HEAT AND ENTROPY FLOWS IN CARNOT GROUPS

LUIGI AMBROSIO AND GIORGIO STEFANI

ABSTRACT. We prove the correspondence between the solutions of the sub-elliptic heat equation in a Carnot group $\mathbb{G}$ and the gradient flows of the relative entropy functional in the Wasserstein space of probability measures on $\mathbb{G}$. Our result completely answers a question left open in a previous paper by N. Juillet, where the same correspondence was proved for $\mathbb{G} = \mathbb{H}^n$, the n-dimensional Heisenberg group.

1. INTRODUCTION

Since the pioneering works [23, 35] and the monograph [4], in the last twenty years there has been an increasing interest in the study of the relation between evolution equations and gradient flows of energy functionals in a large variety of different frameworks, see [5, 7, 13–16, 19, 21, 22, 26, 30, 33, 34, 38, 40].

The prominent case in the literature is represented by the connection between the heat equation and the relative entropy functional. It is well-known that the heat equation

$$\begin{cases} \partial_t u_t = \Delta u_t & \text{in } (0, +\infty) \times \mathbb{R}^n, \\ u_0 = \bar{u} \in L^2(\mathbb{R}^n) & \text{on } \{0\} \times \mathbb{R}^n, \end{cases} \tag{HE}$$

can be seen as the gradient flow in $L^2(\mathbb{R}^n)$ of the Dirichlet energy $\mathsf{D}(u) = \int_{\mathbb{R}^n} |\nabla u|^2 \, dx$ accordingly to the general approach introduced in [11]. If the initial datum $\bar{u} \in L^2(\mathbb{R}^n)$ is such that $\mu_0 = \bar{u} \, \mathscr{L}^n \in \mathscr{P}_2(\mathbb{R}^n)$, where

$$\mathscr{P}_2(\mathbb{R}^n) = \left\{ \mu \in \mathscr{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} |x|^2 \, d\mu(x) < +\infty \right\},$$

then the solution $(u_t)_{t \geq 0}$ of (HE) induces a curve $(\mu_t)_{t \geq 0} \subset \mathscr{P}_2(\mathbb{R}^n)$, $\mu_t = u_t \mathscr{L}^n$. If we endow the set $\mathscr{P}_2(\mathbb{R}^n)$ with its usual Wasserstein distance W, then the curve $(\mu_t)_{t \geq 0}$ is locally absolutely continuous with locally integrable squared W-derivative in $[0, +\infty)$. On the one hand, since $(u_t)_{t \geq 0}$ satisfies (HE), the curve $(\mu_t)_{t \geq 0}$ naturally solves in the weak sense the continuity equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0 & \text{in } (0, +\infty) \times \mathbb{R}^n, \\ \mu_0 = \bar{u} \, \mathscr{L}^n & \text{on } \{0\} \times \mathbb{R}^n, \end{cases} \tag{CE}$$

where the velocity vector field $(v_t)_{t \geq 0}$ is given by $v_t = -\nabla u_t / u_t$. On the other hand, the *relative entropy*

$$\mathsf{Ent}(\mu) = \int_{\mathbb{R}^n} u \log u \, dx, \qquad \text{for } \mu = u \, \mathscr{L}^n \in \mathscr{P}_2(\mathbb{R}^n),$$

computed along a curve $(\mu_t)_{t \geq 0} \subset \mathscr{P}_2(\mathbb{R}^n)$ satisfying (**CE**) for a given velocity field $(v_t)_{t \geq 0}$ is such that

$$\frac{d}{dt}\mathsf{Ent}(\mu_t) = - \int_{\mathbb{R}^n} (\log u_t + 1) \ \mathrm{div}(v_t u_t) \ dx = \int_{\mathbb{R}^n} \left\langle v_t, \frac{\nabla u_t}{u_t} \right\rangle d\mu_t.$$

In analogy with the Hilbertian case, but using Otto's calculus [35] in the interpretation of the right hand side, one says that $(\mu_t)_{t \geq 0}$ is a gradient flow of the entropy in $(\mathscr{P}_2(\mathbb{R}^n), \mathsf{W})$ if and only if the curve $t \mapsto \mathsf{Ent}(\mu_t)$ has maximal dissipation rate. This happens if and only if $v_t = -\nabla u_t / u_t$, i.e. when $(u_t)_{t \geq 0}$ satisfies (**HE**).

Although not fully rigorous, the argument presented above contains all the key tools needed to establish the correspondence between the heat flow and the entropy flow in a general metric measure space $(X, \mathsf{d}, \mathfrak{m})$. Both the heat equation (**HE**) and the continuity equation (**CE**) have been adequately understood in this general context. For (**HE**), one relaxes the Dirichlet energy to the so-called *Cheeger energy*

$$\mathsf{Ch}(u) = \inf \left\{ \liminf_n \int_X |\mathrm{D}u_n|^2 \ d\mathfrak{m} : u_n \to u \text{ in } L^2(X, \mathsf{d}, \mathfrak{m}), \ u_n \in \mathrm{Lip}(X) \right\},$$

where the *local Lipschitz constant* $|\mathrm{D}u|(x) = \limsup_{y \to x} \frac{|u(y) - u(x)|}{\mathsf{d}(x,y)}$ of $u \in \mathrm{Lip}(X)$ plays the same role of the absolute value of the gradient in $\mathbb{R}^n$. It can be shown that the naturally associated Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ is a Banach space (not Hilbertian in general) and that the functional Ch is convex, so that (**HE**) can still be interpreted as its gradient flow in the Hilbert space $L^2(X, \mathsf{d}, \mathfrak{m})$, see [5]. For (**CE**), one introduces an appropriate space $\mathsf{S}^2(X)$ of test functions in $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ and says that $(\mu_t)_{t > 0}$ satisfies the continuity equation with respect to a family of maps $(L_t)_{t > 0} \colon \mathsf{S}^2(X) \to \mathbb{R}$ if $t \mapsto \int_X f \ d\mu_t$ is absolutely continuous for every $f \in \mathsf{S}^2(X)$ with $\frac{d}{dt} \int_X f \ d\mu_t = L_t(f)$ for a.e. $t > 0$, see [20].

The notion of gradient flow of the entropy functional

$$\mathsf{Ent}_{\mathfrak{m}}(\mu) = \int_X \varrho \log \varrho \ d\mathfrak{m}, \qquad \text{for } \mu = \varrho \mathfrak{m} \in \mathscr{P}_2(X),$$

in the Wasserstein space $(\mathscr{P}_2(X), \mathsf{W})$ can be rigorously defined by requiring the validity of the following sharp *energy dissipation inequality*

$$\mathsf{Ent}_{\mathfrak{m}}(\mu_t) + \frac{1}{2} \int_s^t |\dot{\mu}_r|^2 \ dr + \frac{1}{2} \int_s^t |\mathrm{D}^- \mathsf{Ent}_{\mathfrak{m}}|^2(\mu_r) \ dr \leq \mathsf{Ent}_{\mathfrak{m}}(\mu_s)$$

for all $0 \leq s \leq t$, where $t \mapsto |\dot{\mu}_t|$ is the W-derivative of the curve $(\mu_t)_{t > 0} \subset \mathscr{P}_2(X)$ and

$$|\mathrm{D}^- \mathsf{Ent}_{\mathfrak{m}}|(\mu) = \limsup_{\nu \to \mu} \max \left\{ \frac{\mathsf{Ent}_{\mathfrak{m}}(\mu) - \mathsf{Ent}_{\mathfrak{m}}(\nu)}{\mathsf{W}(\mu, \nu)}, 0 \right\}$$

is the so-called *descending slope* of the entropy. Note that this definition is consistent with the standard one in Hilbert spaces, since $u'(t) = -\nabla E(u(t))$ is equivalent to

$$\frac{1}{2}|u'|^2(t) + \frac{1}{2}|\nabla E(u(t))|^2 \leq -\frac{d}{dt} E(u(t))$$

by combining the chain rule with Cauchy–Schwarz and Young's inequalities.

As pointed out in [5, 19], this abstract approach provides a complete equivalence between the two gradient flows if the entropy is *K-convex along geodesics* in $(\mathscr{P}_2(X), \mathsf{W})$ for some

$K \in \mathbb{R}$, that is, if

$$\mathsf{Ent}_\mathfrak{m}(\mu_t) \leq (1-t)\mathsf{Ent}_\mathfrak{m}(\mu_0) + t\,\mathsf{Ent}_\mathfrak{m}(\mu_1) - \frac{K}{2}t(1-t)\mathsf{W}(\mu_0,\mu_1)^2, \qquad t \in [0,1], \qquad \textbf{(K)}$$

holds for a class of constant speed geodesics $(\mu_t)_{t\in[0,1]} \subset \mathscr{P}_2(X)$ sufficiently large to join any pair of points in $\mathscr{P}_2(X)$. The K-convexity (also known as *displacement convexity*) of the entropy heavily depends on the structure of $(X, \mathsf{d}, \mathfrak{m})$ and encodes a precise information about the ambient space: if X is a Riemannian manifold, then $\textbf{(K)}$ is valid if and only if the Ricci curvature satisfies $\mathrm{Ric} \geq K$, see [37]. For this reason, if property $\textbf{(K)}$ holds, then $(X, \mathsf{d}, \mathfrak{m})$ is called a space with *generalized Ricci curvature* bounded from below, or simply a $CD(K, \infty)$ space.

According to this general framework, the correspondence between heat flow and entropy flow has been proved on Riemannian manifolds with Ricci curvature bounded from below, see [14], and on compact *Alexandrov spaces*, see [21, 22, 33]. Alexandrov spaces are considered as metric measure spaces with *generalized sectional curvature* bounded from below (a condition stronger than $\textbf{(K)}$, see [36]).

If $(X, \mathsf{d}, \mathfrak{m})$ is not a $CD(K, \infty)$ space, then the picture is less clear. As stated in [5, Theorem 8.5], the correspondence between heat flow and entropy flow still holds if the descending slope $|\mathsf{D}^-\mathsf{Ent}_\mathfrak{m}|$ of the entropy is an upper gradient of $\mathsf{Ent}_\mathfrak{m}$ and satisfies a precise lower semicontinuity property, basically equivalent to the equality between $|\mathsf{D}^-\mathsf{Ent}_\mathfrak{m}|$ and the so-called *Fisher information*. These assumptions are weaker than $\textbf{(K)}$ but not easy to check for a given non-$CD(K, \infty)$ space.

In [24, 25], it was proved that the *Heisenberg group* $\mathbb{H}^n$ is a non-$CD(K, \infty)$ space in which nevertheless the correspondence between heat flow and gradient flow holds. The Heisenberg group is the simplest non-commutative *Carnot group*. Carnot groups are one of the most studied examples of Carnot–Carathéodory spaces, see [10, 27, 32] and the references therein for an account on this subject. The proof of the correspondence of the two flows in $\mathbb{H}^n$ presented in [25] essentially splits into two parts. The first part shows that a solution of the sub-elliptic heat equation $\partial_t u_t + \Delta_{\mathbb{H}^n} u_t = 0$ corresponds to a gradient flow of the entropy in the Wasserstein space $(\mathscr{P}_2(\mathbb{H}^n), \mathsf{W}_{\mathbb{H}^n})$ induced by the Carnot-Carathéodory distance d_{cc}. The direct computations needed are justified by some precise estimates on the *sub-elliptic heat kernel* in $\mathbb{H}^n$ given in [28, 29]. The second part proves that a gradient flow of the entropy in $(\mathscr{P}_2(\mathbb{H}^n), \mathsf{W}_{\mathbb{H}^n})$ induces a sub-elliptic heat diffusion in $\mathbb{H}^n$. The argument is based on a clever regularization of the gradient flow $(\mu_t)_{t\geq 0}$ based on the particular structure of the Lie algebra of $\mathbb{H}^n$.

An open question arisen in [25, Remark 5.3] was to extend the correspondence of the two flows to any Carnot group. The aim of the present work is to give a positive answer to this problem. To prove that a solution of the sub-elliptic heat equation corresponds to a gradient flow of the entropy, we essentially follow the same strategy of [25]. Since the results of [28, 29] are not known for a general Carnot group, we instead rely on the weaker estimates given in [39] valid in any nilpotent Lie group. To show that a gradient flow of the entropy induces a sub-elliptic heat diffusion, we regularize the gradient flow $(\mu_t)_{t\geq 0}$ both in time and space via convolution with smooth kernels. This regularization does not depend on the structure of the Lie algebra of the group, but nevertheless allows us to preserve the key quantities involved, such as the *continuity equation* and the *Fisher information*. In the presentation of the proofs, we also take advantage of a few results

taken from the general setting of metric measures spaces developed in [5] and in the references therein.

The paper is organized as follows. In Section 2 we collect the standard definitions and well-known facts that are used throughout the work. The precise statement of our main result is given in Theorem 2.4 at the end of this part. In Section 3 we extend the technical results presented in [25, Sections 3 and 4] to any Carnot group with minor modifications and we prove that Carnot groups are non-$CD(K, \infty)$ spaces (see Proposition 3.6), generalizing the analogous result obtained in [24]. Finally, in Section 4, we prove the correspondence of the two flows.

2. Preliminaries

2.1. AC curves, entropy and gradient flows.
Let (X, d) be a metric space, let $I \subset \mathbb{R}$ be a closed interval and let $p \in [1, +\infty]$. We say that a curve $\gamma \colon I \to X$ belongs to $AC^p(I; (X, d))$ if

$$\mathsf{d}(\gamma_s, \gamma_t) \leq \int_s^t g(r) \, dr \qquad s, t \in I, \ s < t, \tag{2.1}$$

for some $g \in L^p(I)$. The space $AC^p_{\mathrm{loc}}(I; (X, d))$ is defined analogously. The case $p = 1$ corresponds to *absolutely continuous curves* and is simply denoted by $AC(I; (X, d))$. It turns out that, if $\gamma \in AC^p(I; (X, d))$, there is a minimal function $g \in L^p(I)$ satisfying (2.1), called *metric derivative* of the curve γ, which is given by

$$|\dot{\gamma}_t| := \lim_{s \to t} \frac{d(\gamma_s, \gamma_t)}{|s - t|} \qquad \text{for a.e. } t \in I.$$

See [4, Theorem 1.1.2] for the simple proof. We call (X, d) a *geodesic* metric space if for every $x, y \in X$ there exists a curve $\gamma \colon [0, 1] \to X$ such that $\gamma(0) = x$, $\gamma(1) = y$ and

$$\mathsf{d}(\gamma_s, \gamma_t) = |s - t| \, \mathsf{d}(\gamma_0, \gamma_1) \qquad \forall s, t \in [0, 1].$$

Let $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$ and let $f \colon X \to \mathbb{R}^*$ be a function. We define the *effective domain* of f as

$$\mathrm{Dom}(f) := \{x \in X : f(x) \in \mathbb{R}\}.$$

Given $x \in \mathrm{Dom}(f)$, we define the *local Lipschitz constant* of f at x by

$$|\mathrm{D}f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{\mathsf{d}(x, y)}.$$

The *descending slope* and the *ascending slope* of f at x are respectively given by

$$|\mathrm{D}^- f|(x) := \limsup_{y \to x} \frac{[f(y) - f(x)]^-}{\mathsf{d}(x, y)}, \qquad |\mathrm{D}^+ f|(x) := \limsup_{y \to x} \frac{[f(y) - f(x)]^+}{\mathsf{d}(x, y)}.$$

Here a^+ and a^- denote the positive and negative part of $a \in \mathbb{R}$ respectively. When $x \in \mathrm{Dom}(f)$ is an isolated point of X, we set $|\mathrm{D}f|(x) = |\mathrm{D}^- f|(x) = |\mathrm{D}^+ f|(x) = 0$. By convention, we set $|\mathrm{D}f|(x) = |\mathrm{D}^- f|(x) = |\mathrm{D}^+ f|(x) = +\infty$ for all $x \in X \setminus \mathrm{Dom}(f)$.

Definition 2.1 (Gradient flow). Let $E \colon X \to \mathbb{R} \cup \{+\infty\}$ be a function. We say that a curve $\gamma \in AC_{\mathrm{loc}}([0, +\infty); (X, \mathsf{d}))$ is a *(metric) gradient flow* of E starting from $\gamma_0 \in \mathrm{Dom}(E)$ if the *energy dissipation inequality (EDI)*

$$E(\gamma_t) + \frac{1}{2} \int_s^t |\dot{\gamma}_r|^2 \, dr + \frac{1}{2} \int_s^t |\mathrm{D}^- E|^2(\gamma_r) \, dr \leq E(\gamma_s) \tag{2.2}$$

holds for all $s, t \geq 0$ with $s \leq t$.

Note that, if $(\gamma_t)_{t \geq 0}$ is a gradient flow of E, then $\gamma_t \in \text{Dom}(E)$ for all $t \geq 0$ and $\gamma \in AC^2_{\text{loc}}([0, +\infty); (X, \mathsf{d}))$ with $t \mapsto |\mathrm{D}^- E|(\gamma_t) \in L^2_{\text{loc}}([0, +\infty))$. Moreover, the function $t \mapsto E(\gamma_t)$ is non-increasing on $[0, +\infty)$ and thus a.e. differentiable and locally integrable.

Remark 2.2. As observed in [5, Section 2.5], if the function $t \mapsto E(\gamma_t)$ is locally absolutely continuous on $(0, +\infty)$, then (2.2) holds as an equality by the chain rule and Young's inequality. In this case, (2.2) is also equivalent to

$$\frac{d}{dt} E(\gamma_t) = -|\dot{\gamma}_t|^2 = -|\mathrm{D}^- E|^2(\gamma_t) \qquad \text{for a.e. } t > 0.$$

2.2. **Wasserstein space.** We now briefly recall some properties of the Wasserstein space needed for our purposes. For a more detailed introduction to this topic, we refer the interested reader to [2, Section 3].

