every $\alpha > 2$.

Besides, the primal objective is at most $2(\mathcal{F}^* + \mathcal{E}_1^* + \mathcal{E}_2^* + \mathcal{E}_3^*)$. Hence, the competitive ratio is $O(\alpha / \ln \alpha)$.

In the rest, we prove the claims.

▸ **Claim A.9.** *Let (s, C, F) be an arbitrary feasible solution of the primal. Then,*

$$L_1(s, C, F, \lambda) \le \frac{1}{(\alpha - 1)^{\frac{1}{\alpha - 1}}} \mathcal{F}^* - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t) > 0\}} dt - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t) > 0\}} dt$$

Proof of claim.   We have

$$L_1(s, C, F, \lambda) := \sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2\frac{P(s(t))}{s(t)} - \delta_j(C_j - r_j) \right) \mathbb{1}_{\{s(t)>0\}} dt$$

$$\leq \sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2\frac{P(s(t))}{s(t)} - \delta_j(t - r_j) \right) \mathbb{1}_{\{s(t)>0\}} dt$$

where the inequality holds because the integral for each job j is taken over $r_j \leq t \leq C_j$.

Let T be the set of time t such that $\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} \leq \frac{P(s^c)}{s^c}$. Then by Lemma A.4, for any $t \in T$

$$\lambda_j - \delta_j(t - r_j) \leq 2\frac{P(s^c)}{s^c}$$

Therefore,

$$\sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2\frac{P(s(t))}{s(t)} - \delta_j(t - r_j) \right) \mathbb{1}_{\{t \in T\}} dt \leq 0 \qquad (10)$$

since $s^c = \arg\min_{z \geq 0} P(z)/z$. Hence,

$$L_1(s, C, F, \lambda)$$

$$\leq \sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\lambda_j - 2\frac{P(s(t))}{s(t)} - \delta_j(t - r_j) \right) \mathbb{1}_{\{s(t)>0\}} \mathbb{1}_{\{F(t)>0\}} \mathbb{1}_{\{t \notin T\}} dt$$

$$\leq \sum_j \int_{r_j}^{C_j} s_j(t) F(t) \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} - \frac{P(s(t))}{s(t)} \right) \mathbb{1}_{\{s(t)>0\}} \mathbb{1}_{\{F(t)>0\}} \mathbb{1}_{\{t \notin T\}} dt$$

$$= \int_0^\infty s(t) \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} - \frac{P(s(t))}{s(t)} \right) \mathbb{1}_{\{s(t)>0\}} \mathbb{1}_{\{F(t)>0\}} \mathbb{1}_{\{t \notin T\}} dt$$

$$\leq \int_0^\infty \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} \bar{s}(t) - P(\bar{s}(t)) \right) \mathbb{1}_{\{s(t)>0\}} \mathbb{1}_{\{F(t)>0\}} \mathbb{1}_{\{t \notin T\}} dt$$

$$\leq \int_0^\infty \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} \bar{s}(t) - P(\bar{s}(t)) \right) \mathbb{1}_{\{F(t)>0\}} \mathbb{1}_{\{s^*(t)>0\}} dt$$

$$\leq \frac{1}{2} \int_0^\infty \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} \bar{s}(t) - P(\bar{s}(t)) \right) \mathbb{1}_{\{F(t)>0\}} dt$$

$$+ \frac{1}{2} \int_0^\infty \left(\frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}} \bar{s}(t) - P(\bar{s}(t)) \right) \mathbb{1}_{\{s^*(t)>0\}} dt$$

The first inequality is due to (10) and note that if $F(t) = 0$ then the contribution of the term inside the integral is 0. The second inequality follows Lemma A.4 and recall that $s^c = \arg\min_{z \geq 0} P(z)/z$. The equality is because $\sum_j s_j(t) F(t) = \sum_j s_j(t) = s(t)$ for t such that $F(t) > 0$ (meaning that $F(t) = 1$). The third inequality is due to the first order derivative and $\bar{s}(t)$ is the solution of $P'(z) = \frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}}$. The fourth inequality holds since the term inside the integral is non-negative and $\{t : t \notin T\} \subset \{t : s^*(t) > s^c\} \subset \{t : s^*(t) > 0\}$.

Recall that $P(z) = z^\alpha + g$. Replacing $\bar{s}(t) = (\alpha-1)^{-1/(\alpha-1)} W(t)^{1/\alpha}$ (solution of $P'(z) = \frac{\alpha}{(\alpha-1)} W(t)^{\frac{\alpha-1}{\alpha}}$), we get:

$$L_1(s, C, F, \lambda)$$

$$\leq \frac{\alpha-1}{(\alpha-1)^{\frac{\alpha}{\alpha-1}}} \int_0^\infty W(t) - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)>0\}} dt - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt$$

$$= \frac{1}{(\alpha-1)^{\frac{1}{\alpha-1}}} \mathcal{F}^* - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{F(t)>0\}} dt - \frac{1}{2} \int_0^\infty g \mathbb{1}_{\{s^*(t)>0\}} dt$$

◀

▶ **Claim A.10.** *Let (s, C, F) be an arbitrary feasible solution of the primal. Define*

$$L_2(F) := \sum_j \int_{r_j}^{C_j} \delta_j(C_j - r_j)s_j(t)F(t)dt + A \int_0^\infty |F'(t)|dt$$

$$+ \frac{1}{2} \int_0^\infty g\mathbb{1}_{\{F(t)>0\}}dt + \frac{1}{2} \int_0^\infty g\mathbb{1}_{\{s^*(t)>0\}}dt$$

Then, $L_2(F) \geq (\mathcal{E}_2^* + \mathcal{E}_3^*)/4$.