Let (X, d) be a Polish space, i.e. a complete and separable metric space. We denote by $\mathscr{P}(X)$ the set of probability Borel measures on X. The *Wasserstein distance* W between $\mu, \nu \in \mathscr{P}(X)$ is given by

$$\mathsf{W}(\mu, \nu)^2 = \inf\left\{ \int_{X \times X} \mathsf{d}(x, y)^2 \, d\pi : \pi \in \Gamma(\mu, \nu) \right\}, \tag{2.3}$$

where

$$\Gamma(\mu, \nu) := \{ \pi \in \mathscr{P}(X \times X) : (p_1)_\# \pi = \mu, \ (p_2)_\# \pi = \nu \}. \tag{2.4}$$

Here $p_i \colon X \times X \to X$, $i = 1, 2$, are the the canonical projections on the components. As usual, if $\mu \in \mathscr{P}(X)$ and $T \colon X \to Y$ is a μ-measurable map with values in the topological space Y, the *push-forward measure* $T_\#(\mu) \in \mathscr{P}(Y)$ is defined by $T_\#(\mu)(B) := \mu(T^{-1}(B))$ for every Borel set $B \subset Y$. The set $\Gamma(\mu, \nu)$ introduced in (2.4) is call the set of *admissible plans* or *couplings* for the pair (μ, ν). For any Polish space (X, d), there exist optimal couplings where the infimum in (2.3) is achieved.

The function W is a distance on the so-called *Wasserstein space* $(\mathscr{P}_2(X), \mathsf{W})$, where

$$\mathscr{P}_2(X) := \left\{ \mu \in \mathscr{P}(X) : \int_X \mathsf{d}(x, x_0)^2 \, d\mu(x) < +\infty \text{ for some, and thus any, } x_0 \in X \right\}.$$

The space $(\mathscr{P}_2(X), \mathsf{W})$ is Polish. If (X, d) is geodesic, then $(\mathscr{P}_2(X), \mathsf{W})$ is geodesic as well. Moreover, $\mu_n \xrightarrow{\mathsf{W}} \mu$ if and only if $\mu_n \rightharpoonup \mu$ and

$$\int_X \mathsf{d}(x, x_0)^2 \, d\mu_n(x) \to \int_X \mathsf{d}(x, x_0)^2 \, d\mu(x) \qquad \text{for some } x_0 \in X.$$

As usual, we write $\mu_n \rightharpoonup \mu$ if $\int_X \varphi \, d\mu_n \to \int_X \varphi \, d\mu$ for all $\varphi \in C_b(X)$.

2.3. **Relative entropy.** Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space, where (X, d) is a Polish metric space and $\mathfrak{m}$ is a non-negative, Borel and σ-finite measure. We assume that the space $(X, \mathsf{d}, \mathfrak{m})$ satisfies the following structural assumption: there exist a point $x_0 \in X$ and two constants $c_1, c_2 > 0$ such that

$$\mathfrak{m}\left(\{x \in X : \mathsf{d}(x, x_0) < r\}\right) \leq c_1 e^{c_2 r^2}. \tag{2.5}$$

The *relative entropy* $\mathsf{Ent}_{\mathfrak{m}}\colon \mathscr{P}_2(X) \to (-\infty, +\infty]$ is defined as

$$\mathsf{Ent}_{\mathfrak{m}}(\mu) := \begin{cases} \displaystyle\iint_X \varrho \log \varrho \; d\mathfrak{m} & \text{if } \mu = \varrho\mathfrak{m} \in \mathscr{P}_2(X), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.6}$$

According to our definition, $\mu \in \mathrm{Dom}(\mathsf{Ent}_{\mathfrak{m}})$ implies that $\mu \in \mathscr{P}_2(X)$ and that the effective domain $\mathrm{Dom}(\mathsf{Ent}_{\mathfrak{m}})$ is convex. As pointed out in [5, Section 7.1], the structural assumption (2.5) guarantees that in fact $\mathsf{Ent}(\mu) > -\infty$ for all $\mu \in \mathscr{P}_2(X)$.

When $\mathfrak{m} \in \mathscr{P}(X)$, the entropy functional $\mathsf{Ent}_{\mathfrak{m}}$ naturally extends to $\mathscr{P}(X)$, is lower semicontinuous with respect to the weak convergence in $\mathscr{P}(X)$ and positive by Jensen's inequality. In addition, if $F\colon X \to Y$ is a Borel map, then

$$\mathsf{Ent}_{F_{\#}\mathfrak{m}}(F_{\#}\mu) \le \mathsf{Ent}_{\mathfrak{m}}(\mu) \qquad \text{for all } \mu \in \mathscr{P}(X), \tag{2.7}$$

with equality if F is injective, see [4, Lemma 9.4.5].

When $\mathfrak{m}(X) = +\infty$, if we set $\mathfrak{n} := e^{-c\,\mathsf{d}(\cdot,x_0)^2}\mathfrak{m}$ for some $x_0 \in X$, where $c > 0$ is chosen so that $\mathfrak{n}(X) < +\infty$ (note that the existence of such $c > 0$ is guaranteed by (2.5)), then we obtain the useful formula

$$\mathsf{Ent}_{\mathfrak{m}}(\mu) = \mathsf{Ent}_{\mathfrak{n}}(\mu) - c\int_X \mathsf{d}(x,x_0)^2 \; d\mu \qquad \text{for all } \mu \in \mathscr{P}_2(X). \tag{2.8}$$

This shows that $\mathsf{Ent}_{\mathfrak{m}}$ is lower semicontinuous in $(\mathscr{P}_2(X), \mathsf{W})$.

2.4. **Carnot groups.** Let $\mathbb{G}$ be a Carnot group, i.e. a connected, simply connected and nilpotent Lie group whose Lie algebra $\mathfrak{g}$ of left-invariant vector fields has dimension n and admits a stratification of step κ,

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_{\kappa}$$

with

$$V_i = [V_1, V_{i-1}] \quad \text{for } i = 1, \ldots, \kappa, \qquad [V_1, V_{\kappa}] = \{0\}.$$

We set $m_i = \dim(V_i)$ and $h_i = m_1 + \cdots + m_i$ for $i = 1, \ldots, \kappa$, with $h_0 = 0$ and $h_{\kappa} = n$. We fix an adapted basis of $\mathfrak{g}$, i.e. a basis $X_1, \ldots, X_n$ such that

$$X_{h_{i-1}+1}, \ldots, X_{h_i} \text{ is a basis of } V_i, \qquad i = 1, \ldots, \kappa.$$

Using exponential coordinates, we can identify $\mathbb{G}$ with $\mathbb{R}^n$ endowed with the group law determined by the Campbell–Hausdorff formula (in particular, the identity $e \in \mathbb{G}$ corresponds to $0 \in \mathbb{R}^n$ and $x^{-1} = -x$ for $x \in \mathbb{G}$). It is not restrictive to assume that $X_i(0) = \mathrm{e}_i$ for any $i = 1, \ldots, n$; therefore, by left-invariance, for any $x \in \mathbb{G}$ we get

$$X_i(x) = dl_x \mathrm{e}_i, \qquad i = 1, \ldots, n, \tag{2.9}$$

where $l_x\colon \mathbb{G} \to \mathbb{G}$ is the left-translation by $x \in \mathbb{G}$, i.e. $l_x(y) = xy$ for any $y \in \mathbb{G}$. We endow $\mathfrak{g}$ with the left-invariant Riemannian metric $\langle \cdot, \cdot \rangle_{\mathbb{G}}$ that makes the basis $X_1, \ldots, X_n$ orthonormal. For any $i = 1, \ldots, n$, we define the gradient with respect to the layer V_i as

$$\nabla_{V_i} f := \sum_{j=h_{i-1}+1}^{h_i} (X_j f)\, X_j \in V_i.$$

We let $H\mathbb{G} \subset T\mathbb{G}$ be the *horizontal tangent bundle* of the group $\mathbb{G}$, i.e. the left-invariant sub-bundle of the tangent bundle $T\mathbb{G}$ such that $H_e\mathbb{G} = \{X(0) : X \in V_1\}$. We use the distinguished notation $\nabla_{\mathbb{G}} := \nabla_{V_1}$ for the *horizontal gradient*.

For any $i = 1, \ldots, n$, we define the degree $d(i) \in \{1, \ldots, \kappa\}$ of the basis vector field X_i as $d(i) = j$ if and only if $X_i \in V_j$. With this notion, the one-parameter family of group dilations $(\delta_\lambda)_{\lambda \geq 0} \colon \mathbb{G} \to \mathbb{G}$ is given by

$$\delta_\lambda(x) = \delta_\lambda(x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda^{d(i)} x_i, \ldots, \lambda^\kappa x_n) \qquad \text{for all } x \in \mathbb{G}. \tag{2.10}$$

The Haar measure of the group $\mathbb{G}$ coincides with the n-dimensional Lebesgue measure $\mathscr{L}^n$ and has the homogeneity property $\mathscr{L}^n(\delta_\lambda(E)) = \lambda^Q \mathscr{L}^n(E)$, where the integer $Q = \sum_{i=1}^\kappa i \dim(V_i)$ is the *homogeneous dimension* of the group.

We endow the group $\mathbb{G}$ with the canonical Carnot–Carathéodory structure induced by $H\mathbb{G}$. We say that a Lipschitz curve $\gamma \colon [0, 1] \to \mathbb{G}$ is a *horizontal curve* if $\dot{\gamma}(t) \in H_{\gamma(t)}\mathbb{G}$ for a.e. $t \in [0, 1]$. The *Carnot–Carathéodory distance* between $x, y \in \mathbb{G}$ is then defined as

$$\mathsf{d}_{\mathsf{cc}}(x, y) = \inf \left\{ \int_0^1 \|\dot{\gamma}(t)\|_{\mathbb{G}} \, dt : \gamma \text{ is horizontal}, \ \gamma(0) = x, \ \gamma(1) = y \right\}.$$

By Chow–Rashevskii's Theorem, the function d_{cc} is in fact a distance, which is also left-invariant and homogeneous with respect to the dilations defined in (2.10), precisely $\mathsf{d}_{\mathsf{cc}}(zx, zy) = \mathsf{d}_{\mathsf{cc}}(x, y)$ and $\mathsf{d}_{\mathsf{cc}}(\delta_\lambda(x), \delta_\lambda(y)) = \lambda \mathsf{d}_{\mathsf{cc}}(x, y)$ for all $x, y, z \in \mathbb{G}$ and $\lambda \geq 0$. The resulting metric space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}})$ is a Polish geodesic space. We let $\mathsf{B}_{\mathbb{G}}(x, r)$ be the d_{cc}-ball centred at $x \in \mathbb{G}$ of radius $r > 0$. Note that $\mathscr{L}^n(\mathsf{B}_{\mathbb{G}}(x, r)) = c_n r^Q$, where $c_n = \mathscr{L}^n(\mathsf{B}_{\mathbb{G}}(0, 1))$. In particular, the metric measure space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ satisfies the structural assumption (2.5).

Let us write $x = (\tilde{x}_1, \ldots, \tilde{x}_\kappa)$, where $\tilde{x}_i := (x_{h_{i-1}+1}, \ldots, x_{h_i})$ for $i = 1, \ldots, \kappa$. As proved in [18, Theorem 5.1], there exist suitable constants $c_1 = 1$, $c_2, \ldots, c_k \in (0, 1)$ depending only on the group structure of $\mathbb{G}$ such that

$$\mathsf{d}_\infty(x, 0) := \max \left\{ c_i |\tilde{x}_i|_{\mathbb{R}^{m_i}}^{1/i} : i = 1, \ldots, \kappa \right\}, \qquad x \in \mathbb{G}, \tag{2.11}$$

induces a left-invariant and homogeneous distance $\mathsf{d}_\infty(x, y) := \mathsf{d}_\infty(y^{-1}x, 0)$, $x, y \in \mathbb{G}$, which is equivalent to d_{cc}.

Let $1 \leq p < +\infty$ and let $\Omega \subset \mathbb{R}^n$ be an open set. The *horizontal Sobolev space*

$$W_{\mathbb{G}}^{1,p}(\Omega) := \{u \in L^p(\Omega) : X_i u \in L^p(\Omega), \ i = 1, \ldots, m_1\} \tag{2.12}$$

endowed with the norm

$$\|u\|_{W_{\mathbb{G}}^{1,p}(\Omega)} := \|u\|_{L^p(\Omega)} + \sum_{i=1}^{m_1} \|X_i u\|_{L^p(\Omega)}$$

is a reflexive Banach space, see [17, Proposition 1.1.2]. By [17, Theorem 1.2.3], the set $C^\infty(\Omega) \cap W_{\mathbb{G}}^{1,p}(\Omega)$ is dense in $W_{\mathbb{G}}^{1,p}(\Omega)$. By a standard cut-off argument, we get that $C_c^\infty(\mathbb{R}^n)$ is dense in $W_{\mathbb{G}}^{1,p}(\mathbb{R}^n)$.

2.5. **Riemannian approximation.** The metric space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}})$ can be seen as the limit in the *pointed Gromov–Hausdorff sense* as $\varepsilon \to 0$ of a family of Riemannian manifolds $\{(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon)\}_{\varepsilon > 0}$ defined as follows, see [12, Theorem 2.12]. For any $\varepsilon > 0$, we define the Riemannian approximation $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon)$ of the Carnot group $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}})$ as the manifold $\mathbb{R}^n$ endowed with the Riemannian metric $g_\varepsilon(\cdot, \cdot) \equiv \langle \cdot, \cdot \rangle_\varepsilon$ that makes orthonormal the vector fields $\varepsilon^{d(i)-1} X_i$, $i = 1, \ldots, n$, i.e. such that

$$\langle X_i, X_j \rangle_\varepsilon = \varepsilon^{2-d(i)-d(j)} \delta_{ij}, \qquad i, j = 1, \ldots, n.$$

We let d_ε be the Riemannian distance induced by the metric g_ε. Note that d_ε is left-invariant and satisfies $\mathsf{d}_\varepsilon \leq \mathsf{d}_{cc}$ for all $\varepsilon > 0$. For any $\varepsilon > 0$, the ε-Riemannian gradient is defined as

$$\nabla_\varepsilon f = \sum_{i=1}^n \varepsilon^{2(d(i)-1)}(X_i f)\, X_i = \sum_{i=1}^\kappa \varepsilon^{2(i-1)} \nabla_{V_i} f.$$

By (2.9), we get that

$$g_\varepsilon(x) = (dl_x)^T D_\varepsilon\, (dl_x), \qquad x \in \mathbb{G},$$

where D_ε is the diagonal block matrix given by

$$D_\varepsilon = \mathrm{diag}(\mathbf{1}_{m_1}, \varepsilon^{-2}\mathbf{1}_{m_2}, \ldots, \varepsilon^{-2(i-1)}\mathbf{1}_{m_i}, \ldots, \varepsilon^{-2(\kappa-1)}\mathbf{1}_{m_\kappa}).$$

As a consequence, the Riemannian volume element is given by

$$\mathrm{vol}_\varepsilon = \sqrt{\det g_\varepsilon}\, dx_1 \wedge \cdots \wedge dx_n = \varepsilon^{n-Q}\mathscr{L}^n.$$

We remark that, for each $\varepsilon > 0$, the n-dimensional Riemannian manifold $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon)$ has Ricci curvature bounded from below. More precisely, there exists a constant $K > 0$, depending only on the Carnot group $\mathbb{G}$, such that

$$\mathrm{Ric}_\varepsilon \geq -K\varepsilon^{-2} \qquad \text{for all } \varepsilon > 0. \tag{2.13}$$

By scaling invariance, the proof of inequality (2.13) can be reduced to the case $\varepsilon = 1$, which in turn is a direct consequence of [31, Lemma 1.1].

In the sequel, we will consider the metric measure space $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon, \mathscr{L}^n)$, i.e. the Riemannian manifold $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon, \mathrm{vol}_\varepsilon)$ with a rescaled volume measure. Both these two spaces satisfy the structural assumption (2.5). Moreover, we have

$$\mathsf{Ent}_{\mathrm{vol}_\varepsilon}(\mu) = \mathsf{Ent}(\mu) + \log(\varepsilon^{Q-n}) \qquad \text{for all } \varepsilon > 0. \tag{2.14}$$

Here and in the following, Ent denotes the entropy with respect to the reference measure $\mathscr{L}^n$.

2.6. **Sub-elliptic heat equation.** We let $\Delta_\mathbb{G} = \sum_{i=1}^{m_1} X_i^2$ be the so-called *sub-Laplacian operator*. Since the horizontal vector fields $X_1, \ldots, X_{h_1}$ satisfy Hörmander's condition, by Hörmander's theorem the *sub-elliptic heat operator* $\partial_t - \Delta_\mathbb{G}$ is *hypoelliptic*, meaning that its fundamental solution $\mathsf{h}\colon (0, +\infty) \times \mathbb{G} \to (0, +\infty)$, $\mathsf{h}_t(x) = \mathsf{h}(t, x)$, the so-called *heat kernel*, is smooth. In the following result, we collect some properties of the heat kernel that will be used in the sequel. We refer the reader to [39, Chapter IV] and to the references therein for the proof.

Theorem 2.3 (Properties of the heat kernel). *The heat kernel* $\mathsf{h}\colon (0, +\infty) \times \mathbb{G} \to (0, +\infty)$ *satisfies the following properties:*

(i) $\mathsf{h}_t(x^{-1}) = \mathsf{h}_t(x)$ *for any* $(t, x) \in (0, +\infty) \times \mathbb{G}$;
(ii) $\mathsf{h}_{\lambda^2 t}(\delta_\lambda(x)) = \lambda^{-Q}\mathsf{h}_t(x)$ *for any* $\lambda > 0$ *and* $(t, x) \in (0, +\infty) \times \mathbb{G}$;
(iii) $\int_\mathbb{G} \mathsf{h}_t\, dx = 1$ *for any* $t > 0$;
(iv) *there exists* $C > 0$, *depending only on* $\mathbb{G}$, *such that*

$$\mathsf{h}_t(x) \leq Ct^{-Q/2} \exp\left(-\frac{\mathsf{d}_{cc}(x, 0)^2}{4t}\right) \qquad \forall(t, x) \in (0, +\infty) \times \mathbb{G}; \tag{2.15}$$

(v) for any $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that

$$\mathsf{h}_t(x) \geq C_\varepsilon t^{-Q/2} \exp\left(-\frac{\mathsf{d}_{\mathsf{cc}}(x,0)^2}{4(1-\varepsilon)t}\right) \qquad \forall (t,x) \in (0,+\infty) \times \mathbb{G}; \tag{2.16}$$

(vi) for every $j,l \in \mathbb{N}$ and $\varepsilon > 0$, there exists $C_\varepsilon(j,l) > 0$ such that

$$|(\partial_t)^l X_{i_1} \cdots X_{i_j} \mathsf{h}_t(x)| \leq C_\varepsilon(j,l) t^{-\frac{Q+j+2l}{2}} \exp\left(-\frac{\mathsf{d}_{\mathsf{cc}}(x,0)^2}{4(1+\varepsilon)t}\right) \qquad \forall (t,x) \in (0,+\infty) \times \mathbb{G}, \tag{2.17}$$

where $X_{i_1} \cdots X_{i_j} \in V_1$.

Given $\varrho \in L^1(\mathbb{G})$, the function

$$\varrho_t(x) = (\varrho \star \mathsf{h}_t)(x) = \int_{\mathbb{G}} \mathsf{h}_t(y^{-1}x)\, \varrho(y)\, dy, \qquad (t,x) \in (0,+\infty) \times \mathbb{G}, \tag{2.18}$$

is smooth and is a solution of the heat diffusion problem

$$\begin{cases} \partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t & \text{in } (0,+\infty) \times \mathbb{G}, \\ \varrho_0 = \varrho, & \text{on } \{0\} \times \mathbb{G}. \end{cases} \tag{2.19}$$

The initial datum is assumed in the L^1-sense, i.e. $\lim_{t \to 0} \varrho_t = \varrho$ in $L^1(\mathbb{G})$. As a consequence of the properties of the heat kernel, if $\varrho \geq 0$ then the solution $(\varrho_t)_{t \geq 0}$ in (2.18) is everywhere positive and satisfies

$$\int_{\mathbb{G}} \varrho_t(x)\, dx = \|\varrho\|_{L^1(\mathbb{G})} \qquad \forall t > 0.$$

In addition, if $\varrho \mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$ then $(\varrho_t \mathscr{L}^n)_{t \geq 0} \subset \mathscr{P}_2(X)$. Indeed, by (2.15), we have

$$C_t := \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2\, \mathsf{h}_t(x)\, dx < +\infty \qquad \forall t > 0.$$

Thus, by triangular inequality, we have

$$(\mathsf{d}_{\mathsf{cc}}(\cdot,0)^2 \star \mathsf{h}_t)(x) = \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(xy^{-1},0)^2\, \mathsf{h}_t(y)\, dy \leq 2\, \mathsf{d}_{\mathsf{cc}}(x,0)^2 + 2C_t,$$

so that, for all $t > 0$, we get

$$\int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2\, \varrho_t(x)\, dx = \int_{\mathbb{G}} (\mathsf{d}_{\mathsf{cc}}(\cdot,0)^2 \star \mathsf{h}_t)(x)\, \varrho(x)\, dx \leq 2 \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2\, \varrho(x)\, dx + 2C_t. \tag{2.20}$$

2.7. **Main result.** We are now ready to state the main result of the paper. The proof is given in Section 4 and deals with the two parts of the statement separately.

Theorem 2.4. *Let $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ be a Carnot group and let $\varrho_0 \in L^1(\mathbb{G})$ be such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathrm{Dom}(\mathsf{Ent})$. If $(\varrho_t)_{t \geq 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t$ with initial datum ϱ_0, then $\mu_t = \varrho_t \mathscr{L}^n$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$ starting from μ_0.*

Conversely, if $(\mu_t)_{t \geq 0}$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$ starting from μ_0, then $\mu_t = \varrho_t \mathscr{L}^n$ for all $t \geq 0$ and $(\varrho_t)_{t \geq 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t$ with initial datum ϱ_0.

3. Continuity equation and slope of the entropy

3.1. **The Wasserstein space on the approximating Riemannian manifold.** Let $(\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon)$ be the Wasserstein space introduced in Section 2.2 relative to the metric measure space $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon, \mathscr{L}^n)$. As observed in Section 2.5, $(\mathbb{G}_\varepsilon, \mathsf{d}_\varepsilon, \mathscr{L}^n)$ is an n-dimensional Riemannian manifold (with rescaled volume measure) whose Ricci curvature is bounded from below. Here we collect some known results taken from [14, 40] concerning the space $(\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon)$. In the original statements, the canonical reference measure is the Riemannian volume. Keeping in mind that $\mathrm{vol}_\varepsilon = \varepsilon^{n-Q}\mathscr{L}^n$ and the relation (2.14), in our statements each quantity is rescaled accordingly. All time-dependent vector fields appearing in the sequel are tacitly understood to be Borel measurable.

Let $\mu \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$ be given. We define the space

$$L_\varepsilon^2(\mu) = \left\{ \xi \in \mathscr{S}(T\mathbb{G}_\varepsilon) : \int_{\mathbb{G}_\varepsilon} \|\xi\|_\varepsilon^2 \, d\mu < +\infty \right\}.$$

Here $\mathscr{S}(T\mathbb{G}_\varepsilon)$ denotes the set of sections of the tangent bundle $T\mathbb{G}_\varepsilon$. Moreover, we define the *'tangent space'* of $(\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon)$ at μ as

$$\mathrm{Tan}_\varepsilon(\mu) = \overline{\{\nabla_\varepsilon \varphi : \varphi \in C_c^\infty(\mathbb{R}^n)\}}^{L_\varepsilon^2(\mu)}.$$

The 'tangent space' $\mathrm{Tan}_\varepsilon(\mu)$ was first introduced in [35]. We refer the reader to [4, Chapter 12] and to [40, Chapters 13 and 15] for a detailed discussion on this space.

Let $\varepsilon > 0$ be fixed. Given $I \subset \mathbb{R}$ an open interval and a time-dependent vector field $v^\varepsilon : I \times \mathbb{G}_\varepsilon \to T\mathbb{G}_\varepsilon$, $(t, x) \mapsto v_t^\varepsilon(x) \in T_x\mathbb{G}_\varepsilon$, we say that a curve $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G}_\varepsilon)$ satisfies the *continuity equation*

$$\partial_t \mu_t + \mathrm{div}(v_t^\varepsilon \mu_t) = 0 \qquad \text{in } I \times \mathbb{G}_\varepsilon \tag{3.1}$$

in the sense of distributions if

$$\int_I \int_{\mathbb{G}_\varepsilon} \|v_t^\varepsilon(x)\|_\varepsilon \, d\mu_t(x) \, dt < +\infty$$

and

$$\int_I \int_{\mathbb{G}_\varepsilon} \partial_t \varphi(t, x) + \langle v_t^\varepsilon(x), \nabla_\varepsilon \varphi(t, x) \rangle_\varepsilon \, d\mu_t(x) \, dt = 0 \qquad \forall \varphi \in C_c^\infty(I \times \mathbb{R}^n).$$

We can thus state the following result, see [14, Proposition 2.5] for the proof. Here and in the sequel, the metric derivative in the Wasserstein space $(\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon)$ of a curve $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G}_\varepsilon)$ is denoted by $|\dot\mu_t|_\varepsilon$.

Proposition 3.1 (Continuity equation in $(\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon)$). *Let $\varepsilon > 0$ be fixed and let $I \subset \mathbb{R}$ be an open interval. If $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; \mathscr{P}_2(\mathbb{G}_\varepsilon))$, then there exists a time-dependent vector field $v^\varepsilon : I \times \mathbb{G}_\varepsilon \to T\mathbb{G}_\varepsilon$ with $t \mapsto \|v_t^\varepsilon\|_{L_\varepsilon^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$ such that*

$$v_t^\varepsilon \in \mathrm{Tan}_\varepsilon(\mu_t) \qquad \text{for a.e. } t \in I \tag{3.2}$$

and the continuity equation (3.1) holds in the sense of distributions. The vector field v_t^ε is uniquely determined in $L_\varepsilon^2(\mu_t)$ by (3.1) and (3.2) for a.e. $t \in I$ and we have

$$\|v_t^\varepsilon\|_{L_\varepsilon^2(\mu_t)} = |\dot\mu_t|_\varepsilon \qquad \text{for a.e. } t \in I.$$

Conversely, if $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G}_\varepsilon)$ is a curve satisfying (3.1) for some $(v_t^\varepsilon)_{t \in I}$ such that $t \mapsto \|v_t^\varepsilon\|_{L_\varepsilon^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$, then $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; (\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon))$ with

$$|\dot\mu_t|_\varepsilon \leq \|v_t^\varepsilon\|_{L_\varepsilon^2(\mu_t)} \qquad \text{for a.e. } t \in I.$$

We can interpret the time-dependent vector field $(v_t^\varepsilon)_{t\in I}$ given by Proposition 3.1 as the 'tangent vector' of the curve $(\mu_t)_{t\in I}$ in $(\mathscr{P}_2(\mathbb{G}_\varepsilon), W_\varepsilon)$. As remarked in [14, Section 2], for a.e. $t \in I$ the vector field v_t^ε has minimal $L_\varepsilon^2(\mu_t)$-norm among all time-dependent vector fields satisfying (3.1). Moreover, this minimality is equivalent to (3.2).

In the following result and in the sequel, $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu)$ denotes the descending slope of the entropy Ent at the point $\mu \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$ in the Wasserstein space $(\mathscr{P}_2(\mathbb{G}_\varepsilon), W_\varepsilon)$.

Proposition 3.2. *Let $\varepsilon > 0$ be fixed and let $\mu = \varrho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$. The following statements are equivalent:*

(i) $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) < +\infty$;
(ii) $\varrho \in W_{\mathrm{loc}}^{1,1}(\mathbb{G}_\varepsilon)$ and $\nabla_\varepsilon\varrho = w^\varepsilon\varrho$ for some $w^\varepsilon \in L_\varepsilon^2(\mu)$.

In this case, $w^\varepsilon \in \mathrm{Tan}_\varepsilon(\mu)$ and $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) = \|w^\varepsilon\|_{L_\varepsilon^2(\mu)}$. Moreover, for any $\nu \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$, we have

$$\mathsf{Ent}(\nu) \geq \mathsf{Ent}(\mu) - \|w^\varepsilon\|_{L_\varepsilon^2(\mu)} W_\varepsilon(\nu, \mu) - \frac{K}{2\varepsilon^2} W_\varepsilon(\nu, \mu)^2, \tag{3.3}$$

where $K > 0$ is the constant appearing in (2.13).

The equivalence part in Proposition 3.2 is proved in [14, Proposition 4.3]. Inequality (3.3) is the so-called *HWI inequality* and follows from [40, Theorem 23.14], see [40, Remark 23.16].

The quantity

$$\mathsf{F}_\varepsilon(\varrho) = \|w^\varepsilon\|_{L_\varepsilon^2(\mu)}^2 = \int_{\mathbb{G}_\varepsilon\cap\{\varrho>0\}} \frac{\|\nabla_\varepsilon\varrho\|_\varepsilon^2}{\varrho} \, d\mathscr{L}^n$$

appearing in Proposition 3.2 is the so-called *Fisher information* of $\mu = \varrho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$. The inequality $F_\varepsilon(\varrho) \leq |\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu)$ holds in the context of metric measure spaces, see [5, Theorem 7.4]. The converse inequality does not hold in such a generality and heavily depends on the lower semicontinuity of the descending slope $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|$, see [5, Theorem 7.6].

3.2. **The Wasserstein space on the Carnot group.** Let $(\mathscr{P}_2(\mathbb{G}), W_\mathbb{G})$ be the Wasserstein space introduced in Section 2.2 relative to the metric measure space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$. In this section, we discuss the counterparts of Proposition 3.1 and Proposition 3.2 in the space $(\mathscr{P}_2(\mathbb{G}), W_\mathbb{G})$. All time-dependent vector fields appearing in the sequel are tacitly understood to be Borel measurable.

Let $\mu \in \mathscr{P}_2(\mathbb{G})$ be given. We define the space

$$L_\mathbb{G}^2(\mu) = \left\{ \xi \in \mathscr{S}(H\mathbb{G}) : \int_\mathbb{G} \|\xi\|_\mathbb{G}^2 \, d\mu < +\infty \right\}.$$

Here $\mathscr{S}(H\mathbb{G})$ denotes the set of sections of the horizontal tangent bundle $H\mathbb{G}$. Moreover, we define the *'tangent space'* of $(\mathscr{P}_2(\mathbb{G}), W_\mathbb{G})$ at μ as

$$\mathrm{Tan}_\mathbb{G}(\mu) = \overline{\{\nabla_\mathbb{G}\varphi : \varphi \in C_c^\infty(\mathbb{R}^n)\}}^{L_\mathbb{G}^2(\mu)}.$$

Given $I \subset \mathbb{R}$ an open interval and a horizontal time-dependent vector field $v^\mathbb{G} \colon I \times \mathbb{G} \to H\mathbb{G}$, $(t, x) \mapsto v_t^\mathbb{G}(x) \in H_x\mathbb{G}$, we say that a curve $(\mu_t)_{t\in I} \subset \mathscr{P}_2(\mathbb{G})$ satisfies the continuity equation

$$\partial_t\mu_t + \mathrm{div}(v_t^\mathbb{G}\mu_t) = 0 \qquad \text{in } I \times \mathbb{G}_\varepsilon \tag{3.4}$$

in the sense of distributions if

$$\int_I \int_\mathbb{G} \|v_t^\mathbb{G}(x)\|_\mathbb{G} \, d\mu_t(x) \, dt < +\infty$$

and

$$\int_I \int_{\mathbb{G}} \partial_t \varphi(t,x) + \left\langle v_t^{\mathbb{G}}(x), \nabla_{\mathbb{G}} \varphi(t,x) \right\rangle_{\mathbb{G}} d\mu_t(x)\, dt = 0 \qquad \forall \varphi \in C_c^{\infty}(I \times \mathbb{R}^n).$$

The following result is the exact analogue of Proposition 3.1. Here and in the sequel, the metric derivative in the Wasserstein space $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$ of a curve $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G})$ is denoted by $|\dot\mu_t|_{\mathbb{G}}$.

Proposition 3.3 (Continuity equation in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$)**.** *Let $I \subset \mathbb{R}$ be an open interval. If $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$, then there exists a horizontal time-dependent vector field $v^{\mathbb{G}} \colon I \times \mathbb{G} \to H\mathbb{G}$ with $t \mapsto \|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$ such that*

$$v_t^{\mathbb{G}} \in \operatorname{Tan}_{\mathbb{G}}(\mu_t) \qquad \text{for a.e. } t \in I \tag{3.5}$$

and the continuity equation (3.4) holds in the sense of distributions. The vector field $v_t^{\mathbb{G}}$ is uniquely determined in $L_{\mathbb{G}}^2(\mu_t)$ by (3.4) and (3.5) for a.e. $t \in I$ and we have

$$\|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} = |\dot\mu_t|_{\mathbb{G}} \qquad \text{for a.e. } t \in I.$$

Conversely, if $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G})$ is a curve satisfying (3.4) for some $(v_t^{\mathbb{G}})_{t \in I}$ such that $t \mapsto \|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$, then $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$ with

$$|\dot\mu_t|_{\mathbb{G}} \le \|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \qquad \text{for a.e. } t \in I.$$

As for Proposition 3.1, we can interpret the horizontal time-dependent vector field $(v_t^{\mathbb{G}})_{t \in I}$ given by Proposition 3.3 as the 'tangent vector' of the curve $(\mu_t)_{t \in I}$ in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$. An easy adaptation of [14, Lemma 2.4] to the sub-Riemannian manifold $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ again shows that for a.e. $t \in I$ the vector field $v_t^{\mathbb{G}}$ has minimal $L_{\mathbb{G}}^2(\mu_t)$-norm among all time-dependent vector fields satisfying (3.4) and, moreover, that this minimality is equivalent to (3.5).

Proposition 3.3 can be obtained applying the general results obtained in [20] to the metric measure space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$. Below we give a direct proof exploiting Proposition 3.1. The argument is very similar to the one of [25, Proposition 3.1] and we only sketch it.

Proof. If $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$, then also $(\mu_t)_t \in AC_{\mathrm{loc}}^2(I; (\mathscr{P}_2(\mathbb{G}_\varepsilon), \mathsf{W}_\varepsilon))$ for every $\varepsilon > 0$, since $\mathsf{d}_\varepsilon \le \mathsf{d}_{\mathsf{cc}}$. Let $v^\varepsilon \colon I \times \mathbb{G}_\varepsilon \to T\mathbb{G}_\varepsilon$ be the time-dependent vector field given by Proposition 3.1. Note that

$$\int_{\mathbb{G}} \|v_t^\varepsilon\|_\varepsilon^2\, d\mu_t = |\dot\mu_t|_\varepsilon^2 \le |\dot\mu_t|_{\mathbb{G}}^2 \qquad \text{for a.e. } t \in I. \tag{3.6}$$

Moreover

$$\|v_t^\varepsilon\|_\varepsilon^2 = \|v_t^{\varepsilon,V_1}\|_1^2 + \sum_{i=2}^{\kappa} \varepsilon^{2(1-i)} \|v_t^{\varepsilon,V_i}\|_1^2 \qquad \text{for all } \varepsilon > 0, \tag{3.7}$$

where v_t^{ε,V_i} denotes the projection of v_t^ε on V_i. Combining (3.6) and (3.7), we find a sequence $(\varepsilon_k)_{k \in \mathbb{N}}$, with $\varepsilon_k \to 0$, and a horizontal time-dependent vector field $v^{\mathbb{G}} \colon I \times \mathbb{G} \to H\mathbb{G}$ such that $v^{\varepsilon_k,V_1} \rightharpoonup v^{\mathbb{G}}$ and $v^{\varepsilon_k,V_i} \to 0$ for all $i = 2, \dots, \kappa$ as $k \to +\infty$ locally in time in the L^2-norm on $I \times \mathbb{G}$ naturally induced by the norm $\|\cdot\|_1$ and the measure $d\mu_t dt$. In particular, $t \mapsto \|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$ and $\|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \le |\dot\mu_t|_{\mathbb{G}}$ for a.e. $t \in I$. To prove (3.4), fix a test function $\varphi \in C_c^{\infty}(I \times \mathbb{R}^n)$ and pass to the limit as $\varepsilon \to 0^+$ in (3.1).