Proof of claim.   Consider the algorithm schedule. An *end-time u* is a moment in the algorithm schedule such that the machine switches from the idle state to the sleep state. Conventionally, the first end-time in the schedule is 0. Partition the time line into phases. A *phase $[u, v)$* is a time interval such that u, v are two consecutive end-times. Observe that in a phase, the schedule has transition cost A and some new job is released in a phase (otherwise the machine would not switch to non-sleep state). We will prove the claim on every phase. In the following, we are only interested in phase $[u, v)$ and define

$$L_2(F)\big|_{[u,v)} := \sum_{j:u \leq r_j < v} \int_{r_j}^{C_j} \delta_j(C_j - r_j)s_j(t)F(t)dt + A \int_u^v |F'(t)|dt$$

$$+ \frac{1}{2} \int_u^v g\mathbb{1}_{\{F(t)>0\}}dt + \frac{1}{2} \int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt$$

By the algorithm, the static energy on machine i during its idle time is A, i.e., $\int_u^v g\mathbb{1}_{\{s^*(t)=0\}}dt = A$. If during $[u, v)$, the schedule induced by solution (s, C, F) makes a transition from non-sleep state to sleep state or inversely then $L_2(F)\big|_{[u,v)} \geq \frac{1}{2} \int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + A$. Hence

$$L_2(F)\big|_{[u,v)} \geq \frac{1}{2}\left(\int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + \int_u^v g\mathbb{1}_{\{s^*(t)=0\}}dt + A\right) = \frac{1}{2}\mathcal{E}_2^*\big|_{[u,v)} + \frac{1}{2}\mathcal{E}_3^*\big|_{[u,v)}.$$

If during $[u, v)$, the schedule induced by solution (x, s, C, F) makes no transition (from non-sleep static to sleep state or inversely) then either $F(t) = 1$ or $F(t) = 0$ for every $t \in [u, v]$. Note that by definition of phases, some job is released during $[u, v)$. We consider cases.

Case 1: $F(t) = 1 \ \forall u \leq t \leq v$.

Then,

$$L_2(F)\big|_{[u,v)}$$
$$\geq \frac{1}{2} \int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + \frac{1}{2} \int_u^v g\mathbb{1}_{\{F(t)=1\}}dt = \frac{1}{2} \int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + \frac{1}{2} \int_u^v gdt$$
$$\geq \frac{1}{2} \int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + \frac{1}{4} \int_u^v g\mathbb{1}_{\{s^*(t)=0\}}dt + \frac{A}{4}$$
$$\geq \frac{1}{4}\left(\int_u^v g\mathbb{1}_{\{s^*(t)>0\}}dt + \int_u^v g\mathbb{1}_{\{s^*(t)=0\}}dt + A\right) = \frac{1}{4}\mathcal{E}_2^*\big|_{[u,v)} + \frac{1}{4}\mathcal{E}_3^*\big|_{[u,v)}$$

where the second inequality follows since the total idle duration in $[u, v)$ incurs a cost A (so the machine switches to sleep state at time v).

Case 2: $F(t) = 0 \ \forall u \leq t \leq v$.

As the machine is in the sleep state during $[u, v)$ in solution (s, C, F), all jobs released in $[u, v)$ will be completed later than v (in the solution s, C, F). However, as at time v the algorithm sets the machine in the sleep state, those jobs are completed before v by the algorithm. Note that the algorithm always balances the total weighted flow-time to the dynamic energy (in case the machine speed is strictly larger than s^c) or the static energy plus dynamic energy (in case the machine speed is s^c). Therefore, the total weighted flow-time of jobs released in $[u, v)$ by solution (s, C, F) is at least the static energy of the algorithm during $[u, v)$. In other words,

$$
\begin{aligned}
L_2(F)\big|_{[u,v)} &\geq \sum_{j: u \leq r_j < v} \int_{r_j}^{C_j} \delta_j(C_j - r_j) s_j(t) F(t) dt + \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t) > 0\}} dt \\
&\geq \int_u^v g \mathbb{1}_{\{s^*(t) = 0\}} dt + \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t) > 0\}} dt \\
&\geq \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t) = 0\}} dt + \frac{1}{2} \int_u^v g \mathbb{1}_{\{s^*(t) > 0\}} dt + \frac{A}{2} = \frac{1}{2} \mathcal{E}_2^*\big|_{[u,v)} + \frac{1}{2} \mathcal{E}_3^*\big|_{[u,v)}
\end{aligned}
$$

where the third inequality is again due to the fact that the total idle duration in $[u, v)$ incurs a static energy at least A.   ◀

    The proofs of the claims complete the theorem proof.   ◀