Conversely, if $(\mu_t)_{t \in I} \subset \mathscr{P}_2(\mathbb{G})$ satisfies (3.4) for some horizontal time-dependent vector field $(v_t^{\mathbb{G}})_{t \in I}$ such that $t \mapsto \|v_t^{\mathbb{G}}\|_{L_\varepsilon^2(\mu_t)} \in L_{\mathrm{loc}}^2(I)$, then we can apply Proposition 3.1 for

$\varepsilon = 1$. By the *superposition principle* stated in [9, Theorem 5.8] applied to the Riemannian manifold $(\mathbb{G}_1, \mathsf{d}_1, \mathscr{L}^n)$, we find a probability measure $\nu \in \mathscr{P}(C(I; (\mathbb{G}_1, \mathsf{d}_1)))$, concentrated on $AC^2_{\mathrm{loc}}(I; (\mathbb{G}_1, \mathsf{d}_1))$, such that $\mu_t = (\mathsf{e}_t)_{\#}\nu$ for all $t \in I$ and with the property that ν-a.e. curve $\gamma \in C(I; (\mathbb{G}_1, \mathsf{d}_1))$ is an absolutely continuous integral curve of the vector field $v^{\mathbb{G}}$. Here $\mathsf{e}_t \colon C(I; (\mathbb{G}_1, \mathsf{d}_1)) \to \mathbb{G}$ denotes the evaluation map at time $t \in I$. Since $v^{\mathbb{G}}$ is horizontal, ν-a.e. curve $\gamma \in C(I; (\mathbb{G}_1, \mathsf{d}_1))$ is horizontal. Therefore, for all $s, t \in I$, $s < t$, we have

$$\mathsf{d}_{\mathsf{cc}}(\gamma(t), \gamma(s)) \leq \int_s^t \|\dot{\gamma}(r)\|_{\mathbb{G}} \, dr = \int_s^t \|v_r^{\mathbb{G}}(\gamma(r))\|_{\mathbb{G}} \, dr$$

and we can thus estimate

$$\mathsf{W}^2_{\mathbb{G}}(\mu_t, \mu_s) \leq \int_{\mathbb{G} \times \mathbb{G}} \mathsf{d}^2_{\mathsf{cc}}(x, y) \, d(\mathsf{e}_t, \mathsf{e}_s)_{\#}\nu(x, y) = \int_{AC^2_{\mathrm{loc}}} \mathsf{d}^2_{\mathsf{cc}}(\gamma(t), \gamma(s)) \, d\nu(\gamma)$$

$$\leq (t - s) \int_{AC^2_{\mathrm{loc}}} \int_s^t \|v_r^{\mathbb{G}}(\gamma(r))\|^2_{\mathbb{G}} \, dr d\nu(\gamma) = (t - s) \int_s^t \int_{\mathbb{G}} \|v_r^{\mathbb{G}}\|^2_{\mathbb{G}} \, d\mu_r dr.$$

This immediately gives $|\dot{\mu}_t|_{\mathbb{G}} \leq \|v_t^{\mathbb{G}}\|_{L^2_{\mathbb{G}}(\mu_t)}$ for a.e. $t \in I$, which in turn proves (3.5). $\qquad\square$

To establish an analogue of Proposition 3.2, we need to prove the two inequalities separately. For $\mu = \varrho \mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$, the inequality $\mathsf{F}_{\mathbb{G}}(\varrho) \leq |\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu)$ is stated in Proposition 3.4 below. Here and in the sequel, $|\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu)$ denotes the descending slope of the entropy Ent at the point $\mu \in \mathscr{P}_2(\mathbb{G})$ in the Wasserstein space $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$.

Proposition 3.4. *Let $\mu = \varrho \mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$. If $|\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu) < +\infty$, then $\varrho \in W^{1,1}_{\mathbb{G}, \mathrm{loc}}(\mathbb{G})$ and $\nabla_{\mathbb{G}}\varrho = w^{\mathbb{G}}\varrho$ for some horizontal vector field $w^{\mathbb{G}} \in L^2(\mu)$ with $\|w^{\mathbb{G}}\|_{L^2_{\mathbb{G}}(\mu)} \leq |\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu)$.*

Proposition 3.4 can be obtained by applying [5, Theorem 7.4] to the metric measure space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$. Below we give a direct proof of this result which is closer in the spirit to the one in the Riemannian setting, see [14, Lemma 4.2]. See also [25, Proposition 3.1].

Proof. Let $V \in C_c^\infty(\mathbb{G}; H\mathbb{G})$ be a smooth horizontal vector field with compact support. Then there exists $\delta > 0$ such that, for any $t \in (-\delta, \delta)$, the flow map of the vector field V at time t, namely

$$F_t(x) := \exp_x(tV), \qquad x \in \mathbb{G},$$

is a diffeomorphism and $J_t = \det(DF_t)$ is such that $c^{-1} \leq J_t \leq c$ for some $c \geq 1$. By the change of variable formula, the measure $\mu_t := (F_t)_{\#}\mu$ is such that $\mu_t = \varrho_t \mathscr{L}^n$ with $J_t \varrho_t = \varrho \circ F_t^{-1}$ for $t \in (-\delta, \delta)$. Let us set $H(r) = r \log r$ for $r \geq 0$. Then, for $t \in (-\delta, \delta)$,

$$\mathsf{Ent}(\mu_t) = \int_{\mathbb{G}} H(\varrho_t) \, dx = \int_{\mathbb{G}} H\left(\frac{\varrho}{J_t}\right) J_t \, dx = \mathsf{Ent}(\mu) - \int_{\mathbb{G}} \varrho \log(J_t) \, dx < +\infty.$$

Note that $J_0 = 1$, $\dot{J}_0 = \mathrm{div}\, V$ and that $t \mapsto \dot{J}_t J_t^{-1}$ is uniformly bounded for $t \in (-\delta, \delta)$. Thus we have

$$\frac{d}{dt}\mathsf{Ent}(\mu_t)\bigg|_{t=0} = -\frac{d}{dt}\int_{\mathbb{G}} \varrho \log(J_t) \, dx\bigg|_{t=0} dx = \int_{\mathbb{G}} -\varrho\frac{\dot{J}_t}{J_t}\bigg|_{t=0} dx = -\int_{\mathbb{G}} \varrho\, \mathrm{div}\, V \, dx.$$

On the other hand, we have

$$\mathsf{W}^2_{\mathbb{G}}(\mu_t, \mu) = \mathsf{W}^2_{\mathbb{G}}((F_t)_{\#}\mu, \mu) \leq \int_{\mathbb{G}} \mathsf{d}^2_{\mathsf{cc}}(F_t(x), x) \, d\mu(x)$$

and so

$$\limsup_{t\to 0} \frac{\mathsf{W}_{\mathbb{G}}^2(\mu_t, \mu)}{|t|^2} \le \int_{\mathbb{G}} \limsup_{t\to 0} \frac{\mathsf{d}_{\mathsf{cc}}^2(F_t(x), x)}{|t|^2} \, d\mu(x) = \int_{\mathbb{G}} \|V\|_{\mathbb{G}}^2 \, d\mu.$$

Hence

$$-\frac{d}{dt}\mathsf{Ent}(\mu_t)\Big|_{t=0} \le \limsup_{t\to 0} \frac{[\mathsf{Ent}(\mu_t) - \mathsf{Ent}(\mu)]^-}{\mathsf{W}_{\mathbb{G}}(\mu_t, \mu)} \cdot \frac{\mathsf{W}_{\mathbb{G}}(\mu_t, \mu)}{|t|} \le |\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu) \left(\int_{\mathbb{G}} \|V\|_{\mathbb{G}}^2 \, d\mu\right)^{\frac{1}{2}}$$

and thus

$$\left|\int_{\mathbb{G}} \varrho \operatorname{div} V \, dx\right| \le |\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu) \left(\int_{\mathbb{G}} \|V\|_{\mathbb{G}}^2 \, d\mu\right)^{\frac{1}{2}}.$$

By Riesz representation theorem, we conclude that there exists a horizontal vector field $w^{\mathbb{G}} \in L_{\mathbb{G}}^2(\mu)$ such that $\|w_{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu)} \le |\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu)$ and

$$-\int_{\mathbb{G}} \varrho \operatorname{div} V \, dx = \int_{\mathbb{G}} \left\langle w^{\mathbb{G}}, V \right\rangle_{\mathbb{G}} \, d\mu \qquad \text{for all } V \in C_c^\infty(\mathbb{G}; H\mathbb{G}).$$

This implies that $\nabla_{\mathbb{G}}\varrho = w^{\mathbb{G}}\varrho$ and the proof is complete. $\qquad\square$

We call the quantity

$$\mathsf{F}_{\mathbb{G}}(\varrho) = \|w^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu)}^2 = \int_{\mathbb{G}\cap\{\varrho>0\}} \frac{\|\nabla_{\mathbb{G}}\varrho\|_{\mathbb{G}}^2}{\varrho} \, d\mathscr{L}^n$$

appearing in Proposition 3.4 the *horizontal Fisher information* of $\mu = \varrho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$. On its effective domain, $\mathsf{F}_{\mathbb{G}}$ is convex and sequentially lower semicontinuous with respect to the weak topology of $L^1(\mathbb{G})$, see [5, Lemma 4.10].

Given $\mu = \varrho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$, it is not clear how to prove the inequality $|\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|^2(\mu) \le \mathsf{F}_{\mathbb{G}}(\varrho)$ under the mere condition $|\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu) < +\infty$. Following [25, Proposition 3.4], in Proposition 3.5 below we show that the condition $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) < +\infty$ for some $\varepsilon > 0$ (and thus any) implies that $|\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|^2(\mu) \le \mathsf{F}_{\mathbb{G}}(\varrho)$.

Proposition 3.5. *Let* $\mu = \varrho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$. *If* $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) < +\infty$ *for some* $\varepsilon > 0$, *then also* $|\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu) < +\infty$ *and moreover* $\mathsf{F}_{\mathbb{G}}(\varrho) = |\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|^2(\mu)$.

Proof. Since $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) < +\infty$, we have $\mathsf{Ent}(\mu) < +\infty$. Since $\mathsf{d}_\varepsilon \le \mathsf{d}_{\mathsf{cc}}$ and so $\mathsf{W}_\varepsilon \le \mathsf{W}_{\mathbb{G}}$, we also have $|\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu) \le |\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu)$. By Proposition 3.4, we conclude that $\varrho \in W_{\mathbb{G},\mathrm{loc}}^{1,1}(\mathbb{G})$ and $\nabla_{\mathbb{G}}\varrho = w^{\mathbb{G}}\varrho$ for some horizontal vector field $w^{\mathbb{G}} \in L_{\mathbb{G}}^2(\mu)$ with $\|w^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu)} \le |\mathrm{D}_{\mathbb{G}}^-\mathsf{Ent}|(\mu)$. We now prove the converse inequality. Since $|\mathrm{D}_\varepsilon^-\mathsf{Ent}|(\mu) < +\infty$, by Proposition 3.2 we have $\mathsf{F}_\varepsilon(\varrho) = |\mathrm{D}_\varepsilon^-\mathsf{Ent}|^2(\mu)$ and

$$\mathsf{Ent}(\nu) \ge \mathsf{Ent}(\mu) - \mathsf{F}_\varepsilon^{1/2}(\varrho)\,\mathsf{W}_\varepsilon(\nu, \mu) - \tfrac{K}{2\varepsilon^2}\,\mathsf{W}_\varepsilon^2(\nu, \mu) \tag{3.8}$$

for any $\nu \in \mathscr{P}_2(\mathbb{G})$. Take $\varepsilon = \mathsf{W}_{\mathbb{G}}(\nu, \mu)^{1/4}$ and assume $\varepsilon < 1$. Since $\mathsf{W}_\varepsilon \le \mathsf{W}_{\mathbb{G}}$, from (3.8) we get

$$\mathsf{Ent}(\nu) \ge \mathsf{Ent}(\mu) - \mathsf{F}_\varepsilon^{1/2}(\varrho)\,\mathsf{W}_{\mathbb{G}}(\nu, \mu) - \tfrac{K}{2\varepsilon^2}\,\mathsf{W}_{\mathbb{G}}^2(\nu, \mu)$$
$$= \mathsf{Ent}(\mu) - \mathsf{F}_\varepsilon^{1/2}(\varrho)\,\mathsf{W}_{\mathbb{G}}(\nu, \mu) - \tfrac{K}{2}\,\mathsf{W}_{\mathbb{G}}^{3/2}(\nu, \mu). \tag{3.9}$$

We need to bound $\mathsf{F}_\varepsilon(\varrho)$ from above in terms of $\mathsf{F}_{\mathbb{G}}(\varrho)$. To do so, observe that

$$\nabla_\varepsilon\varrho = \nabla_{\mathbb{G}}\varrho + \sum_{i=2}^k \varepsilon^{2(i-1)}\nabla_{V_i}\varrho, \qquad \|\nabla_\varepsilon\varrho\|_\varepsilon^2 = \|\nabla_{\mathbb{G}}\varrho\|_{\mathbb{G}}^2 + \sum_{i=2}^k \varepsilon^{2(i-1)}\|\nabla_{V_i}\varrho\|_{\mathbb{G}}^2.$$

In particular, $\frac{\nabla_{V_i}\varrho}{\varrho} \in L^2_{\mathbb{G}}(\mu)$ for all $i = 2, \ldots, k$. Recalling the inequality $(1+r) \leq \left(1 + \frac{r}{2}\right)^2$ for $r \geq 0$, we can estimate

$$\mathsf{F}_\varepsilon(\varrho) = \mathsf{F}_{\mathbb{G}}(\varrho) + \sum_{i=2}^{k} \varepsilon^{2(i-1)} \left\|\frac{\nabla_{V_i}\varrho}{\varrho}\right\|^2_{L^2_{\mathbb{G}}(\mu)} = \mathsf{F}_{\mathbb{G}}(\varrho) \left(1 + \frac{1}{\mathsf{F}_{\mathbb{G}}(\varrho)} \sum_{i=2}^{k} \varepsilon^{2(i-1)} \left\|\frac{\nabla_{V_i}\varrho}{\varrho}\right\|^2_{L^2_{\mathbb{G}}(\mu)}\right)$$

$$\leq \mathsf{F}_{\mathbb{G}}(\varrho) \left(1 + \frac{1}{2\mathsf{F}_{\mathbb{G}}(\varrho)} \sum_{i=2}^{k} \varepsilon^{2(i-1)} \left\|\frac{\nabla_{V_i}\varrho}{\varrho}\right\|^2_{L^2_{\mathbb{G}}(\mu)}\right)^2$$

and thus

$$\mathsf{F}^{1/2}_\varepsilon(\varrho) \leq \mathsf{F}^{1/2}_{\mathbb{G}}(\varrho) \left(1 + \frac{1}{2\mathsf{F}_{\mathbb{G}}(\varrho)} \sum_{i=2}^{k} \varepsilon^{2(i-1)} \left\|\frac{\nabla_{V_i}\varrho}{\varrho}\right\|^2_{L^2_{\mathbb{G}}(\mu)}\right). \tag{3.10}$$

Inserting (3.10) into (3.9), we finally get

$$\mathsf{Ent}(\nu) \geq \mathsf{Ent}(\mu) - \mathsf{F}^{1/2}_{\mathbb{G}}(\varrho)\,\mathsf{W}_{\mathbb{G}}(\nu,\mu) - C\,\mathsf{W}^{3/2}_{\mathbb{G}}(\nu,\mu)$$

for some $C > 0$ independent of ε. This immediately leads to $|\mathsf{D}^-_{\mathbb{G}}\mathsf{Ent}|(\mu) \leq \mathsf{F}^{1/2}_{\mathbb{G}}(\varrho)$. $\qquad\square$

3.3. **Carnot groups are non-$CD(K,\infty)$ spaces.** As stated in [5, Theorem 7.6], if the metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is Polish and satisfies (2.5), then the properties

  (i) $|\mathsf{D}^-\mathsf{Ent}|^2(\mu) = \mathsf{F}(\varrho)$ for all $\mu = \varrho\mathfrak{m} \in \mathrm{Dom}(\mathsf{Ent})$;
  (ii) $|\mathsf{D}^-\mathsf{Ent}|$ is sequentially lower semicontinuous with respect to convergence with moments in $\mathscr{P}(X)$ on sublevels of Ent;

are equivalent. We do not know if property (ii) is true for the space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ and this is why in Proposition 3.5 we needed the additional assumption $|\mathsf{D}^-_\varepsilon\mathsf{Ent}|(\mu) < +\infty$.

By [5, Theorem 9.3], property (ii) holds true if $(X, \mathsf{d}, \mathfrak{m})$ is $CD(K,\infty)$ for some $K \in \mathbb{R}$. As the following result shows, (non-commutative) Carnot groups are not $CD(K,\infty)$, so that the validity of property (ii) in these metric measure spaces is an open problem. Note that Proposition 3.6 below was already known for the Heisenberg groups, see [24].

Proposition 3.6. If $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ is a non-commutative Carnot group, then the metric measure space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ is not $CD(K,\infty)$ for any $K \in \mathbb{R}$.

Proof. By contradiction, assume that $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ is a $CD(K,\infty)$ space for some $K \in \mathbb{R}$. Since the Dirichlet–Cheeger energy associate to the horizontal gradient is *quadratic* on $L^2(\mathbb{G}, \mathscr{L}^n)$ (see [6, Section 4.3] for a definition), by [3, Theorem 6.1] we deduce that $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ is a (σ-finite) $RCD(K,\infty)$ space. By [3, Theorem 7.2], we deduce that $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ satisfies the $BE(K,\infty)$ property, that is,

$$\|\nabla_{\mathbb{G}}(P_t f)\|^2_{\mathbb{G}} \leq e^{-2Kt} P_t(\|\nabla f\|^2_{\mathbb{G}}), \qquad \text{for all } t \geq 0, \ f \in C^\infty_c(\mathbb{R}^n). \tag{3.11}$$

Here and in the rest of the proof, we set $P_t f := f \star \mathsf{h}_t$ for short. Arguing similarly as in the proof of [41, Theorem 1.1], it is possible to prove that (3.11) is equivalent to the following *reverse Poincaré inequality*

$$P_t(f^2) - (P_t f)^2 \geq 2I_{2K}(t)\,\|\nabla(P_t f)\|^2_{\mathbb{G}}, \qquad \text{for all } t \geq 0, \ f \in C^\infty_c(\mathbb{R}^n), \tag{3.12}$$

where $I_K(t) := \frac{e^{Kt}-1}{K}$ if $K \neq 0$ and $I_0(t) := t$. Now, by [8, Propositions 2.5 and 2.6], there exists a constant $\Lambda \in \left[\frac{Q}{2m_1}, \frac{Q}{m_1}\right]$ (where Q and m_1 are as in Section 2.4) such that the

inequality

$$P_t(f^2) - (P_t f)^2 \geq \frac{t}{\Lambda} \|\nabla(P_t f)\|_{\mathbb{G}}^2, \qquad \text{for all } t \geq 0, \ f \in C_c^\infty(\mathbb{R}^n), \qquad (3.13)$$

holds true and, moreover, is sharp. Comparing (3.12) and (3.13), we thus must have that $\Lambda \leq \frac{t}{2I_{2K}(t)}$ for all $t > 0$. Passing to the limit as $t \to 0^+$, we get that $\Lambda \leq \frac{1}{2}$, so that $Q \leq m_1$. This immediately implies that $\mathbb{G}$ is commutative, a contradiction. $\qquad \square$

4. PROOF OF THE MAIN RESULT

4.1. **Heat diffusions are gradient flows of the entropy.** In this section we prove the first part of Theorem 2.4. The argument follows the strategy outlined in [25, Section 4.1].

The following technical lemma will be applied to horizontal vector fields in the proof of Proposition 4.2 below. The proof is exactly the same of [25, Lemma 4.1] and we omit it.

Lemma 4.1. *Let $V \colon \mathbb{R}^n \to \mathbb{R}^n$ be a vector field with locally Lipschitz coefficients such that $|V|_{\mathbb{R}^n} \in L^1(\mathbb{R}^n)$ and $\operatorname{div} V \in L^1(\mathbb{R}^n)$. Then $\int_{\mathbb{R}^n} \operatorname{div} V \ dx = 0$.*

Proposition 4.2 below states that the function $t \mapsto \mathsf{Ent}(\varrho_t \mathscr{L}^n)$ is locally absolutely continuous if $(\varrho_t)_{t \geq 0}$ solves the sub-elliptic heat equation (2.19) with initial datum $\varrho_0 \in L^1(\mathbb{G})$ such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathscr{P}_2(\mathbb{G})$. By [5, Proposition 4.22], this result is true under the stronger assumption that $\varrho_0 \in L^1(\mathbb{G}) \cap L^2(\mathbb{G})$. Here the point is to remove the L^2-integrability condition on the initial datum exploiting the estimates on the heat kernel collected in Theorem 2.3, see also [25, Section 4.1.1].

Proposition 4.2 (Entropy dissipation). *Let $\varrho_0 \in L^1(\mathbb{G})$ be such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathrm{Dom}(\mathsf{Ent})$. If $(\varrho_t)_{t \geq 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t$ with initial datum ϱ_0, then the map $t \mapsto \mathsf{Ent}(\mu_t)$, $\mu_t = \varrho_t \mathscr{L}^n$, is locally absolutely continuous on $(0, +\infty)$ and it holds*

$$\frac{d}{dt} \mathsf{Ent}(\mu_t) = -\int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \ dx \qquad \text{for a.e. } t > 0. \qquad (4.1)$$

Proof. Note that $\mu_t \in \mathscr{P}_2(\mathbb{G})$ for all $t > 0$ by (2.20). Hence $\mathsf{Ent}(\mu_t) > -\infty$ for all $t > 0$. Since $C_t := \sup_{x \in \mathbb{G}} \mathsf{h}_t(x) < +\infty$ for each fixed $t > 0$ by (2.15), we get that $\varrho_t \leq C_t$ for all $t > 0$. Thus $\mathsf{Ent}(\mu_t) < +\infty$ for all $t > 0$.

For each $m \in \mathbb{N}$, define

$$z_m(r) := \min\{m, \max\{1 + \log r, -m\}\}, \qquad H_m(r) := \int_0^r z_m(s) \ ds, \qquad r \geq 0. \qquad (4.2)$$

Note that H_m is of class C^1 on $[0, +\infty)$ with H_m' is globally Lipschitz and bounded. We claim that

$$\frac{d}{dt} \int_{\mathbb{G}} H_m(\varrho_t) \ dx = \int_{\mathbb{G}} z_m(\varrho_t) \Delta_{\mathbb{G}} \varrho_t \ dx \qquad \forall t > 0, \ \forall m \in \mathbb{N}. \qquad (4.3)$$

Indeed, we have $|H_m(\varrho_t)| \leq m \varrho_t \in L^1(\mathbb{G})$ and, given $[a, b] \subset (0, +\infty)$, by (2.17) the function $x \mapsto \sup_{t \in [a,b]} |\Delta_{\mathbb{G}} \mathsf{h}_t(x)|$ is bounded. Thus

$$\sup_{t \in [a,b]} \left| \frac{d}{dt} H_m(\varrho_t) \right| \leq m \sup_{t \in [a,b]} (\varrho_0 \star |\Delta_{\mathbb{G}} \mathsf{h}_t|) \leq m \, \varrho_0 \star \sup_{t \in [a,b]} |\Delta_{\mathbb{G}} \mathsf{h}_t| \in L^1(\mathbb{G}).$$

Therefore (4.3) follows by differentiation under integral sign. We now claim that

$$\int_{\mathbb{G}} z_m(\varrho_t)\Delta_{\mathbb{G}}\varrho_t \, dx = - \int_{\{e^{-m-1}<\varrho_t<e^{m-1}\}} \frac{\|\nabla_{\mathbb{G}}\varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \qquad \forall t > 0, \ \forall m \in \mathbb{N}. \qquad (4.4)$$

Indeed, by Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\frac{\|\nabla_{\mathbb{G}}\varrho_t(x)\|_{\mathbb{G}}^2}{\varrho_t(x)} &\leq \frac{[(\varrho_0 \star \|\nabla_{\mathbb{G}}\mathsf{h}_t\|_{\mathbb{G}})(x)]^2}{(\varrho_0 \star \mathsf{h}_t)(x)} \\
&= \frac{1}{(\varrho_0 \star \mathsf{h}_t)(x)} \left[\int_{\mathbb{G}} \sqrt{\varrho_0(xy^{-1})} \frac{\|\nabla_{\mathbb{G}}\mathsf{h}_t(y)\|_{\mathbb{G}}}{\sqrt{\mathsf{h}_t(y)}} \cdot \sqrt{\varrho_0(xy^{-1})} \sqrt{\mathsf{h}_t(y)} \, dy \right]^2 \\
&\leq \frac{1}{(\varrho_0 \star \mathsf{h}_t)(x)} \left(\int_{\mathbb{G}} \varrho_0(xy^{-1}) \frac{\|\nabla_{\mathbb{G}}\mathsf{h}_t(y)\|_{\mathbb{G}}^2}{\mathsf{h}_t(y)} \, dy \right) \left(\int_{\mathbb{G}} \varrho_0(xy^{-1}) \mathsf{h}_t(y) \, dy \right) \\
&\leq \left(\varrho_0 \star \frac{\|\nabla_{\mathbb{G}}\mathsf{h}_t\|_{\mathbb{G}}^2}{\mathsf{h}_t} \right)(x) \qquad \text{for all } x \in \mathbb{G}.
\end{aligned} \qquad (4.5)$$

Thus, by (2.16) and (2.17), we get

$$\int_{\{e^{-m-1}<\varrho_t<e^{m-1}\}} \frac{\|\nabla_{\mathbb{G}}\varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \leq \int_{\mathbb{G}} \varrho_0 \star \frac{\|\nabla_{\mathbb{G}}\mathsf{h}_t\|_{\mathbb{G}}^2}{\mathsf{h}_t} \, dx = \int_{\mathbb{G}} \frac{\|\nabla_{\mathbb{G}}\mathsf{h}_t\|_{\mathbb{G}}^2}{\mathsf{h}_t} \, dx < +\infty. \quad (4.6)$$

This, together with (4.3), proves that

$$\text{div}(z_m(\varrho_t)\nabla_{\mathbb{G}}\varrho_t) = z_m'(\varrho_t)\|\nabla_{\mathbb{G}}\varrho_t\|_{\mathbb{G}}^2 - z_m(\varrho_t)\Delta_{\mathbb{G}}\varrho_t \in L^1(\mathbb{G}). \qquad (4.7)$$

Thus (4.4) follows by integration by parts provided that

$$\int_{\mathbb{G}} \text{div}(z_m(\varrho_t)\nabla_{\mathbb{G}}\varrho_t) \, dx = 0. \qquad (4.8)$$

To prove (4.8), we apply Lemma 4.1 to the vector field $V = z_m(\varrho_t)\nabla_{\mathbb{G}}\varrho_t$. By (4.7), we already know that $\text{div}\, V \in L^1(\mathbb{G})$, so we just need to prove that $|V|_{\mathbb{R}^n} \in L^1(\mathbb{G})$. Note that

$$\int_{\mathbb{G}} |V|_{\mathbb{R}^n} \, dx \leq m \int_{\mathbb{G}} \varrho_0 \star |\nabla_{\mathbb{G}}\mathsf{h}_t|_{\mathbb{R}^n} \, dx = m \int_{\mathbb{G}} |\nabla_{\mathbb{G}}\mathsf{h}_t|_{\mathbb{R}^n} \, dx,$$

so it is enough to prove that $|\nabla_{\mathbb{G}}\mathsf{h}_t|_{\mathbb{R}^n} \in L^1(\mathbb{G})$. But we have

$$|\nabla_{\mathbb{G}}\mathsf{h}_t(x)|_{\mathbb{R}^n} \leq p(x_1,\dots,x_n)\|\nabla_{\mathbb{G}}\mathsf{h}_t(x)\|_{\mathbb{G}}, \qquad x \in \mathbb{G},$$

where $p\colon \mathbb{R}^n \to [0,+\infty)$ is a function with polynomial growth, because the horizontal vector fields $X_1,\dots,X_{h_1}$ have polynomial coefficients. Since d_{cc} is equivalent to d_∞, where d_∞ was introduced in (2.11), by (2.17) we conclude that $|\nabla_{\mathbb{G}}\mathsf{h}_t|_{\mathbb{R}^n} \in L^1(\mathbb{G})$. This completes the proof of (4.8).

Combining (4.3) and (4.4), we thus get

$$\frac{d}{dt} \int_{\mathbb{G}} H_m(\varrho_t) \, dx = - \int_{\{e^{-m-1}<\varrho_t<e^{m-1}\}} \frac{\|\nabla_{\mathbb{G}}\varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \qquad \forall t > 0, \ \forall m \in \mathbb{N}.$$

Note that

$$t \mapsto \int_{\mathbb{G}} \frac{\|\nabla_{\mathbb{G}}\varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \in L^1_{\text{loc}}(0,+\infty). \qquad (4.9)$$

Indeed, from (2.16) and (2.17) we deduce that

$$t \mapsto \int_{\mathbb{G}} \frac{\|\nabla_{\mathbb{G}} \mathsf{h}_t\|_{\mathbb{G}}^2}{\mathsf{h}_t} \, dx \in L^1_{\mathrm{loc}}(0, +\infty).$$

Recalling (4.5) and (4.6), this immediately implies (4.9). Therefore

$$\int_{\mathbb{G}} H_m(\varrho_{t_1}) \, dx - \int_{\mathbb{G}} H_m(\varrho_{t_0}) \, dx = - \int_{t_0}^{t_1} \int_{\{e^{-m-1} < \varrho_t < e^{m-1}\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx dt$$

for any $t_0, t_1 \in (0, +\infty)$ with $t_0 < t_1$ and $m \in \mathbb{N}$. We now pass to the limit as $m \to +\infty$. Note that $H_m(r) \to r \log r$ as $m \to +\infty$ and that for all $m \in \mathbb{N}$

$$r \log r \le H_{m+1}(r) \le H_m(r) \qquad \text{for } r \in [0, 1] \tag{4.10}$$

and

$$0 \le H_m(r) \le 1 + r \log r \qquad \text{for } r \in [1, +\infty). \tag{4.11}$$

Thus

$$\lim_{m \to +\infty} \int_{\mathbb{G}} H_m(\varrho_t) \, dx = \int_{\mathbb{G}} \varrho_t \log \varrho_t \, dx$$

by the monotone convergence theorem on $\{\varrho_t \le 1\}$ and by the dominated convergence theorem on $\{\varrho_t > 1\}$. Moreover

$$\lim_{m \to +\infty} \int_{t_0}^{t_1} \int_{\{e^{-m-1} < \varrho_t < e^{m-1}\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx dt = \int_{t_0}^{t_1} \int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx dt$$

by the monotone convergence theorem. This concludes the proof. $\qquad \square$

We are now ready to prove the first part of Theorem 2.4. The argument follows the strategy outlined in [25, Section 4.1]. See also the first part of the proof of [5, Theorem 8.5].

Theorem 4.3. *Let $\varrho_0 \in L^1(\mathbb{G})$ be such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathrm{Dom}(\mathsf{Ent})$. If $(\varrho_t)_{t \ge 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t$ with initial datum ϱ_0, then $\mu_t = \varrho_t \mathscr{L}^n$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$ starting from μ_0.*

Proof. Note that $(\mu_t)_{t>0} \subset \mathscr{P}_2(\mathbb{G})$, see the proof of Proposition 4.2. Moreover, $(\mu_t)_{t>0}$ satisfies (3.4) with $v_t^{\mathbb{G}} = \nabla_{\mathbb{G}} \varrho_t / \varrho_t$ for $t > 0$. Note that $t \mapsto \|v_t^{\mathbb{G}}\|_{L^2_{\mathbb{G}}(\mu_t)} \in L^2_{\mathrm{loc}}(0, +\infty)$ by (2.16), (2.17), (4.5) and (4.6). By Proposition 3.3 we conclude that

$$|\dot{\mu}_t|^2 \le \int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \qquad \text{for a.e. } t > 0. \tag{4.12}$$

By Proposition 4.2, the map $t \mapsto \mathsf{Ent}(\mu_t)$ is locally absolutely continuous on $(0, +\infty)$ and so, by the chain rule, we get

$$-\frac{d}{dt}\mathsf{Ent}(\mu_t) \le |\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu_t) \cdot |\dot{\mu}_t|_{\mathbb{G}} \qquad \text{for a.e. } t > 0. \tag{4.13}$$

Thus, if we prove that

$$|\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|^2(\mu_t) = \int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t\|_{\mathbb{G}}^2}{\varrho_t} \, dx \qquad \text{for a.e. } t > 0 \tag{4.14}$$

then, combining this equality with (4.1), (4.12) and (4.13), we find that

$$|\dot{\mu}_t|_{\mathbb{G}} = |\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu_t), \qquad \frac{d}{dt}\mathsf{Ent}(\mu_t) = -|\mathrm{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu_t) \cdot |\dot{\mu}_t|_{\mathbb{G}} \qquad \text{for a.e. } t > 0,$$

so that $(\mu_t)_{t\geq 0}$ is a gradient flow of Ent starting from μ_0 as observed in Remark 2.2.

We now prove (4.14). To do so, we apply Proposition 3.5. We need to check that $|D_\varepsilon^- \mathsf{Ent}|(\mu_t) < +\infty$ for some $\varepsilon > 0$. To prove this, we apply Proposition 3.2. Since $\mathsf{d}_\varepsilon \leq \mathsf{d}_{\mathsf{cc}}$, we have $\mu_t \in \mathscr{P}_2(\mathbb{G}_\varepsilon)$ for all $\varepsilon > 0$. Moreover

$$\mathsf{F}_\varepsilon(\varrho_t) = \mathsf{F}_\mathbb{G}(\varrho_t) + \sum_{i=2}^k \varepsilon^{2(i-1)} \int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{V_i}\varrho_t\|_\mathbb{G}^2}{\varrho_t}\, dx.$$

Since $\mathsf{F}_\mathbb{G}(\varrho_t) < +\infty$, we just need to prove that

$$\int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{V_i}\varrho_t\|_\mathbb{G}^2}{\varrho_t}\, dx < +\infty$$

for all $i = 2, \ldots, \kappa$. Indeed, arguing as in (4.5), by Cauchy–Schwarz inequality we have

$$\frac{\|\nabla_{V_i}\varrho_t\|_\mathbb{G}^2}{\varrho_t} \leq \frac{(\varrho \star \|\nabla_{V_i}\mathsf{h}_t\|_\mathbb{G})^2}{\varrho \star \mathsf{h}_t} = \frac{1}{\varrho \star \mathsf{h}_t} \left[\varrho \star \left(\frac{\|\nabla_{V_i}\mathsf{h}_t\|_\mathbb{G}}{\sqrt{\mathsf{h}_t}} \sqrt{\mathsf{h}_t} \right) \right]^2 \leq \varrho \star \frac{\|\nabla_{V_i}\mathsf{h}_t\|_\mathbb{G}^2}{\mathsf{h}_t}.$$

Therefore, by (2.16) and (2.17), we get

$$\int_{\mathbb{G} \cap \{\varrho_t > 0\}} \frac{\|\nabla_{V_i}\varrho_t\|_\mathbb{G}^2}{\varrho_t}\, dx \leq \int_\mathbb{G} \varrho \star \frac{\|\nabla_{V_i}\mathsf{h}_t\|_\mathbb{G}^2}{\mathsf{h}_t}\, dx = \int_\mathbb{G} \frac{\|\nabla_{V_i}\mathsf{h}_t\|_\mathbb{G}^2}{\mathsf{h}_t}\, dx < +\infty.$$

This concludes the proof. $\qquad\square$

4.2. Gradient flows of the entropy are heat diffusions.

In this section we prove the second part of Theorem 2.4. Our argument is different from the one presented in [25, Section 4.2]. However, as observed in [25, Remark 5.3], the techniques developed in [25, Section 4.2] can be adapted in order to obtain a proof of Theorem 4.8 below for any Carnot group $\mathbb{G}$ of step 2.

Let us start with the following remark. If $(\mu_t)_{t\geq 0}$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_\mathbb{G})$ then, recalling Definition 2.1, we have that $\mu_t \in \mathrm{Dom}(\mathsf{Ent})$ for all $t \geq 0$. By (2.6), this means that $\mu_t = \varrho_t \mathscr{L}^n$ for some probability density $\varrho_t \in L^1(\mathbb{G})$ for all $t \geq 0$. In addition, $t \mapsto |D_\mathbb{G}^-\mathsf{Ent}|(\mu_t) \in L^2_{\mathrm{loc}}([0, +\infty))$ and the function $t \mapsto \mathsf{Ent}(\mu_t)$ is non-increasing, therefore a.e. differentiable and locally integrable on $[0, +\infty)$.

Lemma 4.4 below shows that it is enough to establish (4.15) in order to prove the second part of Theorem 2.4. For the proof, see also the last paragraph of [25, Section 4.2].

Lemma 4.4. *Let $\varrho_0 \in L^1(\mathbb{G})$ be such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathrm{Dom}(\mathsf{Ent})$. Assume $(\mu_t)_{t\geq 0}$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_\mathbb{G})$ starting from μ_0, with $\mu_t = \varrho_t \mathscr{L}^n$ for all $t \geq 0$. Let $(v_t^\mathbb{G})_{t>0}$ and $(w_t^\mathbb{G})_{t>0}$, $w_t^\mathbb{G} = \nabla_\mathbb{G}\varrho_t/\varrho_t$, be the horizontal time-dependent vector fields given by Proposition 3.3 and Proposition 3.4 respectively. If it holds*

$$-\frac{d}{dt}\mathsf{Ent}(\mu_t) \leq \int_\mathbb{G} \left\langle -w_t^\mathbb{G}, v_t^\mathbb{G} \right\rangle_\mathbb{G}\, d\mu_t \qquad \text{for a.e. } t > 0, \tag{4.15}$$

then $(\varrho_t)_{t\geq 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_\mathbb{G}\varrho_t$ with initial datum ϱ_0.

Proof. From Definition 2.1 we get that

$$\mathsf{Ent}(\mu_t) + \frac{1}{2}\int_s^t |\dot\mu_r|_\mathbb{G}^2\, dr + \frac{1}{2}\int_s^t |D^-\mathsf{Ent}|_\mathbb{G}^2(\mu_r)\, dr \leq \mathsf{Ent}(\mu_s)$$

for all $s, t \geq 0$ with $s \leq t$. Therefore

$$-\frac{d}{dt}\mathsf{Ent}(\mu_t) \geq \frac{1}{2}|\dot\mu_t|^2_\mathbb{G} + \frac{1}{2}|\mathrm{D}^-\mathsf{Ent}|^2_\mathbb{G}(\mu_t) \qquad \text{for a.e. } t > 0.$$

By Young's inequality, Proposition 3.3 and Proposition 3.4, we thus get

$$-\frac{d}{dt}\mathsf{Ent}(\mu_t) \geq |\dot\mu_t|_\mathbb{G} \cdot |\mathrm{D}^-_\mathbb{G}\mathsf{Ent}|(\mu_t) \geq \|v^\mathbb{G}_t\|_{L^2_\mathbb{G}(\mu_t)} \cdot \|w^\mathbb{G}_t\|_{L^2_\mathbb{G}(\mu_t)} \qquad \text{for a.e. } t > 0. \quad (4.16)$$

Combining (4.15) and (4.16), by Cauchy–Schwarz inequality we conclude that $v^\mathbb{G}_t = -w^\mathbb{G}_t = -\nabla_\mathbb{G}\varrho_t/\varrho_t$ in $L^2_\mathbb{G}(\mu_t)$ for a.e. $t > 0$. This immediately implies that $(\varrho_t)_{t\geq 0}$ solves the sub-elliptic heat equation $\partial_t\varrho_t = \Delta_\mathbb{G}\varrho_t$ with initial datum ϱ_0 *in the sense of distributions*, i.e.

$$\int_0^{+\infty}\int_\mathbb{G} \partial_t\varphi_t(x) + \Delta_\mathbb{G}\varphi_t(x)\ d\mu_t\,dt + \int_\mathbb{G}\varphi_0(x)\ d\mu_0(x) = 0 \qquad \forall\varphi \in C^\infty_c([0,+\infty)\times\mathbb{R}^n).$$

By well-known results on hypoelliptic operators, this implies that $(\varrho_t)_{t\geq 0}$ solves the sub-elliptic heat equation $\partial_t\varrho_t = \Delta_\mathbb{G}\varrho_t$ with initial datum ϱ_0. $\qquad\square$

To prove Theorem 4.8 below we need some preliminaries. The following two lemmas are natural adaptations of [7, Lemma 2.14] to our setting.

Lemma 4.5. *Let $\mu \in \mathscr{P}(\mathbb{G})$ and $\sigma \in L^1(\mathbb{G})$ with $\sigma \geq 0$. Let $\nu \in \mathscr{M}(\mathbb{G};\mathbb{R}^m)$ be a $\mathbb{R}^m$-valued Borel measure with finite total variation and such that $|\nu| \ll \mu$. Then*

$$\int_\mathbb{G}\left|\frac{\sigma\star\nu}{\sigma\star\mu}\right|^2 \sigma\star\mu\ dx \leq \int_\mathbb{G}\left|\frac{\nu}{\mu}\right|^2 d\mu. \qquad (4.17)$$

In addition, if $(\sigma_k)_{k\in\mathbb{N}} \subset L^1(\mathbb{G})$, $\sigma_k \geq 0$, weakly converges to the Dirac mass δ_0 and $\frac{\nu}{\mu} \in L^2(\mathbb{G},\mu)$, then

$$\lim_{k\to+\infty}\int_\mathbb{G}\left|\frac{\sigma_k\star\nu}{\sigma_k\star\mu}\right|^2 \sigma_k\star\mu\ dx = \int_\mathbb{G}\left|\frac{\nu}{\mu}\right|^2 d\mu. \qquad (4.18)$$

Proof. Inequality (4.17) follows from Jensen inequality and is proved in [7, Lemma 2.14]. We briefly recall the argument for the reader's convenience. Consider the map $\Phi\colon \mathbb{R}^m \times \mathbb{R} \to [0,+\infty]$ given by

$$\Phi(z,t) := \begin{cases} \dfrac{|z|^2}{t} & \text{if } t > 0, \\ 0 & \text{if } (z,t) = (0,0), \\ +\infty & \text{if either } t < 0 \text{ or } t = 0,\ z \neq 0. \end{cases}$$

Then Φ is convex, lower semicontinuous and positively 1-homogeneous. By Jensen's inequality we have

$$\Phi\left(\int_\mathbb{G}\psi(x)\ d\vartheta(x)\right) \leq \int_\mathbb{G}\Phi(\psi(x))\ d\vartheta(x) \qquad (4.19)$$

for any Borel function $\psi\colon \mathbb{G} \to \mathbb{R}^{m+1}$ and any positive and finite measure ϑ on $\mathbb{G}$. Fix $x \in \mathbb{G}$ and apply (4.19) with $\psi(y) = \left(\frac{\nu}{\mu}(y), 1\right)$ and $d\vartheta(y) = \sigma(xy^{-1})d\mu(y)$ to obtain

$$\left|\frac{(\sigma\star\nu)(x)}{(\sigma\star\mu)(x)}\right|^2 (\sigma\star\mu)(x) = \Phi\left(\int_\mathbb{G}\frac{\nu}{\mu}(y)\,\sigma(xy^{-1})\ d\mu(y), \int_\mathbb{G}\sigma(xy^{-1})\ d\mu(y)\right)$$

$$\leq \int_{\mathbb{G}} \Phi\left(\frac{\nu}{\mu}(y), 1\right) \sigma(xy^{-1}) \, d\mu(y) = \int_{\mathbb{G}} \left|\frac{\nu}{\mu}\right|^2 (y) \, \sigma(xy^{-1}) \, d\mu(y),$$

which immediately gives (4.17). The limit in (4.18) follows by the joint lower semicontinuity of the functional $(\nu, \mu) \mapsto \int_{\mathbb{G}} \left|\frac{\nu}{\mu}\right|^2 d\mu$, see Theorem 2.34 and Example 2.36 in [1]. $\quad\square$

In Lemma 4.6 below and in the rest of the paper, we let $f * g$ be the convolution of the two functions f, g with respect to the time variable. We keep the notation $f \star g$ for the convolution of f, g with respect to the space variable.

Lemma 4.6. *Let $\mu_t = \varrho_t \mathscr{L}^n \in \mathscr{P}(\mathbb{G})$ for all $t \in \mathbb{R}$ and let $\vartheta \in L^1(\mathbb{R})$, $\vartheta \geq 0$. If the horizontal time-dependent vector field $v \colon \mathbb{R} \times \mathbb{G} \to H\mathbb{G}$ satisfies $v_t \in L^2_{\mathbb{G}}(\mu_t)$ for a.e. $t \in \mathbb{R}$, then*

$$\int_{\mathbb{G}} \left\|\frac{\vartheta * (\varrho.v.)(t)}{\vartheta * \varrho.(t)}\right\|^2_{\mathbb{G}} \vartheta * \varrho.(t) \, dx \leq \vartheta * \left(\int_{\mathbb{G}} \|v.\|^2_{\mathbb{G}} \, d\mu.\right)(t) \qquad \text{for all } t \in \mathbb{R}. \qquad (4.20)$$

In addition, if $(\vartheta_j)_{j\in\mathbb{N}} \subset L^1(\mathbb{G})$, $\vartheta_j \geq 0$, weakly converges to the Dirac mass δ_0, then

$$\lim_{j\to+\infty} \int_{\mathbb{G}} \left\|\frac{\vartheta_j * (\varrho.v.)(t)}{\vartheta_j * \varrho.(t)}\right\|^2_{\mathbb{G}} \vartheta_j * \varrho.(t) \, dx = \int_{\mathbb{G}} \|v_t\|^2_{\mathbb{G}} \, d\mu_t \qquad \text{for a.e. } t \in \mathbb{R}. \qquad (4.21)$$

Proof. Inequality (4.20) follows from (4.19) in the same way of (4.17), so we omit the details. For (4.21), set $\mu_t^j = \vartheta_j * \mu.(t)$ and $\nu_t^j = \vartheta_j * (v.\mu.)(t)$ for all $t \in \mathbb{R}$ and $j \in \mathbb{N}$. Then $\|\nu_t^j\|_{\mathbb{G}} \ll \mu_t^j$ and $\nu_t^j \rightharpoonup \nu_t = v_t \mu_t$ for a.e. $t \in \mathbb{R}$, so that

$$\liminf_{j\to+\infty} \int_{\mathbb{G}} \left\|\frac{\vartheta_j * (\varrho.v.)(t)}{\vartheta_j * \varrho.(t)}\right\|^2_{\mathbb{G}} \vartheta_j * \varrho.(t) \, dx = \liminf_{j\to+\infty} \int_{\mathbb{G}} \left\|\frac{\nu_t^j}{\mu_t^j}\right\|^2_{\mathbb{G}} d\mu_t^j \geq \int_{\mathbb{G}} \left\|\frac{\nu_t}{\mu_t}\right\|^2_{\mathbb{G}} d\mu_t$$

for a.e. $t \in \mathbb{R}$ by Theorem 2.34 and Example 2.36 in [1]. $\quad\square$

The following lemma is an elementary result relating weak convergence and convergence of scalar products of vector fields. We prove it here for the reader's convenience.

Lemma 4.7. *For $k \in \mathbb{N}$, let $\mu_k, \mu \in \mathscr{P}(\mathbb{G})$ and let $v_k, w_k, v, w \colon \mathbb{G} \to T\mathbb{G}$ be Borel vector fields. Assume that $\mu_k \rightharpoonup \mu$, $v_k\mu_k \rightharpoonup v\mu$ and $w_k\mu_k \rightharpoonup w\mu$ as $k \to +\infty$. If*

$$\limsup_{k\to+\infty} \int_{\mathbb{G}} \|v_k\|^2_{\mathbb{G}} \, d\mu_k \leq \int_{\mathbb{G}} \|v\|^2_{\mathbb{G}} \, d\mu < +\infty \qquad \text{and} \qquad \limsup_{k\to+\infty} \int_{\mathbb{G}} \|w_k\|^2_{\mathbb{G}} \, d\mu_k < +\infty,$$

then

$$\lim_{k\to+\infty} \int_{\mathbb{G}} \langle v_k, w_k \rangle_{\mathbb{G}} \, d\mu_k = \int_{\mathbb{G}} \langle v, w \rangle_{\mathbb{G}} \, d\mu. \qquad (4.22)$$

Proof. By lower semicontinuity, we know that $\lim_{k\to+\infty} \int_{\mathbb{G}} \|v_k\|^2_{\mathbb{G}} \, d\mu_k = \int_{\mathbb{G}} \|v\|^2_{\mathbb{G}} \, d\mu$ and

$$\liminf_{k\to+\infty} \int_{\mathbb{G}} \|tv_k + w_k\|^2_{\mathbb{G}} \, d\mu_k \geq \int_{\mathbb{G}} \|tv + w\|^2_{\mathbb{G}} \, d\mu \qquad \text{for all } t \in \mathbb{R}.$$

Expanding the squares, we get

$$\liminf_{k\to+\infty} \left(2t \int_{\mathbb{G}} \langle v_k, w_k \rangle_{\mathbb{G}} \, d\mu_k + \int_{\mathbb{G}} \|w_k\|^2_{\mathbb{G}} \, d\mu_k\right) \geq 2t \int_{\mathbb{G}} \langle v, w \rangle_{\mathbb{G}} \, d\mu \qquad \text{for all } t \in \mathbb{R}.$$

Choosing $t > 0$, dividing both sides by t and letting $t \to +\infty$ gives the $\liminf$ inequality in (4.22). Choosing $t < 0$, a similar argument gives the $\limsup$ inequality in (4.22). $\quad\square$

We are now ready to prove the second part of Theorem 2.4.

Theorem 4.8. *Let $\varrho_0 \in L^1(\mathbb{G})$ be such that $\mu_0 = \varrho_0 \mathscr{L}^n \in \mathrm{Dom}(\mathsf{Ent})$. If $(\mu_t)_{t \geq 0}$ is a gradient flow of Ent in $(\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}})$ starting from μ_0, then $\mu_t = \varrho_t \mathscr{L}^n$ for all $t \geq 0$ and $(\varrho_t)_{t \geq 0}$ solves the sub-elliptic heat equation $\partial_t \varrho_t = \Delta_{\mathbb{G}} \varrho_t$ with initial datum ϱ_0. In particular, $t \mapsto \mathsf{Ent}(\mu_t)$ is locally absolutely continuous on $(0, +\infty)$.*

Proof. By Lemma 4.4, we just need to show that the map $t \mapsto \mathsf{Ent}(\mu_t)$ satisfies (4.15). It is not restrictive to extend $(\mu_t)_{t \geq 0}$ in time to the whole $\mathbb{R}$ by setting $\mu_t = \mu_0$ for all $t \leq 0$. So from now on we assume $\mu_t \in AC^2_{\mathrm{loc}}(\mathbb{R}; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$. The time-dependent vector field $(v^{\mathbb{G}})_{t > 0}$ given by Proposition 3.3 extends to the whole $\mathbb{R}$ accordingly. Note that $(\mu_t)_{t \in \mathbb{R}}$ is a gradient flow of Ent in the following sense: for each $h \in \mathbb{R}$, $(\mu_{t+h})_{t \geq 0}$ is a gradient flow on Ent starting from μ_h. By Definition 2.1, we get $t \mapsto |\mathsf{D}_{\mathbb{G}}^- \mathsf{Ent}|(\mu_t) \in L^2_{\mathrm{loc}}(\mathbb{R})$, so that $t \mapsto \mathsf{F}_{\mathbb{G}}(\varrho_t) \in L^1_{\mathrm{loc}}(\mathbb{R})$ by Proposition 3.4.

We divide the proof in three main steps.

Step 1: smoothing in the time variable. Let $\vartheta \colon \mathbb{R} \to \mathbb{R}$ be a symmetric smooth mollifier in $\mathbb{R}$, i.e.

$$\vartheta \in C_c^\infty(\mathbb{R}), \quad \mathrm{supp}\,\vartheta \subset [-1, 1], \quad 0 \leq \vartheta \leq 1, \quad \int_{\mathbb{R}} \vartheta(t)\,dt = 1.$$

We set $\vartheta_j(t) := j\,\vartheta(jt)$ for all $t \in \mathbb{R}$ and $j \in \mathbb{N}$. We define

$$\mu_t^j := \varrho_t^j \mathscr{L}^n, \qquad \varrho_t^j := (\vartheta_j * \varrho.)(t) = \int_{\mathbb{R}} \vartheta_j(t - s)\,\varrho_s\,ds \qquad \forall t \in \mathbb{R},\ \forall j \in \mathbb{N}.$$

For any $s, t \in \mathbb{R}$, let $\pi_{s,t} \in \Gamma_0(\mu_s, \mu_t)$ be an optimal coupling between μ_s and μ_t. An easy computation shows that $\pi_t^j \in \mathscr{P}(\mathbb{G} \times \mathbb{G})$ given by

$$\int_{\mathbb{G} \times \mathbb{G}} \varphi(x, y)\,d\pi_t^j(x, y) = \int_{\mathbb{R}} \vartheta_j(t - s) \int_{\mathbb{G} \times \mathbb{G}} \varphi(x, y)\,d\pi_{s,t}(x, y)\,ds,$$

for any $\varphi \colon \mathbb{G} \times \mathbb{G} \to [0, +\infty)$ Borel, is a coupling between μ_t^j and μ_t. Hence we get

$$\mathsf{W}_{\mathbb{G}}(\mu_t^j, \mu_t)^2 \leq \int_{\mathbb{R}} \vartheta_j(t - s)\,\mathsf{W}_{\mathbb{G}}(\mu_s, \mu_t)^2\,ds \qquad \forall t \in \mathbb{R},\ \forall j \in \mathbb{N}.$$

Therefore $\lim_{j \to +\infty} \mathsf{W}_{\mathbb{G}}(\mu_t^j, \mu_t) = 0$ for all $t \in \mathbb{R}$. This implies that $\mu_t^j \rightharpoonup \mu_t$ as $j \to +\infty$ and

$$\lim_{j \to +\infty} \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x, 0)^2\,d\mu_t^j(x) = \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x, 0)^2\,d\mu_t(x) \qquad \forall t \in \mathbb{R}. \tag{4.23}$$

In particular, $(\mu_t^j)_{t \in \mathbb{R}} \subset \mathscr{P}_2(\mathbb{G})$, $\mathsf{Ent}(\mu_t^j) > -\infty$ for all $j \in \mathbb{N}$ and

$$\liminf_{j \to +\infty} \mathsf{Ent}(\mu_t^j) \geq \mathsf{Ent}(\mu_t) \qquad \forall t \in \mathbb{R}. \tag{4.24}$$

We claim that

$$\limsup_{j \to +\infty} \mathsf{Ent}(\mu_t^j) \leq \mathsf{Ent}(\mu_t) \qquad \text{for a.e. } t \in \mathbb{R}. \tag{4.25}$$

Indeed, define the new reference measure $\nu := e^{-c\,\mathsf{d}_{\mathsf{cc}}^2(\cdot, 0)} \mathscr{L}^n$, where $c > 0$ is chosen so that $\nu \in \mathscr{P}(\mathbb{G})$. Since the function $\hat{H}(r) := r \log r + (1 - r)$, for $r \geq 0$, is convex and non-negative, by Jensen's inequality we have

$$\mathsf{Ent}_\nu(\mu_t^j) = \int_{\mathbb{G}} \hat{H}\left(e^{c\,\mathsf{d}_{\mathsf{cc}}^2(\cdot, 0)}\,\vartheta_j * \varrho.(t)\right) d\nu = \int_{\mathbb{G}} \hat{H}\left(\vartheta_j * \left(e^{c\,\mathsf{d}_{\mathsf{cc}}^2(\cdot, 0)}\varrho.\right)(t)\right) d\nu$$

$$\leq \int_{\mathbb{G}} \vartheta_j * \hat{H}\left(e^{c\, \mathsf{d}_{cc}^2(\cdot,0)} \varrho_\cdot\right)(t)\, d\nu = \vartheta_j * \mathsf{Ent}_\nu(\mu_\cdot)(t) \qquad \forall t \in \mathbb{R},\ \forall j \in \mathbb{N}.$$

Therefore $\limsup_{j \to +\infty} \mathsf{Ent}_\nu(\mu_t^j) \leq \mathsf{Ent}_\nu(\mu_t)$ for a.e. $t \in \mathbb{R}$. Thus (4.25) follows by (2.8) and (4.23). Combining (4.24) and (4.25), we get

$$\lim_{j \to +\infty} \mathsf{Ent}(\mu_t^j) = \mathsf{Ent}(\mu_t) \qquad \text{for a.e. } t \in \mathbb{R}. \tag{4.26}$$

Let $(v_t^{\mathbb{G}})_{t \in \mathbb{R}}$ be the horizontal time-dependent vector field relative to $(\mu_t)_{t \in \mathbb{R}}$ given by Proposition 3.3. Let $(v_t^j)_{t \in \mathbb{R}}$ be the horizontal time-dependent vector field given by

$$v_t^j = \frac{\vartheta_j * (\varrho_\cdot v_\cdot)(t)}{\varrho_t^j} \qquad \forall t \in \mathbb{R}. \tag{4.27}$$

We claim that $v_t^j \in L_{\mathbb{G}}^2(\mu_t^j)$ for all $t \in \mathbb{R}$. Indeed, applying Lemma 4.6, we get

$$\int_{\mathbb{G}} \|v_t^j\|_{\mathbb{G}}^2\, d\mu_t^j \leq \vartheta_j * \left(\int_{\mathbb{G}} \|v_\cdot\|_{\mathbb{G}}^2\, d\mu_\cdot\right)(t) = \vartheta_j * |\dot{\mu}_\cdot|_{\mathbb{G}}^2(t) \qquad \forall t \in \mathbb{R}.$$

We also claim that $(\mu_t^j)_{t \in \mathbb{R}}$ solves $\partial_t \mu_t^j + \mathrm{div}(v_t^j \mu_t^j) = 0$ in the sense of distributions for all $j \in \mathbb{N}$. Indeed, if $\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^n)$, then also $\varphi^j := \vartheta_j * \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^n)$, so that

$$\int_{\mathbb{R}} \int_{\mathbb{G}} \partial_t \varphi^j(t,x) + \left\langle \nabla_{\mathbb{G}} \varphi^j(t,x), v_t^j(x) \right\rangle_{\mathbb{G}} d\mu_t^j(x)\, dt =$$

$$= \int_{\mathbb{R}} \vartheta_j * \left(\int_{\mathbb{G}} \partial_t \varphi(\cdot,x) + \left\langle \nabla_{\mathbb{G}} \varphi(\cdot,x), v_\cdot(x) \right\rangle_{\mathbb{G}} d\mu_\cdot(x)\right)(t)\, dt$$

$$= \int_{\mathbb{R}} \int_{\mathbb{G}} \partial_t \varphi(t,x) + \left\langle \nabla_{\mathbb{G}} \varphi(t,x), v_t(x) \right\rangle_{\mathbb{G}} d\mu_t(x)\, dt = 0 \qquad \forall j \in \mathbb{N}.$$

By Proposition 3.3, we conclude that $(\mu_t^j)_t \in AC_{\mathrm{loc}}^2(\mathbb{R}; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$ with $|\dot{\mu}_t^j|_{\mathbb{G}}^2 \leq \vartheta_j * |\dot{\mu}_\cdot|_{\mathbb{G}}^2(t)$ for all $t \in \mathbb{R}$ and $j \in \mathbb{N}$.

Finally, we claim that $\mathsf{F}_{\mathbb{G}}(\varrho_t^j) \leq \vartheta_j * \mathsf{F}_{\mathbb{G}}(\varrho_\cdot)(t)$ for all $t \in \mathbb{R}$ and $j \in \mathbb{N}$. Indeed, arguing as in (4.5), by Cauchy–Schwarz inequality we have

$$\|\nabla_{\mathbb{G}} \varrho_t^j\|_{\mathbb{G}}^2 \leq \left[\vartheta_j * \left(\chi_{\{\varrho_\cdot > 0\}} \sqrt{\varrho_\cdot} \frac{\|\nabla_{\mathbb{G}} \varrho_\cdot\|_{\mathbb{G}}}{\sqrt{\varrho_\cdot}}\right)\right]^2(t) \leq \varrho_t^j\, \vartheta_j * \left(\frac{\|\nabla_{\mathbb{G}} \varrho_\cdot\|_{\mathbb{G}}^2}{\varrho_\cdot} \chi_{\{\varrho_\cdot > 0\}}\right)(t),$$

so that

$$\mathsf{F}_{\mathbb{G}}(\varrho_t^j) \leq \vartheta_j * \left(\int_{\{\varrho_\cdot > 0\}} \frac{\|\nabla_{\mathbb{G}} \varrho_\cdot\|_{\mathbb{G}}^2}{\varrho_\cdot}\, dx\right)(t) = \vartheta_j * \mathsf{F}_{\mathbb{G}}(\varrho_\cdot)(t).$$

Step 2: smoothing in the space variable. Let $j \in \mathbb{N}$ be fixed. Let $\eta\colon \mathbb{G} \to \mathbb{R}$ be a symmetric smooth mollifier in $\mathbb{G}$, i.e. a function $\eta \in C_c^\infty(\mathbb{R}^n)$ such that

$$\mathrm{supp}\, \eta \subset B_1, \quad 0 \leq \eta \leq 1, \quad \eta(x^{-1}) = \eta(x)\ \forall x \in \mathbb{G}, \quad \int_{\mathbb{G}} \eta(x)\, dx = 1.$$

We set $\eta_k(x) := k^Q\, \eta(\delta_k(x))$ for all $x \in \mathbb{G}$ and $k \in \mathbb{N}$. We define

$$\mu_t^{j,k} := \varrho_t^{j,k} \mathscr{L}^n, \qquad \varrho_t^{j,k}(x) := \eta_k \star \varrho_t^j(x) = \int_{\mathbb{G}} \eta_k(xy^{-1}) \varrho_t^j(y)\, dy, \quad x \in \mathbb{G},$$

for all $t \in \mathbb{R}$ and $k \in \mathbb{N}$. Note that

$$\mu_t^{j,k} = \int_{\mathbb{G}} (l_y)_\# \mu_t^j\, \eta_k(y) dy \qquad \forall t \in \mathbb{R},\ \forall k \in \mathbb{N}, \tag{4.28}$$

where $l_y(x) = yx$, $x, y \in \mathbb{G}$, denotes the left-translation.

Note that $(\mu_t^{j,k})_{t \in \mathbb{R}} \subset \mathscr{P}_2(\mathbb{G})$ for all $k \in \mathbb{N}$. Indeed, arguing as in (2.20), we have

$$
\begin{aligned}
\int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2 \, d\mu_t^{j,k}(x) &= \int_{\mathbb{G}} \left(\mathsf{d}_{\mathsf{cc}}(\cdot,0)^2 \star \eta_k \right)(x) \, d\mu_t^j(x) \\
&\leq 2 \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2 \, d\mu_t^j(x) + 2 \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(x,0)^2 \, \eta_k(x) \, dx
\end{aligned}
\tag{4.29}
$$

for all $t \in \mathbb{R}$. In particular, $\mathsf{Ent}(\mu_t^{j,k}) > -\infty$ for all $t \in \mathbb{R}$ and $k \in \mathbb{N}$. Clearly $\mu_t^{j,k} \rightharpoonup \mu_t^j$ as $k \to +\infty$ for each fixed $t \in \mathbb{R}$, so that

$$
\liminf_{k \to +\infty} \mathsf{Ent}(\mu_t^{j,k}) \geq \mathsf{Ent}(\mu_t^j) \qquad \forall t \in \mathbb{R}.
\tag{4.30}
$$

Moreover, we claim that

$$
\limsup_{k \to +\infty} \mathsf{Ent}(\mu_t^{j,k}) \leq \mathsf{Ent}(\mu_t^j) \qquad \forall t \in \mathbb{R}.
\tag{4.31}
$$

Indeed, let $\nu \in \mathscr{P}(\mathbb{G})$ and $\hat{H}$ as in *Step 1*. Recalling (4.28), by Jensen's inequality we get

$$
\mathsf{Ent}_\nu(\mu_t^{j,k}) \leq \int_{\mathbb{G}} \mathsf{Ent}_\nu((l_y)_\# \mu_t^j) \, \eta_k(y) dy.
\tag{4.32}
$$

Define $\nu_y := (l_y)_\# \nu$ and note that $\nu_y = e^{-c \, \mathsf{d}_{\mathsf{cc}}(\cdot, y)} \mathscr{L}^n$ for all $y \in \mathbb{G}$. Thus by (2.7) we have

$$
\mathsf{Ent}_\nu((l_y)_\# \mu_t^j) = \mathsf{Ent}_{\nu_{y^{-1}}}(\mu_t^j) = \mathsf{Ent}(\mu_t^j) + c \int_{\mathbb{G}} \mathsf{d}_{\mathsf{cc}}(yx,0)^2 \, d\mu_t^j(x) \qquad \forall y \in \mathbb{G}.
$$

By the dominated convergence theorem we get that $y \mapsto \mathsf{Ent}_\nu((l_y)_\# \mu_t^j)$ is continuous and therefore (4.31) follows by passing to the limit as $k \to +\infty$ in (4.32). Combining (4.30) and (4.31), we get

$$
\lim_{k \to +\infty} \mathsf{Ent}(\mu_t^{j,k}) = \mathsf{Ent}(\mu_t^j) \qquad \forall t \in \mathbb{R}.
\tag{4.33}
$$

Let $(v_t^j)_{t \in \mathbb{R}}$ be as in (4.27) and let $(v_t^{j,k})_{t \in \mathbb{R}}$ be the horizontal time-dependent vector field given by

$$
v_t^{j,k} = \frac{\eta_k \star (\varrho_t^j v_t^j)}{\varrho_t^{j,k}} \qquad \forall t \in \mathbb{R}, \ \forall k \in \mathbb{N}.
\tag{4.34}
$$

We claim that $v_t^{j,k} \in L_{\mathbb{G}}^2(\mu_t^{j,k})$ for all $t \in \mathbb{R}$. Indeed, applying Lemma 4.5, we get

$$
\int_{\mathbb{G}} \|v_t^{j,k}\|_{\mathbb{G}}^2 \, d\mu_t^{j,k} \leq \int_{\mathbb{G}} \|v_t^j\|_{\mathbb{G}}^2 \, d\mu_t^j \leq |\dot{\mu}_t^j|_{\mathbb{G}}^2 \qquad \forall t \in \mathbb{R}, \ \forall k \in \mathbb{N}.
\tag{4.35}
$$

We also claim that $(\mu_t^{j,k})_{t \in \mathbb{R}}$ solves $\partial_t \mu_t^{j,k} + \mathrm{div}(v_t^{j,k} \mu_t^{j,k}) = 0$ in the sense of distributions for all $k \in \mathbb{N}$. Indeed, if $\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^n)$, then also $\varphi^k := \eta_k \star \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^n)$, so that

$$
\begin{aligned}
\int_{\mathbb{R}} \int_{\mathbb{G}} \partial_t \varphi^k(t,x) &+ \left\langle \nabla_{\mathbb{G}} \varphi^k(t,x), v_t^{j,k}(x) \right\rangle_{\mathbb{G}} d\mu_t^{j,k}(x) \, dt = \\
&= \int_{\mathbb{G}} \eta_k \star \left(\int_{\mathbb{R}} \partial_t \varphi(t,\cdot) + \left\langle \nabla_{\mathbb{G}} \varphi(t,\cdot), v_t^j(\cdot) \right\rangle_{\mathbb{G}} \varrho_t^j(\cdot) \, dt \right)(x) \, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{G}} \partial_t \varphi(t,x) + \left\langle \nabla_{\mathbb{G}} \varphi(t,x), v_t^j(x) \right\rangle_{\mathbb{G}} d\mu_t^j(x) \, dt = 0 \qquad \forall k \in \mathbb{N}.
\end{aligned}
$$

Here we have exploited a key property of the space $(\mathbb{G}, \mathsf{d}_{\mathsf{cc}}, \mathscr{L}^n)$ which cannot be expected in a general metric measure space, that is, the continuity equation in (3.4) is preserved under regularization in the space variable. By Proposition 3.3 and (4.35), we conclude

that $(\mu_t^{j,k})_t \in AC_{\mathrm{loc}}^2(\mathbb{R}; (\mathscr{P}_2(\mathbb{G}), \mathsf{W}_{\mathbb{G}}))$ with $|\dot{\mu}_t^{j,k}|_{\mathbb{G}} \le \|v_t^{j,k}\|_{L_{\mathbb{G}}^2(\mu_t^{j,k})} \le \|v_t^j\|_{L_{\mathbb{G}}^2(\mu_t^j)} \le |\dot{\mu}_t^j|_{\mathbb{G}}$ for all $t \in \mathbb{R}$ and $k \in \mathbb{N}$.

Finally, we claim that $\mathsf{F}_{\mathbb{G}}(\varrho_t^{j,k}) \le \mathsf{F}_{\mathbb{G}}(\varrho_t^j)$ for all $t \in \mathbb{R}$ and $k \in \mathbb{N}$. Indeed, arguing as in (4.5), by Cauchy–Schwarz inequality we have

$$\|\nabla_{\mathbb{G}} \varrho_t^{j,k}\|_{\mathbb{G}}^2 \le \left[\eta_k \star \left(\chi_{\{\varrho_t^j > 0\}} \sqrt{\varrho_t^j} \, \frac{\|\nabla_{\mathbb{G}} \varrho_t^j\|_{\mathbb{G}}}{\sqrt{\varrho_t^j}} \right) \right]^2 \le \varrho_t^{j,k} \, \eta_k \star \left(\frac{\|\nabla_{\mathbb{G}} \varrho_t^j\|_{\mathbb{G}}^2}{\varrho_t^j} \chi_{\{\varrho_t^j > 0\}} \right),$$

so that

$$\mathsf{F}_{\mathbb{G}}(\varrho_t^{j,k}) \le \int_{\mathbb{G}} \eta_k \star \left(\frac{\|\nabla_{\mathbb{G}} \varrho_t^j\|_{\mathbb{G}}^2}{\varrho_t^j} \chi_{\{\varrho_t^j > 0\}} \right) \, dx = \int_{\{\varrho_t^j > 0\}} \frac{\|\nabla_{\mathbb{G}} \varrho_t^j\|_{\mathbb{G}}^2}{\varrho_t^j} \, dx = \mathsf{F}_{\mathbb{G}}(\varrho_t^j).$$

Step 3: truncated entropy. Let $j, k \in \mathbb{N}$ be fixed. For any $m \in \mathbb{N}$, consider the maps $z_m, H_m \colon [0, +\infty) \to \mathbb{R}$ defined in (4.2). We set $\tilde{z}_m(r) = z_m(r) + m$ for all $r \ge 0$ and $m \in \mathbb{N}$. Since $\varrho_t^{j,k} \in \mathscr{P}(\mathbb{G})$ for all $t \in \mathbb{R}$, differentiating under the integral sign we get

$$\frac{d}{dt} \int_{\mathbb{G}} H_m(\varrho_t^{j,k}) \, dx = \int_{\mathbb{G}} \tilde{z}_m(\varrho_t^{j,k}) \, \partial_t \varrho_t^{j,k} \, dx$$

for all $t \in \mathbb{R}$. Fix $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$. Then

$$\int_{\mathbb{G}} H_m(\varrho_{t_1}^{j,k}) \, dx - \int_{\mathbb{G}} H_m(\varrho_{t_0}^{j,k}) \, dx = \int_{t_0}^{t_1} \int_{\mathbb{G}} \tilde{z}_m(\varrho_t^{j,k}) \, \partial_t \varrho_t^{j,k} \, dx\, dt. \tag{4.36}$$

Let $(\alpha_i)_{i \in \mathbb{N}} \subset C_c^\infty(t_0, t_1)$ such that $0 \le \alpha_i \le 1$ and $\alpha_i \to \chi_{(t_0, t_1)}$ in $L^1(\mathbb{R})$ as $i \to +\infty$. Let $i \in \mathbb{N}$ be fixed and consider the function $u_t(x) = \tilde{z}_m(\varrho_t^{j,k}(x)) \, \alpha_i(t)$ for all $(t, x) \in \mathbb{R} \times \mathbb{R}^n$. We claim that there exists $(\psi^h)_{h \in \mathbb{N}} \subset C_c^\infty(\mathbb{R}^{n+1})$ such that

$$\lim_{h \to +\infty} \int_{\mathbb{R}} \int_{\mathbb{G}} |u_t(x) - \psi_t^h(x)|^2 + \|\nabla_{\mathbb{G}} u_t(x) - \nabla_{\mathbb{G}} \psi_t^h(x)\|_{\mathbb{G}}^2 \, dx\, dt = 0. \tag{4.37}$$

Indeed, consider the direct product $\mathbb{G}^* = \mathbb{R} \times \mathbb{G}$ and note that $\mathbb{G}^*$ is a Carnot group. Recalling (2.12), we know that $C_c^\infty(\mathbb{R}^{n+1})$ is dense in $W_{\mathbb{G}^*}^{1,2}(\mathbb{R}^{n+1})$. Thus, to get (4.37) we just need to prove that $u \in W_{\mathbb{G}^*}^{1,2}(\mathbb{R}^{n+1})$ (in fact, the L^2-integrability of $\partial_t u$ is not strictly necessary in order to achieve (4.37)). We have $\varrho^{j,k}, \partial_t \varrho^{j,k} \in L^\infty(\mathbb{R}^{n+1})$, because

$$\|\varrho^{j,k}\|_{L^\infty(\mathbb{R}^{n+1})} \le \|\eta_k\|_{L^\infty(\mathbb{R}^n)}, \qquad \|\partial_t \varrho^{j,k}\|_{L^\infty(\mathbb{R}^{n+1})} \le \|\vartheta_j'\|_{L^1(\mathbb{R})} \|\eta_k\|_{L^\infty(\mathbb{R}^n)}$$

by Young's inequality. Moreover, $\varrho^{j,k} \alpha_i, \partial_t \varrho^{j,k} \alpha_i \in L^1(\mathbb{R}^{n+1})$, because

$$\|\varrho^{j,k} \alpha_i\|_{L^1(\mathbb{R}^{n+1})} = \|\alpha_i\|_{L^1(\mathbb{R})}, \qquad \|\partial_t \varrho^{j,k} \alpha_i\|_{L^1(\mathbb{R}^{n+1})} \le \|\vartheta_j'\|_{L^1(\mathbb{R})} \|\alpha_i\|_{L^1(\mathbb{R})}.$$

Therefore $\varrho^{j,k} \alpha_i, \partial_t \varrho^{j,k} \alpha_i \in L^2(\mathbb{R}^{n+1})$, which immediately gives $u \in L^2(\mathbb{R}^{n+1})$. Now by *Step 2* we have that

$$\int_{\mathbb{G}} \|\nabla_{\mathbb{G}} \varrho_t^{j,k}\|_{\mathbb{G}}^2 \, dx \le \|\varrho_t^{j,k}\|_{L^\infty(\mathbb{R}^n)} \mathsf{F}_{\mathbb{G}}(\varrho_t^{j,k}) \le \|\varrho_t^{j,k}\|_{L^\infty(\mathbb{R}^n)} \mathsf{F}_{\mathbb{G}}(\varrho_t^j) \qquad \forall t \in \mathbb{R},$$

so that by *Step 1* we get

$$\|\nabla_{\mathbb{G}} \varrho^{j,k} \alpha_i\|_{L^2(\mathbb{R}^{n+1})}^2 \le \|\eta_k\|_{L^\infty(\mathbb{R}^n)} \|\alpha_i^2 \cdot \vartheta_j * \mathsf{F}_{\mathbb{G}}(\varrho_.)\|_{L^1(\mathbb{R})}.$$

This prove that $\|\nabla_{\mathbb{G}} u\|_{\mathbb{G}} \in L^2(\mathbb{R}^{n+1})$. The previous estimates easily imply that also $\partial_t u \in L^2(\mathbb{R}^{n+1})$. This concludes the proof of (4.37).

Since $\varrho^{j,k}, \partial_t \varrho^{j,k} \in L^\infty(\mathbb{R}^{n+1})$, by (4.37) we get that

$$\lim_{h \to +\infty} \int_\mathbb{R} \int_\mathbb{G} \psi_t^h \, \partial_t \varrho_t^{j,k} \, dx dt = \int_\mathbb{R} \alpha_i(t) \int_\mathbb{G} \tilde{z}_m(\varrho_t^{j,k}) \, \partial_t \varrho_t^{j,k} \, dx dt \qquad (4.38)$$

and

$$\lim_{h \to +\infty} \int_\mathbb{R} \int_\mathbb{G} \left\langle \nabla_\mathbb{G} \psi_t^h, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt = \int_\mathbb{R} \alpha_i(t) \int_\mathbb{G} \tilde{z}_m'(\varrho_t^{j,k}) \left\langle \nabla_\mathbb{G} \varrho_t^{j,k}, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt \qquad (4.39)$$

for each fixed $i \in \mathbb{N}$. Since $\partial_t \mu_t^{j,k} + \operatorname{div}(v_t^{j,k} \mu_t^{j,k}) = 0$ in the sense of distributions by *Step 2*, for each $h \in \mathbb{N}$ we have

$$\int_\mathbb{R} \int_\mathbb{G} \psi_t^h \, \partial_t \varrho_t^{j,k} \, dx dt = -\int_\mathbb{R} \int_\mathbb{G} \partial_t \psi_t^h \, d\mu_t^{j,k} \, dt = \int_\mathbb{R} \int_\mathbb{G} \left\langle \nabla_\mathbb{G} \psi_t^h, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt.$$

We can thus combine (4.38) and (4.39) to get

$$\int_\mathbb{R} \alpha_i(t) \int_\mathbb{G} \tilde{z}_m(\varrho_t^{j,k}) \, \partial_t \varrho_t^{j,k} \, dx dt = \int_\mathbb{R} \alpha_i(t) \int_\mathbb{G} \tilde{z}_m'(\varrho_t^{j,k}) \left\langle \nabla_\mathbb{G} \varrho_t^{j,k}, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt. \qquad (4.40)$$

Passing to the limit as $i \to +\infty$ in (4.40), we finally get that

$$\int_{t_0}^{t_1} \int_\mathbb{G} \tilde{z}_m(\varrho_t^{j,k}) \, \partial_t \varrho_t^{j,k} \, dx dt = \int_{t_0}^{t_1} \int_\mathbb{G} \tilde{z}_m'(\varrho_t^{j,k}) \left\langle \nabla_\mathbb{G} \varrho_t^{j,k}, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt \qquad (4.41)$$

by the dominated convergence theorem. Combining (4.36) and (4.41), we get

$$\int_\mathbb{G} H_m(\varrho_{t_1}^{j,k}) \, dx - \int_\mathbb{G} H_m(\varrho_{t_0}^{j,k}) \, dx = \int_{t_0}^{t_1} \int_{\left\{ e^{-m-1} < \varrho_t^{j,k} < e^{m-1} \right\}} \left\langle -w_t^{j,k}, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt \qquad (4.42)$$

for all $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$, where $w_t^{j,k} = \nabla_\mathbb{G} \varrho_t^{j,k} / \varrho_t^{j,k}$ in $L_\mathbb{G}^2(\mu_t^{j,k})$ for all $t \in \mathbb{R}$.

We can now conclude the proof. We pass to the limit as $m \to +\infty$ in (4.36) and we get

$$\mathsf{Ent}(\mu_{t_1}^{j,k}) - \mathsf{Ent}(\mu_{t_0}^{j,k}) = \int_{t_0}^{t_1} \int_\mathbb{G} \left\langle -w_t^{j,k}, v_t^{j,k} \right\rangle_\mathbb{G} \, d\mu_t^{j,k} \, dt \qquad (4.43)$$

for all $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$. For the left-hand side of (4.36), recall (4.10) and (4.11) and apply the monotone convergence theorem on $\left\{ \varrho_t^{j,k} \le 1 \right\}$ and the dominated convergence theorem on $\left\{ \varrho_t^{j,k} > 1 \right\}$. For the right-hand side of (4.36), recall that $t \mapsto \|v_t^{j,k}\|_{L_\mathbb{G}^2(\mu_t^{j,k})} \in L_{\mathrm{loc}}^2(\mathbb{R})$ and that $t \mapsto \|w_t^{j,k}\|_{L_\mathbb{G}^2(\mu_t^{j,k})} = \mathsf{F}_\mathbb{G}^{1/2}(\varrho_t^{j,k}) \in L_{\mathrm{loc}}^2(\mathbb{R})$ by *Step 2* and apply Cauchy–Schwarz inequality and the dominated convergence theorem.

We pass to the limit as $k \to +\infty$ in (4.43) and we get

$$\mathsf{Ent}(\mu_{t_1}^j) - \mathsf{Ent}(\mu_{t_0}^j) = \int_{t_0}^{t_1} \int_\mathbb{G} \left\langle -w_t^j, v_t^j \right\rangle_\mathbb{G} \, d\mu_t^j \, dt \qquad (4.44)$$

for all $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$, where $w_t^j = \nabla_\mathbb{G} \varrho_t^j / \varrho_t^j$ in $L_\mathbb{G}^2(\mu_t^j)$ for all $t \in \mathbb{R}$. For the left-hand side of (4.43), recall (4.33). For the right-hand side of (4.43), recall that $t \mapsto \|v_t^j\|_{L_\mathbb{G}^2(\mu_t^j)} \in L_{\mathrm{loc}}^2(\mathbb{R})$ and that $t \mapsto \|w_t^j\|_{L_\mathbb{G}^2(\mu_t^{j,k})} = \mathsf{F}_\mathbb{G}^{1/2}(\varrho_t^j) \in L_{\mathrm{loc}}^2(\mathbb{R})$ by *Step 1*, so that the conclusion follows applying Lemma 4.5, Lemma 4.7, Cauchy–Schwarz inequality and the dominated convergence theorem.

We finally pass to the limit as $j \to +\infty$ in (4.44) and we get

$$\mathsf{Ent}(\mu_{t_1}) - \mathsf{Ent}(\mu_{t_0}) = \int_{t_0}^{t_1} \int_\mathbb{G} \left\langle -w_t^\mathbb{G}, v_t^\mathbb{G} \right\rangle_\mathbb{G} \, d\mu_t \, dt \qquad (4.45)$$

for all $t_0, t_1 \in \mathbb{R} \setminus \mathcal{N}$ with $t_0 < t_1$, where $\mathcal{N} \subset \mathbb{R}$ is the set of discontinuity points of $t \mapsto \mathsf{Ent}(\mu_t)$ and $w_t^{\mathbb{G}} = \nabla_{\mathbb{G}} \varrho_t / \varrho_t$ in $L_{\mathbb{G}}^2(\mu_t)$ for a.e. $t \in \mathbb{R}$ by Proposition 3.4. For the left-hand side of (4.44), recall (4.26). For the right-hand side of (4.44), recall that $t \mapsto \|v_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t)} \in L_{\mathrm{loc}}^2(\mathbb{R})$ and that $t \mapsto \|w_t^{\mathbb{G}}\|_{L_{\mathbb{G}}^2(\mu_t^{j,k})} = \mathsf{F}_{\mathbb{G}}^{1/2}(\varrho_t) \in L_{\mathrm{loc}}^2(\mathbb{R})$, so that the conclusion follows applying Lemma 4.6, Lemma 4.7, Cauchy–Schwarz inequality and the dominated convergence theorem. From (4.45) we immediately deduce (4.15) and we can conclude the proof by Lemma 4.4. In particular, by Proposition 4.2 the map $t \mapsto \mathsf{Ent}(\mu_t)$ is locally absolutely continuous on $(0, +\infty)$. $\qquad\square$

References

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000.
[2] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, Modelling and optimisation of flows on networks, Lecture Notes in Math., vol. 2062, Springer, Heidelberg, 2013, pp. 1–155.
[3] L. Ambrosio, N. Gigli, A. Mondino, and T. Rajala, *Riemannian Ricci curvature lower bounds in metric measure spaces with σ-finite measure*, Trans. Amer. Math. Soc. **367** (2015), no. 7, 4661–4701.
[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[5] L. Ambrosio, N. Gigli, and G. Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math. **195** (2014), no. 2, 289–391.
[6] L. Ambrosio, N. Gigli, and G. Savaré, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Duke Math. J. **163** (2014), no. 7, 1405–1490.
[7] L. Ambrosio and G. Savaré, *Gradient flows of probability measures*, Handbook of differential equations: evolutionary equations. Vol. III, Handb. Differ. Equ., Elsevier/North-Holland, Amsterdam, 2007, pp. 1–136.
[8] F. Baudoin and M. Bonnefont, *Reverse Poincaré inequalities, isoperimetry, and Riesz transforms in Carnot groups*, Nonlinear Anal. **131** (2016), 48–59.
[9] P. Bernard, *Young measures, superposition and transport*, Indiana Univ. Math. J. **57** (2008), no. 1, 247–275.
[10] A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni, *Stratified Lie groups and potential theory for their sub-Laplacians*, Springer Monographs in Mathematics, Springer, Berlin, 2007.
[11] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973 (French). North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).
[12] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson, *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, Progress in Mathematics, vol. 259, Birkhäuser Verlag, Basel, 2007.
[13] E. A. Carlen and J. Maas, *An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy*, Comm. Math. Phys. **331** (2014), no. 3, 887–926.
[14] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 1, 1–23.
[15] M. Erbar and J. Maas, *Gradient flow structures for discrete porous medium equations*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1355–1374.
[16] S. Fang, J. Shao, and K.-T. Sturm, *Wasserstein space over the Wiener space*, Probab. Theory Related Fields **146** (2010), no. 3-4, 535–565.
[17] B. Franchi, R. Serapioni, and F. Serra Cassano, *Meyers-Serrin type theorems and relaxation of variational integrals depending on vector fields*, Houston J. Math. **22** (1996), no. 4, 859–890.
[18] ———, *On the structure of finite perimeter sets in step 2 Carnot groups*, J. Geom. Anal. **13** (2003), no. 3, 421–466.

[19] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), no. 1-2, 101–120.

[20] N. Gigli and B.-X. Han, *The continuity equation on metric measure spaces*, Calc. Var. Partial Differential Equations **53** (2015), no. 1-2, 149–177.

[21] N. Gigli, K. Kuwada, and S.-I. Ohta, *Heat flow on Alexandrov spaces*, Comm. Pure Appl. Math. **66** (2013), no. 3, 307–331.

[22] N. Gigli and S.-I. Ohta, *First variation formula in Wasserstein spaces over compact Alexandrov spaces*, Canad. Math. Bull. **55** (2012), no. 4, 723–735.

[23] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17.

[24] N. Juillet, *Geometric inequalities and generalized Ricci bounds in the Heisenberg group*, Int. Math. Res. Not. IMRN **13** (2009), 2347–2373.

[25] _____, *Diffusion by optimal transport in Heisenberg groups*, Calc. Var. Partial Differential Equations **50** (2014), no. 3-4, 693–721.

[26] B. Khesin and P. Lee, *A nonholonomic Moser theorem and optimal transport*, J. Symplectic Geom. **7** (2009), no. 4, 381–414.

[27] E. Le Donne, *A primer on Carnot groups: homogenous groups, Carnot-Carathéodory spaces, and regularity of their isometries*, Anal. Geom. Metr. Spaces **5** (2017), 116–137.

[28] H.-Q. Li, *Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg*, J. Funct. Anal. **236** (2006), no. 2, 369–394.

[29] _____, *Estimations asymptotiques du noyau de la chaleur sur les groupes de Heisenberg*, C. R. Math. Acad. Sci. Paris **344** (2007), no. 8, 497–502.

[30] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. **261** (2011), no. 8, 2250–2292.

[31] J. Milnor, *Curvatures of left invariant metrics on Lie groups*, Advances in Math. **21** (1976), no. 3, 293–329.

[32] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, vol. 91, American Mathematical Society, Providence, RI, 2002.

[33] S.-i. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math. **131** (2009), no. 2, 475–516.

[34] S.-i. Ohta and K.-T. Sturm, *Heat flow on Finsler manifolds*, Comm. Pure Appl. Math. **62** (2009), no. 10, 1386–1433.

[35] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[36] A. Petrunin, *Alexandrov meets Lott-Villani-Sturm*, Münster J. Math. **4** (2011), 53–64.

[37] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), no. 7, 923–940.

[38] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris **345** (2007), no. 3, 151–154.

[39] N. Th. Varopoulos, L. Saloff-Coste, and T. Coulhon, *Analysis and geometry on groups*, Cambridge Tracts in Mathematics, vol. 100, Cambridge University Press, Cambridge, 1992.

[40] C. Villani, *Optimal transport, old and new*, Fundamental Principles of Mathematical Sciences, vol. 338, Springer-Verlag, Berlin, 2009.

[41] F.-Y. Wang, *Equivalent semigroup properties for the curvature-dimension condition*, Bull. Sci. Math. **135** (2011), no. 6-7, 803–815.

Scuola Normale Superiore, Piazza Cavalieri 7, 56126 Pisa, Italy
E-mail address: `luigi.ambrosio@sns.it`

Scuola Normale Superiore, Piazza Cavalieri 7, 56126 Pisa, Italy
E-mail address: `giorgio.stefani@sns.it`